



TECHNISCHE  
UNIVERSITÄT  
DARMSTADT

FINE-GRANULAR SENSING OF POWER CONSUMPTION —  
A New Sensing Modality  
for an Accurate Detection, Prediction and Forecasting of Higher-Level Contextual  
Information in Smart Environments

Vom Fachbereich Elektrotechnik und Informationstechnik  
der Technischen Universität Darmstadt  
zur Erlangung des akademischen Grades eines  
Doktor-Ingenieurs (Dr.-Ing.)  
genehmigte Dissertation

von

ALAA ALHAMOUD, M.SC.

Geboren am 29. Juni 1985 in Homs, Syrien

Vorsitz: Prof. Dr.-Ing. Helmut F. Schlaak  
Referent: Prof. Dr.-Ing. Ralf Steinmetz  
Korreferent: Prof. Dr.-Ing. Lars Wolf

Tag der Einreichung: 14. September 2016  
Tag der Disputation: 22. November 2016

Hochschulkenziffer D17  
Darmstadt 2016

Dieses Dokument wird bereitgestellt von  
tuprints, E-Publishing-Service der Technischen Universität Darmstadt.  
<http://tuprints.ulb.tu-darmstadt.de>  
[tuprints@ulb.tu-darmstadt.de](mailto:tuprints@ulb.tu-darmstadt.de)

Bitte zitieren Sie dieses Dokument als:

URN: [urn:nbn:de:tuda-tuprints-58396](https://nbn-resolving.org/urn:nbn:de:tuda-tuprints-58396)

URL: <http://tuprints.ulb.tu-darmstadt.de/id/eprint/5839>

Die Veröffentlichung steht unter folgender Creative Commons Lizenz:

*International 4.0 – Namensnennung, nicht kommerziell, keine Bearbeitung*

<https://creativecommons.org/licenses/by-nc-nd/4.0/>



*If you can't explain it simply,  
you don't understand it well enough.*

— Albert Einstein



---

## ABSTRACT

---

INVESTIGATING and extracting users' context has always been a main drive for research in the field of smart environments. Context is defined as the set of conditions that characterize a certain situation in which the user exists. An example of user's context is the activity currently performed by the user along with her/his current indoor location. Identifying users' context paves the road for the realization of a wide variety of context-aware services that improve the quality of life for individuals as well as societies. To extract users' context, researchers have resorted to the approach of deploying a huge number of sensors in the environment where users' context has to be extracted. Examples of these sensors are cameras, microphones, motion and contact sensors, state-change sensors and other types of sensors that monitor every aspect of users' context. However, this approach has always been criticized as causing a huge deployment and maintenance overhead for researchers. Moreover, it has been perceived as an intrusive approach by users because they feel themselves surrounded by all possible types of visible sensors.

Recent years have seen an increasing adoption of smart metering technologies along with the manufacturing of new appliance-level power sensors that are able to measure the fine-granular power consumption of individual devices in smart environments. As a result, fine-granular sensing of power consumption has emerged as a new sensing modality that avoids the afore-mentioned problems of other sensing modalities. One of the main goals of this thesis is to develop intelligent models that infer and predict several challenging and essential aspects of users' context only based on fine-granular sensing of power consumption.

Activities of daily living (ADL) represent an essential part of users' context that has always motivated researchers in the field of smart environments. Context-aware services such as energy conservation in smart environments and ambient assisted living can be realized based on the recognition of users' current activity. In this thesis, we develop *SMARTENERGY.KOM*, an intelligent hardware/software platform for recognizing activities of users in single-user environments. We build an activity recognition model and evaluate its performance based on a dataset we collect by deploying *SMARTENERGY.KOM* in two single-user apartments.

As an essential part of this thesis, we conduct an in-depth analytical study on the dataset collected by *SMARTENERGY.KOM* with three main contributions, namely modeling of user's daily behavior, indoor localization based on fine-granular power consumption data, and profiling of user's hourly power consumption. As users tend to follow a daily routine in performing their activities, identifying behavioral patterns of users helps to improve the predictive performance of activity recognition models. We develop an approach that identifies such patterns of user's behavior. Evaluation results show that feeding these patterns into the model of activity recognition leads to a significant improvement in its predictive performance. Indoor location represents another important aspect of users' context that has its potential benefits in realizing location-aware services. We develop a localization model that is able to determine the indoor location of users based on their fine-granular power

consumption. Building a profile that characterizes average hourly power consumption of users has its potential benefits in increasing their awareness of the power they consume as well as in detecting abnormal consumption patterns. Driven by this motivation, we develop an approach that identifies and builds this profile based on *SMARTENERGY.KOM* dataset.

As more than one user tend to live, work and reside in one common place, activity recognition models need to cope with the fact that parallel and overlapping activities of several users have to be recognized and assigned to their respective users. This fact has always represented a great challenge for researchers in the field of activity recognition. In this thesis, we address this challenge by developing *ML-SMARTENERGY.KOM*, our platform for activity recognition in multi-user environments. We develop our model for recognizing activities based on the concept of multi-label classification, which exploits label dependency and temporal relations between activities.

Forecasting fine-granular power consumption of individual consumers represents another aspect of users' context that has its potential benefits for consumers as well as electric utilities. In this thesis, we develop state-of-the-art forecasting models that are able to forecast hourly, daily, and monthly power consumption of individual buildings based on building characteristics, demographic features of residents, available appliances as well as historical power consumption values. We evaluate these models using a dataset collected by Commission for Energy Regulation (CER) in Ireland.

---

## KURZFASSUNG

---

**D**IE Erfassung und Analyse von Nutzerkontext ist ein großer Bestandteil der Forschung auf dem Gebiet der intelligenten Umgebungen. Kontext wird als die Menge von Bedingungen definiert, die eine bestimmte Situation charakterisieren, in der der Nutzer existiert. Ein Beispiel für Kontext ist die Aktivität, die gerade durch den Nutzer an seinem aktuellen Standort durchgeführt wird. Die Identifizierung des Nutzerkontexts ebnet den Weg für die Realisierung einer Vielzahl von kontextsensitiven Diensten, die die Lebensqualität des Einzelnen als auch ganzer Gesellschaften verbessern können. Um den Kontext des Nutzers zu extrahieren, wird in verwandten Arbeiten der Ansatz verfolgt, eine große Anzahl von Sensoren in der Umgebung einzusetzen, in der der Nutzerkontext erfasst werden soll. Beispiele für solche Sensoren sind Kameras, Mikrofone, Bewegungsmelder, Kontaktsensoren, Zustandsänderungssensoren und andere Arten von Sensoren, die jeden Aspekt des Nutzerkontexts überwachen. Allerdings wurde dieser Ansatz immer kritisiert, weil er großen Einsatz- und Wartungsaufwand verursacht. Darüber hinaus wird dieser Ansatz als invasive Methode von den Nutzern wahrgenommen, weil sie sich umgebend von allen möglichen Arten von sichtbaren Sensoren unwohl fühlen.

Die Einführung von Smart-Metering-Technologien einhergehend mit der Innovation von Stromsensoren, die in der Lage sind den fein-granularen Stromverbrauch einzelner Geräte in intelligenten Umgebungen zu messen, hat in den letzten Jahren stark zugenommen. Als Ergebnis meiner Arbeit hat sich die fein-granulare Erfassung des Stromverbrauchs als neue Erfassungsmodalität herausgestellt, die die oben genannten Probleme der anderen Erfassungsmodalitäten vermeidet. Eines der wichtigsten Ziele dieser Arbeit war es, intelligente Modelle zu entwickeln, die in der Lage sind, herausfordernde und wesentliche Aspekte des Nutzerkontexts - sowohl für einzelne Nutzer als auch für mehrere Nutzer und sowohl für einzeln als auch parallel durchgeführte Aktivitäten - anhand der fein-granularen Erfassung des Stromverbrauchs zu erkennen und vorherzusagen.

Aktivitäten des täglichen Lebens stellen einen wesentlichen Teil des Nutzerkontexts dar. Kontextbewusste Dienste wie beispielsweise zur Energieeinsparung in intelligenten Umgebungen und umgebungsunterstütztes Leben können auf Basis der Erkennung der aktuellen Nutzeraktivität realisiert werden. In dieser Arbeit wird *SMARTENERGY.KOM* konzipiert und entwickelt, eine intelligente Hardware/Software Plattform für die Erkennung von Nutzeraktivitäten in Einzelnutzer-Umgebungen. Es wird ein Aktivitätserkennungsmodell erstellt und basierend auf einem Datensatz, der durch den Einsatz von *SMARTENERGY.KOM* in zwei Einzelnutzer-Wohnungen gesammelt wird, wird seine prädiktive Performanz bewertet. Als ein wesentlicher Teil dieser Arbeit wird eine gründliche analytische Studie über den *SMARTENERGY.KOM*-Datensatz mit drei wichtigen Beiträgen durchgeführt, das sind die Modellierung des täglichen Benutzerverhaltens, Lokalisierung in Gebäuden basierend auf fein-granularen Stromverbrauchsdaten und die Profilierung des stündlichen Stromverbrauchs. Menschen neigen dazu, eine tägliche Routine bei der Durchführung ihrer Aktivitäten zu verfolgen, die Identifizierung von Verhaltensmustern hilft dabei, die

prädiktive Performanz der Aktivitätserkennungsmodelle zu verbessern. Daher wird ein Ansatz entwickelt, der solche Verhaltensmuster identifizieren kann. Die Auswertungsergebnisse zeigen, dass die Integration dieser Muster in das Aktivitätserkennungsmodell zu einer deutlichen Verbesserung der prädiktiven Performanz dieses Modells führt. Lokalisierung in Gebäuden ist ein weiterer wichtiger Aspekt des Nutzerkontexts, der die Realisierung von standortbezogenen Dienste ermöglicht. Es wird ein Lokalisierungsmodell entwickelt, welches in der Lage ist, den Aufenthaltsort der Nutzer in Gebäuden basierend auf ihrem fein-granularen Stromverbrauch zu bestimmen. Der Aufbau eines Profils, das den durchschnittlichen, stündlichen Stromverbrauch charakterisiert, ermöglicht es, das Bewusstsein für den Stromverbrauch zu erhöhen, sowie ungewöhnliche Konsummuster zu entdecken. Angetrieben von dieser Motivation, wird in dieser Arbeit der Ansatz verfolgt, dieses Profil auf Basis von *SMARTENERGY.KOM*-Datensatz zu identifizieren und zu erstellen.

Neben überwiegend einzeln genutzten Wohnungen gibt es viele Orte, an denen gemeinsam gelebt bzw. gearbeitet wird. Damit müssen Aktivitätserkennungsmodelle auch überlappende Aktivitäten mehrerer Nutzer erkennen und ihrem jeweiligen Nutzern zuordnen. Dies stellt eine große Herausforderung auf dem Gebiet der Aktivitätserkennung dar. Diese Arbeit befasst sich daher mit dieser Herausforderung durch die Entwicklung von *ML-SMARTENERGY.KOM*, unsere Plattform für die Aktivitätserkennung in Multi-Nutzer-Umgebungen. Das Aktivitätserkennungsmodell wird basierend auf dem Konzept der Multi-Label-Klassifikation entwickelt, das die Abhängigkeit von Klassen und die zeitlichen Beziehungen zwischen den Aktivitäten ausnutzt.

Die Vorhersage des fein-granularen Stromverbrauchs einzelner Verbraucher stellt einen weiteren Aspekt des Nutzerkontexts dar, der verschiedene potenzielle Vorteile für die Verbraucher als auch Stromversorger bietet. In dieser Arbeit werden Vorhersagemodelle entwickelt, die in der Lage sind, stündliche, tägliche, und monatliche Stromverbräuche der einzelnen Gebäude auf Basis von Gebäudeeigenschaften, demographischen Merkmalen der Bewohner, verfügbaren Geräten sowie historischen Stromverbrauchswerten vorherzusagen. Diese Modelle werden auf Basis des von der Commission for Energy Regulation (CER) in Irland gesammelten Datensatzes evaluiert.



---

## CONTENTS

---

1	INTRODUCTION .....	1
1.1	Motivation .....	1
1.2	Research Goals and Contributions .....	2
1.3	Structure of the Thesis .....	3
2	BACKGROUND .....	5
2.1	Machine Learning .....	5
2.1.1	Single-Label Classification .....	6
2.1.2	Multi-Label Classification .....	10
3	RELATED WORK .....	17
3.1	Activity Recognition and Behavioral Modeling .....	17
3.1.1	Single-User Environments .....	18
3.1.2	Multi-User Environments .....	20
3.2	Forecasting of Power Consumption .....	22
3.3	Summary and Discussion .....	26
4	ACTIVITY RECOGNITION IN SINGLE-USER ENVIRONMENTS .....	29
4.1	Activities of Daily Living (ADLs) .....	29
4.2	SMARTENERGY.KOM Activity Recognition Platform .....	29
4.3	Deployments and Experimental Setup .....	30
4.4	Data Preprocessing and Feature Extraction .....	32
4.5	Results and Evaluation .....	33
4.5.1	Optimal Time Slot Length .....	34
4.5.2	Power and Environmental Sensors .....	35
4.5.3	Power Sensors .....	37
4.6	Summary .....	39
5	BEHAVIORAL MODELING OF USERS IN INDOOR ENVIRONMENTS .....	41
5.1	Indoor Localization Based on Power Consumption .....	41
5.1.1	Dataset .....	42
5.1.2	Evaluation .....	42
5.2	Patterns in User's Daily Life and Their Effect on the Accuracy of Activity Recognition .....	44
5.2.1	Identification of Temporal Patterns .....	45
5.2.2	Evaluation .....	48
5.3	Power Profiling .....	48
5.3.1	Extracting Hourly Power Consumption .....	49
5.3.2	Similarity Analysis .....	53
5.3.3	Further Analysis .....	55
5.4	Summary .....	57
6	ACTIVITY RECOGNITION IN MULTI-USER ENVIRONMENTS .....	59
6.1	ML-SMARTENERGY.KOM Activity Recognition Platform .....	59
6.2	Data Preprocessing and Feature Extraction .....	61
6.3	Binary Relevance Model of Activity Recognition .....	62
6.3.1	Evaluation .....	67
6.4	Label Dependency in ML-SMARTENERGY.KOM Dataset .....	72

6.4.1	Intra-User Label Dependency .....	73
6.4.2	Inter-User Label Dependency .....	74
6.5	ML-SMARTENERGY.KOM Activity Recognition Model .....	76
6.5.1	Evaluation .....	79
6.6	Summary .....	81
7	HYBRID LARGE SCALE FORECASTING OF POWER CONSUMPTION .....	83
7.1	Dataset .....	83
7.2	Long-term Forecasting of Power Consumption .....	84
7.2.1	Building-Performance Multiple Regression Model .....	85
7.2.2	Hybrid Model .....	87
7.3	Short-term Forecasting of Power Consumption .....	91
7.3.1	Grouping of Buildings .....	92
7.3.2	Hourly Consumption Pattern Matching (HCPM) Model .....	98
7.3.3	Total Consumption Pattern Matching (TCPM) Model .....	106
7.4	Summary .....	109
8	SUMMARY AND OUTLOOK .....	111
8.1	Summary .....	111
8.2	Outlook .....	113
	 BIBLIOGRAPHY .....	 115
A	AUTHOR'S PUBLICATIONS .....	125
A.1	Main Publications .....	125
A.2	Co-Authored Publications .....	125
B	CURRICULUM VITÆ .....	127
C	SUPERVISED STUDENTS .....	129
D	ERKLÄRUNG LAUT §9 DER PROMOTIONSORDNUNG .....	131

---

## INTRODUCTION

---

### 1.1 MOTIVATION

**F**INE-GRANULAR sensing of power consumption has become a new sensing modality after the wide spread of smart metering technologies [18]. Moreover, the emergence of new appliance-level power sensors made it possible to obtain power consumption data from several single devices in smart environments [88]. With electricity virtually present in a majority of homes, power consumption has become an obvious modality to infer context information. The availability of such fine-granular measurements paves the road for an accurate extraction and utilization of users' context in smart environments solely based on power measurements and without the need for any other sensing modalities. Anind K. Dey has provided a general, clear and widely accepted definition of context [35, Page 2]:

“Context is any information that can be used to characterize the situation of an entity. An entity is a person, place, or object that is considered relevant to the interaction between a user and an application, including the user and applications themselves.”

The identification, recognition and prediction of users' context have always been considered as main enablers for a wide variety of context-aware systems [97] such as ambient assisted living [83][19][37], energy saving in smart spaces [20] and user-oriented communication services [45].

As parts of users' context, we focus in this work on the recognition of daily activities in single-user and multi-user environments, modeling of users' daily behavior, indoor localization of users, power profiling and forecasting of long-term and short-term power consumption.

Activity recognition represents an essential basis of several intelligent services that can be realized in the context of smart environments. By recognizing activities of users, we can realize services that help potential groups of people in their daily life such as elderly care [94] and energy conservation services [5]. Researchers in the field have always used a huge amount of sensors in order to recognize daily activities of users [105]. We argue that with the availability of fine-granular information about power consumption, we can build powerful predictive models that can recognize human daily activities without the need of installing any other additional sensors. Moreover, multi-user environments where more than one user resides and performs co-temporal activities represent a great challenge for researchers in the field of activity recognition. The challenge lies in the fact that parallel activities of several users have to be recognized and assigned to their respective users. With the help of fine-granular power sensing and the techniques of multi-label classification, this challenge can be solved as we show in Chapter 6.

Getting insights into user lifestyle and whereabouts helps in providing users with personalized intelligent services that assist them in their everyday life [83]. Fine-

granular power measurements offer a great potential for extracting such insights. One example is the indoor localization of users. It represents an important aspect of any smart environment that aims to provide users with location-aware services. Indoor localization is usually realized using motion sensors that sense and report movement in the environment. In our work, we show that it can be realized based on power measurements and without the need for installing any other sensing modalities [6]. Apart from that, human beings tend to perform their daily activities in a daily routine which repeats itself every day. Identifying insights about such routines can be of a great benefit in improving the predictive accuracy of activity recognition models.

Forecasting of hourly power consumption of individual buildings represents another aspect that can be realized with the help of fine-granular measurements of power consumption. Forecasting of such data can be of a great benefit to consumers as well as electric utilities. By knowing this information, consumers can become aware of their future expected consumption values based on their historical consumption. This awareness can lead consumers to adjust their consumption so that they avoid expensive power prices in peak hours. Electric utilities can also benefit from such information in adjusting their power generation so that they avoid shortage and surplus of generated power.

## 1.2 RESEARCH GOALS AND CONTRIBUTIONS

This work has the main goal of utilizing fine-granular power consumption data as the main modality for inferring and predicting different aspects of users' context in smart environments. This thesis provides a list of contributions to achieving this goal and addressing the motivation presented in the previous section. These contributions are summarized in the following list and their dependencies are depicted in Figure 1:

- The development of *SMARTENERGY.KOM*, an intelligent hardware/software platform for activity recognition in single-user environments based on the installation of appliance-level power sensors.
- The deployment of *SMARTENERGY.KOM* in two single-user apartments and the collection of a dataset which we use in the development of an accurate predictive model for recognizing activities in single-user environments.
- An in-depth analytical study of the dataset collected by *SMARTENERGY.KOM* with the following main contributions:
  - Identifying and extracting behavioral patterns in users' daily life.
  - Utilizing the extracted patterns for improving the predictive power of our single-user activity recognition model. The proposed model achieves a very good predictive performance with an average f-measure value of 94.5% in recognizing users' activities solely based on their fine-granular power consumption data combined with their identified behavioral patterns.
  - The development of an indoor localization method that extracts the indoor location of users based on their fine-granular power consumption data.

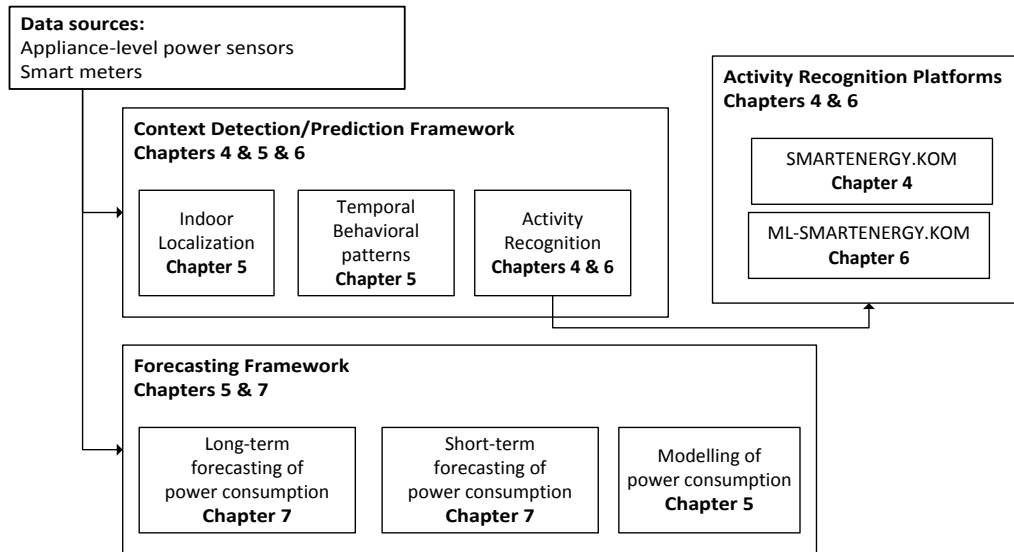


Figure 1: Research Contributions

- Building a profile of user’s hourly power consumption with the goal of creating a model that characterizes the normal average hourly power consumption of users in smart environments. Such model can increase users’ awareness of the power consumed and detect abnormal power consumption values.
- The development of *ML-SMARTENERGY.KOM*, an extension of *SMARTENERGY.KOM* which provides activity recognition in multi-user environments.
- The deployment of *ML-SMARTENERGY.KOM* in a two-user apartment where a multi-label activity recognition dataset has been collected.
- The development of an accurate multi-label activity recognition model using the techniques of multi-label classification. This model represents one of the first activity recognition models that extract, represent and exploit label dependency and temporal relations between activities for improving the predictive performance of activity recognition in multi-user environments. The proposed model achieves a very good predictive performance with an average f-measure value of 91%.
- The development of short-term and long-term power forecasting models based on a fine-granular power consumption dataset collected for thousands of buildings by Commission for Energy Regulation (CER) in Ireland [25].

### 1.3 STRUCTURE OF THE THESIS

This work is structured as follows. In Chapter 2, we introduce the background knowledge necessary to understand further parts of this work. We discuss several research projects related to the contributions of this work in Chapter 3. We present

the strengths and weaknesses of each research project and clarify its relation to our work. In Chapter 4, we introduce *SMARTENERGY.KOM*, our platform for activity recognition in single-user environments. This chapter presents the hardware as well as the software components of the platform. Moreover, it describes the deployment of *SMARTENERGY.KOM* in two single-user apartments and the collection process of the dataset required to the development of activity recognition model. Chapter 5 presents an in-depth analytical study covering three main contributions. Firstly, it explains the process of modeling user's behavior and improving the predictive performance of activity recognition models by feeding behavioral patterns of the user into them. Secondly, it presents our approach of indoor localization based on fine-granular power consumption data. Thirdly, it introduces our approach of building daily power consumption profiles of users in smart environments. Chapter 6 introduces *ML-SMARTENERGY.KOM*, our platform for activity recognition in multi-user environments. This chapter introduces our concept for extracting, representing, and utilizing label dependency and temporal relations between activities in multi-user environments to enhance the predictive performance of activity recognition models. Chapter 7 introduces our approaches of long-term and short-term forecasting of power consumption. It covers an extensive evaluation study of these approaches based on a fine-granular dataset of power consumption measurements collected by CER in Ireland. We conclude this work and present an overview of potential future research work in Chapter 8.

---

## BACKGROUND

---

**I**N this chapter, we introduce the background knowledge necessary to understand the main contributions of this thesis. We start by defining the field of machine learning, its main sub-fields, and the different learning strategies used to build a machine learning model. Afterwards, we explain in details two types of supervised learning problems. We start by introducing single-label classification, its important algorithms and the main performance metrics used to evaluate single-label classification problems. After introducing the problem of single-label classification, we further introduce multi-label classification (MLC) as an extension to it. We present the two main approaches used to handle multi-label classification problems, namely problem transformation and algorithm adaptation. Moreover, we introduce the metrics used to extract the multi-label characteristics of a dataset as well as the main performance metrics used in evaluating multi-label classification problems.

### 2.1 MACHINE LEARNING

Machine learning can be defined as the process in which a computer program progressively and continuously learns to perform a task normally done by a human agent by introducing it to a series of examples related to this task. Tom M. Mitchell has provided a general and clear definition of machine learning [73, Page 2]:

“A computer program is said to learn from experience  $E$  with respect to some class of tasks  $T$  and performance measure  $P$ , if its performance at tasks in  $T$ , as measured by  $P$ , improves with experience  $E$ ”.

An example of a task  $T$  handled by machine learning is the problem of face recognition. In this task, a computer program has to learn with respect to a performance measure  $P$  how to correctly recognize faces and assign a correct label to each individual face presented to it. Each face will be represented as a feature vector which can contain for example the image pixel representation of it. An experience  $E$  in this scenario can be a set of faces represented as feature vectors where a correct label is assigned to each vector telling the machine learning model what the correct labeling is. This set of labeled examples is called a training set and it is used to train the machine learner to perform the task in a way that optimizes the value of the performance measure  $P$ . Each feature vector in the training set is called a training instance to which a correct label must be assigned in this use case.

Machine learning deals with four main types of learning paradigms based on the experience “ $E$ ” presented to the machine learner. These paradigms are supervised, unsupervised, semi-supervised and reinforcement learning. Each of these paradigms has its different learning strategy and learning objective. Supervised learning as the name implies works with a training set completely labeled with the correct output. This means that the process of learning is supervised by an external entity which provides the machine learner with the correct output for a set of instances. This set

has to be used for building a machine learning model that is supposed to provide the correct output for a set of unlabeled examples. The problem of face recognition we mentioned before is an example of a supervised learning problem. Supervised learning mainly deals with two types of problems, namely classification and regression problems [14]. On one hand, a problem is said to be a classification problem if it has a categorical output such as “true” and “false”. On the other hand, a regression problem is characterized by a numerical output where the machine learning model has to predict a numerical value. In this work, we mainly work with classification problems where human activities have to be recognized based on a series of sensor readings obtained from an indoor sensor deployment.

Unsupervised learning is characterized by a different learning goal as the experience  $E$  presented to the machine learner consists only of unlabeled instances. Therefore, a direct labeling of instances by the machine learner is not possible in this paradigm. Unsupervised learning is used when similarities or hidden structures have to be found in unlabeled datasets. An example problem of unsupervised learning can be the grouping of a set of articles into several sub-groups based on similarities between words in the articles. These sub-groups can be different text categories such as political, religious, movies and box office, sport and weather. Another paradigm is the semi-supervised learning which is characterized by having a dataset with both labeled and unlabeled examples.

As the fourth paradigm, reinforcement learning is characterized by the notion of a reward function. An intelligent agent has to perform a sequence of decisions so that it maximizes a cumulative reward function. There are no explicit pairs of the form (feature vector, label) as in supervised learning. The agent learns from the environment in a trial and error fashion as it receives a certain reward that reflects its success in performing the desired sequence of decisions.

In the following sections, we mainly focus on two important types of supervised learning, namely single-label and multi-label classification.

### 2.1.1 *Single-Label Classification*

As we have mentioned before, in a supervised learning setting each feature vector in the training set should be labeled with its correct label. In this setting, the supervised learning problem can be mathematically formulated as follows:

Given a training set consisting of pairs of instances  $(x_i, y_i)$  where  $i \in [1, N]$  and  $y_i = f(x_i)$ , the machine learning model has to learn a function  $\hat{f}$  that approximates the true real function  $f$  [92].

In classical machine learning problems,  $y_i$  corresponds to a single value that represents one and only one label. For example, each face in the problem of face recognition should belong to one person i.e. each feature vector has exactly one label. However, there are certain problems that impose the existence of more than one label per feature vector. Classification of movies into different genres represents such a problem where one movie can belong to more than one category. For example, most of the action movies belong to the categories crime, and thriller. In this section, we focus on single-label classification. We introduce the problem of multi-label classification in details in Section 2.1.2.



Depending on label type, single-label classification can be divided into three different categories:

- One-class (unary) classification: it focuses on the recognition of objects belonging to a specific class among a group of objects [75]. The training set in this setting contains only instances that are labeled with one specific class. The task of the classifier is to recognize the objects belonging to this class. An example application scenario of one-class classification is outlier detection. This can be achieved by training the classifier on instances only representing the normal behavior in the system. This can be very helpful in situations where data about abnormal behavior is very rare or not available as is the case in building a classifier which detects abnormal behavior in the operational state of a nuclear plant.
- Binary classification: it focuses on problems where the label takes one of only two possible values. Diagnostic systems are example application scenarios for binary classification where the presence or absence of a certain disease has to be detected. In contrast to unary classification, the training set contains instances belonging to both classes.
- Multi-class classification: it focuses on problems where the objects to be classified belong to more than two classes. The training set in this setting contains examples from each class and the task of the classifier is to predict the correct class for a certain instance presented to it. An example scenario is the classification of images into different classes where each class refers to the object shown in the image.

As mentioned before, each classification object has to be represented as a feature vector. Features represent discriminative object characteristics that distinguish objects belonging to different classes from each other. These features can be for example pixel values of an image or the words contained in a text document. Features can take the form of numerical or categorical variables [118]. Determining the relevant features for a classification problem represents a challenge that essentially affects the predictive performance of a classification model.

In order to create the training set necessary to build a classification model, each feature vector should be annotated with its correct label. The annotation process is mostly done manually by a domain expert who has to assign the correct label for each feature vector based on her/his own judgment. In order to evaluate the predictive performance of a classifier, the set of annotated instances has to be divided into a training and testing set. The training set is used to build the classification model whereas the testing set is used to evaluate its predictive performance by predicting the labels of testing instances and comparing them with ground truth labels. There are different methodologies used in dividing the data into training and testing sets. These methodologies, their advantages and disadvantages are the topic of Section 2.1.1.2.

#### 2.1.1.1 *Single-label Classification Algorithms*

Single-label classification models are broadly categorized into two main categories, namely generative and discriminative models [46]. Generative models focus on modeling the distribution from which the data has been generated. More precisely, it

models the joint probability distribution  $p(x, y)$  of the feature vectors  $x$  and the class  $y$ . Discriminative models, as their name implies, focus on directly mapping the objects into their respective classes without modeling the joint distribution of the data itself [76]. As generative models require the joint probability distribution  $p(x, y)$  to be explicitly modeled, it is more compelling to use discriminative models [111].

Naive Bayes classifier represents an example of generative models. It uses Bayes theorem [14] to compute the posterior probability  $p(y|x)$  of a feature vector  $x$  having the label  $y$ . It is based on the naive assumption that the features are independent of each other. For a given feature vector  $x_i$ , naive Bayes classifier computes the posterior probability of each possible outcome  $y_i$  to be the correct label of  $x_i$  [77]. As a result, the label with highest posterior probability is chosen as the outcome label of  $x_i$ .

Apart from generative models, decision trees build a tree-based discriminative model which is traversed based on the values of feature vectors to determine the correct labels [89]. The root node in a decision tree model represents the most discriminative feature. Less discriminative features follow in the lower levels of the tree. Different approaches can be used to determine the discriminative power of a specific feature in a decision tree. These approaches can either be heuristic-based or statistical. Given a feature vector  $x_i$ , the tree is traversed starting from the root node. In each node, a decision is made based on the value of the respective feature which it represents. This is continued until a leaf node is reached. A leaf node determines the correct label of a feature vector. One special type of decision trees extensively used in the course of this work is conditional inference trees (ctrees) [51]. Ctrees follow a statistical approach in selecting the important features for building the decision tree. A null hypothesis of independence is assumed between the response variable i.e. the label and each of the predictor variables i.e. the features. A statistical association test is performed to estimate the association between the response variable and each of the predictor variables. If an association exists, the null hypothesis of independence is rejected and the feature with the highest association is chosen as the next node of the tree.

Artificial neural networks (ANNs) are discriminative models that work very efficiently in modeling nonlinear and complex structures in the data. An artificial neural network is composed of an input layer, one or more hidden layers and an output layer [47]. Each layer is composed of a number of neurons which are connected to the ones in preceding layer as well as in succeeding layer. Each neuron aggregates the input signals delivered to it from preceding layer. The aggregated signal is forwarded to the succeeding layer when it is above a determined threshold. New types of neural networks have been proposed with new topological architectures such as convolutional and recurrent neural networks [44]. This is attributed to the emergence of deep learning where deep architecture with several hidden layers have been proposed and proved to achieve very good predictive performance results [44].

Support Vector Machines (SVMs) build kernel-based discriminative models [32]. They define a hyperplane that separates the instances of different classes in a way that maximizes the distance between borderline instances that represent the support vectors. SVMs achieve very good predictive performance results as they avoid overfitting even in extreme cases when the number of features i.e. the dimensionality of the dataset exceeds the number of available training instances. The problem of overfitting happens when the classifier tends to model the random noise in the data instead of modeling the abstract relation between input and output.

### 2.1.1.2 Evaluation of Predictive Performance

The ability of a certain classifier to generalize beyond its training set represents the essential criterion in evaluating its predictive performance. Different metrics are used to evaluate the predictive performance of a classifier. Each metric has its own meaning and importance based on the application scenario and requirements. All evaluation metrics are built and calculated based on the confusion matrix which visualizes the correctness of a classifier in a tabular view [80]. In Table 1, we show the confusion matrix for binary classification where the instances are classified either as positives or as negatives. The rows refer to the actual class while the columns refer to the predicted class. Each cell in the table refers to the value of the respective category. TP, for example, represents the number of positive instances that have been correctly classified as positives. The sum of all cells is always equal to  $N$  which is the total number of instances [103].

Table 1: Confusion matrix (TP = "True Positive" = Correctly classified as positive; FN = "False Negative" = Wrongly classified as negative; FP = "False Positive" = Wrongly classified as positive; TN = "True Negative" = Correctly classified as negative)

		Classified As		Sum
		<i>positive</i>	<i>negative</i>	
Actual Class	<i>positive</i>	TP	FN	TP + FN
	<i>negative</i>	FP	TN	FP + TN
Sum		TP + FP	FN + TN	$N$

Accuracy is defined as the percentage of positive and negative instances that are correctly classified.

$$\text{Accuracy} = \frac{TP + TN}{N} \quad (1)$$

Precision is defined as percentage of true positives among all objects that are classified as positives.

$$\text{Precision} = \frac{TP}{TP + FP} \quad (2)$$

Recall is defined as the percentage of positive instances that are correctly classified as positives.

$$\text{Recall} = \frac{TP}{TP + FN} \quad (3)$$

The importance of each measure is closely tied to the application scenario. In a diagnostic system which detects if a person suffers from cancer, it is of great importance to detect all positive instances that suffer from cancer. Therefore, a high recall value has to be achieved even at the expense of misclassifying some negative instances as positives. This leads to a low precision value.

A measure which represents the harmonic mean of precision and recall is the F-Measure.

$$\text{F-Measure} = \frac{2 * \text{Precision} * \text{Recall}}{\text{Precision} + \text{Recall}} \quad (4)$$

Kappa statistic measures the difference between two quantities, namely the agreement we observe between our classifier and the ground truth, and the agreement we expect to have in case of a random classifier given the confusion matrix we have

[112]. Accuracy as defined in Eq. 1 represents the observed agreement. Expected agreement is calculated as follows:

$$e = \frac{TP + FN}{N} * \frac{TP + FP}{N} + \frac{FP + TN}{N} * \frac{FN + TN}{N}. \quad (5)$$

After defining expected agreement  $e$ , we can define kappa statistic as follows:

$$k = \frac{\text{Accuracy} - e}{1 - e} : \quad (6)$$

### 2.1.2 Multi-Label Classification

Multi-label classification represents an emerging research topic in machine learning where classical single-label classification approaches are being extended to cover more realistic scenarios in which an object is allowed to have multiple labels [108]. The need for multi-label classification comes from the fact that most of the objects in the real world can be described in different ways and can belong to several categories based on our subjective interpretation. The image shown in Figure 2 can be described for example using three different labels, namely "Tree", "Winter" and "Blue Sky". Movies can also be seen as examples of objects that always belong to more than one category. For example, a movie can be both an action and sci-fi movie.



Figure 2: An example of multi-label classification: the labels "Tree", "Winter", and "Blue Sky" can be assigned to this image.

Table 2 shows a synthetic example of a multi-label dataset. As we can see from the table, an instance can take up to five different labels. The task of a multi-label classifier is to find out all possible labels that an instance takes. This can be mathematically formulated as follows [108]: For each  $n$ -dimensional feature vector  $x_i$ , we define a label set  $y_i \subseteq L$  where  $L = (\lambda_1, \lambda_2, \dots, \lambda_n)$  represents the set of all defined labels an instance may take.

Instance	Feature vectors	Label set
1	$f_1$	$\lambda_1, \lambda_3$
2	$f_2$	$\lambda_3, \lambda_4$
3	$f_3$	$\lambda_1, \lambda_2$
4	$f_4$	$\lambda_2, \lambda_4$
5	$f_4$	$\lambda_3$

Table 2: An example of a multi-label dataset

One important characteristic of multi-label datasets is label dependency. It refers to the fact that labels in multi-label datasets are correlated and associated with each other and not independent of each other. This means that the presence or absence of a certain label for a specific instance can imply with a certain probability the presence or absence of another label based on the type and degree of association these two labels have. Exploiting label dependency for enhancing the predictive performance of multi-label classification has become a very active research topic in the field of multi-label classification [33].

As single-label classifiers are only designed to work with problems in which an instance has one and only one label, new methodologies need to be developed for solving multi-label classification problems. Two main methodologies have been proposed by the research community, namely problem transformation and algorithm adaptation [108]. We present both of these methodologies in the next two sections with example approaches of each one of them.

#### 2.1.2.1 Problem Transformation

Problem Transformation represents one of the first methodologies that have been proposed to handle multi-label classification problems. It is based on the conversion of multi-label datasets into single-label datasets on which single-label classifiers can be used.

Binary Relevance (BR) is one of the first and basic problem transformation approaches. As its name implies, BR builds one single-label dataset for each label  $y_i \in L$ . Therefore, for an  $n$ -label dataset, BR creates  $n$  single-label datasets. A single-label classifier can then be applied to each of the  $n$  datasets and the results of all classifiers are merged to get the final classification result. BR has the main disadvantage of ignoring label dependencies. As a separate dataset is built for each label, BR ignores and loses all the information related to label associations. This affects its predictive performance in comparison to other problem transformation approaches that extract and utilize label dependency to enhance their predictive performance.

Table 3 shows the results of applying BR on the dataset shown in Table 2. As we can see from the table, a new dataset is built for each one of the labels in which only its presence or absence is encoded.

Ranking by Pairwise Comparison (RPC) represents another problem transformation approach which takes label dependency into account by creating a new dataset for each pair of labels [52]. This results in the creation of  $q(q-1)/2$  datasets for a set of  $q$  labels. Each of these datasets contains instances for which only one of its two

Instance	Label	Instance	Label	Instance	Label	Instance	Label
1	$\lambda_1$	1	$\neg\lambda_2$	1	$\lambda_3$	1	$\neg\lambda_4$
2	$\neg\lambda_1$	2	$\neg\lambda_2$	2	$\lambda_3$	2	$\lambda_4$
3	$\lambda_1$	3	$\lambda_2$	3	$\neg\lambda_3$	3	$\neg\lambda_4$
4	$\neg\lambda_1$	4	$\lambda_2$	4	$\neg\lambda_3$	4	$\lambda_4$
5	$\neg\lambda_1$	5	$\neg\lambda_2$	5	$\lambda_3$	5	$\neg\lambda_4$

Table 3: The result of applying BR on the multi-label dataset shown in Table 2

labels occurs but not both of them. This results in a binary classification problem for each of the new datasets and therefore transforms the multi-label classification problem into a set of binary classification problems which can be solved using the classical techniques of single-label classification. Table 4 shows the result of applying RPC on the dataset shown in Table 2.

Instance	Label	Example	Label	Example	Label
1	$\lambda_{1,-2}$	2	$\lambda_{-1,3}$	1	$\lambda_{1,-4}$
4	$\lambda_{-1,2}$	3	$\lambda_{1,-3}$	2	$\lambda_{-1,4}$
		5	$\lambda_{-1,3}$	3	$\lambda_{1,-4}$
				4	$\lambda_{-1,4}$

Example	Label	Example	Label	Example	Label
1	$\lambda_{-2,3}$	2	$\lambda_{-2,4}$	1	$\lambda_{3,-4}$
2	$\lambda_{-2,3}$	3	$\lambda_{2,-4}$	4	$\lambda_{-3,4}$
3	$\lambda_{2,-3}$			5	$\lambda_{3,-4}$
4	$\lambda_{2,-3}$				
5	$\lambda_{-2,3}$				

Table 4: The result of applying RPC on the multi-label dataset shown in Table 2

Another problem transformation approach which takes label dependency into consideration is Label Powerset (LP) [107]. It is based on the idea of creating a new set of labels in which each of existing basic label combinations is defined as a new label. The result of applying LP on our example dataset is shown in Table 5. LP has two main disadvantages. On one hand, it results in a dataset which only contains examples for the label combinations available in the training set. Therefore, it cannot predict new label combinations for which no new labels have been created. On the other hand, creating a new label for each unique combination of labels leads to an exponential increase in the number of labels where each label is only associated with a few number of instances. These two problems affect the predictive performance of any single-label classification model that has to be applied to the resulting dataset.

RAKEL (Random k-Labelsets) [109] is an extension of LP. It avoids the problems of LP while taking label dependency into account by creating an ensemble of LP classifiers where each classifier is based on a random subset of the label set. By

Example	Label
1	$\lambda_{1,3}$
2	$\lambda_{3,4}$
3	$\lambda_{1,2}$
4	$\lambda_{2,4}$
5	$\lambda_3$

Table 5: The result of applying LP on the dataset shown in Table 2

reducing the number of labels on which each LP classifier is built, RAKEL tries to avoid the main two disadvantages of LP.

Another important problem transformation approach is classifier chains [85]. It exploits label dependency by using the predictions obtained for a certain label as a feature in predicting another label. This leads to the creation of a classifier chain where each classifier predicts only one label as it is the case using BR. The main difference to BR is the inclusion of label predictions into the feature vectors.

#### 2.1.2.2 Algorithm Adaptation

Algorithm adaptation represents the second and more complicated methodology for dealing with multi-label problems. It is based on adapting single-label classification models so that they directly work on multi-label data. Several algorithms have been adapted to work directly on multi-label data. An example algorithm is k-nearest neighbors which has been adapted into multi-label k-nearest neighbors (MLknn) [125], Binary Relevance k-nearest neighbors (BRknn) [104] and DMLkNN[122]. Researchers have also adapted other algorithms such as SVM [38], C4.5 [24], AdaBoost.MH [96], AdaBoost.MR [96] and the perceptron algorithm [30].

#### 2.1.2.3 Dataset Metrics

In order to quantify the multi-label nature of datasets, researchers designed a set of metrics which check different multi-label aspects [108][84]. In the following, we present a set of important metrics that have to be calculated for each multi-label dataset in order to understand its multi-label characteristics and based on that to decide about the best classification methodology to use. We define  $D$  as a multi-label dataset,  $n$  as the number of instances,  $L$  as the set of labels,  $q$  as the number of labels and  $Y_i$  as the correct label set of an instance  $X_i$ .

Label cardinality (LCard) measures the average number of labels per instance in a dataset.

$$\text{LCard}(D) = \frac{1}{n} \sum_{i=1}^n |Y_i| \quad (7)$$

Label density (LDens) quantifies label cardinality with respect to the number of labels in a dataset.

$$\text{LDens}(D) = \frac{1}{n} \sum_{i=1}^n \frac{|Y_i|}{q} = \frac{\text{LCard}(D)}{q} \quad (8)$$

Distinct labelsets (DL) calculates the number of all label combinations existing in a dataset.

$$DL(D) = |\mathcal{Y} \subseteq L | \exists (x, Y) \in D| \quad (9)$$

The proportion of distinct labelsets (PDL) calculates the ratio of distinct label sets with respect to the total number of instances in a dataset.

$$PDL(D) = \frac{DL(D)}{n} \quad (10)$$

The proportion of unique label combinations (PUniQ) calculates the ratio of unique label sets i.e. label sets that have only one instance with respect to the total number of instances in a dataset. A high value for PUniQ shows a dataset with irregular labeling pattern as it indicates the existence of many label sets with only one training instance which makes the classification process more complicated.

$$PUniQ(D) = \frac{|\mathcal{Y} \subseteq L | \exists! x : (x, Y) \in D|}{n} \quad (11)$$

The PMax metric focuses on the most frequent label set in a dataset. It calculates the ratio of instances associated with this label set with respect to the total number of instances in a dataset. A high value for PMax indicates a dataset that is skewed towards its dominant label set.

$$PMax(D) = \max_{Y \subseteq L} \frac{\text{count}(Y, D)}{n} \quad (12)$$

#### 2.1.2.4 Evaluation Metrics

This section presents the evaluation metrics used in evaluating the predictive performance of multi-label classification algorithms. As more than one label per instance has to be predicted, different evaluation metrics than the ones presented in Section 2.1.1.2 have to be used. A classification model may predict part of a label set correctly and not all labels. Therefore, metrics which are able to quantify the partial predictive performance with regard to each label as well as to each instance have to be used. There are two different types of evaluation metrics for multi-label classification, namely example-based and label-based metrics.

Example-based metrics as their name implies are calculated with respect to all instances in the dataset. We define  $D$  as a multi-label dataset,  $n$  as the number of instances,  $L$  as the set of labels,  $q$  as the number of labels and  $Y_i, Z_i$  as the correct and predicted label sets of an instance  $X_i$  respectively.

Subset accuracy represents an example-based metric that calculates the number of instances for which all labels have been correctly predicted. It considers only instances with an exact match between predicted and correct label set. Therefore, it is considered to be a strict metric as it does not take into account any partially correct prediction.

$$0/1\text{subsetaccuracy} = \frac{1}{n} \sum_{i=1}^n \mathbb{I}[Z_i = Y_i] \quad (13)$$

To take partially correct predictions into consideration, another example-based metric has been proposed, namely Hamming Loss. It calculates the number of mis-



matches i.e. the distance between predicted and correct label sets. A low value of hamming loss indicates a good predictive performance.

$$\text{HammingLoss} = \frac{1}{n} \sum_{i=1}^n \frac{1}{q} |Z_i \Delta Y_i| \quad (14)$$

As in single-label classification, the classical metrics of precision, recall, accuracy and f-measure play an important role in quantifying the predictive performance of multi-label classification models. As example-based metrics, they are defined as follows [43]:

Recall calculates the fraction of positive labels for each instance that have been correctly predicted averaged over all instances in the dataset.

$$\text{Recall} = \frac{1}{n} \sum_{i=1}^n \frac{|Z_i \cap Y_i|}{|Y_i|} \quad (15)$$

Precision calculates the fraction of positive labels that have been predicted for each instance and are correctly positives averaged over all instances in the dataset.

$$\text{Precision} = \frac{1}{n} \sum_{i=1}^n \frac{|Z_i \cap Y_i|}{|Z_i|} \quad (16)$$

Accuracy calculates the fraction of positive and negative labels that have been correctly predicted for each instance averaged over all instances in the dataset.

$$\text{Accuracy} = \frac{1}{n} \sum_{i=1}^n \frac{|Z_i \cap Y_i|}{|Z_i \cup Y_i|} \quad (17)$$

F-measure calculates the harmonic mean of precision and recall values.

$$\text{F-Measure} = \frac{1}{n} \sum_{i=1}^n \frac{2|Z_i \cap Y_i|}{|Z_i| + |Y_i|} \quad (18)$$

In contrast to example-based metrics, label-based metrics are calculated with respect to each label and then averaged over all labels. Using this methodology, we can compute classical single-label evaluation metrics with respect to each label individually as a first step. Thereafter, we perform the averaging process over all labels as a second step. The averaging process can be performed in two different ways, namely macro averaging and micro averaging [120]. Define  $M(tp, tn, fp, fn)$  as a label-based evaluation metric and  $tp_a$ ,  $fp_a$ ,  $tn_a$  and  $fn_a$  as true positives, false positives, true negatives and false negatives of label  $a$ . The macro and micro averaged  $M$  can be calculated as follows:

$$M_{\text{macro}} = \frac{1}{q} \sum_{a=1}^q M(tp_a, fp_a, tn_a, fn_a). \quad (19)$$

$$M_{\text{micro}} = M\left(\sum_{a=1}^q tp_a, \sum_{a=1}^q fp_a, \sum_{a=1}^q tn_a, \sum_{a=1}^q fn_a\right). \quad (20)$$

As we can see from Eq. 19, we compute the macro-averaged evaluation metric  $M$  by summing up the individual  $M$  metrics of each label and dividing them by the total number of labels. For computing the micro-averaged  $M$ , we first sum up the values of  $tp$ ,  $fp$ ,  $tn$  and  $fn$  for all labels and then calculate the evaluation metric based on these aggregated values as shown in Eq. 20.



---

## RELATED WORK

---

RECOGNITION and prediction of human context has always been an important research topic in the area of smart environments. In this chapter, we explore several research projects with the main goal of extracting, detecting and predicting user's context in smart environments based on a wide variety of sensing modalities. This work focuses on two main aspects of users' context, namely the recognition of users' daily activities and the prediction of long-term and short-term power consumption of individual buildings. Therefore, we restrict the research projects presented in this chapter to these two areas. We begin in Section 3.1 by presenting a group of research projects that handle the problem of activity recognition in single-user as well as multi-user smart environments. For each project, we discuss the following list:

- Application scenario and monitored environment.
- Used sensing modalities.
- Monitored activities.
- Predictive performance of resulting activity recognition models.

Thereafter, we present in Section 3.2 a group of state-of-the-art research projects in the field of power consumption forecasting. We introduce for each project the entities for which power consumption is predicted, type of used forecasting models and their achieved predictive performance.

### 3.1 ACTIVITY RECOGNITION AND BEHAVIORAL MODELING

As mentioned before, recognizing humans' daily activities has its potential benefits in several application scenarios such as elderly care in ambient assisted living, power saving and comfort increasing in smart homes, to name a few. As a result of such benefits, researchers have designed several systems that are able to recognize humans' daily activities in different settings and environments based on their application scenarios. To recognize users' activities, their environments have to be monitored with sensing modalities that are able to gather enough information about the environments as well as about the users. Wireless sensor nodes that take battery lifetime into consideration paved the road for building efficient monitoring systems that collect the required environmental parameters for building an accurate activity recognition model while maintaining a long battery lifetime which guarantees seamless system operations [62][86].

Single-user environments where one user is assumed to live alone and to perform only one activity at a given time represent the most used setting in activity recognition research [121]. The researchers impose such a restriction to avoid the complexity

generated by allowing users to perform parallel and overlapping activities. However, such an assumption is contradicted with real-world scenarios where users tend to live together and to perform more than one activity at a given time. Therefore, we start by presenting projects that impose this restriction and then move to more advanced projects where more than one user are assumed to live in a common place and to perform parallel activities.

### 3.1.1 *Single-User Environments*

As we mentioned before, most of the research conducted in the field of activity recognition impose the single-user single-activity restriction on the environment to be monitored. This means that the environment is supposed to have only one resident performing a single activity at a given time. Several research projects handled the problem of activity recognition from this perspective. Researchers have deployed a wide variety of sensing modalities to achieve their goal of building accurate activity recognition models. Examples are Radio-frequency identification (RFID) sensors, cameras, microphones, wearable sensors, motion detectors and environmental sensors, to name a few.

Based on the deployed sensor nodes, activity recognition can be classified into three main categories, namely audio/video-based models [15][78][41], infrastructure-based models[110][105] and wearable-based models [102][23][81][68][94][93]. Video and audio-based activity recognition models deploy a group of cameras and microphones in the monitored environments. To collect the required data for building and evaluating the model of activity recognition, the researchers save the collected data and then perform a manual annotation process in which each data instance is annotated with its corresponding activity. Audio/video-based models achieve very good predictive performance due to the rich amount of information provided by cameras and microphones. However, such models are confronted by very low acceptance of users due to their privacy concerns. Moreover, with cameras and microphones deployed all over the environment, users feel themselves strictly monitored and start to follow a behavior which is different from their normal behavior.

Wearable-based models depend on sensors that can be mounted on the bodies of monitored persons to collect information about their current context. Examples of such sensors are acceleration, orientation and pulse sensors, to name a few. As they depend on body-mounted sensors, most of the wearable-based models focus on activities that are related to body movements. Examples of such activities are walking, running, sitting and biking. The main problem of wearable-based models lies in the fact that several electronic devices have to be mounted on the body of monitored person. People do not prefer electronic devices to be mounted on their bodies except in experimental settings and therefore these models are not suitable for real-world deployments. Moreover, activities of daily living such as watching TV, reading, preparing a meal cooking cannot be recognized purely based on the information collected from wearable sensors. With the advancements in the production of smartphones and smartwatches, these devices started to contain sensors such as accelerometers and gyroscopes that can be used to collect information about the people holding them. As a result of these advancements, several researchers developed activity recognition models that can recognize human activities based on sensors

readings collected from their smartphones [63] [11] and smartwatches [70][115]. The main problem of smartphone-based models lies in the fact that people tend to leave their mobile phones away from them for most of the time [79]. Moreover, they are restricted to the same set of movement-related activities. Smartwatches are worn rather the whole day; however, smartwatch-based models still impose the restriction of movement-related activities.

Infrastructure-based activity recognition models utilize sensors that can be deployed in the environment without being perceived by the residents as a threat to their privacy as it is the case with cameras and microphones. Examples of such sensors are motion detectors, temperature, brightness and humidity sensors as well as simple state-change sensors. Kasteren et al. [110] designed a system which can recognize human activities based on the information provided by a set of simple state-change sensors. In their experimental setup, the authors have positioned state-change sensors on a wide variety of objects in the monitored environments. A group of 14 sensors has been attached to the cupboards, doors, fridge and toilet flush. By deploying these sensors in a single-person apartment, the authors collected a dataset of 28 days with an overall number of 245 activities and 2120 sensor readings. For annotating the collected data with ground truth labels, the authors designed a speech recognition system so that the user can provide her/his currently performed activity with a Bluetooth microphone. Seven different activities have been monitored with the proposed system, namely making dinner, making breakfast, preparing a beverage, toileting, showering, sleeping, and leaving home. When no activity is provided by the user as a feedback, the stream of sensor readings is annotated with the label "Idle". To build their predictive models, the authors use conditional random fields (CRFs) and hidden Markov models (HMMs) as they are able to model the temporal patterns contained in the data. Moreover, they designed three different types of features for constructing the required dataset. The first type of features is directly built using the raw sensor readings. The second type takes into consideration state transitions of sensors where the feature representing each sensor takes the value of one when this sensor changes its state during the respective time slot. The third type of features is built by giving the value of one for the feature which represents the sensor lastly changed its value. This feature continues to take the value of one until another sensor changes its value, and then it is switched to zero. The predictive performance of both HMM and CRF models is evaluated with respect to the three different types of features. The best performance has been achieved with a feature set consisting of both second and third types of features. With this set of features, HMM and CRF models reached a class accuracy of 79.4% and 70.8% respectively. Another experimental aspect of this work is the size of the training set required to achieve the best predictive performance. The authors showed that a training set of 12 days is enough for building an accurate activity recognition model.

Another infrastructure-based activity recognition platform is proposed by Tapia et al. in [105]. The authors designed a system based on state-change sensors to recognize activities in single-user environments. To collect the required training and testing datasets, the authors deployed their system in two single-person apartments where 77 and 88 state-change sensors were installed in the first and second apartment respectively. The sensors have been mounted to different objects in the monitored apartments such as doors, windows, electrical appliances, lamps and light switches, to name a few. The residents of both apartments were ordinary users who

have no technical experience with the deployed system. The system monitors a set of 35 activities where the users are supposed to provide the feedback about their currently performed activity using a software installed on a personal digital assistant (PDA) given to each one of them. After collecting required data, the authors noticed that several activities have very few numbers of instances which are not enough for training an accurate prediction model. As a result, they decided to keep only eight activities while excluding all other activities. In contrast to [110], the authors allow for parallel activities performed by the same user to exist. This is done by building a set of binary naive Bayesian classifiers where each classifier is responsible for recognizing only one activity. By evaluating the predictive performance of all naive Bayesian classifiers, the authors reached an average class accuracy of 55% with regard to all eight monitored activities.

As the first step in our work, we deal with the problem of activity recognition from a single-user perspective where we do not allow for any parallel or overlapping activities. We design a system which is able to recognize users' activities in single-user environments only based on the measurements of power consumed by each available appliance in the monitored environment. In contrast to the aforementioned projects, our system does not require the installation of any sensing modality apart from the appliance-level power sensors which have the advantage of being perceived by the users as normal electrical parts that do not threaten their privacy. As a second main part of this work, we improve our activity recognition platform so that it works in multi-user environments where intra-user and inter-user parallel and overlapping activities are allowed to happen.

### 3.1.2 *Multi-User Environments*

Recognizing activities in multi-user environments represents a difficult challenge for researchers in the field of smart environments [13]. This is due to two main reasons. On one hand, parallel and overlapping activities performed by one or more users have to be recognized and differentiated from each other. On the other hand, users performing each of the running activities have to be identified so that each activity is assigned to its respective user.

Crandall et al. [31] proposed an approach for user identification in multi-user environments. They designed a mapping procedure in which triggered sensor events are mapped to the users who have generated them. By identifying the users triggering certain sensor events, we can accurately assign each running activity to its respective user. In this work, the authors use two types of sensors. On one hand, they deploy motion sensors to monitor the presence of users in the monitored environment. On the other hand, they record the interactions between users and light switches. To achieve an accurate mapping procedure, the authors differentiate between users' behaviors by taking temporal features that characterize each triggered sensor event into consideration. As users tend to perform the same activities in different time periods, having temporal information about each triggered sensor event simplifies the process of user identification. Therefore, the following features have been collected regarding each triggered sensor event:

- Hour.
- Day.

- Day segment i.e. morning, afternoon and evening.
- Day type i.e. business day, weekend, or holiday.

As a result, each generated sensor event has a feature vector consisting of its temporal features combined with its serial number and the generated event message. Each of these feature vectors is labeled with the user ID. In the training period, users have been asked to provide their presence information by pushing a special button whenever they enter or leave the monitored environment. The authors utilized a naive Bayes classifier to build the classification model which maps sensor events to users. Evaluation of models that includes sensors with temporal features show good results in differentiating the users. The dataset has been randomly divided into a 90% training set and 10% testing set. Through an extensive evaluation study, the authors were able to reach an accuracy of 95% for the mapping process. This work has the main disadvantage of not considering parallel activities performed by more than one user into consideration. This is due to the fact that it is not possible for the system to differentiate between users with similar temporal behavior.

To build an activity recognition model that can recognize parallel activities of multiple users, we need to allow an instance to have more than one activity as its label. The main problem which faces researchers when building a model for activity recognition in multi-user environments is the rarity of powerful multi-label classification algorithms that can handle instances with more than one label as it is the case in multi-label classification problems. Most of the available multi-label algorithms are problem transformation approaches that convert multi-label classification problems into single-label problems as we showed in Section 2.1.2. As a result, most of the researchers follow an approach in which they combine all concurrent activities into an artificial joint label that represents all of them [36][22].

To recognize activities in clinical context, Doryab et al. [36] designed an approach that creates an artificial joint label for any combination of concurrent activities. The goal of this work is to develop a system which is able to recognize concurrent surgical activities performed in operating rooms. The researchers deployed a group of wearable and embedded sensors in operating rooms so that they get information about clinicians' locations, positions of objects and the way in which they are used by the clinicians in the operating room. Ten minimally invasive surgeries (MISs) have been monitored to collect the required dataset for building and testing an activity recognition model. Six clinicians took part in each of the surgeries. After transforming all sensor data into the binary form, the researchers used it to build a model that recognizes base single activities. Around 70% of instances in the dataset have more than two labels indicating the occurring of more than two base activities at a given time. The researchers used Apriori algorithm [1] which extracts all frequently occurring patterns from a dataset to identify base activities that frequently occur together. An artificial joint label is assigned to each frequent set of concurrent activities so that the collected data is converted into a single-label dataset. The researchers assume their dataset to have a time series nature in which a temporal relation exists between subsequent sets of actions. These relations can be incorporated into the activity recognition model so that they enhance their predictive performance. To capture such relations, they utilize a virtual evidence boosting algorithm. To build the joint activity recognition models, the researchers chose to use CRFs in which the identified temporal relations are incorporated. They built three different models,

namely model *A*, *B* and *C*. Model *A* is characterized by creating an artificial joint label for all available combinations of base activities. However, due to the big number of newly created joint labels, this model has shown poor predictive performance with an average accuracy of 43% accompanied with computational inefficiency. To reduce the number of newly created labels, the researchers separated activities of anesthesia and operating teams. A separate CRF chain is built for the activities of each team with the same set of observations. With accuracy increasing to reach 63%, model *B* achieved relatively good predictive performance in comparison to model *A*. However, this model ignores the dependencies between actions of both teams. Therefore, a model *C* is built which models the concurrent activities between both teams. This model achieved an accuracy of 62% which is almost similar to the accuracy of model *B*.

To solve the problem of concurrent activity recognition in multi-resident environments, Chen et al. [22] proposed a model which works in two different stages. In the first stage, the researchers transform all multi-label instances into single-label instances by creating combined label states based on the available data associations. During this stage, the model learns combined label states. In the second stage, the model recognizes multiple activities of home residents. The researchers used HMMs and CRFs to model the data where CASAS dataset of Washington state university [26] is used to build and evaluate the activity recognition model. They utilized metrics of multi-label classification to evaluate the predictive performance of the proposed models. The evaluation results showed that HMM and CRF models achieved a good average accuracy of about 75%.

Another approach for recognizing concurrent activities in multi-user environments was proposed by Wu et al. in [124]. In this work, the researchers built two models for activity recognition using Factorial CRFs (FCRFs) and Linear Chain CRFs (LCRFs). They used *House\_n* dataset collected by Massachusetts Institute of Technology (MIT) for building and evaluating their models. This dataset contains readings of switch sensors, light sensors and current sensors deployed in a house. The collection duration was four hours, namely from 9 AM to 1 PM. The annotation process was done manually by the person who performed the activities where he provided information about his activity and location. As a first step, the researchers clustered the total number of activities performed by the user, namely 89 activities into six clusters that are allowed to overlap with each other. Thereafter, they designed FCRF model so that it takes temporal relations between activities into consideration. Evaluation results have shown poor predictive performance for both FCRF and LCRF models with f-measure values of 49.8% and 41.7 % respectively. The main reason behind these results is the very small size of the dataset used in this work.

### 3.2 FORECASTING OF POWER CONSUMPTION

Forecasting future power consumption of individual buildings has always represented an important research topic for science and industry. The importance of this topic lies in the benefits it provides especially for electric utilities. On one hand, forecasting of short-term future power consumption of individual customers provides electric utilities with deep insights into expected consumption behavior so that they can efficiently schedule and control their resources to avoid any surpluses or short-



falls in power provisioning [2]. On the other hand, forecasting of long-term future power consumption allows the electric utilities to accurately plan and expand their infrastructure so that they meet the predicted increase in power consumption efficiently and without wasting any resources. Researchers have utilized different types of models for achieving an accurate forecasting of future power consumption. Based on the utilized modeling approach, we can classify these models into different five categories, namely averaging models [27][29][28], regression models [10][40][55][50], time series models [99][17][69], artificial intelligence models [119][59][72][91] as well as hybrid models [74][100].

Averaging models represent one of the simplest models existing for forecasting of power consumption. Their simplicity lies in the fact that they make their prediction by taking into consideration only the average consumption values of similar points in time. As a result, these models only require the historical data of previous power consumption as input. A comprehensive evaluation study of several averaging models for short-term forecasting of power consumption is presented by Aman et al. in [9]. In this work, the authors have compared three averaging models, namely Time of the Week (ToW) averaging model, Southern California Edison (CASCE) model [27][29] and New York ISO (NY ISO) model [28]. Power consumption is predicted in short 15-minute intervals. ToW model predicts the power consumption of a certain 15-minute interval  $x$  by taking the average power consumed during the same interval  $x$  with regard to all weeks in the available history. ToW model has the main advantage of being able to capture the difference in power consumption between different daily periods based on the activities and schedules of residents. Moreover, it can easily capture the different power consumption behaviors in weekend and business days. However, it has limited capability in capturing seasonal variations in power consumption. Time of Year (ToY) models can capture such seasonal effects. However, ToY models are usually outperformed by ToW models.

Regression models represent another approach with which the researchers handle the problem of forecasting power consumption. With regression models, the power consumption is treated as a response variable of a set of predictor variables such as weather conditions, demographics of residents and buildings' characteristics, to name a few. Researchers use two types of regression models to model the relation between power consumption and its predictor variables. On one hand, they use simple linear regression models where they assume the existence of linear relations between response and predictor variables. On the other hand, they use more advanced models where non-linear relations are assumed to exist between response and predictor variables. Regression trees represent an example of such advanced models. A multiple linear regression model for forecasting hourly electric load has been proposed by Hong et al. in [50]. In this work, the authors modeled electric load as a linear response of several predictor variables, namely temperature, month of the year, day of the week and hour of the day. As these variables interact with each other, the authors modeled their cross-effects by creating additional variables that represent their multiplications. To model and forecast the hourly electric load, the authors built nine different models based on different combinations of the predictor variables and their modeled cross-effects. To evaluate the predictive performance of these models, they used a four-year dataset of hourly electric load and temperature values provided by a US utility. This dataset covers the period from 2005 to 2008. As a training set, they used the years 2005-2007 while 2008 has been used as a testing set. They selected

their optimal model through an iterative trial-and-error process where they used goodness-of-fit statistics, namely adjusted R-square, mean absolute percentage error (MAPE) and standard deviation of the absolute percentage error (STDAPE) to evaluate the model performance on the training dataset. Moreover, they evaluated the predictive performance on the testing set using accuracy statistics, namely MAPE, STDAPE, mean absolute error (MAE) and standard deviation of the absolute error (STDAE). Their optimal model reached a MAPE value of 4.558% on the testing set. As most of the relations between power consumption and its predictor variables are of complex and non-linear nature, it is recommended to utilize non-linear approaches that are able to model such complex relations. Aman et al. in [10] introduced such an approach based on regression trees to model and forecast the daily as well as fine-granular 15-min power consumption on campus- and building-level for a university campus microgrid. As predictor variables, the authors utilized indirect indicators of power consumption that are related to academic environments, namely day of the week, semester as well as holiday. Moreover, they utilized buildings' characteristics as well as weather information, namely humidity and average and maximum temperature as direct indicators of power consumption. To build and evaluate their forecasting models, they used a three-year dataset of smart meters' readings collected from 170 buildings at the campus of University of Southern California (USC) with 15-minute granularity. The dataset covers the period from 2008 to 2010. To evaluate the predictive performance of their models, the authors used the coefficient of variation of the root mean squared error (CV-RMSE). As a first step, they built a campus-level regression tree model which forecasts the daily power consumption of the whole campus based on holiday, day of week, semester and maximum temperature as a feature set. The years 2008 and 2009 have been used as a training set while 2010 has been used as a testing set. This model achieved a CV-RMSE value of 7.45%. To quantify the relative importance of single features, the authors built a set of models based on different combinations of them. As a result, they considered "day of the week" to be the most important feature. As a second step, they built another regression tree model to forecast the fine-granular 15-min power consumption of the whole campus based on the same previous set of features combined with campus humidity. Using the same previous evaluation setting, this model achieved a CV-RMSE value of 13.70% with temperature as the most important feature. As a final step, they built a regression tree to forecast the power consumption of individual buildings. Building characteristics combined with the previous feature set has been used to build the model. A dataset belonging to 23 buildings has been used to train and evaluate the model. This model achieved CV-RMSE values of 11.77%, 12.09% and 19.32% based on building's type.

Artificial intelligence techniques have also attracted the attention of researchers as potential methods for forecasting power consumption. Several research projects applied such techniques for predicting short-term power consumption [72]. The power consumed by a certain entity can have linear and non-linear complex relationships with a wide spectrum of external variable such as weather conditions, demographic features of residents as well as behavioral patterns, to name a few. As a result of this complexity, it is required to utilize a modeling methodology that is capable of accurately representing such a combination of linear and non-linear relations. Due to their powerful capability in representing complicated non-linear functions, artificial neural networks (ANNs) represent one of the widely used AI techniques in the field

of power consumption forecasting [49]. An example ANN-based model for forecasting short-term power consumption of campus buildings is introduced by Wan et al. in [114]. In this work, the authors developed an ANN which predicts the power consumption of campus buildings based on two different types of predictor variables. On one hand, they used variables related to weather conditions such as outside temperature and humidity, air pressure, wind speed, dew point temperature and rainfall rate. In addition to these predictors, they utilized the current power consumption as an input to the ANN. The network predicts the next 15-minute power consumption as its output. To determine the optimal network architecture, the authors designed two ANNs with different network configurations. The first ANN consists of one hidden layer in which nine neurons exist. The second one consists of two hidden layers with the first one containing ten neurons and the second one containing 5 neurons. The dataset used in this work represents a one-month consumption data of an administration building at the California Polytechnic State University. It consists of 2878 sample measurements taken at 15-minute sampling granularity. The dataset has been divided so that 2000 instances have been used as a training set and the rest as a testing set. Evaluating the predictive performance of both networks has shown that the first one achieves a root mean square (RMS) value of 5.66% while the second one achieves an RMS value of 10.76%. Both RMS values are computed for the testing error. These results show that the first ANN outperforms the second one even though it contains less number of hidden layers. A comparative evaluation study for identifying the main factors that influence the performance of ANNs in predicting short-term power consumption is provided by Rui et al. in [91].

Another well-known approach in the field of power consumption forecasting is time series based modeling. The main advantage of time series models lies in the fact that they require only historical consumption data for building a model which can predict the future consumption. They do not depend on any external factors such as weather conditions or demographic and behavioral data of residents. A wide variety of time series models has been used by researchers to forecast long-term as well as short-term power consumption. Among these models are Moving Average (MA) models, Auto-Regressive Moving Average (ARMA) models and Auto-Regressive Integrated Moving Average (ARIMA) models [17]. Pattern sequence matching is another approach used in analyzing time series. It is based on the idea that every time series contains temporal patterns that repeat themselves over time. By identifying and extracting such patterns, we can forecast the future behavior of a certain series. Martinez Alvarez et al. introduced an approach for forecasting day-ahead power consumption based on pattern sequence matching in [69]. In this work, the authors developed the algorithm of Pattern Sequence-based Forecasting (PSF). As a first step, PSF algorithm creates a 24-hour power consumption vector for each day in the historical data. As a result, each day is represented as a point in 24-dimensional Euclidean space. As a second step, PSF utilizes k-means clustering algorithm to cluster all these points into a set of homogeneous clusters where each cluster represents a group of days that follow a similar 24-hour power consumption pattern. PSF utilizes clustering validity indexes in order to determine the optimal number of clusters. A label is assigned to each cluster so that all days belonging to it get labeled with this label. In order to predict the day-ahead power consumption, PSF extracts the sequence of labels representing the  $w$  days preceding the day to be predicted. The number of preceding days  $w$  is referred to as the window size. Thereafter, it searches for this

pattern in the historical data. In case no match is found, it reduces window size to be  $w - 1$  and restarts the search process. PSF calculates the power consumption of the day to be predicted as the average power consumption of each day succeeding each found pattern. The authors applied their algorithm on three different datasets collected by three electricity market operators, namely in Spain, New York and Australia for the year 2006. Each dataset consists of the total hourly power consumed in the respective market for the year 2006. As a result of an exhaustive evaluation study, PSF achieved a predictive performance of 4.99%, 2.87% and 3.43% for New York, Spain and Australia datasets respectively in terms of Mean Error Relative (MER). Apart from forecasting day-ahead power consumption, the authors utilized PSF for predicting electricity prices in the same three markets with good results.

To exploit the advantages of different modeling methodologies in forecasting future power consumption, several researchers developed hybrid forecasting models in which multiple modeling techniques are combined to deal with complex different relations between the power consumption and its driving factors. Mori et al. [74] developed a hybrid model in which they combine regression trees with relevance vector machines (RVMs) [106] to predict the day-ahead maximum load. As an initial step, regression trees are used to divide the training dataset into a set of homogeneous groups. Thereafter, an RVM is built for each terminal node to forecast the maximum day-ahead load of the group this terminal represents. RVMs has the main advantage of providing the distribution of its predicted value and therefore its upper and lower limits. To build their forecasting model, the authors utilized the following set of features:

- Average and minimum humidity of the next day
- Average, minimum and maximum temperature of the next day
- Daylight and discomfort index of the next day
- Maximum load demand of the current day.

As weather features of the next day are not available, they need to be estimated. Therefore, the authors built a combined CART-RVM forecasting model for each weather feature to predict its value for the next day based on its current value. After estimating the values of these features, they are used in combination with the maximum load of current day for forecasting next day's maximum load. For training and evaluating their model, the authors used an 11-year dataset which is provided by a Japanese utility and covers the period from 1991 to 2001. Ten years from 1991 to 2000 were used for training purposes while the remaining year was used for testing purposes. By evaluating the proposed model, it has achieved a good predictive performance with a maximum error of 6.04%.

### 3.3 SUMMARY AND DISCUSSION

In this chapter, we presented a detailed overview of the state-of-the-art research projects in the fields of activity recognition and behavioral modeling as well as power consumption forecasting. We started in Section 3.1.1 by introducing a group of research projects that worked on activity recognition in single-user environments.

Thereafter, we presented more advanced projects that handled the problem of activity recognition in multi-user environments in Section 3.1.2. For each project, we presented in details its sensing modalities, deployment process, monitored environments and activities, modeling techniques as well as its achieved predictive performance. In Section 3.2, we introduced a categorization of research projects in the field of power consumption forecasting based on their modeling methodologies. We presented five different categories of forecasting approaches with one or more example projects. For each of the presented projects, we clarified in details its modeling and evaluation methodology, its experimental dataset used to build and evaluate the model as well as its achieved predictive performance.

In our work, we handle the problem of activity recognition in multi-user environments from a new and different perspective. Instead of transforming it into a single-label classification problem, we build an ensemble of multivariate conditional inference tree classifiers where each classifier takes a set of correlated and dependent labels as its multivariate response. To identify these sets of labels, we develop an innovative approach which identifies dependency relations between concurrent labels using the algorithm of conditional inference trees. We identify two types of dependency, namely intra-user and inter-user dependencies that study dependency relations within the activities of a single user and between the activities of all users respectively. Moreover, we take temporal relations between subsequent activities into account to build the final activity recognition model. Our work is distinguished by using appliance-level fine-granular measurements of power consumption as its only sensing modality. Our final activity recognition model works without using any information provided by other sensing modalities such as motion, wearable, environmental or state-change sensors. Moreover, it achieves a very good average f-measure value of 91% which outperforms the predictive performance of all multi-user activity recognition models presented in this section.

Concerning the forecasting of long-term and short-term power consumption, we follow in our work a hybrid methodology in which we develop several models for forecasting long-term and short-term power consumption of individual buildings as well as of a portfolio of buildings. In contrast to most of the aforementioned research projects, this work focuses on forecasting the detailed power consumption of individual buildings and not only on aggregated power consumption of a community such as a city or a country. Our novel models for forecasting total and detailed 24-hour day-ahead power consumption of individual buildings have shown very good predictive performance even when working with a heterogeneous set of buildings where different behavioral patterns of residents make it difficult for other forecasting approaches to achieve an accurate predictive performance.



---

## ACTIVITY RECOGNITION IN SINGLE-USER ENVIRONMENTS

---

ACTIVITY recognition represents an important research field in the area of smart environments [21]. By realizing accurate and rigorous solutions for activity recognition, we pave the road for a wide spectrum of IT services that enhance human beings' quality of life. In this chapter, we present *SMARTENERGY.KOM*, our platform for activity recognition in single-user environments, where we utilize fine-granular power consumption data to extract users' context and recognize their current activities. We start by giving a general definition of user activities in Section 4.1. In Section 4.2, we introduce software and hardware components of *SMARTENERGY.KOM*. We present our experimental deployments in Section 4.3. In Section 4.4, we clarify the processes of data preprocessing and feature extraction. We evaluate the predictive performance of our solution for activity recognition in Section 4.5. We summarize the chapter in Section 4.6. The research contributions presented in this chapter were the main topics of our paper in [5].

### 4.1 ACTIVITIES OF DAILY LIVING (ADLS)

As defined by Merriam-Webster [71], "an activity is the state of being active: behavior or actions of a particular kind". It can also be defined as "something that is done as work or for a particular purpose". ADLs represent an important type of activities which has been the subject of activity recognition literature. This type involves all kinds of activities that we perform in our daily life in order to maintain good health conditions [117]. ADLs are defined as "the things we normally do...such as feeding ourselves, bathing, dressing, grooming, work, homemaking, and leisure" [117].

Human activities can be categorized into atomic and non-atomic [95]. Atomic activities consist of only one event which represents a primitive action such as opening a door or pouring water into the teacup. Non-atomic activities consist of multiple sub-events such as making tea or preparing a meal in the kitchen. In our work, we focus on non-atomic activities of daily living such as eating, watching TV, reading and so on.

### 4.2 SMARTENERGY.KOM ACTIVITY RECOGNITION PLATFORM

Our main research goal in this chapter is to recognize users' context based on their fine-granular power consumption data. In order to achieve this goal, we built a system of sensor nodes which collects appliance-level power consumption data of all appliances at home. Moreover, it collects other environmental parameters, namely temperature, brightness and motion in the environment. The different components of the system are shown in Figure 3.

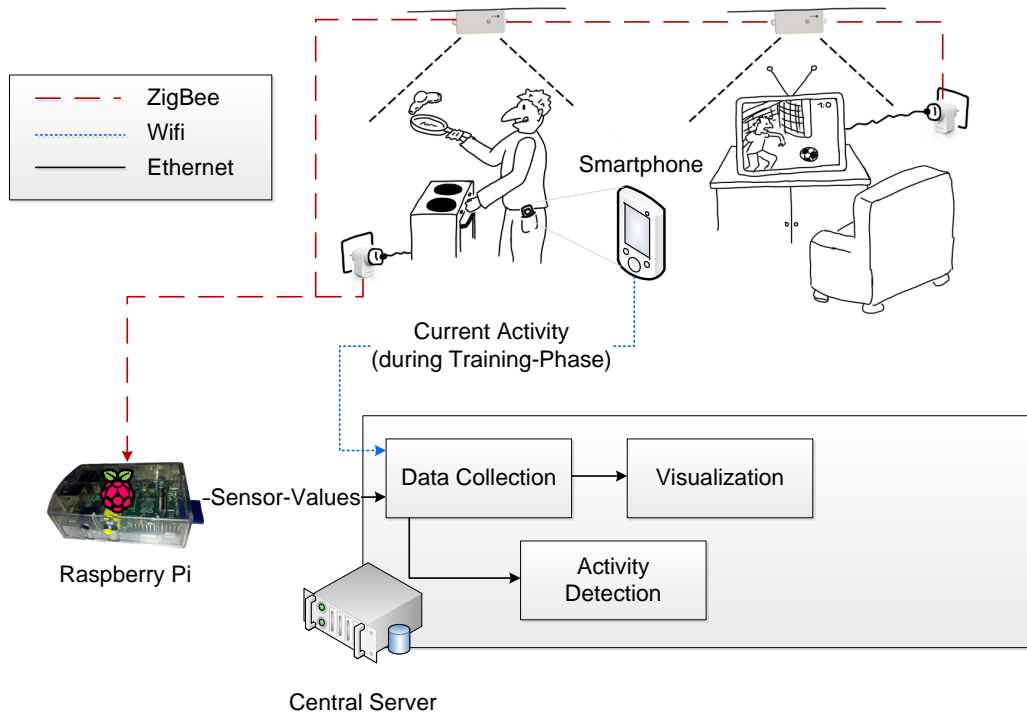


Figure 3: System architecture and network topology

**sensor nodes** Our platform utilizes two types of sensor nodes. On one hand, Plugwise appliance-level power sensors<sup>1</sup> are used to measure the power consumption of each electrical appliance. On the other hand, Pikkerton environmental sensors<sup>2</sup> are used to measure the temperature, brightness, and to detect motion in the environment. Figure 3 shows the environmental sensors mounted to the ceiling as well as the power sensors connected to the TV and oven as an example.

**raspberry pi** Raspberry Pi<sup>3</sup> plays the role of a gateway which collects all sensor readings and forwards them to the control server.

**smartphone** The user’s smartphone represents the feedback channel in the system as it is used by the user to tag her/his ongoing activities.

**control server** The locally deployed control server is responsible for collecting and storing the data. Moreover, it represents the smartness of the platform by pre-processing the data and constructing of the activity recognition model.

#### 4.3 DEPLOYMENTS AND EXPERIMENTAL SETUP

In order to collect the required training data for building and evaluating the activity recognition model, we deployed our platform in two different homes which we

<sup>1</sup> <http://www.plugwise.com/>

<sup>2</sup> <http://www.pikkerton.com/>

<sup>3</sup> <http://www.raspberrypi.org/>



<i>House A</i>	<i>House B</i>
<i>Cooking</i>	<i>CuttingBread</i>
<i>WatchingTV</i>	<i>WatchingTV</i>
<i>WorkingAtPc</i>	<i>ListenToRadio</i>
<i>Eating</i>	<i>Eating</i>
<i>MakingCoffee</i>	<i>MakingTea</i>
<i>CleaningDishes</i>	<i>Ironing</i>
<i>Reading</i>	<i>Reading</i>
<i>Sleeping</i>	<i>Sleeping</i>
<i>OutOfHome</i>	<i>OutOfHome</i>
<i>Ignore</i>	<i>Ignore</i>

Table 6: List of monitored activities in each house

call in the context of this chapter *House A* and *House B*. *House A* was occupied by a researcher who knows the technical details of the system whereas *House B* was occupied by a person who has no expertise in the field. Table 6 shows the lists of monitored activities in each of the houses. These lists have been designed based on our discussion with the residents who have provided us with the required information about their usual daily behavior and the activities they normally perform on a daily basis. Table 7 shows the rooms that we monitor in each of the houses as well

<b>Room</b>	<b>In <i>House A</i></b>	<b>In <i>House B</i></b>
Kitchen	Coffee machine, Radio, Electric kettle, Lamp above the hob, Lamp above the kitchen sink, Oven, Fridge	Radio, Electric kettle, Electric iron, Cooking mixer, Electric bread cutter
Living room	Projector, Audio system, Lamp	Television, Satellite receiver, Lamp
Office room	PC, PC accessories (screen, loudspeaker, etc.)	-
Sleeping room	-	Lamp, Radio alarm clock

Table 7: List of available electrical appliances at each house

as the available electrical appliances in each room. We connect each appliance to a Plugwise sensor. Moreover, we deploy an environmental sensor in each of the rooms apart from the sleeping room in *House B*.

To preserve the private sphere of users, we give them the choice to report their current activity as *Ignore*. In this way, we enable the users to perform the activities

that they do not want to be monitored without interrupting or affecting the system operations. All sensor readings related to the feedback *Ignore* are deleted from the datasets in the preprocessing step. We have deployed the system for 82 days in *House A* where around 22.5 million of data points have been collected. The duration of deployment was shorter in *House B* with 62 days and about 20 million collected data points.

#### 4.4 DATA PREPROCESSING AND FEATURE EXTRACTION

Our platform follows a supervised learning approach to build a machine learning model whose task is to learn the correlation between sensor readings and user activities from the collected datasets. As mentioned before, our data consists of around 42.5 million data points. In order to extract meaningful training and testing instances that are necessary to build the machine learning model, we follow a windowing approach in which we divide our series of data into time slots where each time slot represents an instance of sensor readings labeled with the accompanied activity. We extract the features in each time slot as shown in Figure 4 where the maximum reading of each sensor during this time slot is taken as a representative feature for it. The reason behind taking the maximum value as a feature is that it ignores all zero readings of the respective sensor during time slots in which the device connected to this sensor is turned on after the beginning of the respective time slot which causes zero readings to be initially reported. Other features such as the average will be affected by such zero values and therefore can cause confusion for machine learning models. Eq. 21 shows the feature vector consisting of maximum sensor readings as well as the timestamp at which the activity is happening. The timestamp is only represented by the hour because minutes and seconds cause the model to overfit the training data.

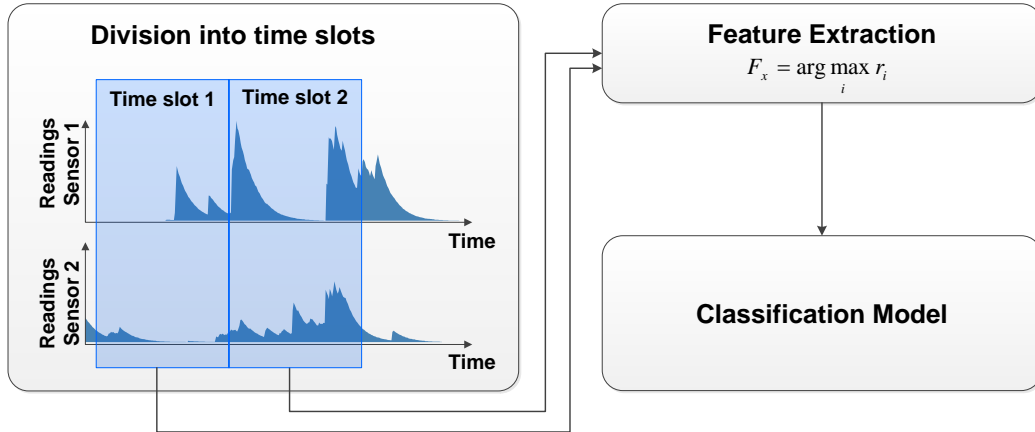


Figure 4: The process of feature extraction

$$\text{Feature vector : } F(t) = \langle mS_1(t), mS_2(t), \dots, mS_N(t), H \rangle \quad (21)$$

Where:

- $mS_i(t)$  : maximum value of sensor  $s_i$  during timeslot  $t$

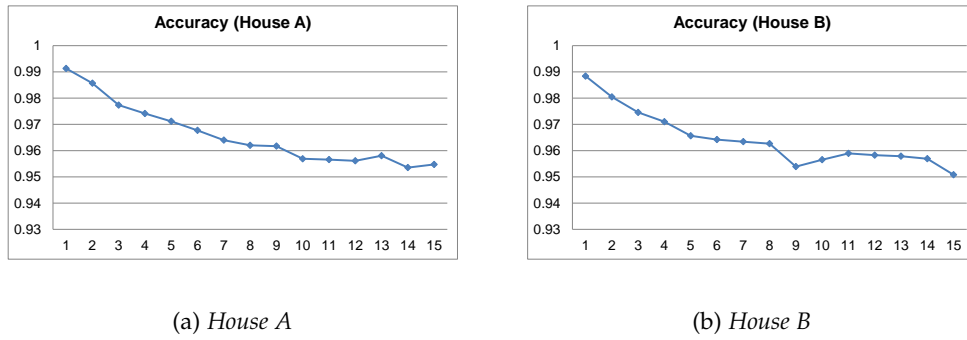


Figure 5: Length of the time slot vs. accuracy of the activity recognition platform for *House A* and *House B*

- $N$  : number of sensors
- $H$  : timestamp represented by the hour

The optimal length of the time slot has to be specified empirically. In Section 4.5, we run an experiment to specify this optimal value.

#### 4.5 RESULTS AND EVALUATION

To evaluate the predictive performance of *SMARTENERGY.KOM*, we design a set of experiments that cover several aspects of the collected datasets. As we are using a windowing technique to divide our data into time slots, we conduct an experiment which determines the optimal length of the windowing time slot. After determining this value, we build our model for activity recognition based on the datasets collected from both houses. As our dataset contains both power and environmental data, we conduct two different experiments. In the first one, we incorporate the values of environmental and power sensors into the feature vector presented in Eq. 21. In the second experiment, we exclude environmental sensors from the feature vector and build our model for activity recognition solely based on power consumption data. The hour is used as a feature in both models. The goal of the second experiment is to study the effect of excluding environmental sensors on the predictive accuracy of activity recognition models.

We use the random forest classifier [16] for all experiments as it produces one of the most robust and accurate machine learning models. Moreover, it has proven to be the most suitable classifier for other datasets that have the same characteristics as our dataset [39] [87]. We divide the data into 70% training set and 30% testing set. For building random forest models, we use the caret library [61] in R [82] by applying a 10-fold cross-validation [12] on the training set. Using training set, caret determines the optimal set of parameters for building a random forest model and returns a model based on this set of parameters.

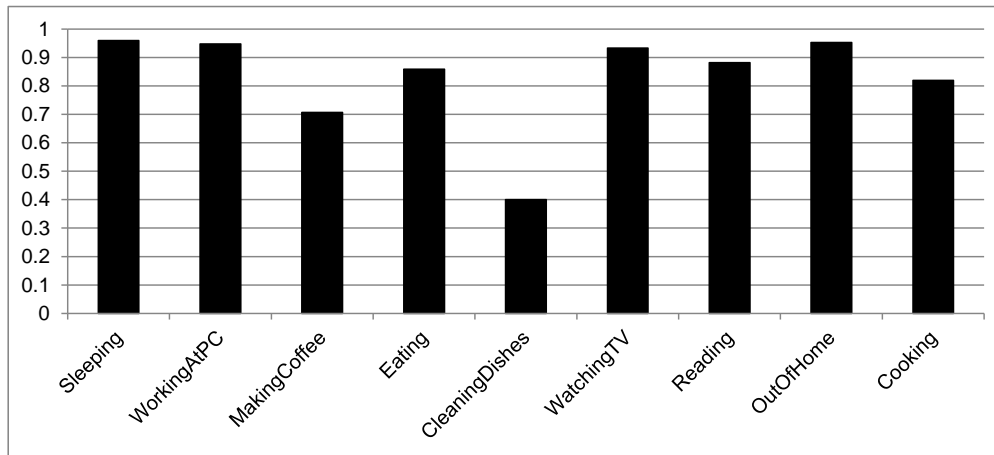


Figure 6: F-measure values of environmental-power model with regard to all activities in *House A*

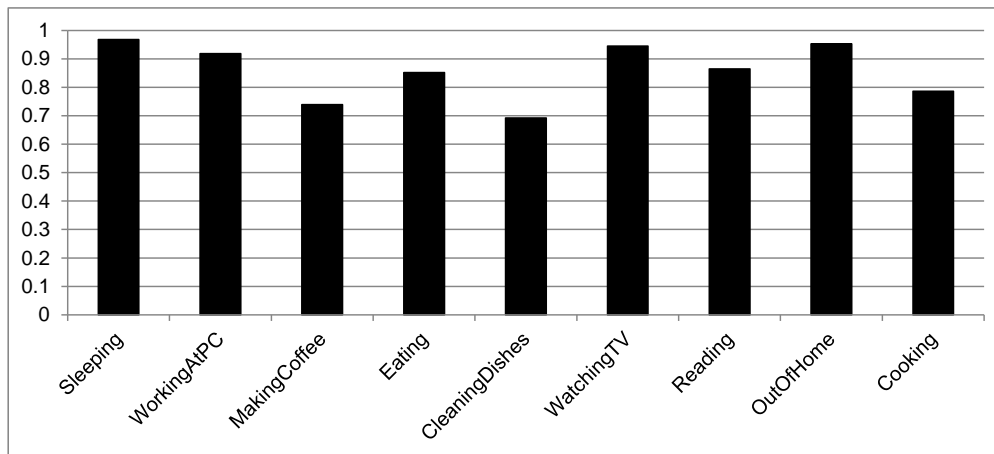


Figure 7: Accuracy values of environmental-power model with regard to all activities in *House A*

#### 4.5.1 Optimal Time Slot Length

The length of windowing time slot plays an essential role in determining the predictive performance of the activity recognition model. We assume a scenario in which the user is allowed to perform one activity at a time. Therefore, a long time slot may cause the instances of two different activities to be combined in one time slot as some activities such as *MakingCoffee* lasts for 3 to 5 minutes. Having two activities in one time slot renders it useless as we assume only a single activity to be performed at a given time. Moreover, a long time slot leads to a decreasing number of training and testing instances which has a negative effect on the performance of activity recognition model.

Figures 5a and 5b show the accuracy of activity recognition models with regard to the length of the time slot. As we can see from both figures, the accuracy decreases as the length of time slot increases. This can be explained by the increased number of slots with more than one activity as we clarified before. A 1-minute time slot leads to the best accuracy value. However, we choose the time slot length to be 2 minutes

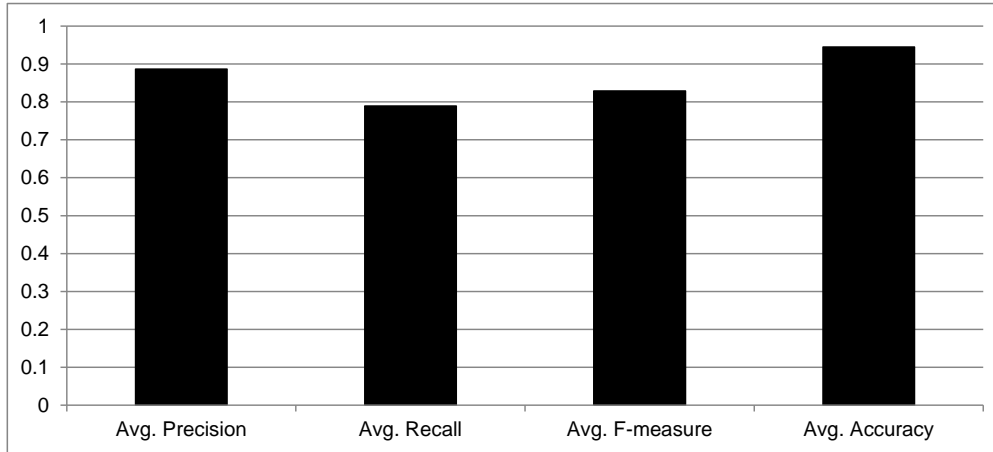


Figure 8: Overall performance of environmental-power model with regard to *House A* in terms of average f-measure, accuracy, precision and recall

as a 1-minute slot is very short for all sensors to change their values according to the new activity in case of activity change. This is because users tend to give their feedback before they start performing the new activity.

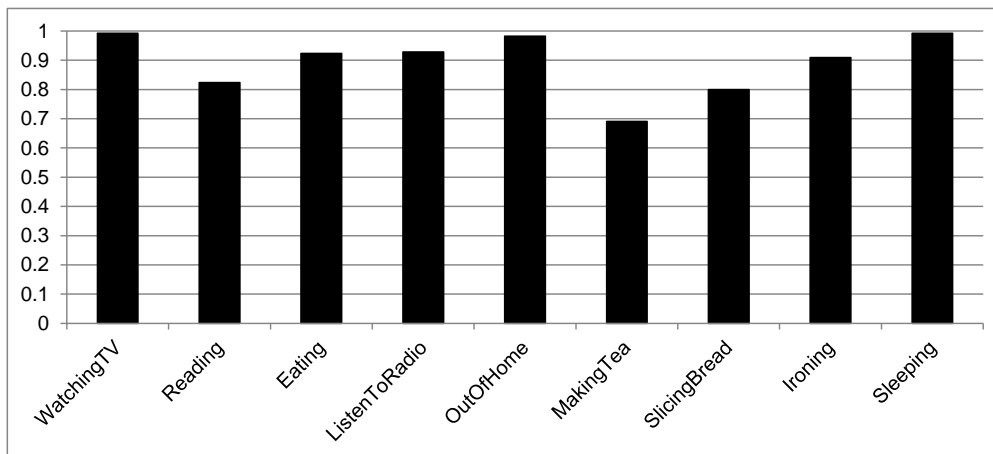


Figure 9: F-measure values of environmental-power model with regard to all activities in *House B*

#### 4.5.2 Power and Environmental Sensors

In this section, we evaluate the predictive performance of our model for activity recognition based on power and environmental sensors. Using a time slot of 2 minutes, we build the feature vectors based on the readings of both power and environmental sensors as clarified in Figure 4. We refer to this model as “environmental-power model”. As mentioned before, we use caret library in R to train a random forest classifier using 70% of the data as a training set. After getting the optimal classifier using a 10-fold cross-validation in caret, we test it on the remaining 30% of the dataset. We present our evaluation results for both deployments, namely *House A* and *House B*.

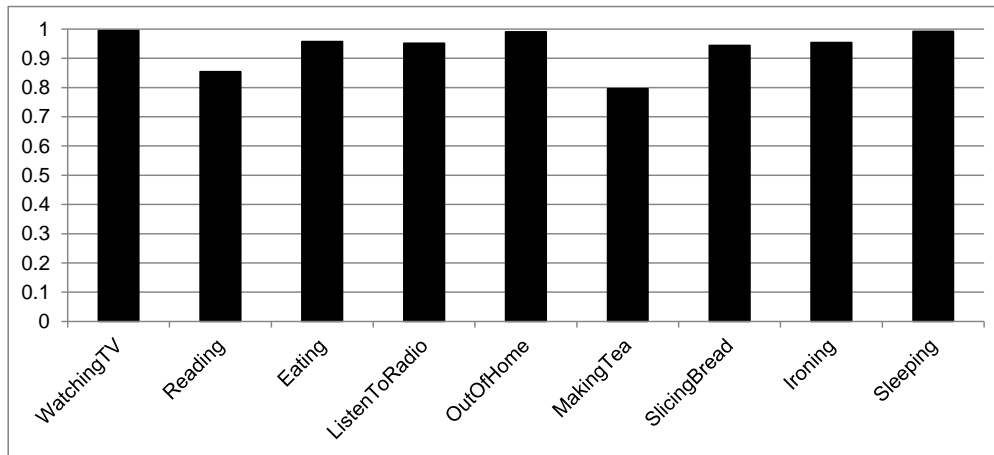


Figure 10: Accuracy values of environmental-power model with regard to all activities in *House B*

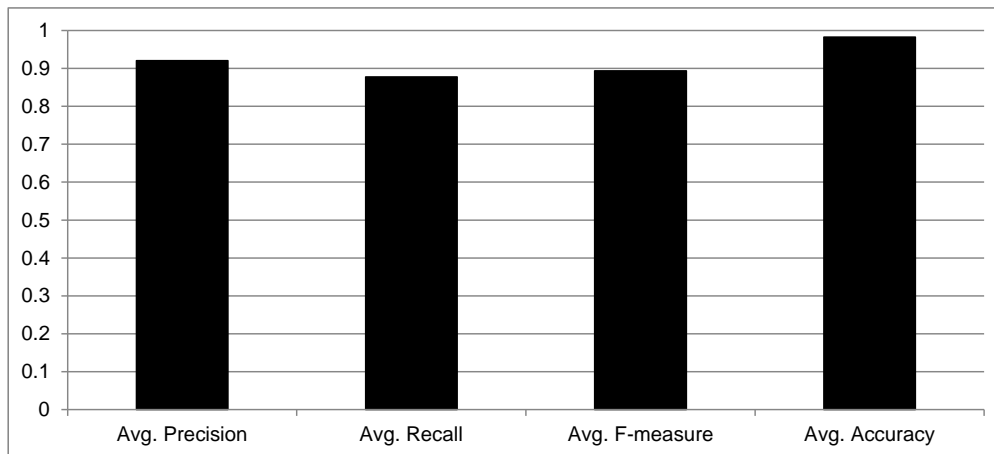


Figure 11: Overall performance of environmental-power model with regard to *House B* in terms of average f-measure, accuracy, precision and recall

Figures 6 and 7 show the predictive performance in terms of f-measure and accuracy values achieved in recognizing each of the activities in *House A*. We notice from both figures that our activity recognition model is able to recognize most of the activities with very good predictive performance represented by f-measure values of up to 95%. However, for activities such as *CleaningDishes* and *MakingCoffee* the model reached f-measure values of 40% and 70% respectively. The reason behind this poor performance in recognizing both activities is their short duration which leads to a very few number of instances available for training the model. We present a solution for this problem in Chapter 5 by incorporating temporal relations between subsequent activities into the feature space of our activity recognition model. Figure 8 presents the overall average performance of *House A*'s model in terms of f-measure, accuracy, precision and recall.

The results obtained after evaluating the predictive performance on *House B*'s dataset follow the same direction as shown in Figures 9 and 10. As we see from both figures, our model is able to recognize activities such as *Sleeping*, *OutOfHome* and *WatchingTV* with f-measure values of up to 99%. Moreover, it shows a very

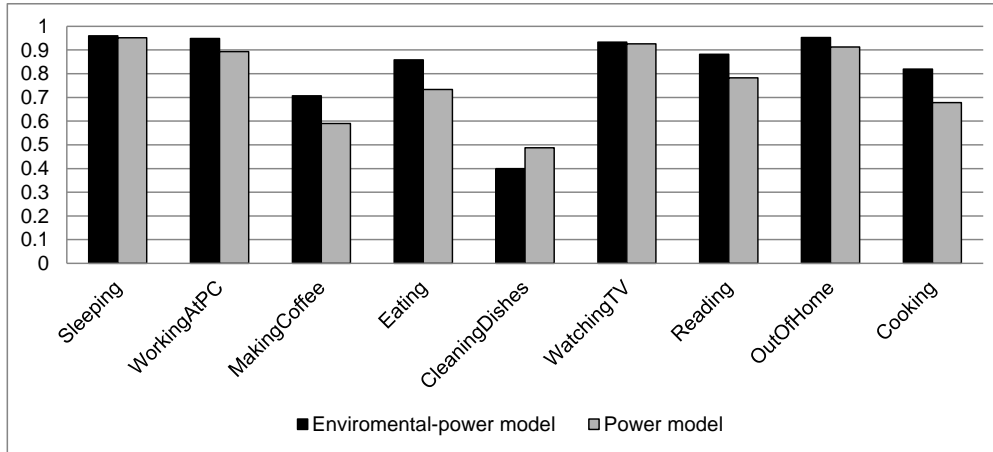


Figure 12: Comparison between power-only and environmental-power models in terms of f-measure values for activities of *House A*

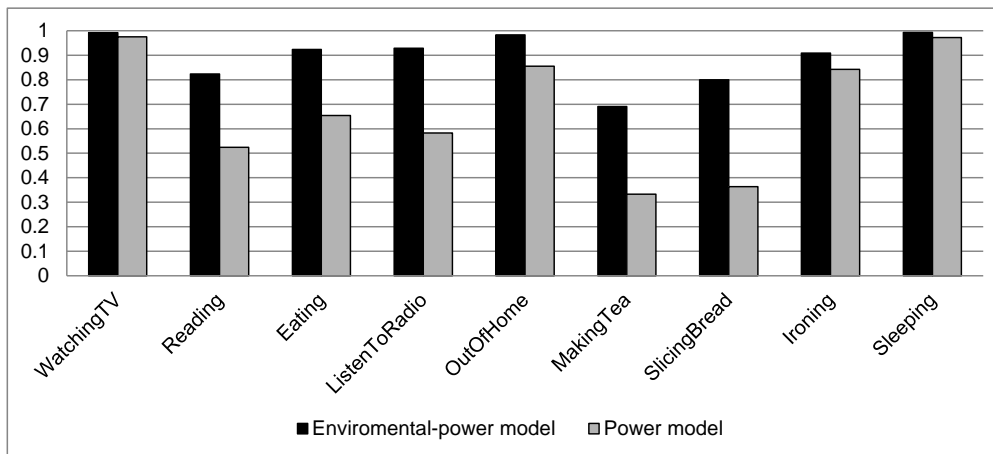


Figure 13: Comparison between power-only and environmental-power models in terms of f-measure values for activities of *House B*

good performance in recognizing the activities of *Reading*, *Eating*, *ListenToRadio* and *Ironing*. However, it recognizes the activities of *MakingTea* and *SlicingBread* with f-measure values of 69% and 80% respectively. As explained before, this is because the few number of instances these activities have due to their short durations. We present our solution for this problem in Chapter 5.

#### 4.5.3 Power Sensors

In this section, we study the effect of excluding environmental sensors on the predictive performance of our model for activity recognition. In this evaluation setting, feature vectors consist of only readings of power sensors combined with the hour as activity's timestamp. We repeat the same evaluation setup in which we build a random forest model using R's caret library with 70% of the data as a training set and 30% as a testing set. We refer to this model as "power-only" model.

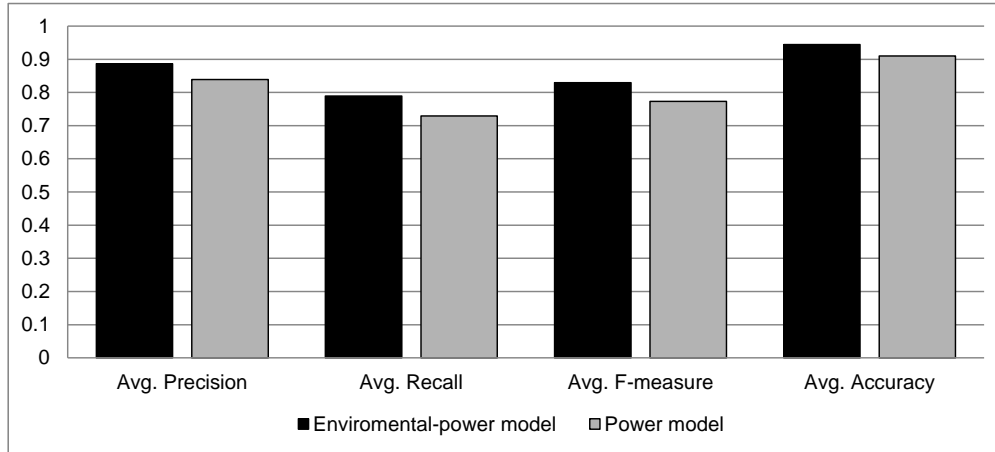


Figure 14: Comparison between power-only and environmental-power models in terms of average f-measure, accuracy, precision and recall achieved for *House A*

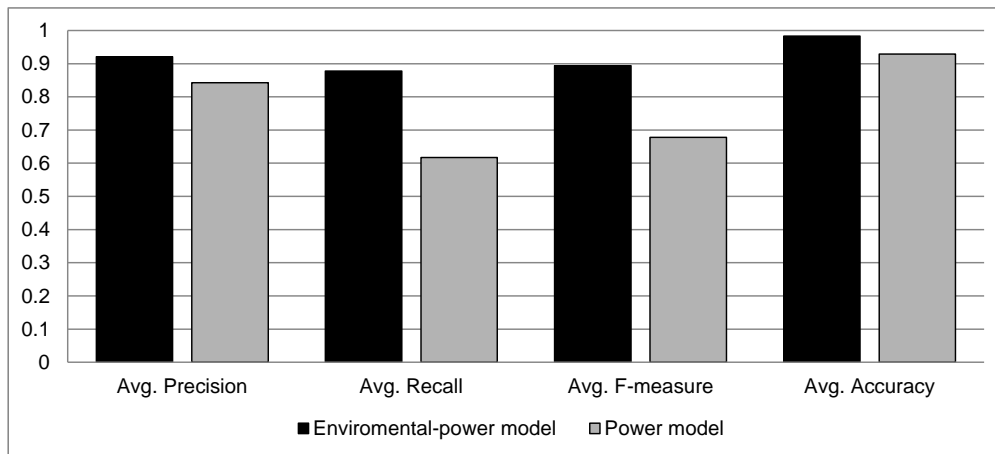


Figure 15: Comparison between power-only and environmental-power models in terms of average f-measure, accuracy, precision and recall achieved for *House B*

Figure 12 presents a comparison between the f-measure values achieved in recognizing the activities of *House A* using power-only and power-environmental models. We notice from this figure 10% to 14% decrease in f-measure values for the activities *MakingCoffee*, *Eating*, *Reading* and *Cooking*. No serious decrease can be noticed for the activities *Sleeping*, *WorkingAtPc*, *WatchingTV* and *OutOfHome*. Moreover, we notice an 8% increase in the f-measure value of the activity *CleaningDishes*. A larger decrease in f-measure values can be noticed in recognizing activities of *House B* as shown in Figure 13. We notice 30% to 40% decrease in recognizing the activities *Reading*, *Eating*, *ListenToRadio*, *MakingTea* and *SlicingBread*.

Figure 14 and 15 present an overall comparison between environmental-power and power-only models in terms of average f-measure, accuracy, precision and recall achieved in recognizing the activities of *House A* and *House B* respectively. As can be seen from Figure 14, f-measure and recall values for the model of *House A* have slightly decreased by 5% and 6% respectively. However, Figure 15 shows that f-measure and recall values for the model of *House B* have decreased by 22% and 26% respectively.



The obtained results show a decrease in predictive performance of our activity recognition model after the exclusion of environmental sensors. However, we present in Chapter 5 a solution that enhances the predictive performance of power-only models so that they outperform environmental-power models. This solution is based on incorporating temporal relations between subsequent activities into the feature space of power-only models.

#### 4.6 SUMMARY

In this chapter, we presented *SMARTENERGY.KOM*, our platform for activity recognition in single-user environments. In Section 4.2, we introduced the hardware components of the platform. Moreover, we presented the deployment of *SMARTENERGY.KOM* in two single-user apartments in Section 4.3. In Section 4.4, we explained the processes of data preprocessing and feature extraction. We introduced the experimental setup of constructing activity recognition models in Section 4.5. We clarified the used windowing technique and identified the optimal time slot length in Section 4.5.1. In Section 4.5.2, we evaluated the predictive performance of environmental-power activity recognition models. We studied the effect of excluding environmental sensors on the predictive performance of activity recognition models in Section 4.5.3.



---

## BEHAVIORAL MODELING OF USERS IN INDOOR ENVIRONMENTS

---

**H**UMANS tend to follow daily routines in performing their everyday activities. Identifying and extracting such routines pave the road for the realization of a wide variety of intelligent context-aware services that enhance many aspects of people's life. Examples of such services are energy-conservation and comfort-increasing services.

As users' activities in smart environments are strongly associated with their power consumption, fine-granular sensing of power consumption represents an important sensing modality for identifying and extracting insights into users' everyday behavior. In this chapter, we cover three essential aspects of users' behavior in indoor environments. First, we present our approach for indoor localization of users based on their fine-granular power consumption in Section 5.1. Thereafter, we introduce our approach for identifying behavioral patterns that are followed by users in performing their activities in Section 5.2. Furthermore, we analyze the effect of these patterns on improving the predictive performance of activity recognition models. In Section 5.3, we conduct an in-depth analysis of users' hourly power consumption with the goal of building an hourly power consumption profile of users. Such a profile can be used for increasing users' awareness of their power consumption and for identifying and detecting abnormal power consumption behaviors. We evaluate all of the approaches presented in this chapter based on the dataset we collected using *SMARTENERGY.KOM* framework as clarified in Chapter 4. The research contributions presented in this chapter were the main topics of our papers in [6][7].

### 5.1 INDOOR LOCALIZATION BASED ON POWER CONSUMPTION

Localization of users in indoor environments has always represented a challenge for researchers in the field of smart environments [67]. Many approaches have been developed to tackle this challenge. They have utilized a wide variety of sensing modalities to achieve this goal. Examples are Passive Infrared sensors (PIRs) [101], Radio-frequency identification (RFID) tags [60], wireless transceiver [4][98][113], cameras [123] and so on. Our research goal in this section is to design an accurate approach for indoor localization only based on fine-granular sensing of power consumption and without the need for deploying any other sensing modality.

In order to test and evaluate the predictive performance of our approach, we need a dataset in which fine-granular measurements of power consumption are labeled with user location in an indoor environment. We build this dataset by modifying *SMARTENERGY.KOM* dataset so that it contains users' indoor location as explained in the next section.

### 5.1.1 Dataset

As our research goal is to accurately localize users in indoor environments based on their fine-granular power consumption, we require a dataset which combines this information. In Chapter 3, we presented the dataset we collected by deploying *SMARTENERGY.KOM* in two single-user apartments. Each feature vector in this dataset is labeled with user's current activity. As the users in both deployments perform each of their activities in one known place which does not change throughout the deployment, we can infer user location based on her/his current activity. Therefore, we modify both datasets so that each feature vector gets labeled with the location at which the activity happens instead of the activity itself. As a result, we get a dataset in which each instance combines appliance-level power measurements with user location inside the house as shown in Eq. 22. Table 8 shows the location of each activity performed by users in *House A* and *House B*. It can be seen from the table that *House A* has only one room, namely living room in which the user lives and sleeps. After constructing the required dataset, we build the classification model and evaluate its performance in the next section.

$$I_t = \langle mS_1(t), mS_2(t), \dots, mS_N(t), H, \text{Loc}(t) \rangle \quad (22)$$

Where:

- $mS_i(t)$  : maximum value of sensor  $s_i$  during timeslot  $t$
- $N$  : number of sensors
- $H$  : timestamp represented by the hour
- $\text{Loc}(t)$  : location of user during timeslot  $t$

Location	In House A	In House B
Kitchen	<i>Cooking, Eating, Making-Coffee, CleaningDishes</i>	<i>Eating, ListenToRadio, MakingTea, SlicingBread, Ironing</i>
Living room	<i>Sleeping, WatchingTV, Reading</i>	<i>WatchingTV</i>
Office room	<i>WorkingAtPc</i>	-
Sleeping room	-	<i>Sleeping, Reading</i>
Outside	<i>OutOfHome</i>	<i>OutOfHome</i>

Table 8: Locations of activities

### 5.1.2 Evaluation

This section presents the evaluation results of our indoor localization models with regard to both *House A* and *House B*. Following the same evaluation settings presented in Chapter 4, we use R's caret library to build two random forest models for

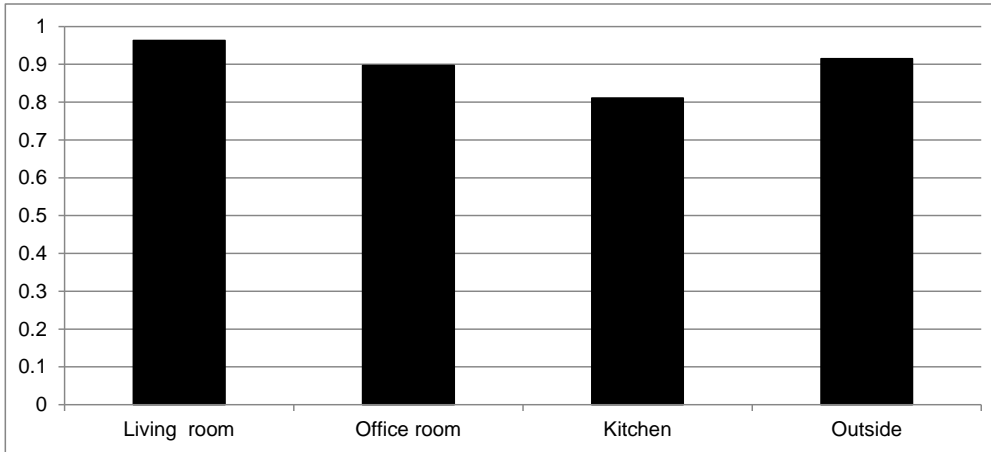


Figure 16: F-measure values of indoor localization model with regard to *House A*

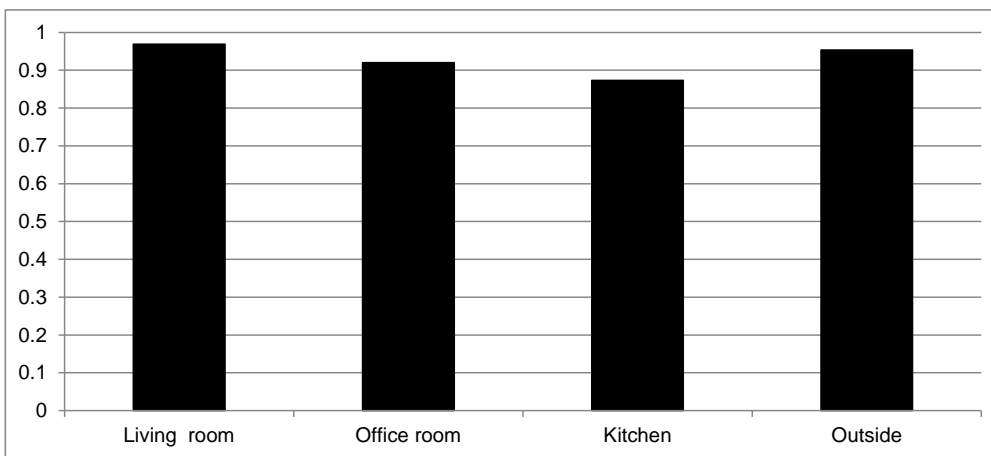


Figure 17: Accuracy values of indoor localization model with regard to *House A*

each house. We utilize 70% of the dataset as a training set to construct the optimal random forest models using 10-fold cross validation. We test the resulting models on the remaining 30% of the dataset.

Figures 16 and 17 show the values of f-measure and accuracy achieved in recognizing each location in *House A*. Both figures confirm the very good predictive performance achieved in recognizing each location. The model achieved f-measure values of 96%, 89% and 91% in recognizing the locations “*living room*”, “*office room*” and “*outside*”. A lower f-measure value, namely 81% has been achieved in recognizing the location “*kitchen*”. This is due to the fewer number of instances this location has as the user spent most of his time in living room

Figure 18 shows that our model is able to recognize the locations with an average f-measure value of 89% and an average accuracy value of 92%.

The same applies for *House B* as shown in Figures 19, 20 and 21.

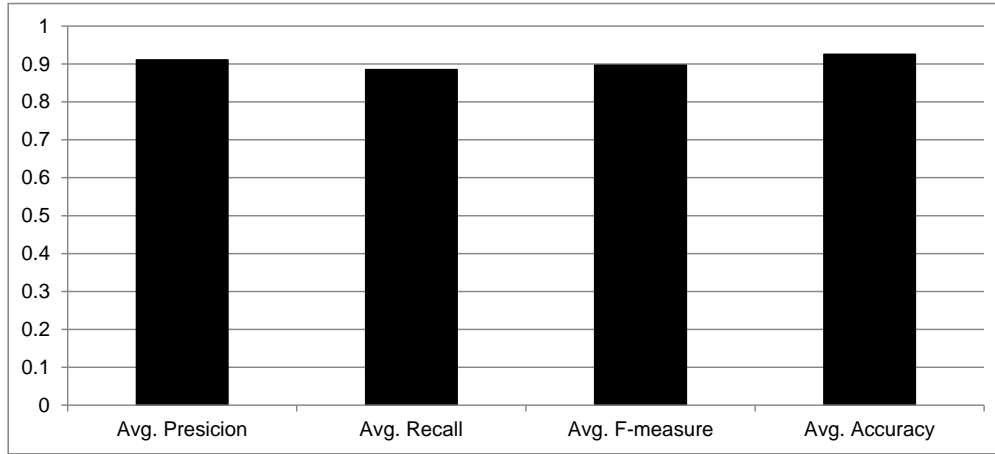


Figure 18: Overall performance of indoor localization model with regard to *House A* in terms of average f-measure, accuracy, precision and recall

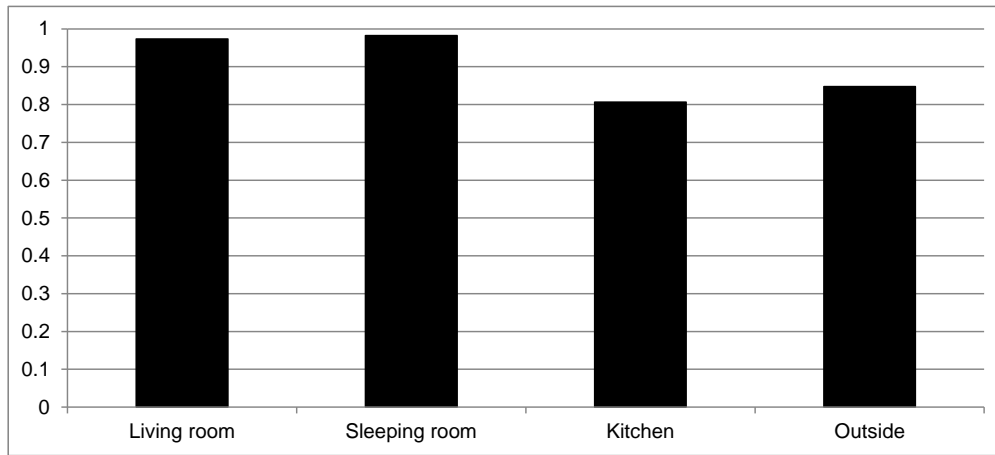


Figure 19: F-measure values of indoor localization model with regard to *House B*

## 5.2 PATTERNS IN USER'S DAILY LIFE AND THEIR EFFECT ON THE ACCURACY OF ACTIVITY RECOGNITION

Humans tend to perform their daily activities in certain routines that repeat themselves over the days. Logically, we perform the activity of *Cooking* before we start *Eating*. Moreover, we tend to watch TV or read a book before going to sleep. Introducing such temporal patterns between subsequent activities should be beneficial in enhancing the predictive performance of activity recognition models. In this section, we identify such patterns in the dataset we collected using *SMARTENERGY.KOM* platform as introduced in Chapter 4. Furthermore, we present our approach for incorporating these patterns into our model for activity recognition presented in Chapter 4. Finally, we evaluate the predictive performance of our model for activity recognition after incorporating the temporal patterns and we present the obtained results.



Figure 20: Accuracy values of indoor localization model with regard to *House B*

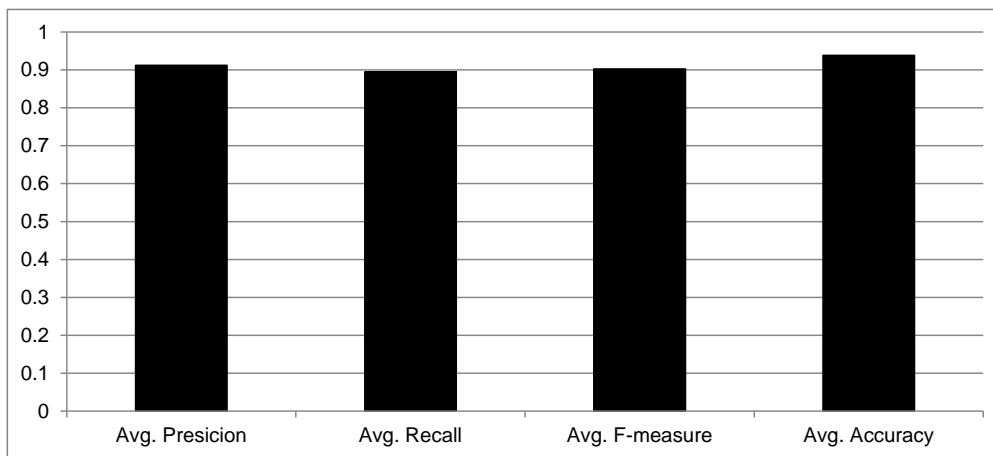


Figure 21: Overall performance of indoor localization model with regard to *House B* in terms of average f-measure, accuracy, precision and recall

### 5.2.1 Identification of Temporal Patterns

J. F. Allen and G. Ferguson [8] defined three types of temporal relations between two activities  $X$  and  $Y$ , namely  $X$  happens after  $Y$ ,  $X$  happens before  $Y$  or  $X$  overlaps with  $Y$ . In Chapter 4, we assumed the user to perform only a single activity at a given time. Therefore, we are interested only in the first two temporal relations as no overlapping activities are allowed in our dataset. We use Apriori algorithm [1] for identifying frequent activity sequences i.e. temporal patterns in our dataset. Before presenting the approach of identifying temporal patterns using Apriori algorithm, we need to define two basic terms, namely episode and sequence.

- **Episode:** D. Lymberopoulos et al. [66] defined an episode as a sequence of one or more activities characterized by begin and end timestamps. An episode represents the main unit of interest in mining temporal patterns. Therefore, we choose our episode to be a whole day starting at 00:00:00 and ending at 23:59:59. We build an episode dataset for *House A* and *House B*. This results in two datasets comprised of 64 and 61 episodes for *House A* and *House B*

respectively. Eq. 23 represents the mathematical representation of an episode comprised of  $n$  activities.

$$\text{Episode}_i = \langle A_1(T_1), A_2(T_2), \dots, A_n(T_n) \rangle \quad i \in [1, d] \quad (23)$$

Where:

- $A_j$  : activity  $j$
  - $T_j$  : the timestamp associated with activity  $j$
  - $n$  : number of activities during day  $i$
  - $d$  : number of days
- **Sequence:** it is defined as a set of two or more activities that successively happen. A sequence defined as  $\langle \text{Cooking}, \text{Eating} \rangle$  implies that the activity of eating directly happens after the activity of cooking.

As mentioned before, we utilize Apriori algorithm for identifying temporal patterns in our episode datasets. The main goal of Apriori algorithm is to extract frequent sequences from an episode dataset based on two main values, namely support and confidence of these sequences. The support of a certain sequence is defined as the number of episodes that contain this sequence divided by the total number of episodes as shown in Eq. 24. A sequence is considered as a frequent sequence when its support is larger than a predefined threshold which we refer to as *minSupp*.

$$\text{Support}(\langle X, Y \rangle) = \frac{\# \text{episodesContaining} \langle X, Y \rangle}{\# \text{episodes}} \quad (24)$$

Confidence is defined as the probability of activity  $Y$  occurring given that activity  $X$  has already occurred. It is computed by dividing the support of sequence  $\langle X, Y \rangle$  over the support of sequence  $\langle X \rangle$  as shown in Eq. 25.

$$\text{Confidence}(\langle X, Y \rangle) = \frac{\text{Support}(\langle X, Y \rangle)}{\text{Support}(\langle X \rangle)} \quad (25)$$

The result of applying Apriori algorithm on an episode dataset is a set of rules that take the form  $A \xrightarrow{\text{confidence}} B$ . Such a rule implies a high support value for both sequences  $\langle A \rangle$  and  $\langle A, B \rangle$ . Moreover, it implies a high confidence for the sequence  $\langle A, B \rangle$  meaning that the activity  $B$  occurs with high probability given that activity  $A$  has occurred. However, requiring high support values for these sequences lead to a problem in our use case. The reason behind this problem is the existence of activities that are rarely performed by the user. Such activities are characterized by having low support values. However, they might be part of other sequences that have very high confidence values. These sequences will be discarded due to the low support values of rare activities. For example, the activity *Reading* has a small support value, namely 4.7% since it is rarely performed by the user. However, the sequence  $\langle \text{Reading}, \text{Sleeping} \rangle$  is characterized by having a very high confidence value which implies that any occurrence of reading activity will be followed by sleeping with a high probability. To solve this problem, we follow [65] by utilizing the approach of multiple minimum supports which assigns a different miniSupp value for each item by multiplying its support value by the global miniSupp value shown in Eq. 26. This



results in individual miniSupp values that are proportional to the individual support values.

$$\text{miniSupp}_i = \text{global\_miniSupp} \times \text{support}(A_i) \quad (26)$$

By applying Apriori algorithm on our episode datasets, we identify all frequent sequences of activities that regularly occur after each other. Table 9 shows the list of these sequences.

Temporal Relations	
Previous Activity	Current Activity
<b>Deployment 1</b>	
Sleeping	WorkingAtPC
WorkingAtPC	Eating
MakingCoffee	Eating
Eating	WorkingAtPC
WatchingTV	Sleeping
OutOfHome	WorkingAtPC
Cooking	Eating
<b>Deployment 2</b>	
Reading	Sleeping
ListenToRadio	Eating
MakingTea	Eating

Table 9: Temporal patterns identified in our episode datasets

After the identification of all frequent temporal patterns in the datasets, we incorporate this knowledge into the feature space of our activity recognition model. This is done by expanding the feature vector so that it contains besides sensor readings, the previous performed activity as well as the most probable current activity based on the identified temporal patterns. Eq. 27 shows the new feature vector structure.

$$\text{Feature vector : } F(t) = \langle mS_1(t), mS_2(t), \dots, mS_N(t), H, A_{t-1}, A_t \rangle \quad (27)$$

Where:

- $mS_i(t)$  : maximum value of sensor  $s_i$  during timeslot  $t$
- $N$  : number of sensors
- $H$  : timestamp represented by the hour
- $A_{t-1}$  : the activity performed in the previous slot
- $A_t$  : the most probable current activity

After transforming all instances of our datasets based on the new feature vector structure, we conduct our evaluation study to measure the effect of extracted temporal knowledge on the predictive performance of activity recognition models.

### 5.2.2 Evaluation

This section presents the results obtained by evaluating the predictive performance of our activity recognition model after incorporating temporal knowledge into the feature space. As presented in Chapter 4, *House A* and *House B* datasets contain readings of power and environmental sensors. We incorporate the temporal knowledge only to the power models as our main goal is to recognize users' activities solely based on their fine-granular power consumption data. Therefore, readings of environmental sensors are excluded from the feature vector presented in Eq. 27. We compare the obtained results to the predictive performance of environmental-power and power-only activity recognition models presented in Chapter 4. We build one random forest model for each deployment using R's caret library with 70% of the data as a training set and 30% as a testing set.

Figure 22 compares the f-measure values achieved in recognizing activities of *House A* by environmental-power, power-only and power-temporal activity recognition models. It is clear from the figure that incorporating temporal knowledge significantly enhanced the predictive performance of the power-only model. Moreover, power-temporal model outperforms environmental-power model in recognizing all activities. An overall comparison between all models in terms of average f-measure, accuracy, precision and recall achieved in recognizing activities of *House A* is shown in Figure 23 which clearly shows that power-temporal model outperforms environmental-power and power-only models with respect to all four measures. We conduct the same comparison for *House B* dataset as shown in Figures 24 and 25. It is clear from both figures that power-temporal model outperforms both other models for the dataset of *House B* as well.

The obtained results lead to the conclusion that users' activities can be recognized with very good predictive performance solely based fine-granular sensing of power consumption and without the need for deploying any other sensing modality which proves our hypothesis. Even though the predictive performance of activity recognition model has decreased after the exclusion of environmental sensors, we were able to significantly increase it by incorporating temporal knowledge into the feature space. The final power-temporal model outperformed environmental-power model in terms of all performance measures, namely f-measure, accuracy, precision and recall as shown in Figure 26.

## 5.3 POWER PROFILING

As humans are used to performing the same activities every day, their daily power consumption tends to have a pattern that repeats itself over the days. Our goal in this section is to identify such a pattern and to build a daily power consumption profile of users by conducting an in-depth analysis of their daily power consumption. Identifying a daily power profile of users has its potential benefits in several application scenarios. On one hand, it helps users understanding their detailed daily power consumption and thereby being more aware of the power they consume. On the other hand, daily power consumption profiles of consumers help electric utilities in planning their resources and recommending more suitable tariffs to consumers based on their power consumption behavior.

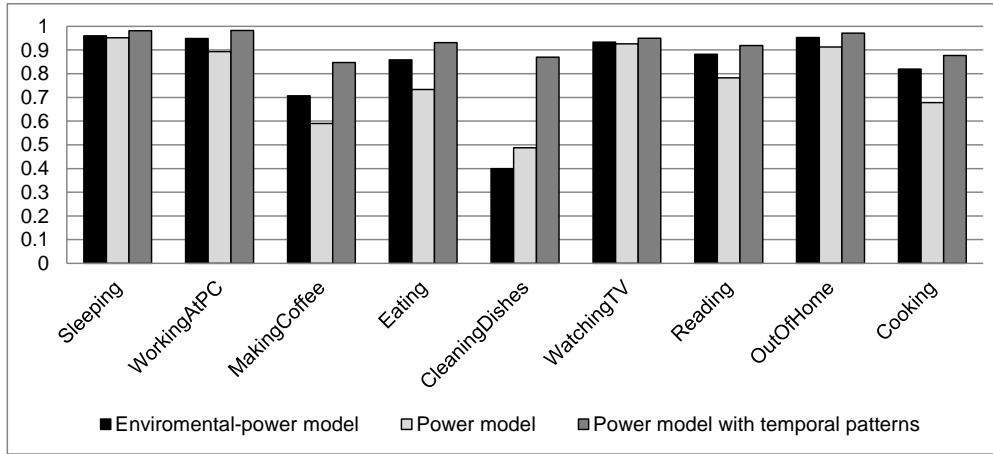


Figure 22: Comparison between the models: power-environmental, power-only and power with temporal patterns in terms of f-measure values for activities of *House A*

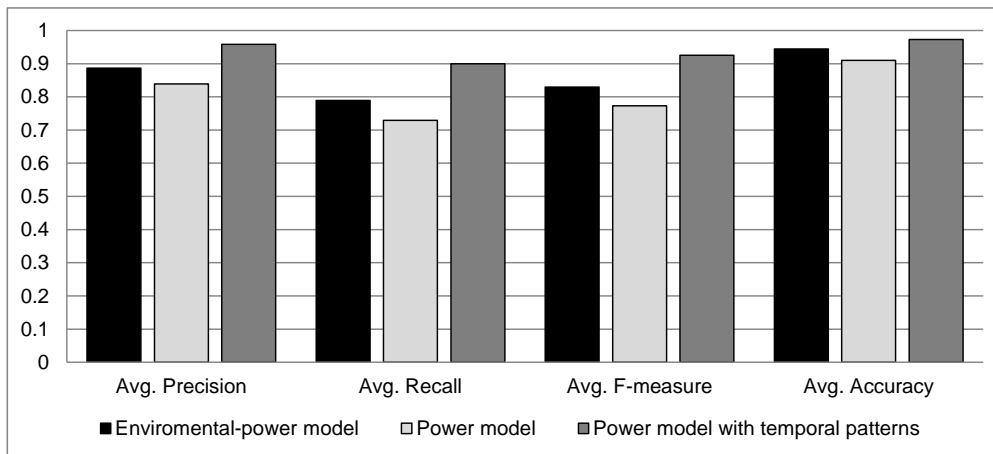


Figure 23: Comparison between the models: power-environmental, power-only and power with temporal patterns in terms of average f-measure, accuracy, precision and recall achieved for *House A*

We use *SMARTENERGY.KOM* dataset to conduct our analytical study in this section. Based on the appliance-level power measurements, we build a detailed time series for each day that gives the power consumed during each hour of the respective day. Section 5.3.1 presents in details the construction process of the dataset.

### 5.3.1 Extracting Hourly Power Consumption

In this section, we explain the process of constructing the dataset required to conduct our analysis of users' daily power consumption. As *SMARTENERGY.KOM* dataset contains power measurements for each appliance in *House A* and *House B*, we compute the power consumed during each hour as follows:

- For each time slot belonging to the respective hour, we calculate the average value of the readings of each Plugwise sensor.

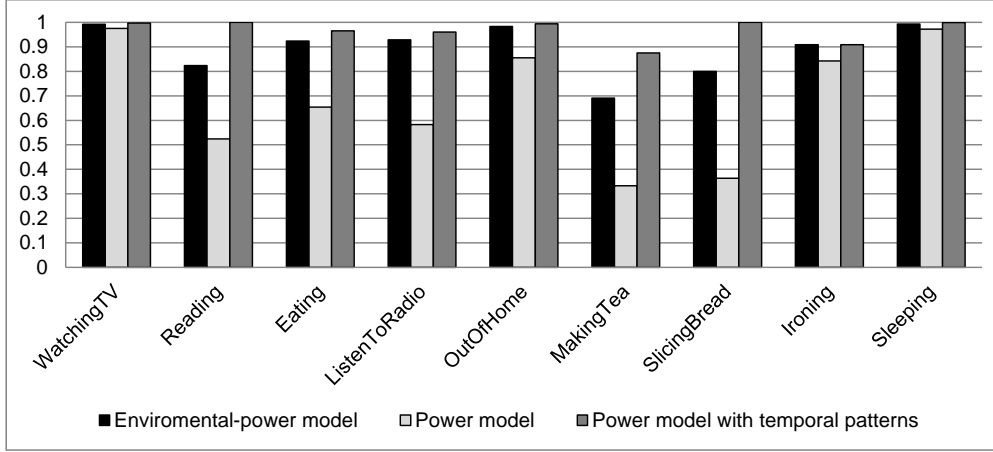


Figure 24: Comparison between the models: power-environmental, power-only and power with temporal patterns in terms of f-measure values for activities of *House B*

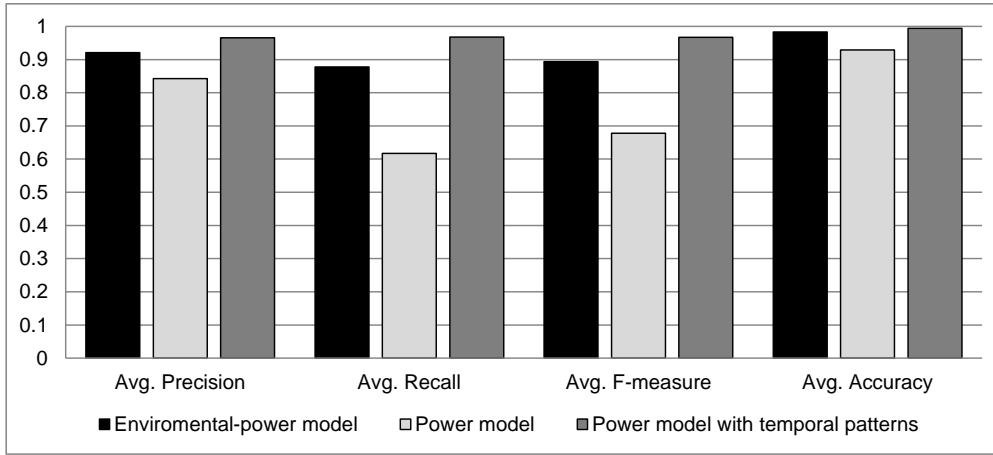


Figure 25: Comparison between the models: power-environmental, power-only and power with temporal patterns in terms of average f-measure, accuracy, precision and recall achieved for *House B*

- We divide the obtained average value by 30 so that we convert “Watt” readings of Plugwise sensor into “Wh (Watt-hour)”
- We sum up the Wh values obtained from all Plugwise sensors in the time slot.
- We obtain the hourly power consumption by summing up the power consumed during each time slot of the hour. Eq. 28 and Eq. 29 show mathematically the computation of power consumed during a time slot and during an hour respectively.

$$P_{slot_i} = \frac{\sum_{j=1}^n S_j(Slot_i)}{30} \quad (28)$$

$$P_h = \sum_{i=1}^m P_{slot_i} \quad h \in [1, 24] \quad (29)$$

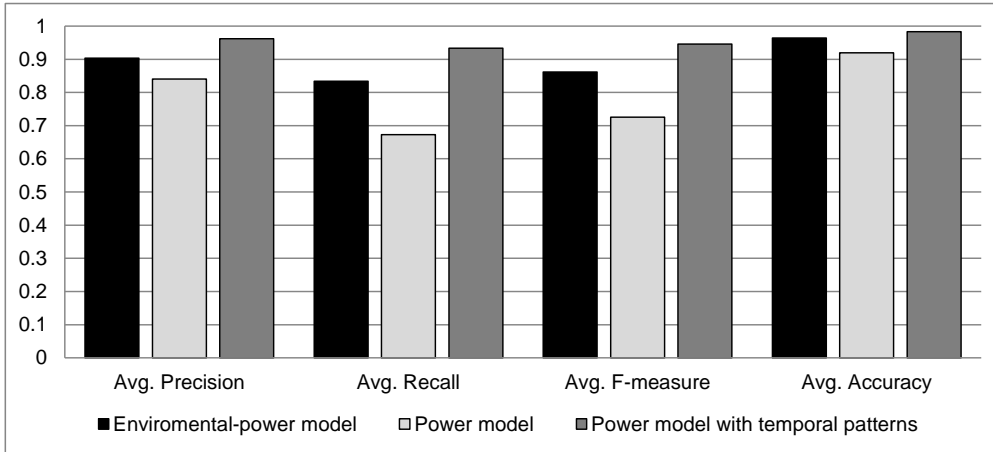


Figure 26: Comparison between the models: power-environmental, power-only and power with temporal patterns in terms of average f-measure, accuracy, precision and recall achieved for both houses

Where:

- $P_{slot_i}$  : power consumed during time slot  $i$
- $S_j(Slot_i)$  : average of sensor  $j$  readings during time slot  $i$
- $n$  : number of sensors
- $P_h$  : power consumed during hour  $h$  where  $h \in [1, 24]$
- $m$  : number of time slots : in hour  $h$

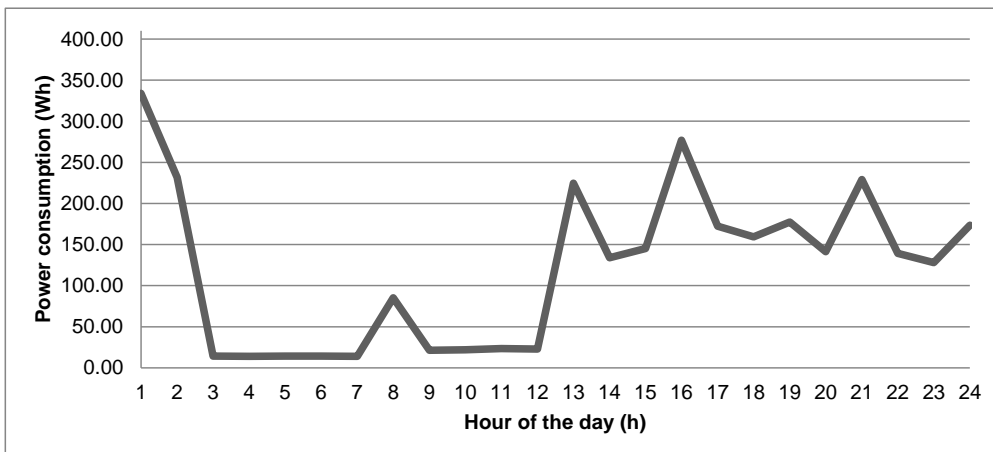


Figure 27: Hourly power consumption of House A on “2013-04-06”

Using these two equations, we compute the hourly power consumption for each house during the whole deployment period. Figure 27 shows a time series representing the hourly power consumed on Saturday “2013-04-06” by House A. We notice from this figure that the user consumes more power from afternoon till midnight. The power consumption after midnight is close to zero as the user is sleeping. Moreover, we notice a moderate consumption around 08:00 am as the user tends to wake

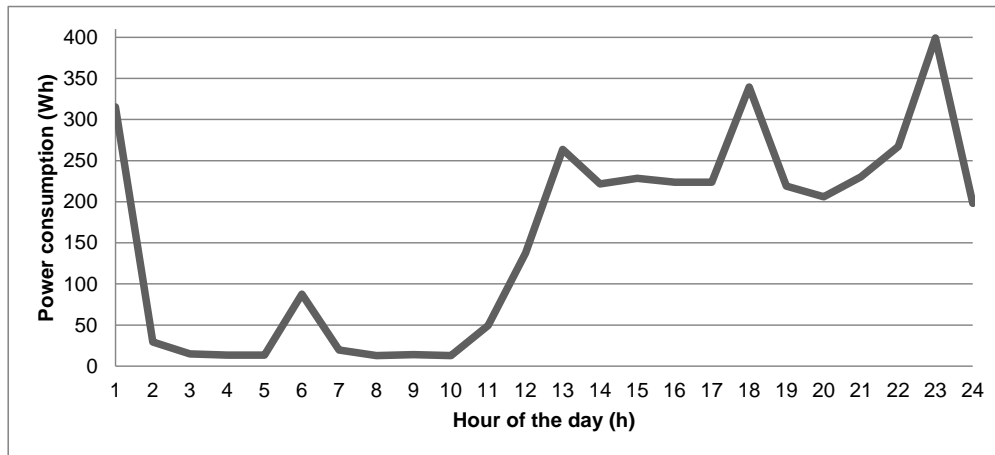


Figure 28: Hourly power consumption of *House A* on "2013-05-21"

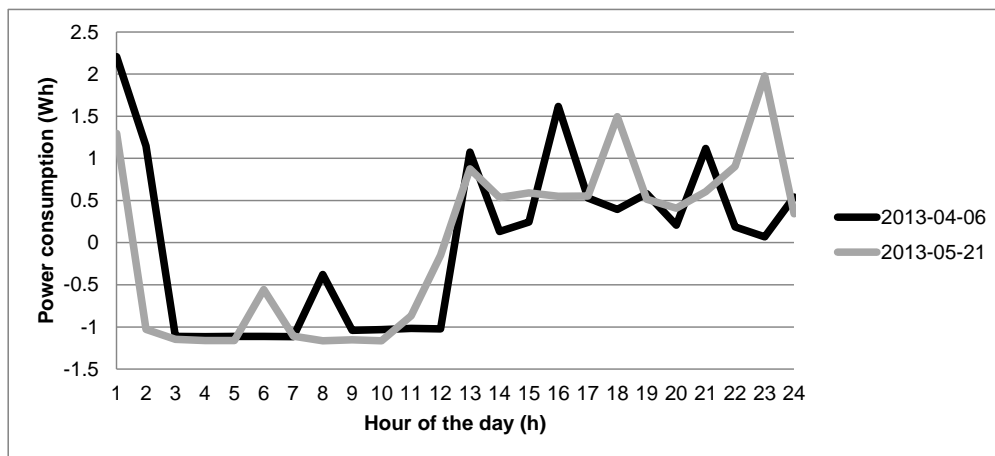


Figure 29: Comparison between consumption behaviors on the days "2013-04-06" and "2013-05-21" for *House A* after standardizing hourly consumption values

up and take a breakfast at this time. The same consumption behavior can be noticed on Tuesday "2013-05-21" as shown in Figure 28 with a time shift of 2 hours as the user starts his day around 06:00 am. This is because the user tends to sleep more during weekends and therefore starts his day two hours later on Saturday. However, he follows the same consumption pattern in which the power is more consumed between afternoon and midnight. Figure 29 clearly shows this similarity. As can be seen in this figure, we standardized hourly consumption values so that each day has a mean of 0 and standard deviation of 1. The standardization process has been performed for all days of both deployments as it is a prerequisite for the similarity analysis presented in the next section.

Figure 30 presents two time series for the normalized consumption values of *House B* on Monday "2013-05-27" and Wednesday "2013-06-05". This figure clearly shows that the user in *House B* as well follows a power consumption pattern which repeats itself every day.

By normalizing and visualizing the 24-hour consumption patterns for certain days, we are able to visually assume the existence of a user-specific consumption pattern which is daily followed by each user. In the next section, we provide a mathematical

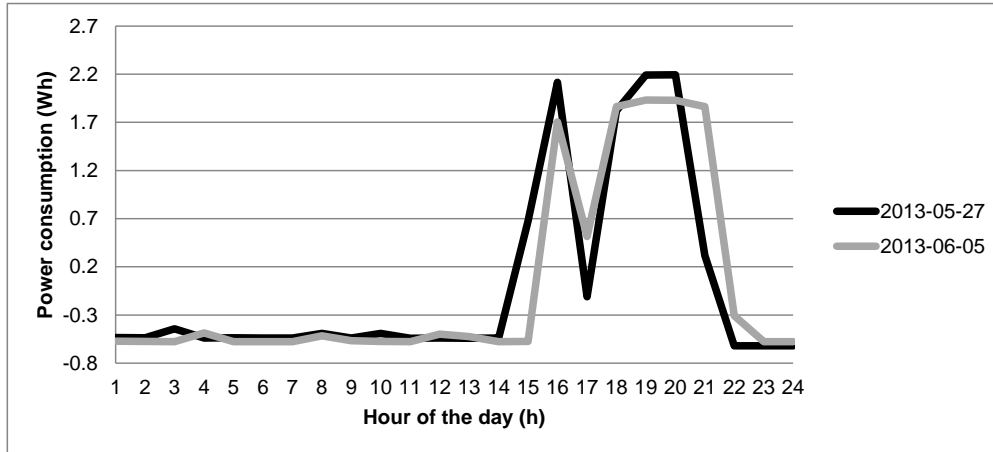


Figure 30: Comparison between consumption behaviors on the days “2013-05-27” and “2013-06-05” for *House B* after standardizing hourly consumption values

proof of the existence of such pattern using the algorithm of Dynamic Time Warping (DTW) [58].

### 5.3.2 Similarity Analysis

In order to decide if a user follows the same power consumption behavior every day, we need to compute a similarity metric between the time series that represent the hourly power consumption during each day. There are several algorithms for computing similarities between time series. Calculating the Euclidean distance represents one of the simplest approaches. Another well-known algorithm is symbolic representation where each time series is represented by a set of symbols [64] [57]. However, these two approaches have the main disadvantage of not being able to handle the problem of time shifting. If two time series are completely identical except the fact that one is shifted in time, these two approaches will consider them as different time series with no similarity between them. This problem can be noticed in Figure 29 where we see that the same consumption pattern is followed in both days but with a 2-hour shift in time. Moreover, both approaches require the time series to be equal in size for a similarity metric to be computed.

The DTW algorithm [58] provides a solution for this problem as it is able to handle shifts in time while computing similarities between time series of equal or different sizes. To compute the similarity between two time series  $A = (a_1, a_2, \dots, a_n)$  and  $B = (b_1, b_2, \dots, b_m)$ , DTW utilizes the following working principle:

- As a first step, DTW computes the Euclidean distance between each element  $a_i$  of  $A$  and all the elements of  $B$ . This results in  $n * m$  distance matrix.
- The main goal of DTW is to find the optimal warping paths between  $A$  and  $B$  where each warping path represents a mapping between  $A$  and  $B$  and has an associated warping cost which is calculated as the sum of all Euclidean distances along it divided by its length. Figure 31 shows a distance matrix with an example warping path. The optimal path between two time series is the one

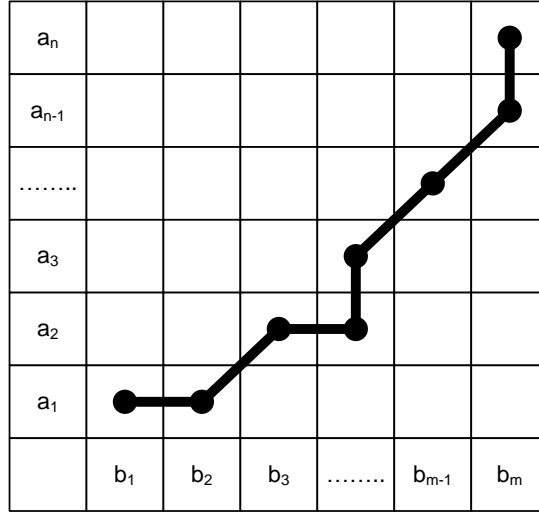


Figure 31: Distance matrix with an example warping path associated with it

associated with the minimum warping cost and is calculated according to the iterative process shown in Eq. 31.

- For a warping path to be valid, it should fulfill the boundary condition where it starts at  $(1,1)$  and ends at  $(n,m)$ , the monotonic condition where it is not allowed to go back in time and the continuity condition where it is allowed to move only one step forward.

$$\text{Optimal Warping Path : } P(x, y) = D(x, y) + \min(P(x-1, y), P(x-1, y-1), P(x, y-1)) \quad (30)$$

$$\text{Warping Score : } WS = \frac{P(n, m)}{k} \quad (31)$$

Where:

- The calculation is initialized with  $P(1, 1) = D(a_1, b_1)$
- $D(x, y)$  : Euclidean distance between  $a_x$  and  $b_y$
- $k$  : length of the optimal warping path

To compute the similarity between two time series  $A$  and  $B$  based on DTW, we find the optimal warping path which aligns  $A$  to  $B$  and then we compute its warping score  $WS$  following Eq. 31. The similarity score between  $A$  and  $B$  can be computed according to Eq 32 [56].

$$\text{Similarity - Score} = \frac{1}{1 + WS} \quad (32)$$

We apply DTW on all 24-hour power consumption time series for *House A* and *House B* respectively where we calculate for each house the minimum, average and maximum similarity scores as shown in Table 10. As we can see from the table, average similarity scores of 90.32% and 93.50% are obtained for the time series of *House A* and *House B* respectively. These results prove that both users follow a power



Table 10: Results of similarity comparison of all 24-hour power consumption time series in each deployment

	<i>House A</i>	<i>House B</i>
Minimum Similarity	81.72%	88.13%
Maximum Similarity	97.13%	97.68%
Average Similarity	90.32%	93.50%

consumption pattern which repeats itself on a daily basis. To further verify these results, we present in the next section a detailed analytical study in which we identify potential outliers in power consumption that can affect our similarity comparison.

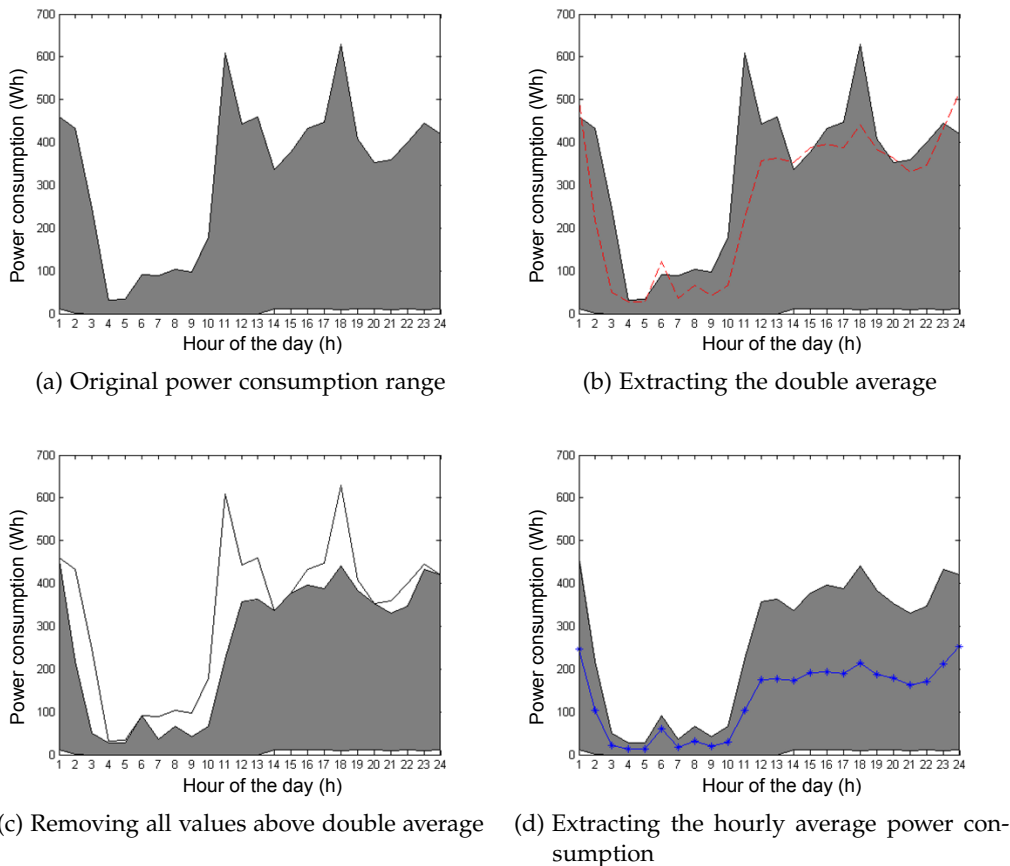


Figure 32: The process of identifying and excluding outliers from hourly power consumption time series of *House A*

### 5.3.3 Further Analysis

This section presents our approach for identifying potential outliers in all 24-hour power consumption time series that can affect our assumption of similarity in the previous section. To achieve that, we extract for each hour the minimum and maximum power consumed during it over the whole deployment period for each of the houses. As a result, we get for each hour the complete range of possible power

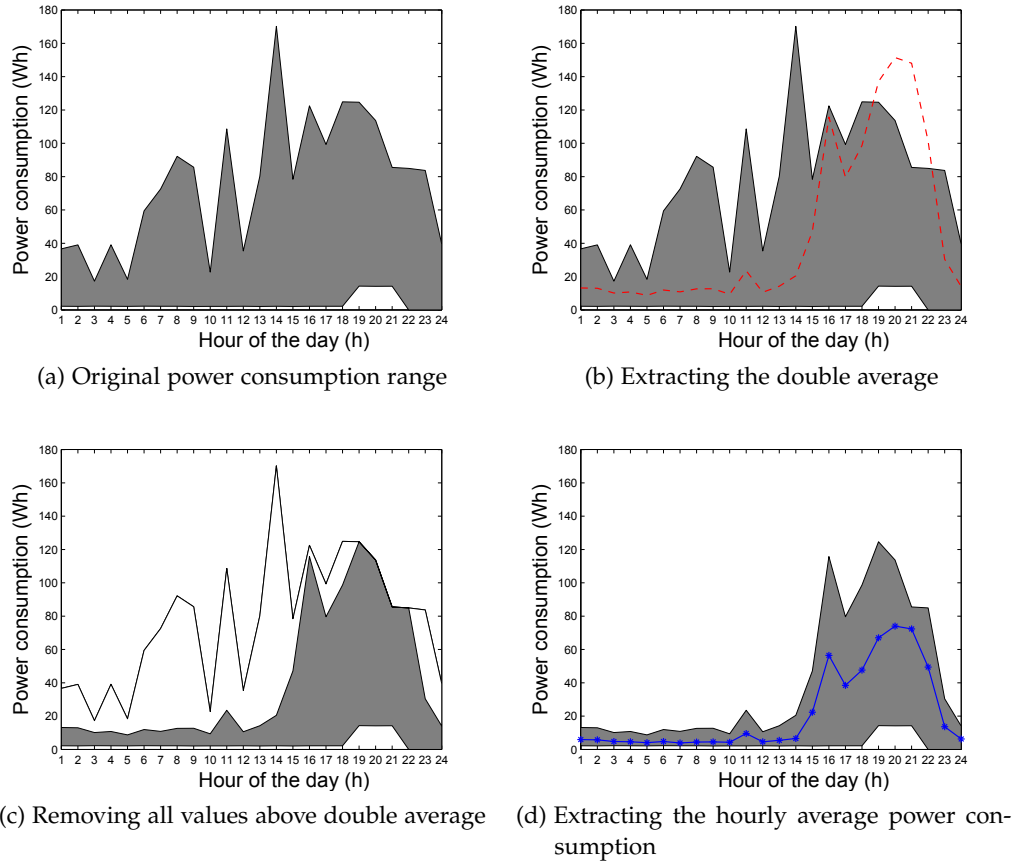


Figure 33: The process of identifying and excluding outliers from hourly power consumption time series of *House B*

consumption values during the whole deployment period as shown in Figures 32a and 33a. We notice in Figure 33a a significant fluctuation at hour 14:00 where the power consumed in *House B* ranges from 0Wh to 160Wh. The same applies for *House A* where we see a fluctuation between 0Wh and 600Wh for the hour 10:00. Such fluctuations can happen due to abnormal power consumption values in certain days. To prove the existence of such outliers, we compute the double average of hourly power consumption for both houses. The result is shown in Figures 32b and 33b where the red dashed line represents the values of double average. For identifying potential outliers, we consider these values of double average as empirical thresholds. Any value of hourly power consumption that exceeds its respective threshold is considered as an outlier and excluded. The results of this exclusion process are shown in Figures 32c and 33c. Both figures show that most of the peaks are caused by the existence of outliers that have been excluded based on the values of double average. To further identify the normal power consumption pattern in each of the houses, we compute the hourly average of consumption values after the exclusion of outliers. Figures 32d and 33d show these values as a dotted line. Both average lines indicate power consumption behaviors that are characterized by low consumption values during night and morning. The consumption continues to be low until 14:00 o'clock for *House B*. Starting from the afternoon, we notice a significant increase in power consumption for both houses. This is because the users in both houses tend

to stay at home and watch TV starting from this time as indicated by their feedback. Moreover, they perform other types of activities such as cooking, eating and working at PC during this period. The power consumption drops again to low values during the night where both users are sleeping.

#### 5.4 SUMMARY

In this chapter, we introduced three novel approaches for identifying and predicting several aspects of users' behavior in indoor environments. We started in Section 5.1 by presenting our novel approach for indoor localization in smart environments solely based on fine-granular measurements of power consumption. We evaluated this approach in Section 5.1.2. The evaluation results showed that we are able to recognize indoor locations of users with f-measure values ranging from 81% to 96%. In Section 5.2.1, we introduced our approach of identifying behavioral patterns of users in indoor environments. We introduced our concept for incorporating such patterns in the building process of activity recognition models so that we enhance their predictive performance. Evaluation results in Section 5.2.2 showed that incorporating these patterns into the feature space led to a clear enhancement in the predictive performance of our activity recognition models. In Section 5.3, we explained our approach for the identification of daily power consumption patterns that are followed by users in their everyday lives. In Sections 5.3.2 and 5.3.3, we mathematically proved the existence of such patterns and identified them for the dataset collected by *SMARTENERGY.KOM* platform.



---

## ACTIVITY RECOGNITION IN MULTI-USER ENVIRONMENTS

---

As any activity recognition model is supposed to work in real-world scenarios, it must take into consideration the following facts:

- people live in multi-user environments where more than one user perform similar or different activities at the same time.
- A person tends to perform more than one activity at the same time which leads to parallel and overlapping activities.
- In a multi-user environment, it is of a great importance for the system to identify the person who is performing the activity.

In Chapter 4, we introduced *SMARTENERGY.KOM*, our platform for activity recognition in single-user environments. Although *SMARTENERGY.KOM* has shown a very good predictive performance in recognizing user activities, it was restricted to single-user environments where the user is allowed to perform only one activity at a given time. As these restrictions do not take into consideration the facts mentioned above, we developed *ML-SMARTENERGY.KOM*, our platform for activity recognition in multi-user environments. It represents an improvement of *SMARTENERGY.KOM* as it is able to recognize parallel and overlapping activities for a group of users while identifying which user is performing which activity.

In this chapter, we focus on three main aspects. Firstly, we show that human activities in multi-user environments can be accurately recognized solely based on the measurements of fine-granular appliance-level power consumption in the respective environment. Secondly, we prove the existence of temporal relations between subsequent activities and we show that exploiting such relations can notably improve the predictive performance of activity recognition models in multi-user environments. Thirdly, we introduce the concept of dependency between concurrent activities. We show that observed activities in multi-user environments exhibit dependency characteristics. We identify and exploit these characteristics to enhance the predictive performance of our activity recognition model. The research contributions presented in this chapter were the main topics of our paper in [3].

### 6.1 ML-SMARTENERGY.KOM ACTIVITY RECOGNITION PLATFORM

Building upon our platform for activity recognition in single-user environments, we designed *ML-SMARTENERGY.KOM* by following the same research goal of extracting user context based on her/his fine-granular power consumption data. Therefore, we followed the same hardware architecture we used for *SMARTENERGY.KOM* as it has been shown in Figure 3 in Chapter 4. The core parts of the platform are power sensor nodes which collect the fine-granular power consumption of every electrical

appliance. Moreover, environmental sensors have been used as well to monitor temperature, brightness, and movement in the environment. A Raspberry Pi module has been used to collect all the readings sent by the sensors and send them to the central database server where the data is stored and analyzed to build the required machine learning model. Table 11 shows the list of power devices which have been connected to power sensors.

Lamp	Monitor
User 1's PC	User 2's PC
Oven	Stove
Vacuum cleaner	Sound system
TV	Water heater

Table 11: List of power devices which have been connected to power sensors

In order to evaluate the predictive performance of *ML-SMARTENERGY.KOM*, we deployed it in a studio apartment shared by two students. The layout of the apartment is shown in Figure 34. We designed the list of activities to be monitored after a discussion with the students in which they told us about their daily behavior and the activities they perform during their normal day. Table 12 shows the list of monitored activities. This list contains 11 different activities besides the *Ignore* activity which we have included to preserve the users' privacy and to give them a chance to perform activities which they do not want to be monitored. The platform has been deployed for a period of 23 days. By the end of the deployment, we were able to collect a dataset of 335000 sensor readings combined with 677 user feedbacks. In section 6.2, we explain the process with which we extract the features required for building the activity recognition model.

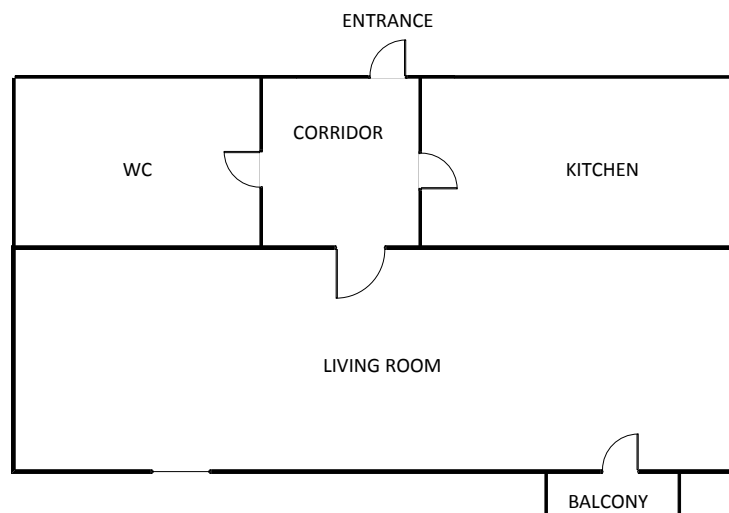


Figure 34: Layout of the apartment

<i>Eating</i>	<i>Reading</i>
<i>WorkingAtPc</i>	<i>WatchingMovie</i>
<i>WatchingTV</i>	<i>Cooking</i>
<i>Cleaning</i>	<i>ListenToMusic</i>
<i>Sleeping</i>	<i>OutOfHome</i>
<i>MakingTea</i>	<i>Ignore</i>

Table 12: List of monitored activities

## 6.2 DATA PREPROCESSING AND FEATURE EXTRACTION

Extracting the correct features from the raw collected dataset plays a crucial role in building a predictive model which can generalize beyond the training data. In this section, we follow the same windowing technique we presented in Section 4.4. As a time slot value, we choose a duration of 2 minutes as it has proven to achieve the best predictive performance in *SMARTENERGY.KOM* as shown in Section 4.5.1. After dividing the dataset into time slots of 2 minutes, we choose the maximum value of each sensor in a time slot to be the feature representing this sensor in this time slot as shown in Eq. 33. The reason behind using the maximum value as a feature has been clarified in Section 4.4.

Activity Recognition in multi-user environments represents an example of multi-label classification (MLC) problems as more than one activity i.e. label can belong to the same feature vector i.e. instance. This is because two users are performing activities at the same time where each user can perform more than one activity at a time. Therefore, each feature vector will have a label vector representing the labels associated with it. Eq. 34 shows the mathematical representation of the label vector. It is represented as a binary vector with “L” dimensions where “L” is the number of distinctive labels in the dataset. Each label takes the value “1” if it associated with the respective feature vector and “0” otherwise. “L” is equal to  $2 * 11 = 22$  in our dataset as each user can perform 11 different activities.

$$\text{Feature vector : } F(t) = \langle mS_1(t), mS_2(t), \dots, mS_N(t) \rangle \quad (33)$$

$$\text{Label vector : } Y(t) = \langle y_1(t), y_2(t), \dots, y_L(t) \rangle \quad (34)$$

Where:

- $mS_i(t)$  : maximum value of sensor  $s_i$  during timeslot  $t$
- $y_i(t) \in \{0, 1\}$  : value of label  $y_i$  during timeslot  $t$
- $N$  : number of features
- $L$  : number of labels

Table 13 shows the multi-label characteristics of our dataset. As explained in Section 2.1.2.3, label cardinality, label density, PMax, and PDL give a deep insight into the multi-label nature of a certain dataset. The table shows these characteristics for the label set of each user alone as well as for the whole label set combining the labels of both users. As we can see from the table, we get a label cardinality of around 2.3 by combining the label sets of both users. A label cardinality of more than 2 indicates that in most of the instances we get a label vector with more than one label.

Dataset	Label Cardinality	Label Density	PMax	PDL
<i>User 1</i>	1.21816	0.101513	0.31014	0.00199074
<i>User 2</i>	1.14584	0.0954868	0.335700	0.001584
Both users	2.33553	0.0973138	0.324116	0.0095757

Table 13: Multi-label characteristics of the dataset

### 6.3 BINARY RELEVANCE MODEL OF ACTIVITY RECOGNITION

As explained in the previous section, activity recognition in multi-user environments represents an instance of MLC problem. In Section 2, we introduced several approaches that handle MLC problems. One of these approaches is binary relevance in which the multi-label dataset is divided into  $L$  single-label datasets so that each label has its own dataset. To handle our activity recognition problem, we decide as a first step to follow the approach of binary relevance as shown in Figure 35. Therefore, we build a model for each user activity so that we have 22 models. As seen in Figure 35, we use the algorithm of conditional inference trees (ctree) to build the individual prediction models. We have chosen ctree algorithm due to two reasons. On one hand, ctree follows a rigorous statistical framework in which the predictors are selected based on their statistical correlation with the response variable. On the other hand, ctree is suitable for our problem as for each activity the user performs, only a subset of sensors are related while other sensors have no effect on it.

For each label, we build five different ctree models based on the selected predictor variables. As explained in Section 6.2, the maximum value of each sensor in a time slot is used as a predictor variable. Moreover, we utilize the time during which an activity happens as an extra predictor variable. This is due to the fact that human beings follow a daily pattern in which they repeat their activities on a daily basis where each activity approximately occurs during the same time period every day. Figure 36 shows the five individual ctree models we build for each label. As we can see in the figure, we use the label for which we build the model as the response variable of the ctree. As stated before in addition to using maximum sensor readings as features, we use time-related features. The five models resulting from the different combinations of features are as follows:

- The “WithoutTime-Model” which uses only sensor readings as features with no time-related features.
- The “Hour-Model” which uses the hour at which the activity has occurred as an extra feature.



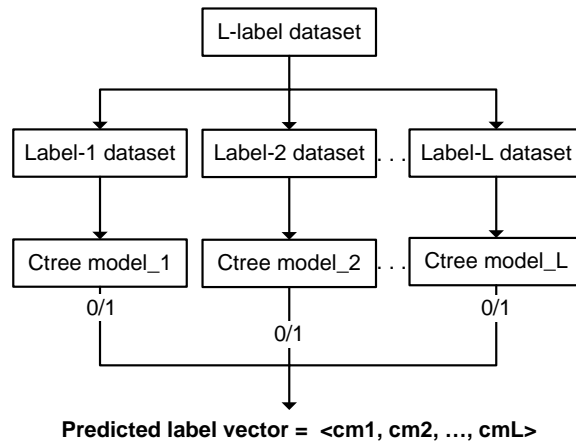


Figure 35: Binary relevance model

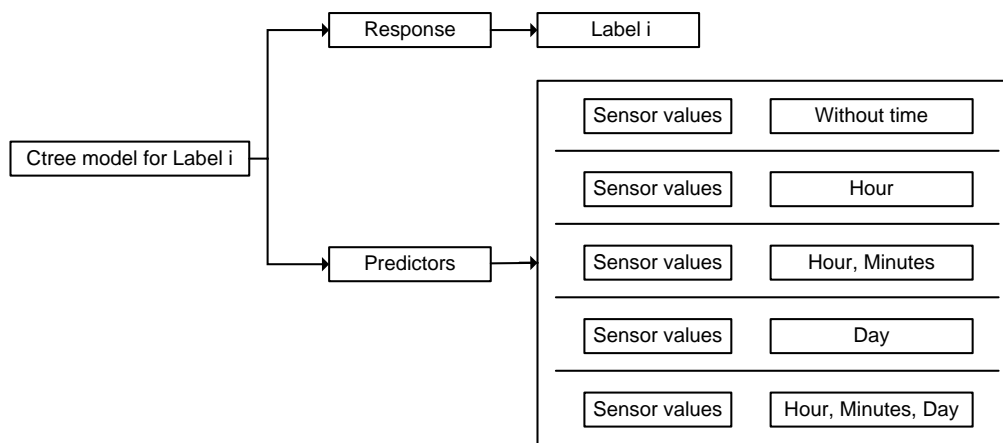


Figure 36: Individual ctree model

- The “Hour+Minute-Model” which uses the hour and minute at which the activity has occurred as extra features.
- The “Day-Model” which uses the day at which the activity has occurred as an extra feature.
- The “Hour+Minute+Day-Model” which uses the hour, minute, and day at which the activity has occurred as extra features.

Figure 37 shows the ctree model for *Reading* activity of *User 2*. This tree has been built using the “Hour-Model”. As we can see in the figure, the ctree was able to isolate the positive from the negative examples based on the Watt readings of lamp power sensor and the kWh readings of the sensors connected to sound system and PC of *User 1*. It is clear from this ctree that no reading activity happens when no power is consumed by the lamp i.e. it is switched off. Another ctree model is shown in Figure 38 which represents the *Cooking* activity of *User 1*. It has been generated using the “Hour-Model” as well.

The models built for each label have been evaluated using the collected dataset to measure their predictive performance. In the next section, we present an exhaustive evaluation study in which we compute a set of predictive metrics to evaluate the overall performance of our approach as well as to select the combination of predictors which achieve the best predictive performance.

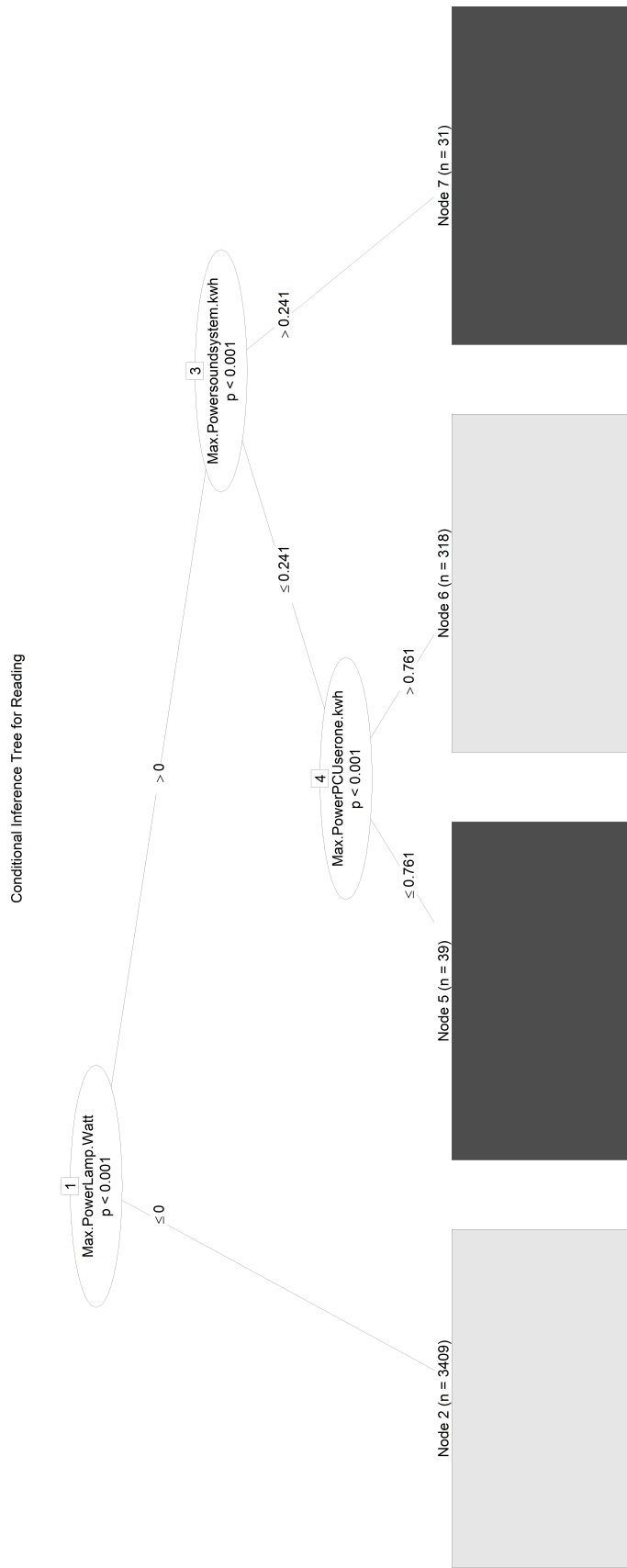


Figure 37: User 2: Ctree model of the activity *Reading*. Only hour has been used as a time predictor.

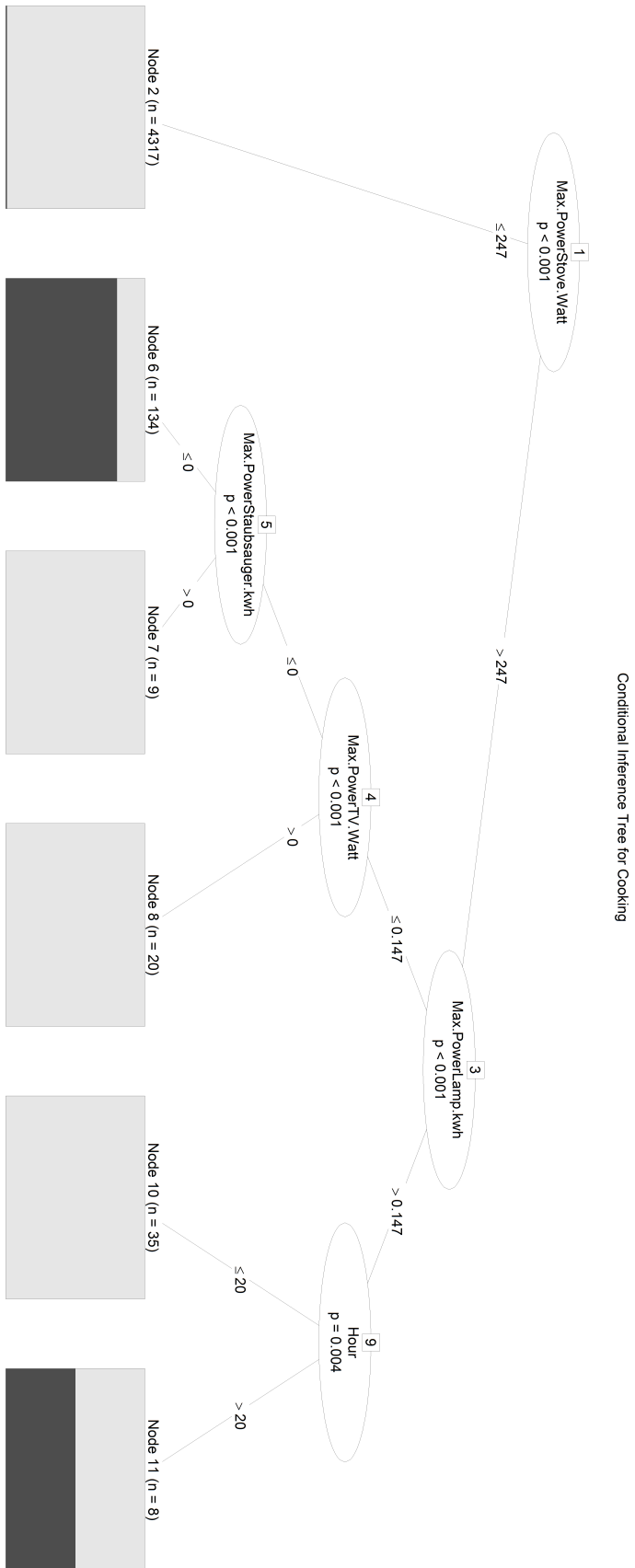


Figure 38: *User 1*: Ctree model of the activity *Cooking*. Only hour has been used as a time predictor.

### 6.3.1 Evaluation

In this section, we evaluate the predictive performance of our binary relevance approach. As explained in the previous section, each label will be modeled using five different models. Our research goal is to recognize human activities in multi-user environments based on the fine-granular power consumption values of the environment. Two types of sensors have been used in our deployment, namely power and environmental sensors. In the next two sections, we present our evaluation results when using power sensors alone as features and when using them in combination with environmental sensors. The goal of this evaluation strategy is to show that power sensors alone are enough for achieving high predictive performance.

In order to start our evaluation, we divide the dataset into training and testing sets by taking the first two weeks of collection as a training set and the last week as a testing set. The evaluation study is structured as follows. Section 6.3.1.1 evaluates the predictive performance of binary relevance approach when using environmental and power sensors as features. The effect of temporal relations between activities on the performance of our activity recognition models is studied in Section 6.3.1.2. Section 6.3.1.3 evaluates the effect of excluding environmental sensors from feature vectors on the predictive performance of our models. It proves that fine-granular power measurements are enough to build a power prediction model that is able to recognize human activities.

#### 6.3.1.1 Power and Environmental Sensors

In this evaluation setting, we consider the feature vector of each time slot to contain the maximum readings of all sensors including the environmental sensors. As stated before, we use the ctree algorithm as the classifier in building binary relevance models. We build five different ctree models for each label using the training set. Each of these models is tested using the testing set and four performance metrics have been calculated, namely recall, precision, f-measure and accuracy. Figures 39, 40 show the values of f-measure and accuracy for each label with regard to *User 1*. The same is shown for *User 2* in Figures 41, 42.

As we can see from the figures, binary relevance models were able to recognize the activities of both users with different predictive performance. No model was able to predict the activity of *MakingTea* for both users. This is due to the fact that the number of instances for this activity is very low in comparison with other activities. Therefore, it is not possible for the machine learning algorithm to build a model that can recognize it. The low number of instances has also affected the activity of *Eating* which has been predicted with low predictive performance.

Figure 43 summarizes the average values of recall, precision, f-measure and accuracy for all models with respect to all activities of both users. As we can see from this figure, "Hour" model has achieved the best values for recall, f-measure and accuracy. This can be explained by the fact that humans tend to follow a daily pattern in which they perform their activities in the same order within the same time period every day. "Hour+Minute" model has achieved the same results as the "Hour" model which means that adding minutes to the hour as a timestamp does not improve the predictive performance. Therefore, we decide to use the hour as the only feature that represents the timestamp of an activity in the following analysis.

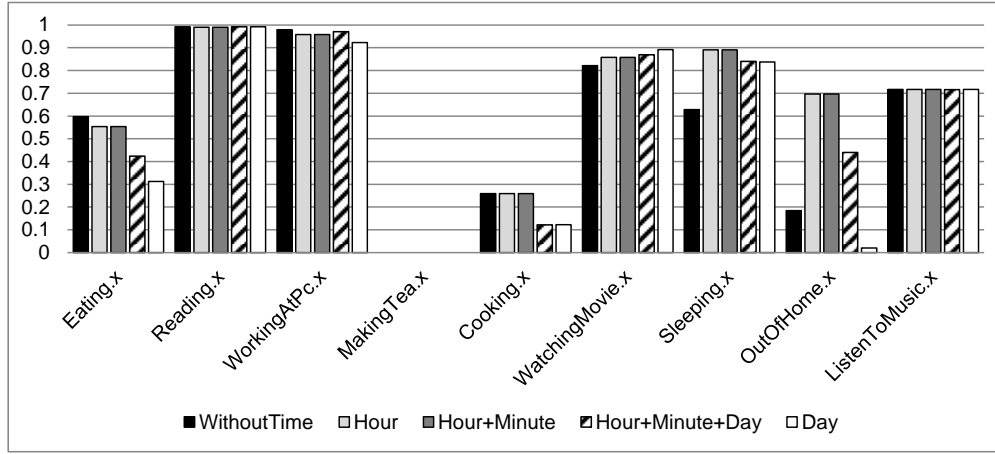


Figure 39: F-measure values of environmental-power binary relevance models with regard to *User 1*

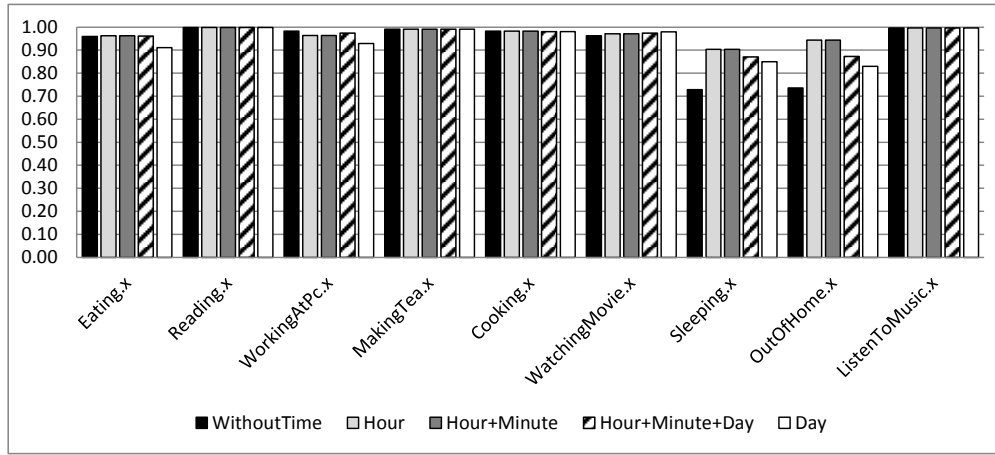


Figure 40: Accuracy values of environmental-power binary relevance models with regard to *User 1*

### 6.3.1.2 Temporal Relations Between Activities

As users tend to follow a daily routine in performing their everyday activities, providing any information about previous activities performed by the users to activity recognition models has a potential in improving their predictive performance. In this section, we analyze and evaluate this potential. We extend feature vectors previously explained in Section 6.2 so that they contain the activities performed by the user in the previous timeslot as extra features. Eq. 35 shows the new mathematical representation of feature vectors where  $Act(t-1)$  represents the previous user's activities performed in time slot  $t-1$ .

$$\text{Feature vector : } F(t) = \langle mS_1(t), mS_2(t), \dots, mS_N(t), Act(t-1) \rangle \quad (35)$$

To study the effect of temporal relations on activity recognition models, we conduct an evaluation study in which we re-evaluate the predictive performance of binary relevance models with the new feature vectors containing previous activities

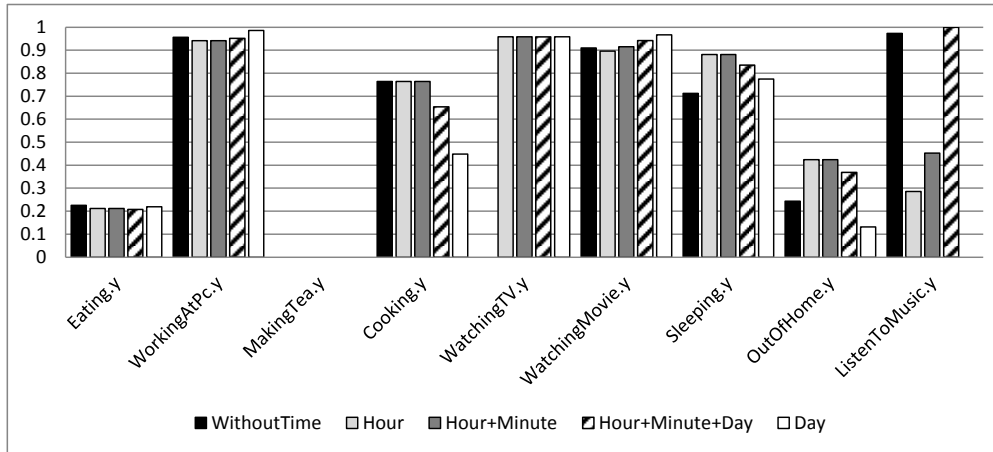


Figure 41: F-measure values of environmental-power binary relevance models with regard to *User 2*

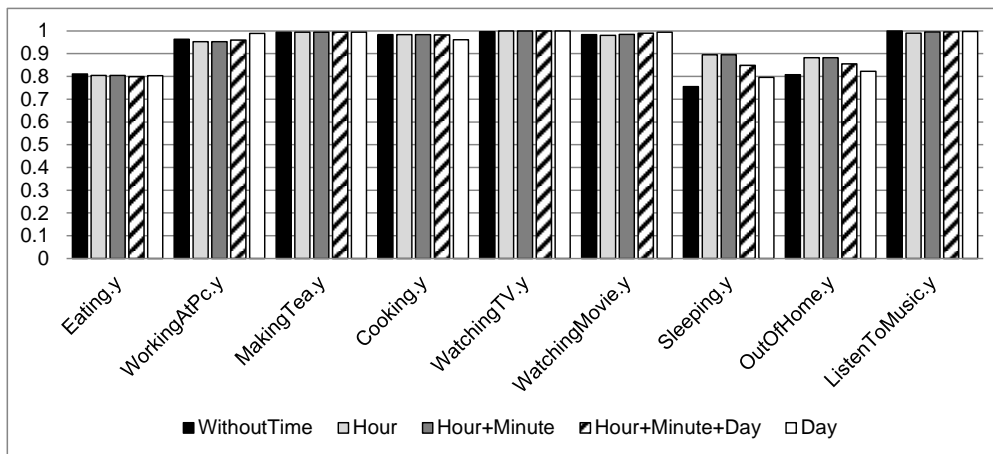


Figure 42: Accuracy values of environmental-power binary relevance models with regard to *User 2*

as features. We follow the same previously used evaluation setting in which we take the first two weeks of data as a training set and the third week as a testing set.

Figure 44 shows the improvement in f-measure values achieved in recognizing each of *User 1*'s activities after adding previous activities as features. As we can see in the figure, temporal relations have improved the predictive performance in recognizing all activities especially the ones with low f-measure values. We see an improvement of about 34% and 37% for the activities *Cooking* and *Eating* respectively. However, it is still not possible for the models to recognize any instance of the activity *MakingTea*. We present a solution to this problem in Section 6.4 by identifying intra-user and inter-user dependencies between activities. The same applies for *User 2* as shown in Figure 45. We see an improvement of f-measure values for all activities apart from the activity of *MakingTea* for which no instance has been predicted. This improvement has reached the values of 60% and 56% for the activities of *ListenToMusic* and *Eating* respectively.

Figure 46 compares the average overall performance of temporal and non-temporal models in terms of accuracy, f-measure, recall and precision. The average value of

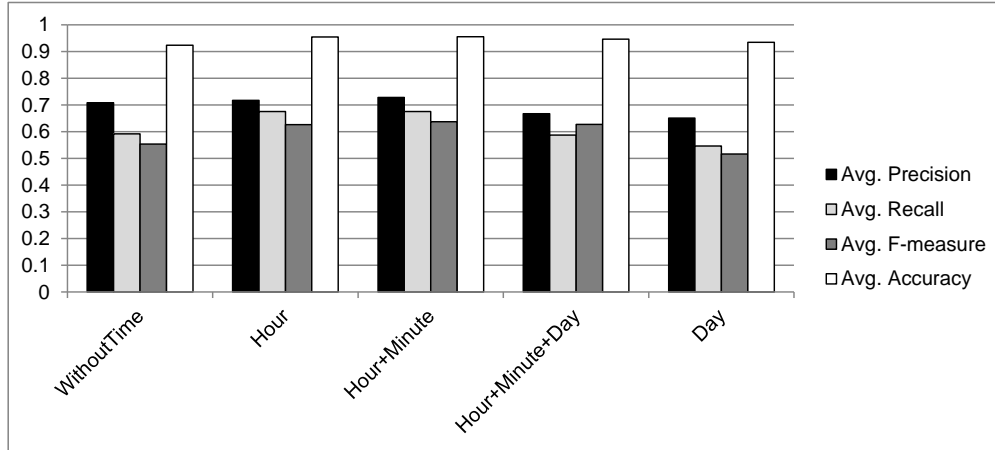
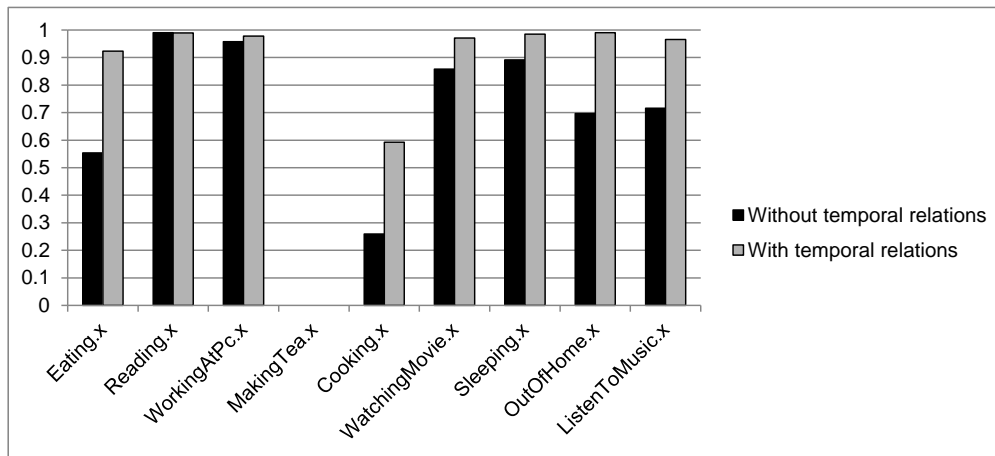


Figure 43: Comparison between the five time models

Figure 44: The improvement in f-measure values achieved by encoding temporal relations as extra features for the activities of *User 1*

each measure shown in the figure is computed with respect to all activities of both users. As we can see from the figure, an improvement in all performance measures can be seen. These results prove the existence of a temporal pattern in which users perform their daily activities. Moreover, it shows that providing such pattern into the model of activity recognition leads to a remarkable improvement in its predictive performance. In the following evaluation studies, previous activity will always be part of feature vectors as clarified in Eq. 35.

### 6.3.1.3 Power Sensors

In this section, we evaluate the predictive performance of binary relevance models by only using fine-granular measurements of power consumption as information for recognizing users' current activities. Therefore, we exclude the values of environmental and motion sensors from feature vectors and repeat the previous experiments. We keep temporal patterns as part of the feature vectors as we stated before. Our goal is to prove that fine-granular power consumption data provides enough information for an accurate recognition of users' activities in multi-user environments.



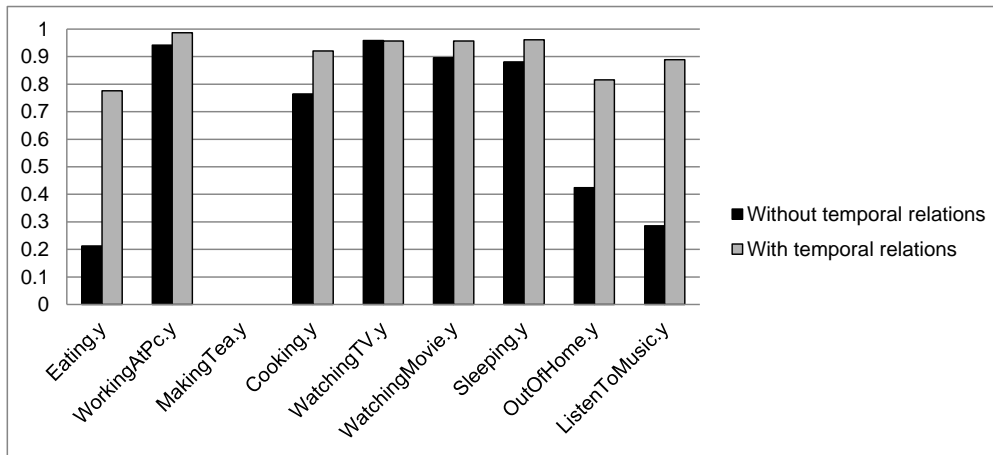


Figure 45: The improvement in f-measure values achieved by encoding temporal relations as extra features for the activities of *User 2*

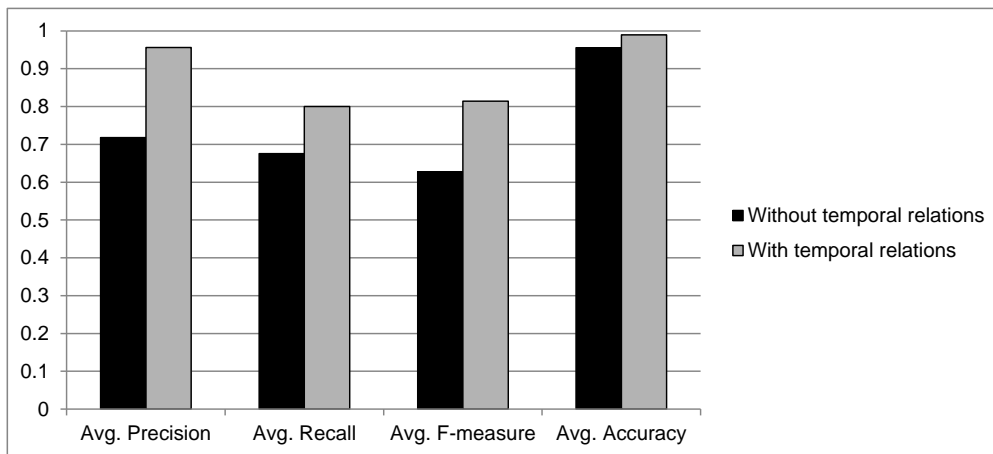


Figure 46: Comparison between temporal and non-temporal models in terms of average f-measure, accuracy, precision and recall

Figure 47 compares the predictive performance of power-environmental and power-only models in terms of f-measure values for the activities of *User 1*. As we can see from this figure, the exclusion of environmental and motion sensors has not much affected the predictive performance in recognizing all activities except the activity of *Reading* which has seen a decrease of about 40% in f-measure value. The reason of this decrease can be explained by the fact that the absence of motion sensors has caused confusion between the activities of *Reading* and *Sleeping* as both of them take place in the sleeping room where the user tends to read a bit before falling asleep.

Figure 48 shows the previous comparison for the activities of *User 2*. In this figure, we can notice that the exclusion of environmental and motion sensors has not much affected the f-measure values. It has even improved the predictive performance in recognizing activities *OutOfHome* and *Eating*.

An overall comparison between power-only and power-environmental models in terms of average accuracy, f-measure, recall and precision is shown in Figure 49. We have computed the average values of all measures with respect to the activities of both users. This figure shows that the exclusion of environmental and motion sensors

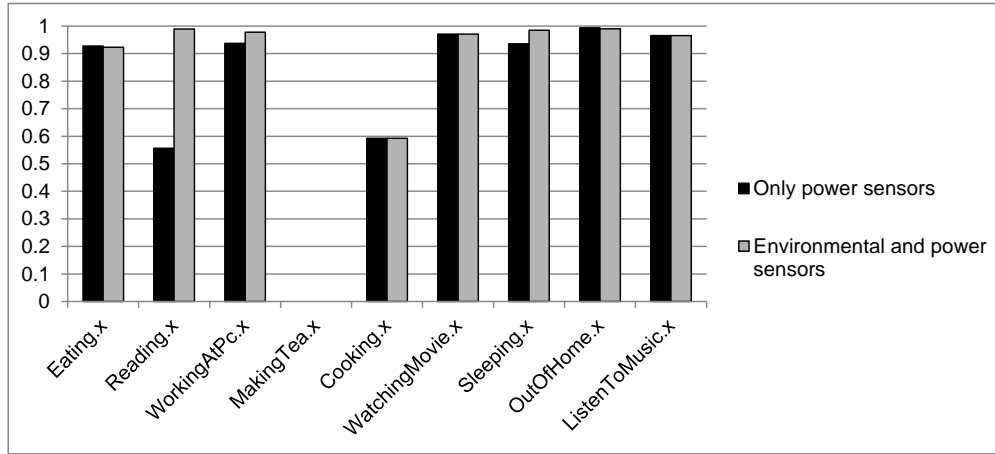


Figure 47: Comparison between power-only and power-environmental activity recognition models in terms of f-measure values for activities of *User 1*

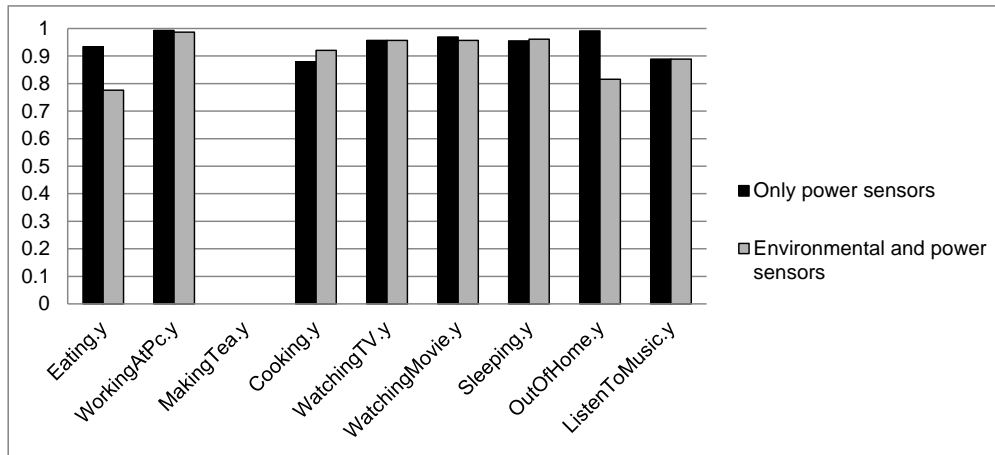


Figure 48: Comparison between power-only and power-environmental activity recognition models in terms of f-measure values for activities of *User 2*

has no notable effect on the predictive performance of activity recognition models. It proves the ability of our platform to recognize users' activities in multi-user environments solely based on fine-granular measurements of power consumption without the need for any other sensing modality.

The problem of predicting the activity of *MakingTea* still represents a difficult challenge for all models evaluated so far. We present a solution for this problem in the following section where we develop a multi-label activity recognition model which identifies and utilizes dependency relations between activities for realizing an accurate predictive performance in recognizing users' activities.

#### 6.4 LABEL DEPENDENCY IN ML-SMARTENERGY.KOM DATASET

Most of multi-label datasets have the important feature that the labels are correlated and dependent on each other [34]. The dependency in this context means that the values each label takes are statistically correlated with the values of some other labels

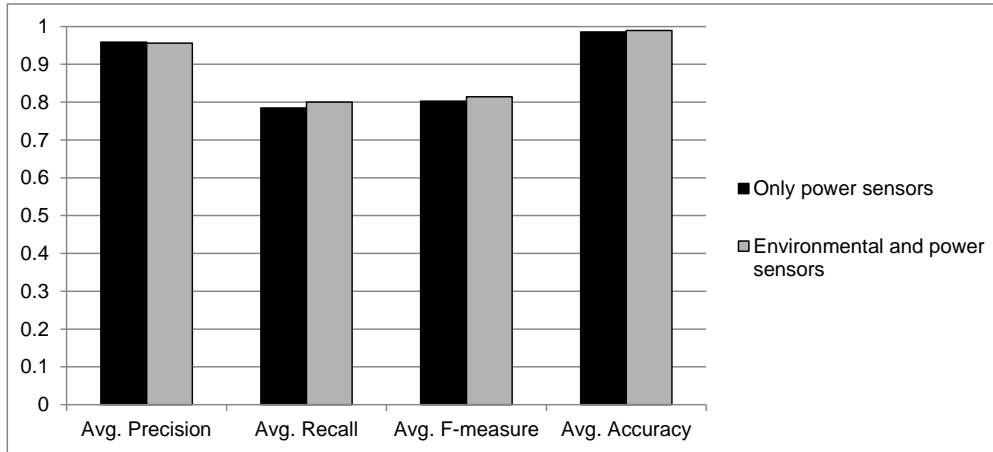


Figure 49: Comparison between power-only and power-environmental activity recognition models in terms of average f-measure, accuracy, precision and recall

in the label set. As a result of this dependency relation, we can predict the value of a certain label with a higher predictive performance if we already know the values of some other labels correlated with it. This leads us to the fact that predicting each label alone in a binary relevance setting will cause a loss of information related to this dependency. Therefore, it is of great importance for any multi-label machine learning model to be able to find this dependency relation and to utilize it in a way that enhances its predictive performance.

In this section, we examine in details the dependency relations between the labels in *ML-SMARTENERGY.KOM* dataset. In order to extract all dependency relations between the labels, we follow a rigorous mathematical framework, namely the ctrees algorithm which mathematically proves the existence of such relations if they exist. As explained in Section 2.1.1.1, the ctrees algorithm examines the null hypothesis of independence between its response variable and all the predictor variables. It studies the statistical associations between each one of the predictors and its response variable. In order to use ctrees algorithm for finding the dependencies between labels, we have to model our problem in a way that one label will be the response of the other labels that represent the predictors in the ctrees. Using this methodology, ctrees algorithm extracts the association between the response label and each of the predictor labels. In order to realize this methodology, we create a new dataset that only contains the label vectors from *ML-SMARTENERGY.KOM* dataset.

Dependency relations can exist between the labels of one user i.e. intra-user dependency as well as between the labels of both users i.e. inter-user dependency. In the next two sections, we present in details these two types of dependency.

#### 6.4.1 Intra-User Label Dependency

Intra-user dependency refers to the associations that exist between the labels of an individual user. An example of this dependency can be a user who performs the activities of *Eating* and *WatchingTV* most of the time simultaneously. Being able to predict one of these two activities with a very good predictive performance will make it easier to predict the other one. In this section, we extract all dependency relations

that are exhibited by the label set of each individual user in our dataset. Figure 50 clarifies the process of extracting dependencies. As mentioned in the previous section, each of the labels in the label set will be taken as a response of a ctree whose predictors are all other labels in the label set.

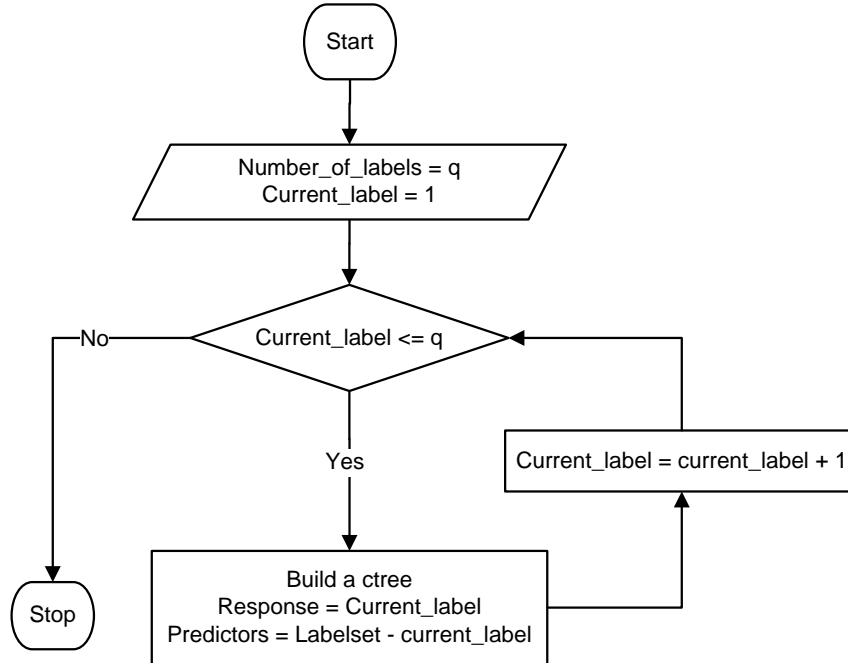


Figure 50: The methodology of extracting label dependency in intra-user scenario

Figure 51 shows an example dependency ctree for *Cooking* activity of *User 1*. The label *Sleeping* represents the root node for this ctree which implies a strong association i.e. dependency between these two activities. This can be explained by the fact that no activity occurs during sleeping which implies a strong negative association between *Sleeping* and any other activity. We notice that by inspecting “Node 21” in the ctree which contains 3002 examples for them labels *Sleeping* and *Cooking* take the values of 1 and 0 respectively.

#### 6.4.2 Inter-User Label Dependency

People who reside in a common place tend to perform their activities in accordance with each other. They follow a common pattern in which they take into consideration the activities of others while performing their own activities. As an example, people who live with each other tend to go to sleep at the same time, to eat dinner and watch TV together. Our goal in this section is to extract such common patterns so that we know which activities are simultaneously performed by both users in our dataset and which are not. We refer to these patterns as inter-user label dependency.

Figure 52 introduces the details of our approach for identifying inter-user label dependency. To identify it for label  $x$  belonging to *User 1*, we build a ctree in which

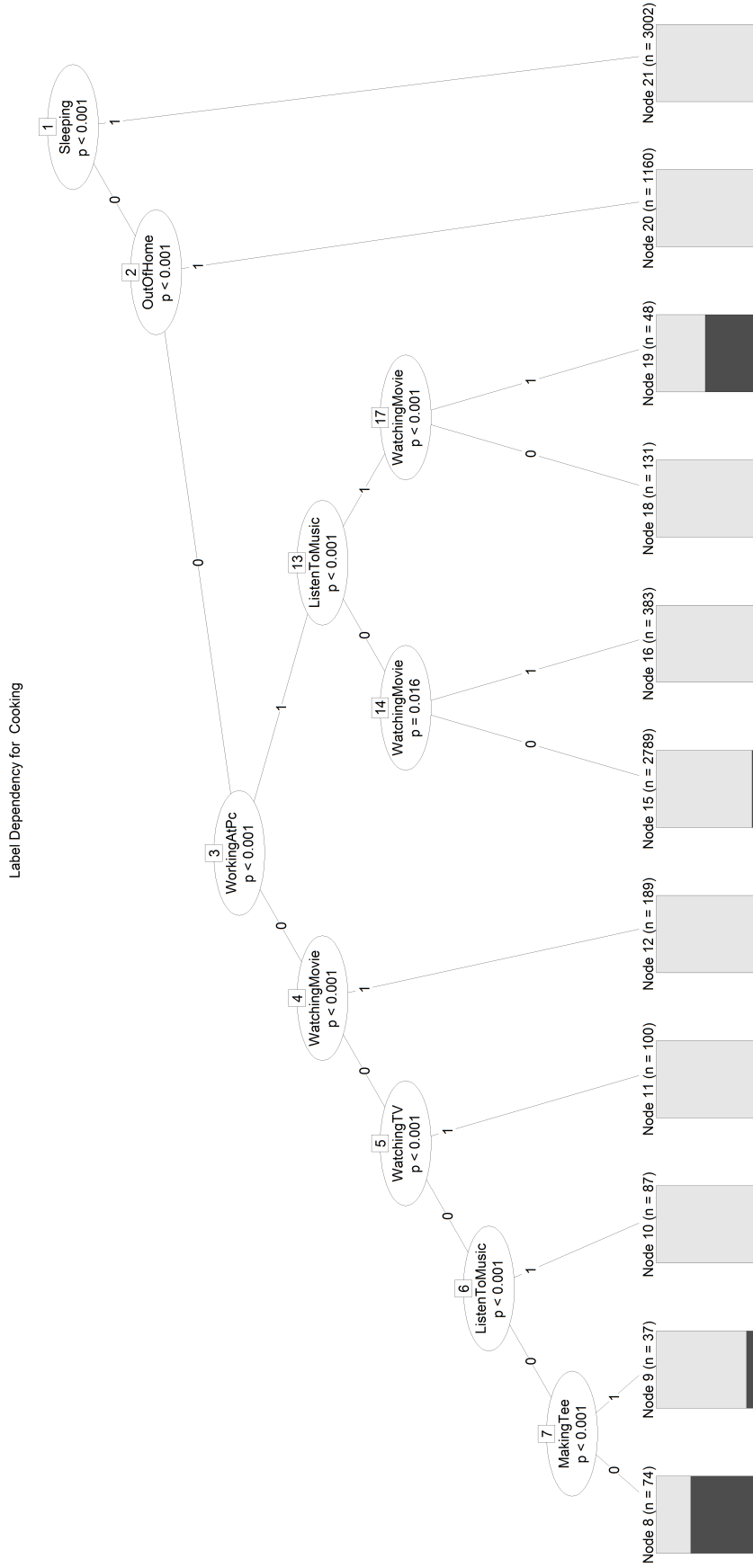


Figure 51: Dependency tree for Cooking activity of User 1 with respect to all his other activities

label  $x$  represents the response variable while the predictor variables are all the labels of *User 2*. The individual steps of this process are clarified in the flowchart shown in Figure 52. An example inter-user dependency tree for the activity *OutOfHome* of *User 1* is shown in Figure 53. The ctree shows that this activity is strongly correlated with *OutOfHome* activity of *User 2* which means that both users tend to be outside during the same time period. Figure 54 presents another example ctree for *Sleeping* activity of *User 2*. It is clear from this ctree that both users tend to go sleep at the same time. Moreover, we notice in the tree a strong negative association between this activity and the activities *WorkingAtPc* and *WatchingTV* of *User 1*.

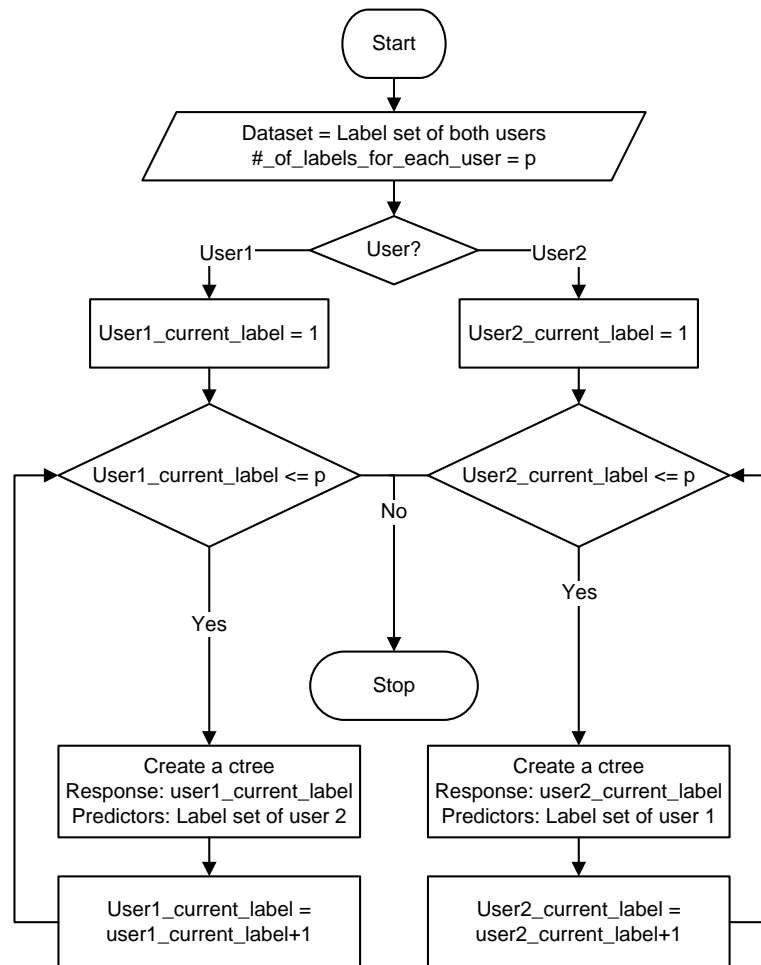


Figure 52: The methodology of extracting label dependency in inter-user scenario

## 6.5 ML-SMARTENERGY.KOM ACTIVITY RECOGNITION MODEL

After the identification of intra- and inter-user dependency relations, we present in this section our approach for utilizing these relations to build a multivariate model for activity recognition in multi-user environments. For each label in our dataset, we

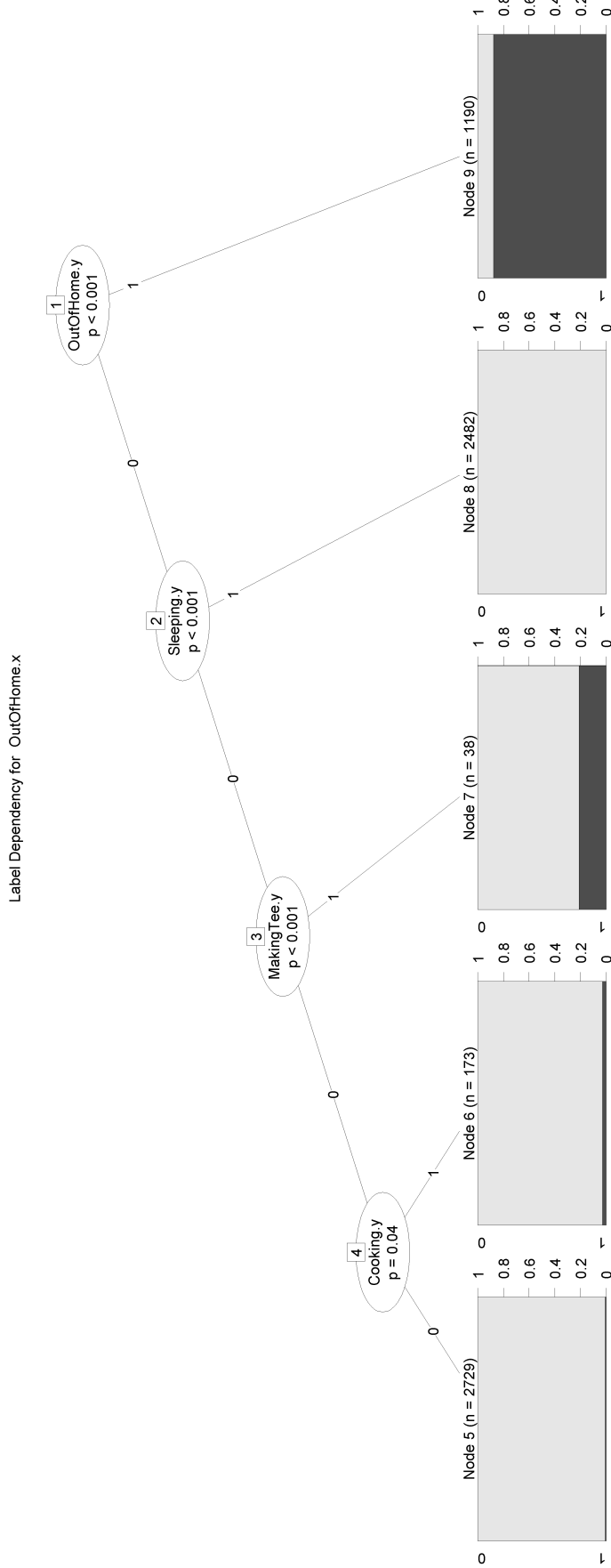


Figure 53: Inter-user dependency tree for *OutOfHome* activity of *User 1* with respect to all activities of *User 2*

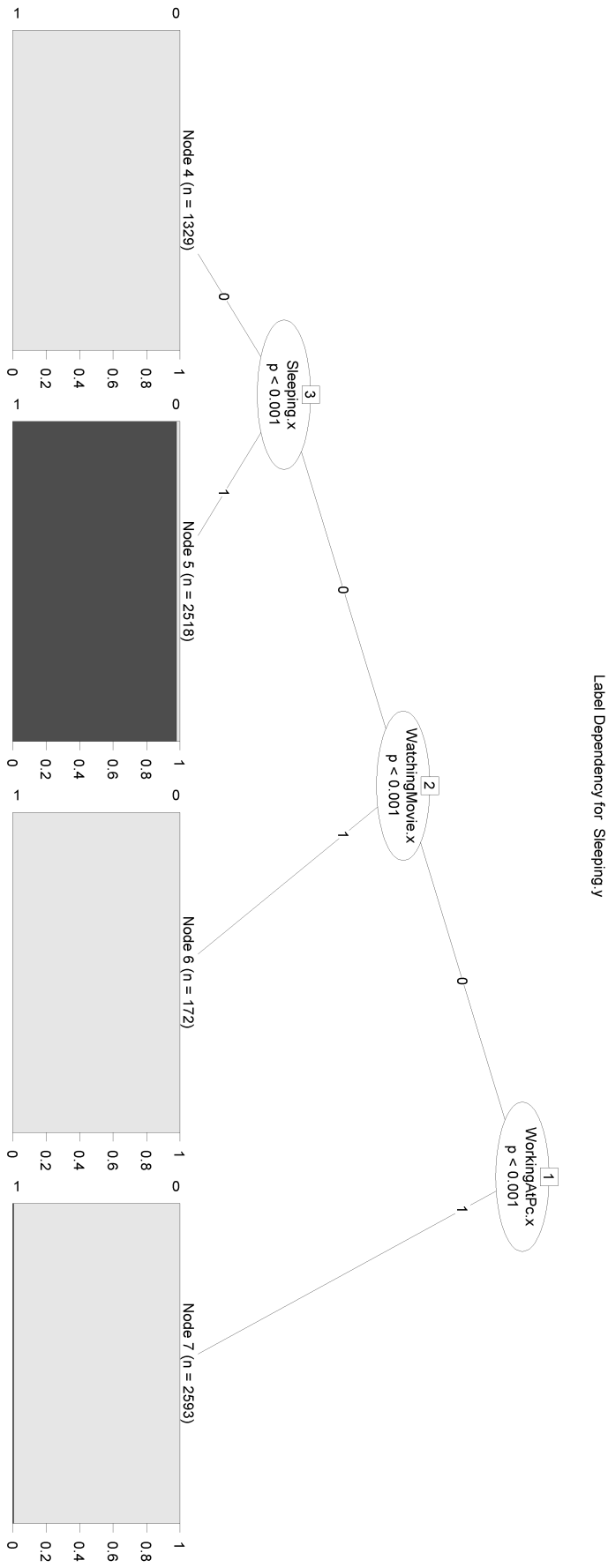


Figure 54: Inter-user dependency tree for Sleeping activity of User 2 with respect to all activities of User 1



build intra- and inter-user dependency ctrees. By interpreting each of these trees, we obtain the following dependency relations:

- Each of the activities *Sleeping*, *Reading*, *WatchingTV*, *Cooking*, *OutOfHome*, *Eating* and *WatchingMovie* is generally performed by both users at the same time.
- Each of the users tends to perform the activity *ListenToMusic* while he is *WorkingAtPc*.
- A notable negative association exists between the activities *OutOfHome*, *Sleeping* and *MakingTea*.

In order to utilize the information encoded in dependency patterns, we build a set of multivariate activity recognition models. Each of these models has a combination of dependent labels as its multivariate response. As the ctrees algorithm accepts multivariate response variables, we choose it as a base classifier. Figure 55 shows the combination of labels for which a ctrees model is built. The resulting *ML-SMARTENERGY.KOM* activity recognition model is an ensemble of multivariate ctrees models.

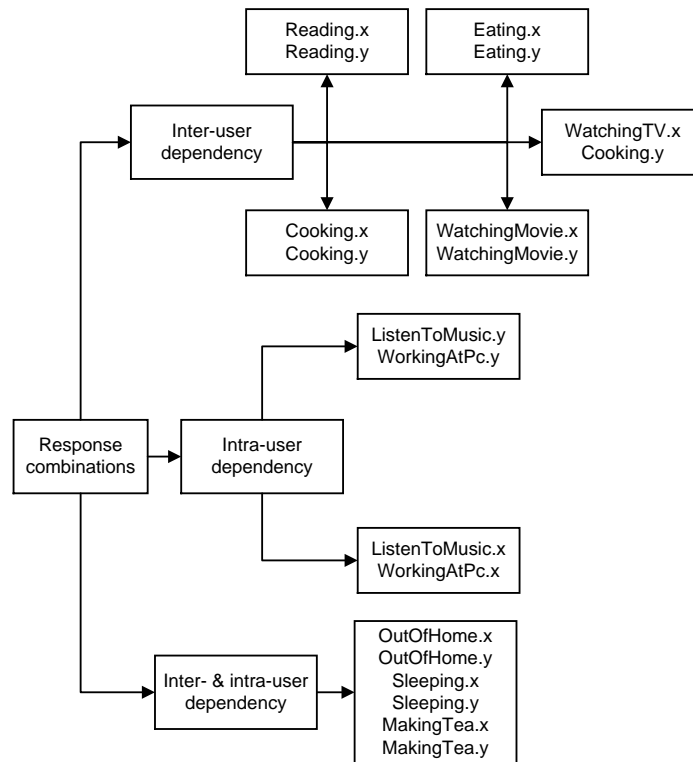


Figure 55: The set of label combinations used to build the multivariate ctrees models

### 6.5.1 Evaluation

In this section, we present the improvement achieved in the predictive performance of activity recognition models by combining correlated labels into a set of multivariate

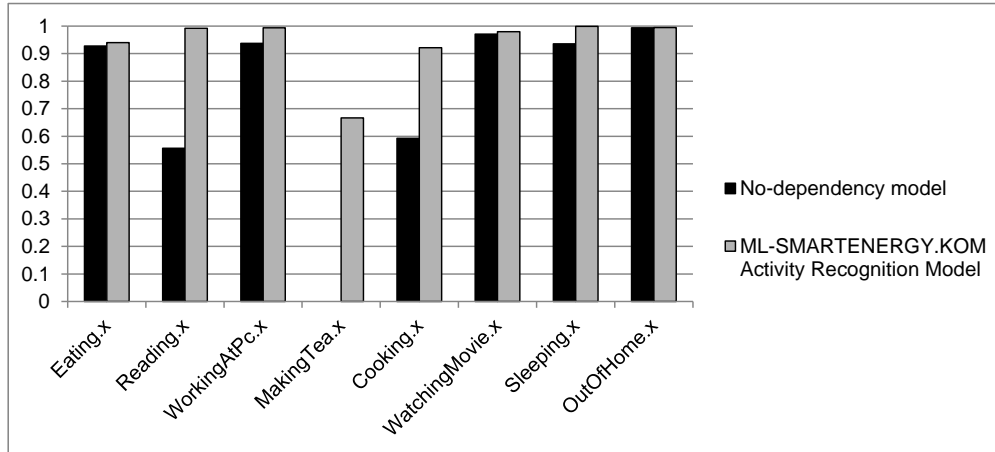


Figure 56: Comparison between *ML-SMARTENERGY.KOM* and no-dependency activity recognition models in terms of f-measure values for activities of *User 1*

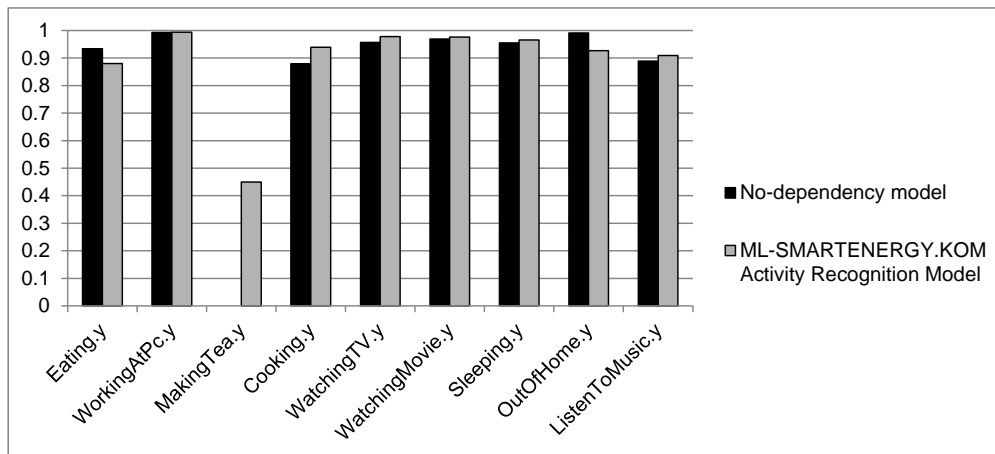


Figure 57: Comparison between *ML-SMARTENERGY.KOM* and no-dependency activity recognition models in terms of f-measure values for activities of *User 2*

ate responses. Based on the results presented in Section 6.5, we showed that activities performed in multi-user environments expose dependency patterns that provide a potential benefit for improving the predictive performance of activity recognition models. We compare the predictive performance of *ML-SMARTENERGY.KOM* model to the performance of binary relevance approach presented in Section 6.3.

Figure 56 compares the values of f-measures achieved using binary relevance approach and *ML-SMARTENERGY.KOM* model in predicting the activities of *User 1*. The figure shows a remarkable improvement in predicting the activities of *Reading* and *Cooking*. Moreover, we notice that *ML-SMARTENERGY.KOM* model is able to predict the activity of *MakingTea* with a predictive performance of 64% in terms of f-measure value. Figure 57 shows that the same applies for *User 2* as our model was able to predict his activity of *MakingTea* as well.

Figure 58 presents an overall comparison between the approach of binary relevance and *ML-SMARTENERGY.KOM* model. We notice from this figure an increase of 11% and 13% in values of f-measure and recall respectively. This noticeable improvement in results proves the importance of dependency patterns between concur-

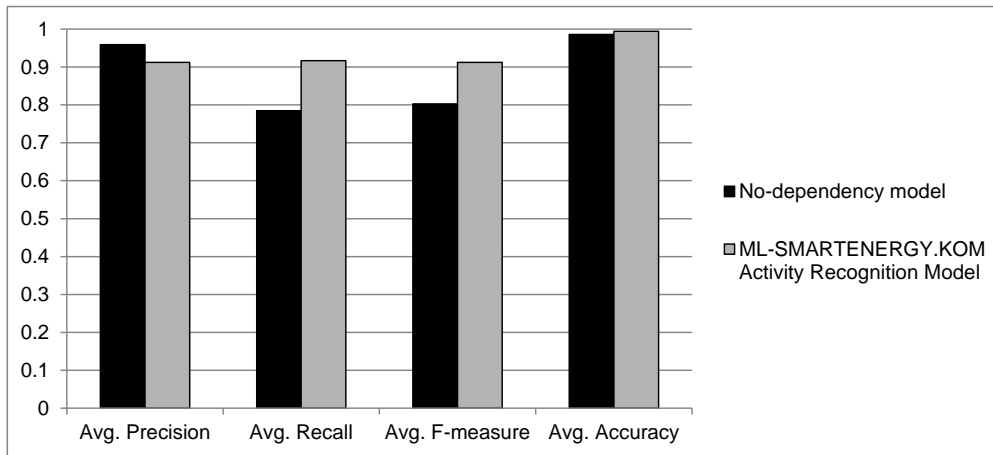


Figure 58: Comparison between *ML-SMARTENERGY.KOM* and no-dependency activity recognition models in terms of average f-measure, accuracy, precision and recall

rent activities in improving the predictive performance of activity recognition models in multi-user environments.

## 6.6 SUMMARY

In this chapter, we presented *ML-SMARTENERGY.KOM*, our platform for activity recognition in multi-user environments. We started by introducing the hardware and software components of the platform in Section 6.1. We explained the deployment of *ML-SMARTENERGY.KOM* in a two-user apartment. Then, we highlighted the data collection process and presented the set of monitored devices and activities. Thereafter, we clarified the process of feature extraction and data preprocessing in Section 6.2 where we also quantified the multi-label characteristics of the collected dataset. In Section 6.3, we presented our basic model for activity recognition in multi-user environments, namely the binary relevance model. We evaluated the predictive performance of this model in Section 6.3.1. By conducting a comprehensive evaluation study, we showed the important effect of temporal relations between subsequent activities on enhancing the predictive performance of activity recognition models in multi-user environments in Section 6.3.1.2. Moreover, we proved the adequacy and sufficiency of fine-granular power consumption measurements as a sensing modality for building an accurate activity recognition model in Section 6.3.1.3. To further improve the predictive performance of activity recognition models in multi-user environments, we introduced our approach for identifying intra- and inter-user label dependency between concurrent activities in Sections 6.4.1 and 6.4.2 respectively. Based on the identified label dependency relations, we presented our multivariate *ML-SMARTENERGY.KOM* activity recognition model in Section 6.5. We evaluated the predictive performance of this model in Section 6.5.1 where we showed how it outperforms binary relevance models by exploiting intra- and inter-user label dependency.



---

## HYBRID LARGE SCALE FORECASTING OF POWER CONSUMPTION

---

FORECASTING of long-term and short-term power consumption has always been an important research topic for academia and industry [48][116]. It has its benefits for consumers and providers of electric power. On one hand, it provides electric utilities with an overview about expected power consumption of individuals and communities in near and far future. This allows utilities to plan their short-term and long-term provisioning of power so that no surplus or shortage can happen. Moreover, this information makes it easier for utilities to plan their infrastructure so that they meet the expected future demand of their customers. On the other hand, this information increases consumers' awareness of their power consumption. It allows them to plan their consumption so that they reduce the cost of the power they consume by avoiding peak hours and following different available tariffs introduced by utilities.

The emergence and increased adoption of smart metering technologies paved the road for having a very detailed overview of power consumption of individual buildings as well as complete geographical areas. By providing measurements on a minute scale, smart meters open new perspectives for research in the field of power consumption forecasting. Due to the availability of such fine-granular measurements, it has become possible to accurately forecast daily and hourly consumption of individual buildings. In this chapter, we start by introducing our model for the average monthly power consumption of individual buildings. Thereafter, we present our model for forecasting long-term daily power consumption of individual buildings. Moreover, we present two models for forecasting short-term daily and hourly power consumption. We evaluate our work based on a fine-granular dataset of power consumption measurements collected by the Commission for Energy Regulation (CER) in Ireland [25].

### 7.1 DATASET

To start the adoption process of smart metering technologies in Ireland, the CER has started a project in which they installed about 5000 smart meters in Irish residential buildings as well as small and medium enterprise (SME) buildings. The project has covered eight urban areas and three villages for duration of one and a half year starting from July 2009 and ending in 2010. All participants have contributed to pre-trial and post-trial surveys. Residential participants provided information about the following aspects in their pre-trial survey:

- Demographic features of residents such as number of people living in the house, age groups, household income and employment status.

- Physical characteristics of the house such as floor size, house type, number of bedrooms, heating type and insulation.
- Type and number of available electrical appliances in the house.
- Behavioral features of residents such as their usage patterns of electrical appliances as well as their awareness degree of the power each appliance consumes.

Different questions have been provided to the owners of SME buildings in the pre-trial survey. They have answered questions related to:

- Business sector of their establishment such as entertainment, office, industrial and retail.
- Number of employees.
- Availability of Internet access.
- Hours and days of operation.

Based on this dataset and the pre-trial survey, we build and evaluate our models for long-term and short-term forecasting of power consumptions as presented in the next sections. Two types of features are used in building our forecasting models, namely time-independent and time-dependent features:

- With time-independent features, we refer to the features that affect power consumption but do not change with time. Examples are demographic features of residents, building characteristics, available appliances and their usage pattern.
- With time-dependent features, we refer to the features that change with time leading to fluctuations in mean power consumption. Weather conditions represent an example of time-dependent features. Another example is type of the day i.e. business day, weekend, or holiday.

Time-independent features contribute to the modeling of mean power consumed by buildings irrespective of the effect of time-dependent features. Floor size is an example time-independent feature where buildings with large floor size tend to consume more power than small buildings. Time-dependent features explain the seasonal fluctuations that happen in power consumption. For example, high summer temperatures lead to more power consumption by the same building due to the need for air conditioning.

## 7.2 LONG-TERM FORECASTING OF POWER CONSUMPTION

Forecasting of long-term power consumption has its potential benefits on the strategic future development of electric utilities' infrastructure so that they meet the increasing demand of their residential and industrial customers. By foreseeing the future development of customers' needs, electric utilities can accurately plan the required future expansion of their facilities including Transmission and Distribution (T&D) equipment.

In this section, we present the building-performance multiple regression model which uses time-independent features to model the average monthly power consumption of residential buildings. Thereafter, we introduce our hybrid model for

forecasting long-term daily power consumption based on time-independent and time-dependent features. We build and evaluate our models using only the data of residential buildings.

In order to build and evaluate our models, the dataset has to be divided into training and testing sets. To do that, we design a novel statistical approach which ensures that training and testing sets follow the same distribution and represent the total potential variety in the data. We build a conditional inference tree (ctree) that models the total power consumption of each building based on its time-independent features. Our goal from building such a tree is to divide our dataset into several homogeneous groups of buildings that have similar total power consumption patterns. We divide each of these groups into 80% training set and 20% testing set. As a result, we obtain training and testing sets that represent all available buildings in the dataset.

Figure 59 shows the resulting ctree. Apart from dividing the data into homogeneous groups, this ctree reflects the importance of several time-independent features in modeling total power consumption. As we can see from the figure, these features are number of bedrooms, floor size, people description, numbers of game consoles, dishwashers and tumble dryers. After dividing the data into training and testing sets, we present our approach for modeling average monthly power consumption of buildings in Section 7.2.1.

### 7.2.1 Building-Performance Multiple Regression Model

In this section, we introduce our approach for modeling average monthly power consumption of residential buildings based on their time-independent features. To build our model, we follow the technique of multiple linear regression. We build a linear regression model following Eq. 36. This model takes as a response the average monthly power consumption of a building. As predictors, it takes demographic features of residents, building characteristics, heating sources and number of available appliances. We refer to this model as building-performance multiple regression model as it is only based on the characteristics of buildings and their residents. This model achieves two purposes. On one hand, it estimates the effect of each time-independent feature on average monthly power consumption. As a result, we obtain a list of important features that can be used in further modeling steps. On the other hand, it helps in detecting buildings with abnormal power consumption as we explain later in this section.

$$y_i = \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_p x_{ip} + e_i \quad (36)$$

Where:

- $y_i$  : average monthly power consumption of building  $i$
- $x_{ij}$  : predictor variable  $j$  where  $j \in [1 : p]$
- $p$  : number of predictor variables
- $\beta_1 \dots \beta_p$  : regression coefficients
- $e_i$  : estimation error of building  $i$

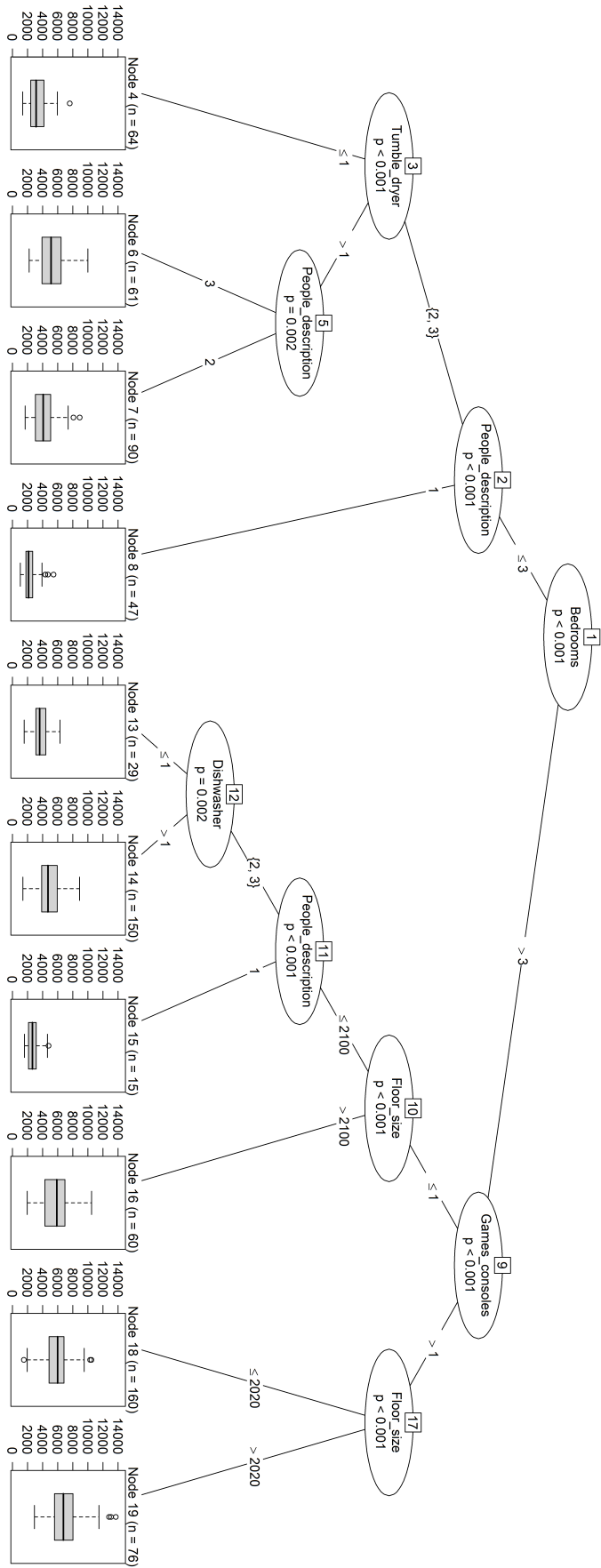


Figure 59: Dividing buildings data into homogeneous groups using the tree algorithm and based on time-independent features



Some predictors are linearly correlated. For example, floor size is correlated with the number of bedrooms as a larger floor size means more bedrooms in the house. This leads to a multicollinearity problem that can cause a false interpretation of feature importance as the importance level of certain features can be underestimated or overestimated. To handle this problem, we follow an iterative backward elimination approach to identify the most accurate regression model and delete all predictors that have no effect on the response variable. We start our modeling approach with 48 predictors representing time-independent features taken from the pre-trial survey. Backward elimination produces a model with only the predictors that contribute to the improvement of model's predictive performance. We utilize this model for identifying buildings with abnormal consumption patterns. Eq. 37 computes the standardized residual for each building in terms of actual and predicted average power consumption. Any building with an absolute standardized residual  $|\hat{z}_i| > 2$  is considered to be an outlier and has to be excluded from the modeling process [90, page 420][54]. After excluding all outliers, we fit the model again. We continue this iterative process as long as the predictive performance of the model is being improved.

$$\hat{z}_i = \frac{y_i - \hat{y}_i}{\hat{\sigma}} \quad (37)$$

Where:

- $y_i$  : actual average monthly power consumption of building  $i$
- $\hat{y}_i$  : predicted average monthly power consumption of building  $i$
- $\hat{z}_i$  : standardized residual for building  $i$
- $\hat{\sigma}$  : standard error

Table 14 shows the coefficients of the final model which we obtained after the application of backward elimination and the deletion of all outliers. We notice from this table that features such as people description, number of bedrooms and number of laptop and desktop computers are very important in determining the average monthly power consumption. The values that are taken by "People description" as a feature are obtained from the answers to the following question from the pre-trial survey:

What best describes the people you live with?

- 1            I live alone
- 2            All people in my home are over 15 years of age
- 3            Both adults and children under 15 years of age live in my home

### 7.2.2 Hybrid Model

This section introduces our model for forecasting long-term daily power consumption of residential buildings based on their time-independent and time-dependent features. To build this model, we follow a hybrid approach in which we combine the ctree algorithm with multiple linear regression. As a first step, we divide our

Coefficient	Estimate	Std. Error	t-value	Pr(>  t )
(Intercept)	-2.995e+03	3.775e+02	-7.934	8.48e-15
People description	6.079e+02	8.455e+01	7.189	1.69e-12
Floor size	2.836e-01	7.365e-02	3.851	0.000128
Bedrooms	4.378e+02	7.111e+01	6.157	1.25e-09
Water central heating system	-2.323e+02	1.479e+02	-1.571	0.116663
Water electric (immersion)	3.509e+02	9.609e+01	3.652	0.000280
Water heating (Gas)	-5.435e+02	1.423e+02	-3.819	0.000146
Water heating (Oil)	-2.729e+02	1.200e+02	-2.274	0.023293
Water heating (Other)	-1.628e+03	8.723e+02	-1.866	0.062394
Cook	-2.634e+02	7.533e+01	-3.497	0.000501
Tumble dryer	4.529e+02	1.174e+02	3.857	0.000125
Dishwasher	4.296e+02	1.281e+02	3.355	0.000837
Electric heater plug in	1.622e+02	6.928e+01	2.341	0.019511
Stand-alone freezer	3.134e+02	8.788e+01	3.566	0.000388
TV greater 21	1.972e+02	5.469e+01	3.605	0.000334
Desktop computers	5.562e+02	8.141e+01	6.832	1.83e-11
Laptop computers	3.146e+02	5.755e+01	5.467	6.39e-08
Games consoles	2.612e+02	6.441e+01	4.056	5.56e-05

Table 14: The coefficients of building-performance model where “Estimate” refers to the estimated value of each coefficient, Std. Error refers to its standard deviation, t-statistic tests the null hypothesis  $H_0 : \beta_i = 0$  against the alternative hypothesis  $H_1 : \beta_i \neq 0$ ,  $Pr(> |t|)$  gives the corresponding p-value.

heterogeneous set of buildings into a set of homogeneous groups. This is done by building a ctree model which takes the daily power consumption of a building as its response variable. As predictors, this ctree model takes time-independent features, namely demographic features of residents, building characteristics, heating sources as well as type and number of available electric appliances. The obtained ctree models the mean daily power consumption of a building based on its time-independent features.

As a second step, we build a multiple linear regression model for each of the obtained homogeneous groups of buildings. This model takes as a response the daily power consumption of a building. As predictors, it combines time-dependent features, namely temperature and day type i.e. holiday, business day or weekend with a part of time-independent features, namely people description, number of bedrooms, floor size, construction year and home description. We reuse time-independent features as they contribute to the modeling of base power consumption. By utilizing time-dependent features, we model the fluctuations that happen in power consumption due to seasonal effects, temperature changes and different behavioral patterns of users according to the day type.

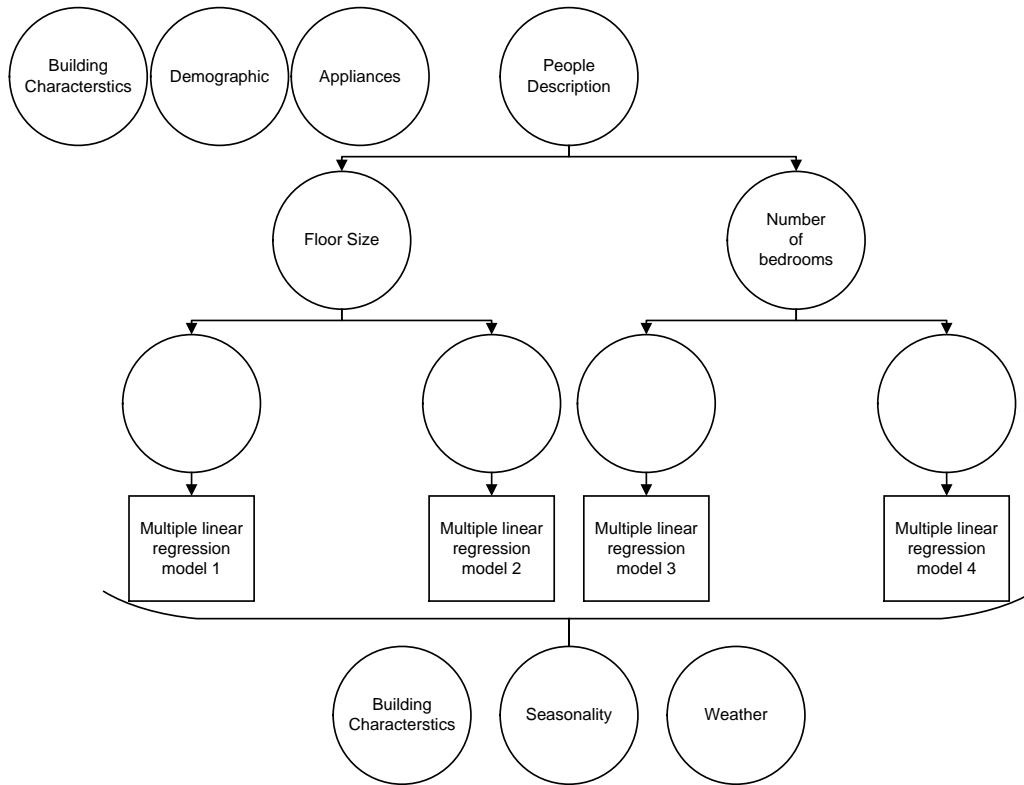


Figure 60: The methodology of building the hybrid model for forecasting long-term daily power consumption

Figure 60 shows both modeling steps. We notice in this figure that a linear regression model is fitted in each node in the resulting ctree. To build and evaluate the proposed hybrid model, we use power consumption data of residential buildings after dividing it into training and testing sets as clarified in Figure 59.

#### 7.2.2.1 Evaluation

We present in this section the evaluation results of our hybrid model for forecasting long-term daily power consumption of residential buildings. As a first step, we use building-performance model to remove buildings with abnormal power consumption. As a result, we obtain a training set of 753 buildings and an out-of-sample testing set of 139 buildings. We train and test our model on two six-month datasets extracted from training and testing sets.

To evaluate the predictive performance of our hybrid model, we forecast the daily power consumption of each building in the testing group during the test period from August to December. We divide our evaluation study into two parts. In the first part, we evaluate the ability of our model to predict the total daily power consumed by all buildings in the testing set. This is done by predicting the daily power consumption of each test building and then summing all these values up as shown in Eq. 38. In the

second part, we compute the predictive performance achieved by the hybrid model in forecasting the daily power consumption of each building individually.

$$\hat{p}_{\text{total}}^j = \sum_{i=1}^{i=139} \hat{p}_i^j \quad (38)$$

Where:

- $\hat{p}_{\text{total}}^j$  : predicted total power consumption of all buildings for day j
- $\hat{p}_i^j$  : predicted power consumption of building i for day j

Figure 61 compares the predicted total sum of daily consumption values of all buildings to the actual total consumption. As we can see from this figure, our hybrid model is able to accurately predict the total daily power consumed by all buildings in the testing set. Figures 62 and 63 show the predictive performance achieved when using a random forest model and ctree model respectively. The goal of these two figures is to compare our hybrid model to two rigorous machine learning models, namely random forests and ctree algorithms. As we can see from the figures, both algorithms achieve good results in predicting the total daily power consumption. However, our hybrid model outperforms both of them in terms of Mean Absolute Percentage Error (MAPE) and Mean Absolute Error (MAE) as shown in Table 15. MAPE evaluates the predictive performance of a forecasting model by calculating the mean absolute percentage error as shown in Eq. 39. MAE calculates the mean absolute error as shown in Eq. 40. MAPE is more expressive than MAE as we can use it to judge the predictive performance of a model without having any previous knowledge about the scale of actual values to be predicted.

$$\text{MAPE} = \frac{100}{N} \sum_{h=1}^N \left| \frac{x_h - \hat{x}_h}{x_h} \right| \quad (39)$$

$$\text{MAE} = \frac{1}{N} \sum_{h=1}^N |x_h - \hat{x}_h| \quad (40)$$

Our hybrid model can be generalized so that it predicts the long-term daily power consumption of new constructions where no information is available with regard to residents' demographic data and number of available appliances. Therefore, we build the generalized hybrid model based on the same evaluation settings used to build the hybrid model after removing demographic data and available appliances from the set of predictors. The generalized hybrid model achieves good predictive performance as shown in Figure 64. However, we notice from Table 15 an increase in the values of MAPE and MAE in comparison to the values of the original hybrid model. This increase is the main result caused by the removal of demographic data and available appliances from the set of predictors.

Table 16 further presents the predictive performance of all previous models in terms of individual buildings. In this evaluation setting, we compare the values of predicted and actual power consumption for each building individually. As we can see from the table, hybrid model outperforms all other models. However, it achieves low predictive performance when compared to the performance achieved in forecasting the power consumption of a portfolio of buildings. This can be explained

Model	Portfolio of buildings	
	MAPE	MAE
Ctree	4.84%	176.44
Random Forest	5.38%	215.63
Hybrid Model	2.43%	89.41
Generalized Hybrid Model	3.43%	123.11

Table 15: Comparison between hybrid, generalized hybrid, random forests and ctree models in terms of MAPE and MAE values achieved for a portfolio of buildings

Model	Individual buildings	
	MAPE	MAE
Ctree	58.65%	10.51
Random Forest	52.34%	9.65
Hybrid Model	49.01%	8.82
Generalized Hybrid Model	50.67%	9.00

Table 16: Comparison between hybrid, generalized hybrid, random forests and ctree models in terms of MAPE and MAE values achieved for individual buildings

by the fact that building a single model for forecasting power consumption of several buildings leads to an averaging process where individual behavioral patterns of single buildings are not taken into consideration. To further improve the predictive performance in forecasting daily power consumption and to forecast hourly power consumption, we introduce two main models in the next sections, namely Hourly Consumption Pattern Matching (HCPM) model and Total Consumption Pattern Matching (TCPM) model.

### 7.3 SHORT-TERM FORECASTING OF POWER CONSUMPTION

Short-term forecasting of day-ahead power consumption plays an essential role for electric utilities. It allows them to reliably and accurately manage their power generation so that they avoid any shortfalls or surpluses that can have very bad effects on the company. Moreover, it has a great potential in increasing consumers' awareness of the power they are going to consume.

In this section, we introduce two models for predicting short-term day-ahead power consumption of individual buildings with two different granularities. On one hand, we present the Hourly Consumption Pattern Matching (HCPM) model with the main goal of forecasting the detailed 24-hour day-ahead power consumption of individual buildings. On the other hand, we present the Total Consumption Pattern Matching (TCPM) model whose main goal is to predict the total day-ahead power consumption of individual buildings.

To build both models and evaluate their predictive performance, we use the data of SME buildings. This data has in total 225 Buildings belonging to four different

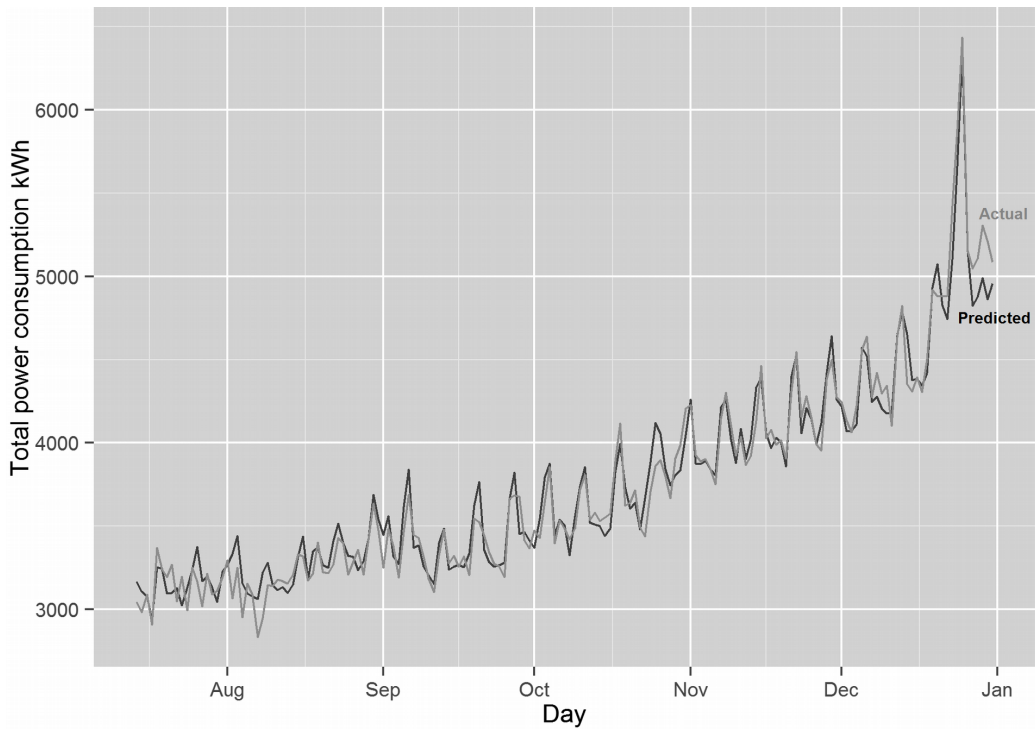


Figure 61: Predictive performance of the hybrid model

business sectors, namely retail, industrial, entertainment and office buildings. For creating an accurate forecasting model, it is always recommended to have a homogeneous training set in which a common pattern exists between all training objects. Therefore, we cluster the heterogeneous set of SME buildings into a set of homogeneous building groups where each group follows common seasonal, weekly and daily consumption patterns. For this purpose, we design a set of discriminative temporal features that will be clarified in Section 7.3.1.

### 7.3.1 Grouping of Buildings

The evaluation results in Section 7.2.2.1 has shown that using time-independent features for creating homogeneous groups of buildings did not lead to an accurate prediction model. Therefore, we designed a new set of features that use historical consumption information to distinguish the buildings based on their seasonal, weekly and daily patterns of power consumption. This set is constructed by computing the following list of values for each building:

- The normalized value of the total power consumed during the whole trial period. The normalization is performed according to Eq 41.
- The percentages of power consumed during Saturdays and Sundays.
- The percentage of power consumed during each of the following six day segments: early morning, morning, early afternoon, afternoon, early night and late night.

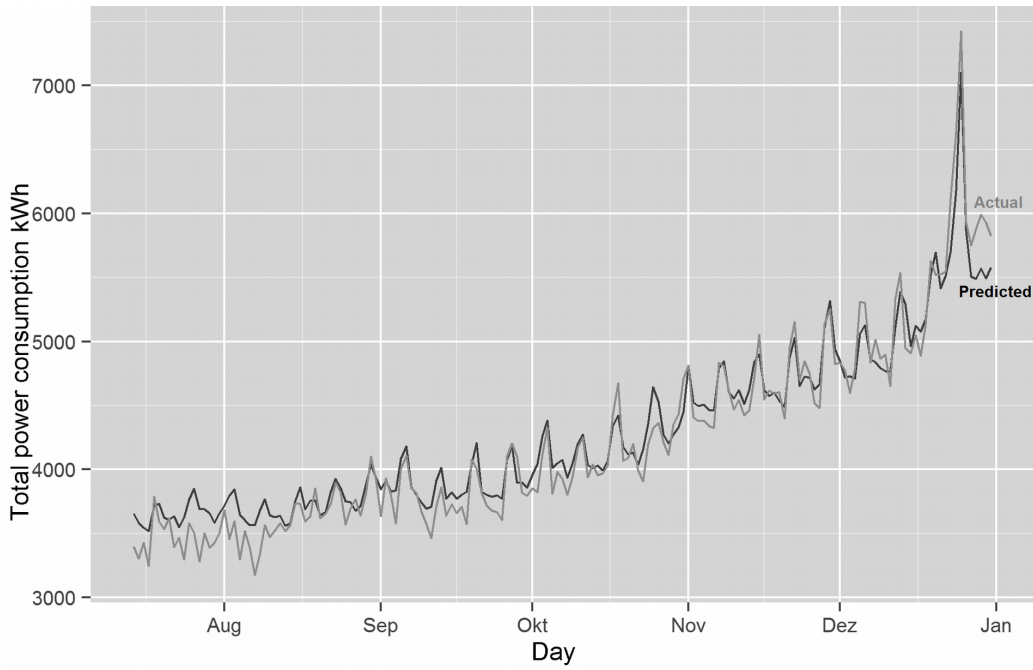


Figure 62: Predictive performance of the random forest model

$$p_{tn}^i = \frac{p_t^i - p_{min}}{p_{max} - p_{min}} \quad (41)$$

Where:

- $p_{tn}^i$  : normalized total power consumption of building  $i$
- $p_t^i$  : total power consumption of building  $i$
- $p_{min}$  : minimum value of total power consumption over all buildings
- $p_{max}$  : maximum value of total power consumption over all buildings

The total power consumed during the whole trial period distinguishes between buildings with different business sectors. Industrial buildings, for example, tend to consume more power in general than office buildings. To differentiate between buildings in the same sector, we use the percentages of power consumed during Saturdays and Sundays. These features distinguish buildings that belong to the same sector but follow different paradigms of working hours during weekends. To further cluster the buildings based on the daily variation in their consumption patterns, we utilize the percentage of power consumed during each day segment. All these features have their values in the range  $[0, 1]$ . However, we do not consider them to have the same importance in the clustering process. As the total amount of power consumed during the whole trial period is a major discriminative feature for distinguishing between business sectors, we assign a high importance level to it by multiplying it by a factor of 8. For all remaining features, we assign the same importance level, namely 1.

To obtain the optimal clustering results, we utilize two robust algorithms, namely k-means and hierarchical clustering. Figure 65 demonstrates the different steps of the clustering process. As a first step, we aggregate power consumption data so that

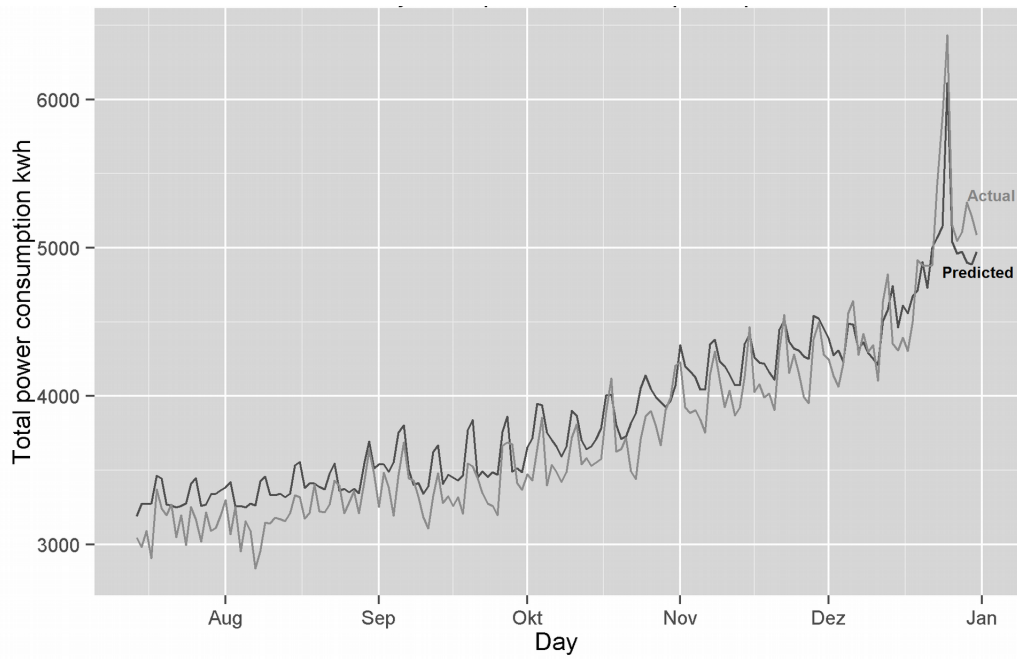


Figure 63: Predictive performance of the ctree model

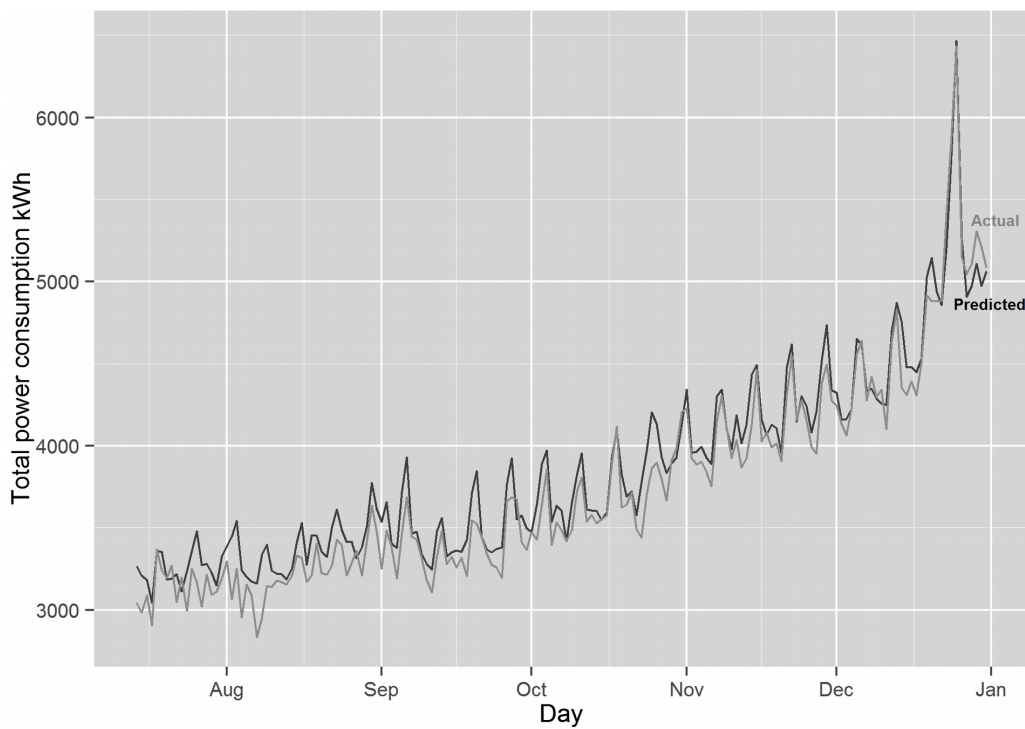


Figure 64: Predictive performance of the generalized hybrid model

total, monthly, daily and hourly consumption values are computed for each building. Based on the aggregated data, we compute the values of each feature vector that are later normalized to be in the range  $[0, 1]$ . Thereafter, we assign for each feature its importance level by multiplying it with a scaler. We run k-means and hierarchical



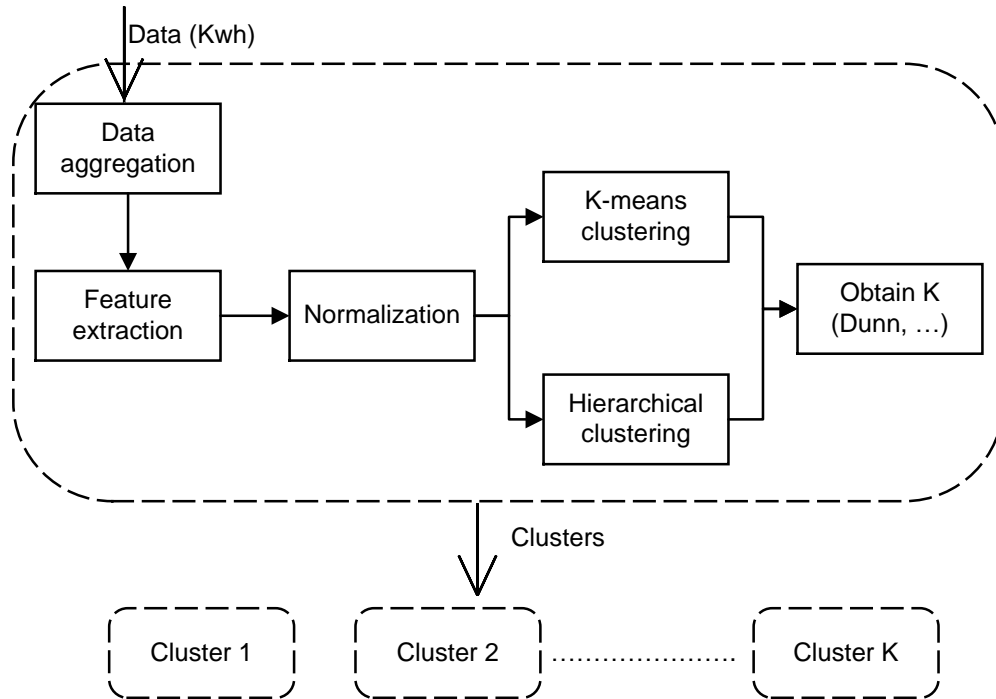


Figure 65: Workflow diagram of the clustering process

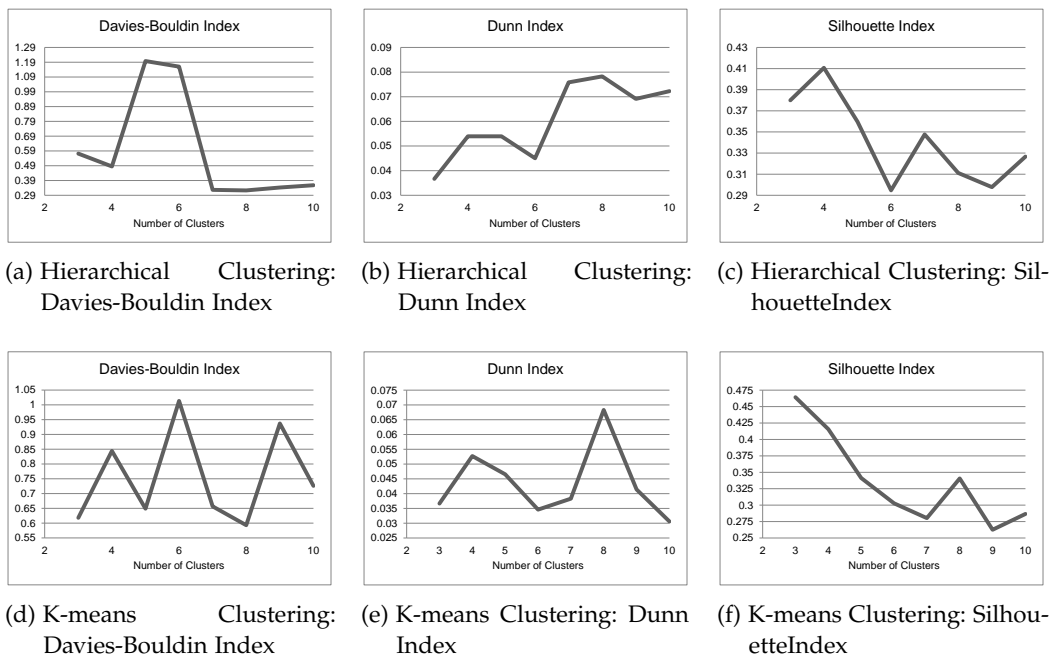


Figure 66: Values of clustering validity indexes with different numbers of clusters

clustering using the obtained normalized feature vectors and for a number of clusters in the range [3, 10]. Due to the fact that only 225 buildings are available in our dataset,

K	K-means clustering			Hierarchical clustering		
	Silhouette	Davies-Bouldin	Dunn	Silhouette	Davies-Bouldin	Dunn
3	0.464	0.618	0.036	0.379	0.572	0.036
4	0.415	0.844	0.052	0.410	0.485	0.053
5	0.341	0.648	0.046	0.359	1.197	0.053
6	0.302	1.013	0.034	0.294	1.160	0.045
7	0.280	0.656	0.038	0.347	0.326	0.075
8	0.340	<b>0.593</b>	<b>0.068</b>	0.311	<b>0.323</b>	<b>0.078</b>
9	0.262	0.937	0.041	0.297	0.342	0.069
10	0.286	0.726	0.030	0.326	0.357	0.072

Table 17: Values of clustering validity indexes with different numbers of clusters

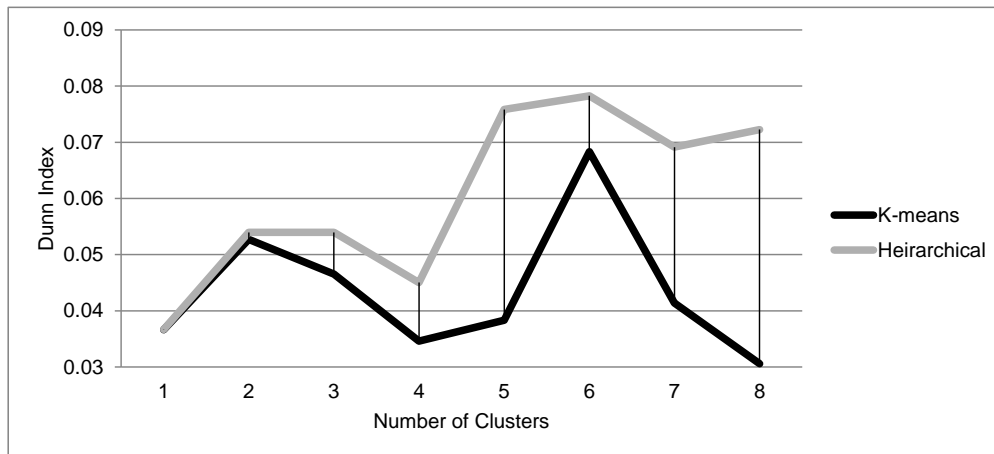


Figure 67: Comparison between the values of Dunn index for different number of clusters using k-means and hierarchical clustering

we choose the maximum value which can be taken by the number of clusters to be 10. Allowing the number of clusters to take higher values than 10 results in clusters that have only a small and statistically insignificant subset of buildings. As a final step, we utilize clustering validity indexes, namely Silhouette, Dunn and Davies-Bouldin to determine the optimal number of clusters as well as the optimal clustering algorithm. A number  $k$  of clusters is considered to be the optimal number if at least two indexes reach their optimal values with it. With an optimal number of clusters, Silhouette and Dunn indexes must reach their highest values while Davies-Bouldin index must reach its minimum value.

Figure 66 and Table 17 show the values of each clustering validity index obtained by applying the algorithms of k-means and hierarchical clustering with numbers of clusters in the range [3 – 10]. As we notice, two validity indexes, namely Dunn and Davies-Bouldin reach their optimal values for both algorithms with a number of clusters  $K = 8$ . Therefore, we consider the optimal number of clusters to be 8. By further inspecting the values of validity indexes when  $k = 8$ , we notice that they reach better

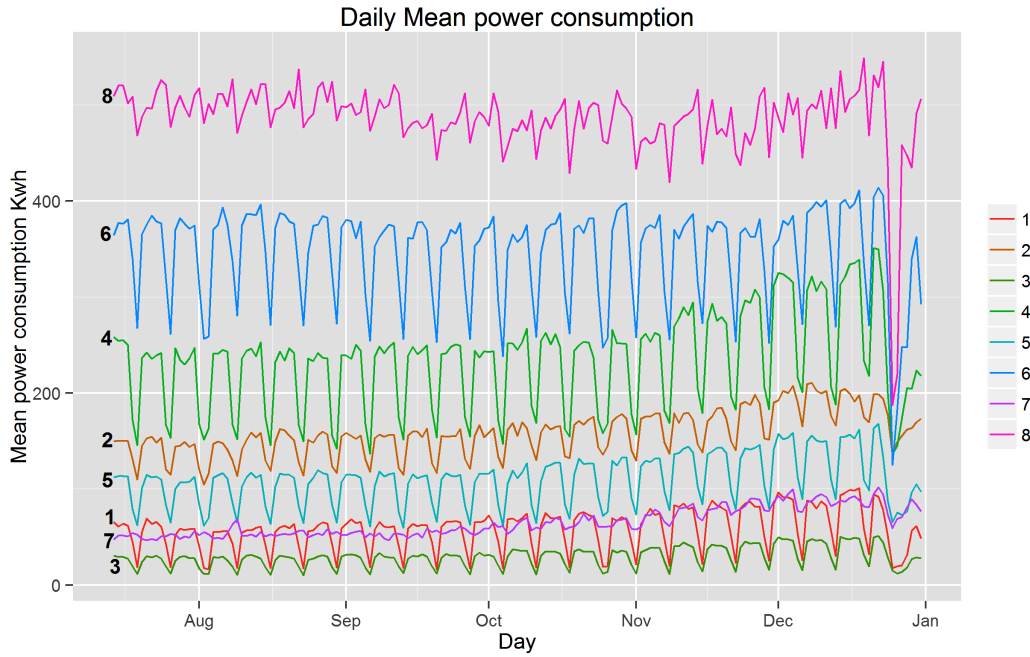


Figure 68: Comparison between the mean daily power consumed by each of the eight obtained clusters using the algorithm of hierarchical clustering

values with hierarchical clustering in comparison to k-means. Furthermore, Figure 67 shows that Dunn index always reaches better values with hierarchical clustering for all numbers of clusters. Therefore, we choose the results obtained by using hierarchical clustering with a number of clusters  $k = 8$  as the optimal final clustering results.

Figure 68 compares the mean daily power consumed by each cluster created using the algorithm of hierarchical clustering. As we notice from the figure, clusters 2, 3, 4, 5, 6 and 8 expose medium to large differences between their mean daily power consumption values. Therefore, the feature “total power consumed during the whole trial period” can be considered as an important discriminative feature between these clusters as it represents the sum of power consumed during all days in the trial. However, this feature has less impact in distinguishing between the clusters 1 and 7 as they tend to have similar mean daily power consumption. To differentiate between these two clusters, the percentage of power consumed during weekends plays an essential role as we can see in Figure 69. This figure shows that cluster 7 consumes around 14 % of its power during the weekend while cluster 1 consumes around 8%. Moreover, Figure 70 demonstrates the percentage of power consumed during each of the six day segments by each cluster. It is clear from this figure that cluster 7 has its consumption peak in the afternoon while cluster 1 consumes most of its power in the morning period.

For each of the obtained 8 homogeneous clusters of buildings, an HCPM model will be built. Each of these models is responsible for forecasting the detailed 24-hour day-ahead consumption of SME buildings into the respective cluster using the techniques of clustering and pattern sequence matching [69].

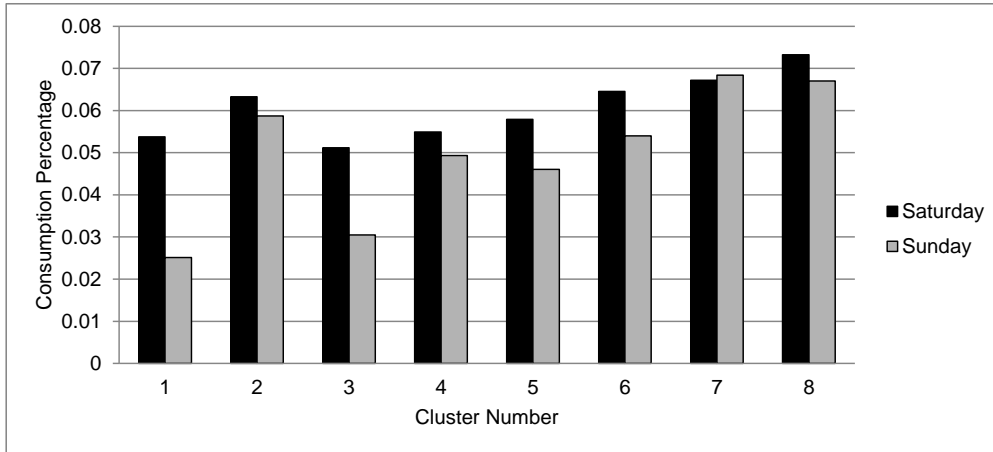


Figure 69: Percentage of power consumed during weekend days by each cluster

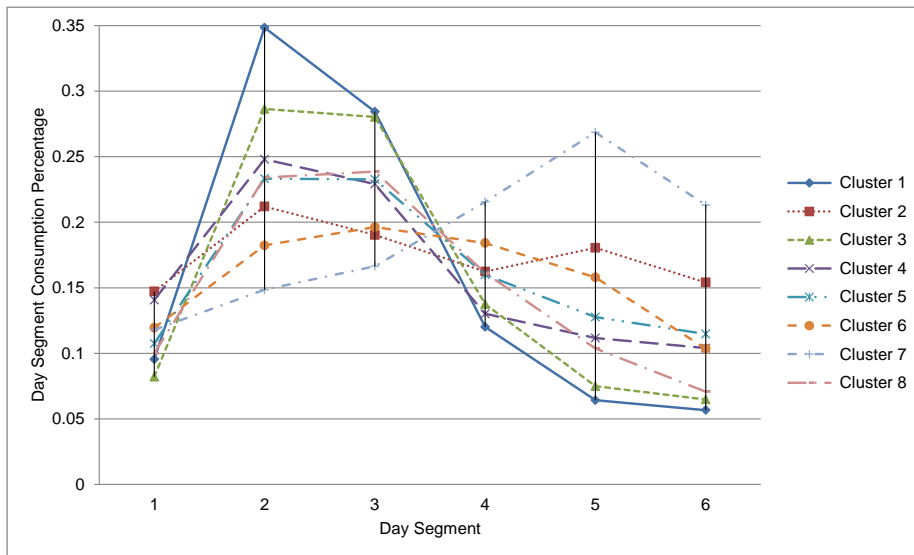


Figure 70: Percentage of power consumed during each day segment by each cluster

### 7.3.2 Hourly Consumption Pattern Matching (HCPM) Model

As its name implies, HCPM focuses on forecasting the detailed 24-hour day-ahead power consumption of individual buildings. As mentioned before, a main prerequisite for building an accurate forecasting model is to have a set of training objects that follow a certain pattern which can be modeled. Therefore, a separate HCPM model is built for each of the eight clusters that are obtained using the algorithm of hierarchical clustering.

Figure 71 demonstrates the workflow of HCPM approach. As an input, HCPM takes one homogeneous cluster of buildings. As a first step, it introduces a new clustering process in which all days belonging to its input are clustered based on their detailed 24-hour power consumption. As a result, a set of clusters will be produced where each cluster represents a different daily 24-hour consumption pattern.

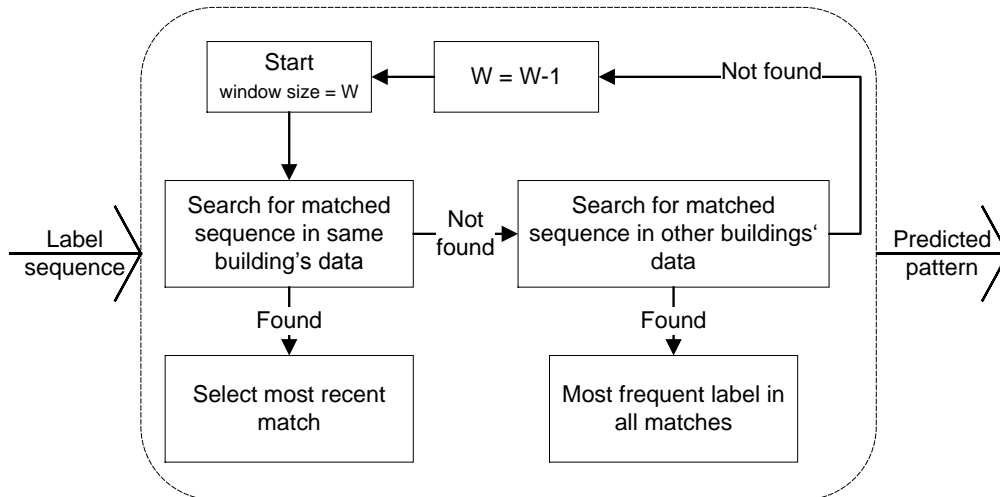


Figure 71: Workflow diagram of HCPM forecasting model

The produced clusters are labeled so that all the days belonging to a certain cluster are labeled with its label. The input of HCPM is therefore transformed into a series of labels where each label corresponds to the 24-hour consumption pattern of the respective day.

HCPM utilizes the approach of pattern sequence matching to forecast the detailed 24-hour day-ahead consumption of a certain building. It extracts the sequence of labels representing the preceding days of the day to be forecast. The number of preceding days is referred to as the window size  $W$ . The value of  $W$  significantly affects the predictive performance of the forecasting model as we will see in the evaluation study in Section 7.3.2.1. A found sequence is accepted only if its succeeding day represents the same day of the week as the day to be forecast. For example, if the day to be forecast is Saturday, then we accept only sequences whose succeeding day is Saturday. After extracting the respective sequence of labels, HCPM starts to look for it in its input series of labels. The search is initially restricted to the days belonging to the same building. When more than one match is found, the most recent one is taken as the consumption behavior tends to be similar in near past. The label succeeding a chosen matching sequence is used as the forecasting result for day-ahead power consumption. In case no match is found in same building's data, HCPM looks for the sequence in the days belonging to other buildings. In case more than one match is found, the most frequent succeeding label of all sequences is used as a prediction for the day-ahead power consumption. In case no match at all is found, HCPM repeats the whole process with a window size  $W - 1$ . HCPM keeps reducing  $W$  by 1 until a match is found. Figure 71 introduces the detailed workflow of HCPM.

### 7.3.2.1 Evaluation

In this section, we evaluate the predictive performance of HCPM model. For the evaluation, we choose a group of 58 buildings representing the third cluster obtained via hierarchical clustering as explained in Section 7.3.1. Figure 72 shows the distribution

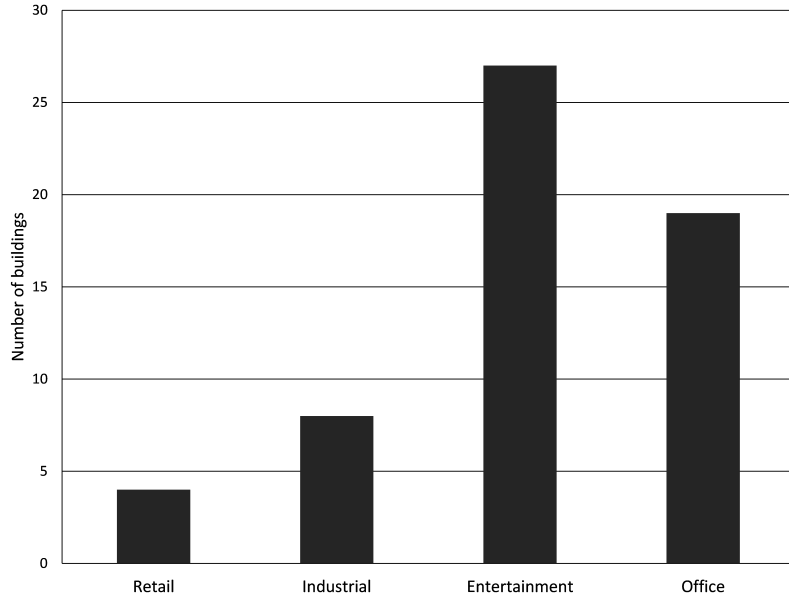


Figure 72: Distribution of buildings in the evaluation group with respect to each business sector

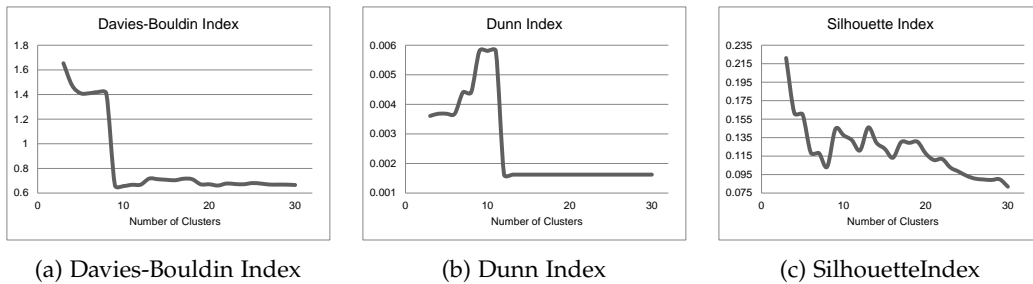


Figure 73: Values of clustering validity indexes with different numbers of clusters

of buildings in this group with respect to each business sector. We use a dataset representing five months of detailed 24-hour daily power consumption for this group of buildings. A subset of four months is used as an input to build HCPM model. This data has to be transformed into a series of labels. The remaining one-month data is used as a testing group.

As an input, HCPM is provided with a set of instances where each instance is a vector of 24 features representing the detailed 24-hour power consumption of the respective day. To cluster these instances, we use the algorithm of hierarchical clustering with a cluster number  $K \in [3 - 30]$ . As we did before, we evaluate clustering quality for each number of clusters using three validity indexes, namely Silhouette, Dunn and Davies-Bouldin. Figure 73 shows the values taken by the three indexes for different numbers of clusters. We choose  $K = 9$  and  $K = 10$  to be the best numbers of clusters as all validity indexes reach relatively optimal values with them.

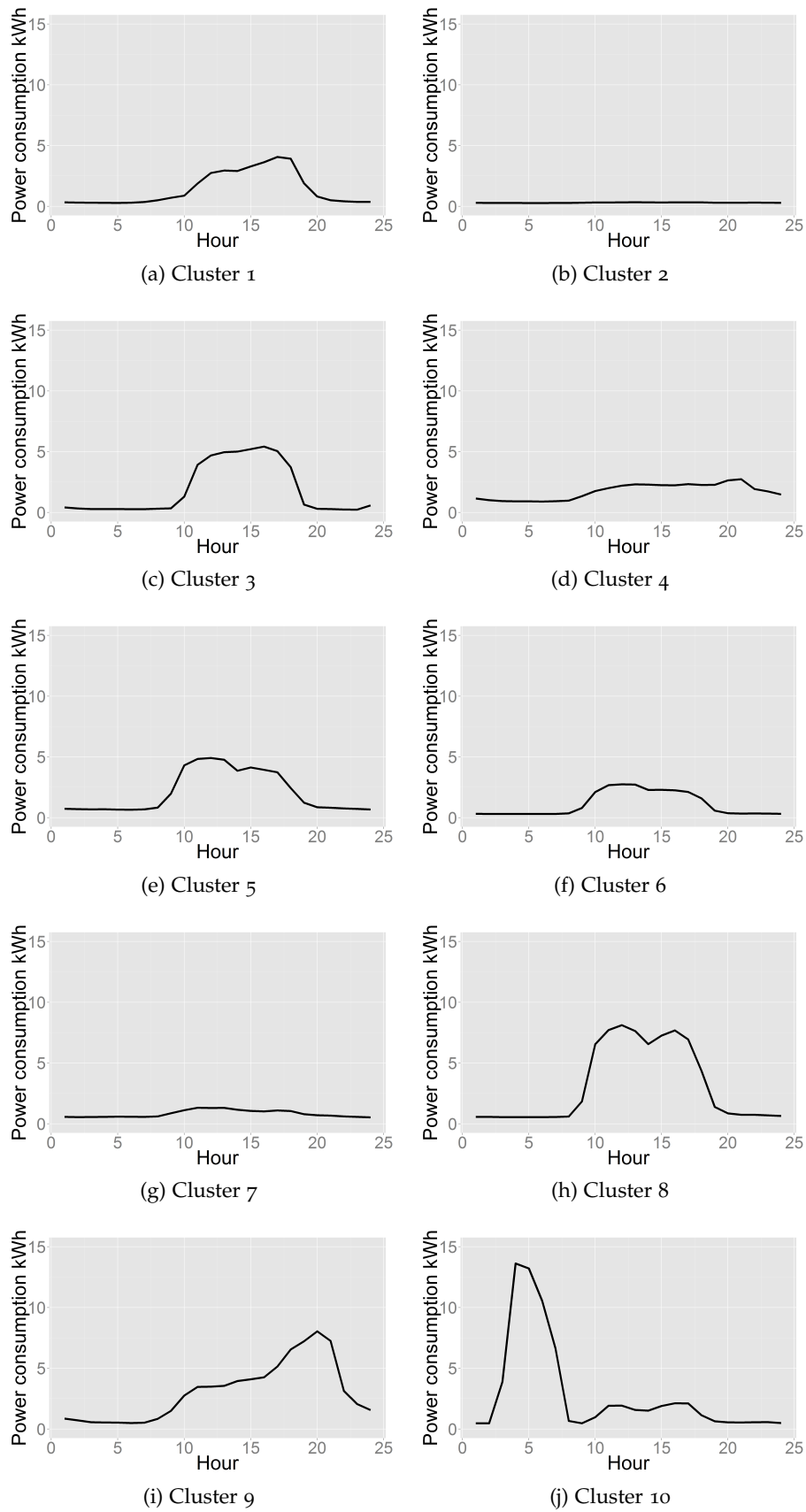


Figure 74: Mean detailed 24-hour power consumption for each of the 10 clusters

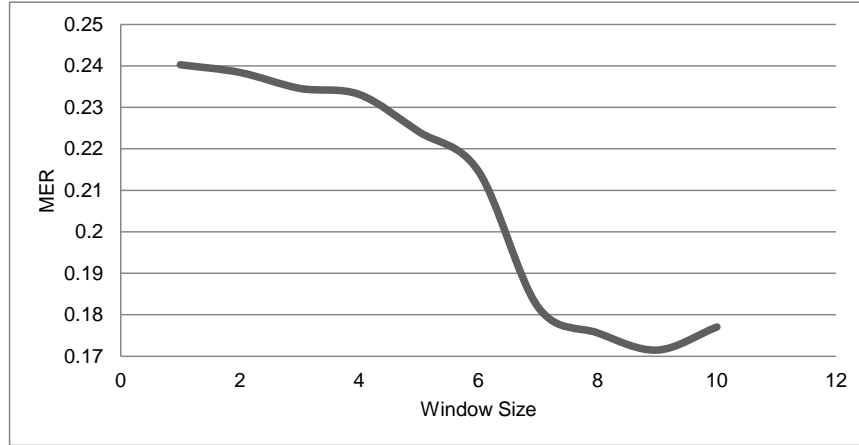


Figure 75: Predictive performance of HCPM model in terms of MER for different values of window size

Figure 74 shows the average power consumed by each cluster over 24 hours. It is clear from the figure that each cluster represents a distinctive daily consumption pattern. It is important to notice from this figure that clusters 2 and 7 show special cases with approximately constant and low power consumption patterns as it is the case during weekend days.

As we mentioned before, window size has an essential effect on the predictive performance of HCPM model. We demonstrate this fact by evaluating HCPM model with different values for the window size. Mean Error Relative (MER) is used to evaluate the predictive performance of HCPM model with  $W \in [1, 10]$ . For each actual value of the response variable, MER [99] as shown in Eq. 42 calculates the percentage of its difference to its predicted counterpart with respect to the mean of all actual values that can be taken by the response variable.

$$\text{MER} = 100 * \frac{1}{D} \sum_{h=1}^D \frac{|p_h - \hat{p}_h|}{\bar{p}} \quad (42)$$

Where:

- $p_h$  : actual power consumed at hour h
- $\hat{p}_h$  : predicted power consumed at hour h
- $\bar{p}$  : mean hourly power consumption in respective day
- D : number of hours predicted in respective day

Figure 75 illustrates the predictive performance achieved by HCPM model in terms of MER and with respect to different values of window size. This figure clearly shows the significant effect of window size on MER values. An MER value of 24% has been achieved with a window size of 1. This value has been significantly decreased to reach 17.1% for a window size of 9. This means, an increasing value of the window size leads to a better predictive performance of HCPM model. However, when the window size is increased to 10, MER value starts to increase. This is due to the



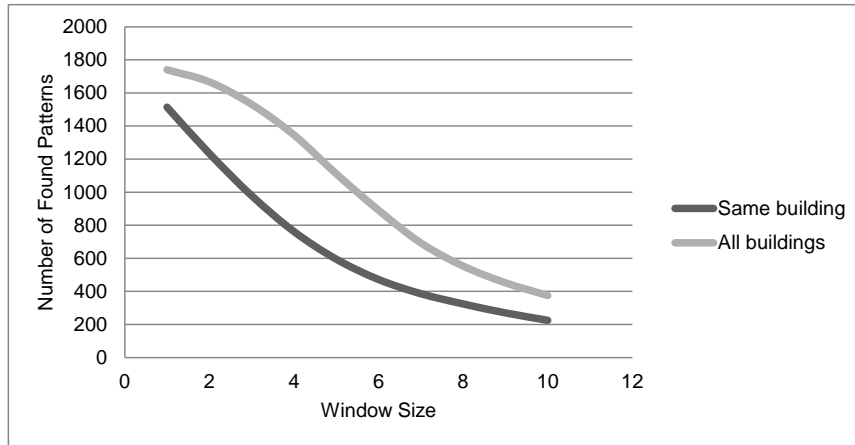


Figure 76: Number of matches found with different values of window size

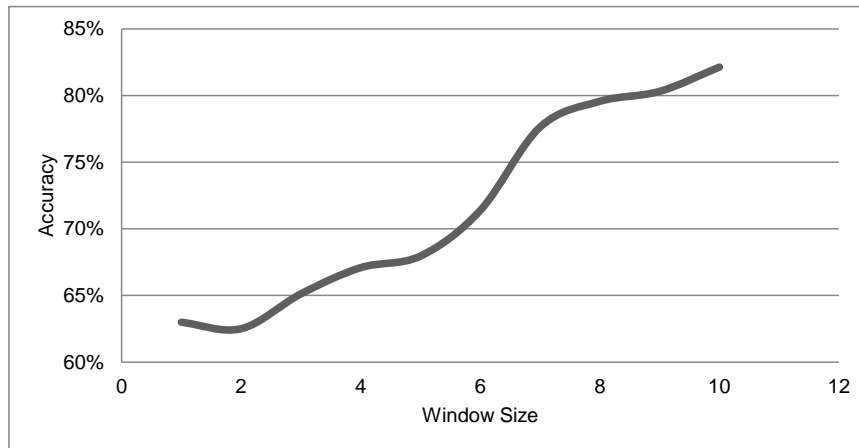


Figure 77: Accuracy of HCPM model for different values of window size.

fact that less number of matched patterns is found with high values of window size which negatively affects the predictive performance of HCPM model. Figure 76 illustrates this fact by presenting the relation between window size and total number of found patterns in all buildings as well as in the same building to which the day to be forecast belongs. It is clear from the figure that fewer matches are found with larger window size.

To further evaluate the predictive performance of HCPM, we compute its accuracy in assigning a correct label to each of the days to be forecast. Accuracy is calculated by dividing the number of correctly labeled instances to the total number of instances. Figure 77 shows accuracy values achieved by HCPM for different window sizes. This figure confirms the results obtained by computing MER values as the model accuracy also increases with an increasing window size. The model started with an accuracy of 62.5% for a window size of 1 and reached an accuracy of 82.1% for a window size of 10.

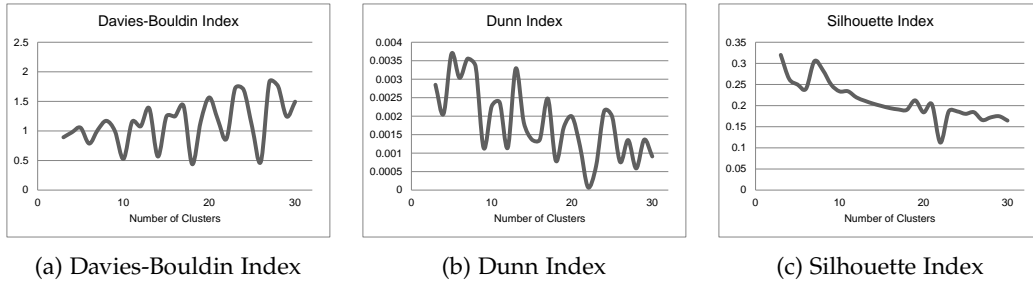


Figure 78: PSF algorithm: values of validity indexes obtained for different numbers of clusters using k-means algorithm

Cluster Number	Average Hourly Power Consumption (kWh)
1	6.62
2	15.14
3	18.73
4	35.66
5	1.32
6	10.03
7	4.46

Table 18: PSF algorithm: Average amount of power consumed during an hour by each of the obtained clusters

### 7.3.2.2 Comparison of HCPM and PSF

This section presents the results of comparing HCPM model against Pattern Sequence-based Forecasting (PSF) [69] as an algorithm which is designed to achieve the same goal of predicting the detailed 24-hour day-ahead power consumption. PSF works by directly clustering the days belonging to all buildings into a set of clusters based on their detailed 24-hour consumption pattern. In contrast to our HCPM model, PSF assumes homogeneity in its data and does not involve any initial clustering step in which the buildings are clustered into homogeneous groups based on a specific set of features related to the characteristics of the building itself or its occupants. For the purpose of comparison, we apply PSF on our dataset by first utilizing the algorithm of k-means to create a set of clusters where each cluster corresponds to a certain 24-hour consumption pattern. To determine the optimal number of clusters, we utilize clustering validity indexes namely, Davies-Bouldin, Dunn and Silhouette indexes. Figure 78 shows the different values obtained by each clustering index for a number of clusters  $k \in [3, 30]$ . By following a majority vote, we choose  $k = 7$  as the optimal number of clusters.

Figure 79 shows the average 24-hour consumption pattern for each of the obtained clusters using  $k = 7$  as the optimal number of clusters. Table 18 compares each of these clusters in terms of average hourly power consumption. This table shows a big difference in the average hourly power consumed in each of the clusters.

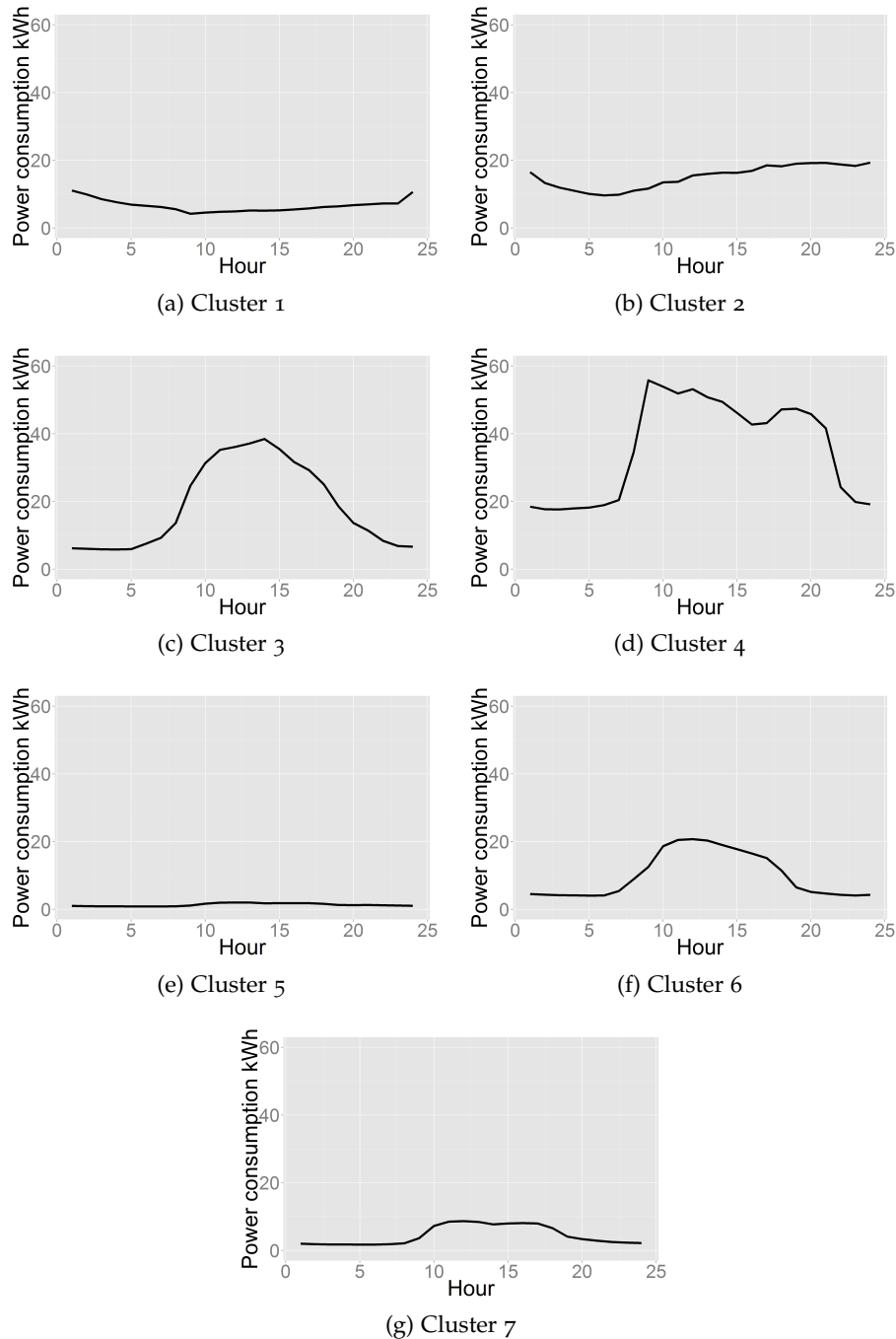


Figure 79: PSF algorithm: the average 24-hour consumption pattern for each of the obtained clusters using  $k = 7$  as the optimal number of clusters.

Figure 79 and Table 18 clearly show that the clustering results produced by  $k$ -means are mainly based on the average hourly power consumed during a day. By inspecting the days assigned to each cluster, we notice that most of the days belonging to the same building are assigned to the same cluster. This is due to the fact that average hourly power consumption plays the main role in building the different clusters and not the daily consumption pattern where the power consumed during each segment of the day distinguishes the clusters from each other.

Window Size	MER	
	PSF	HCPM
W		
7	46.73 %	24.02%
8	45.11%	23.84%
9	43.15%	17.14%
10	39.91%	17.70%

Table 19: Comparison between the MER values achieved by PSF and HCPM

After assigning a label to each day of the dataset, PSF predicts the day-ahead power consumption of a certain day  $x$  by looking for a matching pattern in its historical dataset. As no initial clustering process is performed by PSF, the search space for a matching pattern expands to include the days belonging to all buildings in the dataset. As a result, we notice with PSF an execution time of up to 5 days with a window size of 10 compared to an execution time of 2 minutes for HCPM. To compare the predictive performance of both algorithms, we run both of them with a window size  $w$  in  $[7, 10]$ . Evaluation results in terms of achieved MER values are shown in Table 19. As we can see from this table, HCPM clearly outperforms PSF in terms of MER values and for all different window sizes. This is mainly because PSF relies on average hourly power consumption to distinguish between the clusters without taking daily segments into account. This leads to more found matches in the historical data. However, these matches will produce faulty 24-hour consumption patterns.

### 7.3.3 Total Consumption Pattern Matching (TCPM) Model

This section introduces TCPM, our model for forecasting day-ahead total power consumption of individual SME buildings. In contrast to HCPM, this model forecasts the total daily power consumption and not the detailed 24-hour consumption. As the predictive performance of HCPM model can be affected by the chosen optimal number of clusters and the averaging process caused by combining several days of power consumption into a certain cluster, it is not recommended to use it for forecasting the total daily power consumption. It is more efficient and accurate to build a separate model for achieving this purpose and therefore we build the TCPM model. The main difference between TCPM and HCPM models lies in the fact that TCPM does not perform the clustering process performed by HCPM. As shown in Figure 80, TCPM follows the same workflow of HCPM except that it takes as input the total power consumed during each day of the dataset. No labels are assigned to the days based on their 24-hour power consumption. To determine whether two days, namely  $D_1$  and  $D_2$  are similar or not, we first compute the percentage change of their power consumption as shown in Eq. 43. We consider two days to be similar if their percentage change is less than 10%.

$$PD(P_1, P_2) = \left| \frac{P_1 - P_2}{P_1} \right| \quad (43)$$

Where:

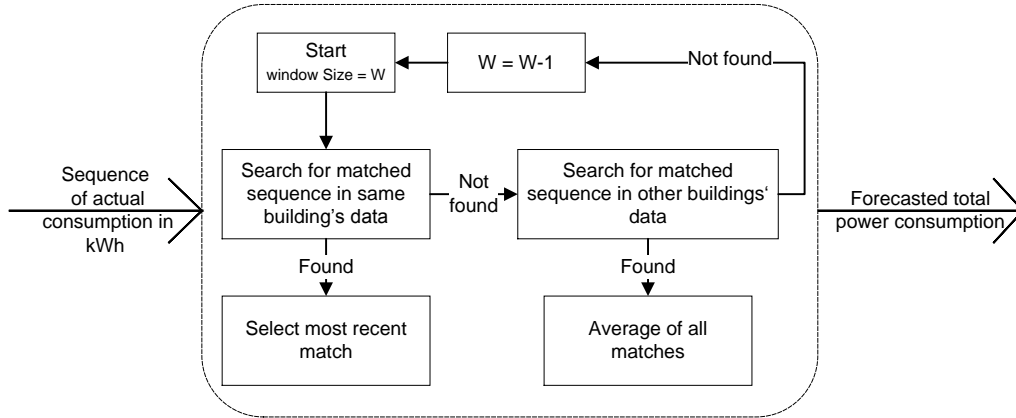


Figure 80: Workflow diagram of TCPM forecasting model

- $PD(P_1, P_2)$  : percentage change between  $D_1$  and  $D_2$
- $P_1$  : total power consumed during  $D_1$
- $P_2$  : total power consumed during  $D_2$

For forecasting the power consumption of day  $x$ , TCPM follows the same steps of HCPM. It starts by searching for all day sequences that match the day sequence preceding  $x$ . The length of sequence is determined by the window size  $w$  as it is the case with HCPM. As a first step, TCPM searches only in the data belonging to the same building. In case more than one match are found, it chooses the most recent match and uses it for predicting the power consumption of  $x$ . In case no match is found, it starts searching in the data belonging to other buildings. When several matches are found, TCPM takes their average as the predicted power consumption of  $x$ . If no match at all is found, TCPM reduces the window size  $w$  by 1 and repeats the same previous search procedure.

### 7.3.3.1 Evaluation

In this section, we evaluate the predictive performance of TCPM in terms of MER as expressed in Eq. 42. We follow the same evaluation setup of HCPM in which the group of 58 SME buildings representing the third cluster is used for building and evaluating the model. We divide the same HCPM five-month dataset into a four-month training set and 1-month testing set. We mainly study the effect of window size  $w$  on the achieved MER values as shown in Figure 81. As we can see from this figure, increasing window size has a positive effect on MER value. TCPM starts with an MER value of 22.6% for a window size of 1. This value decreases to 9.1% for a window size of 8. As these results imply, larger window size leads to a better predictive performance. This is mainly due to the increasing knowledge embedded in larger sequences. However, large window size leads to a less number of found matches in same building's dataset as well as in the complete repository of all buildings' data. Figure 82 clearly shows this effect as the number of found matches gradually decreases with larger values of window size. Less number of found matches can have a

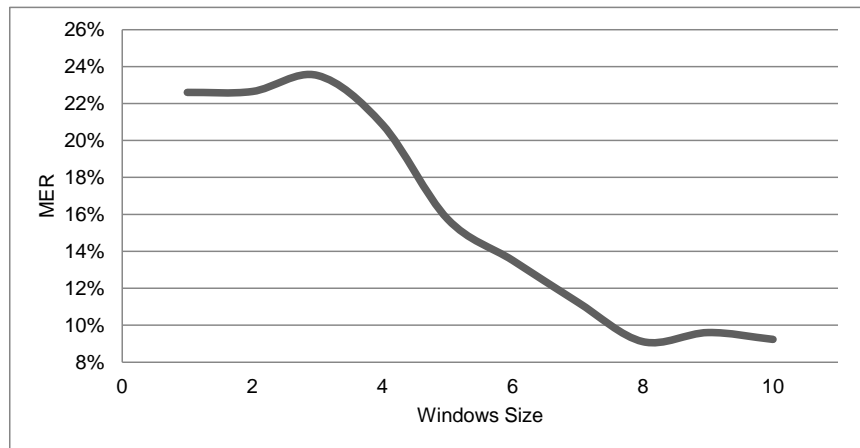


Figure 81: The effect of window size on MER for TCPM.

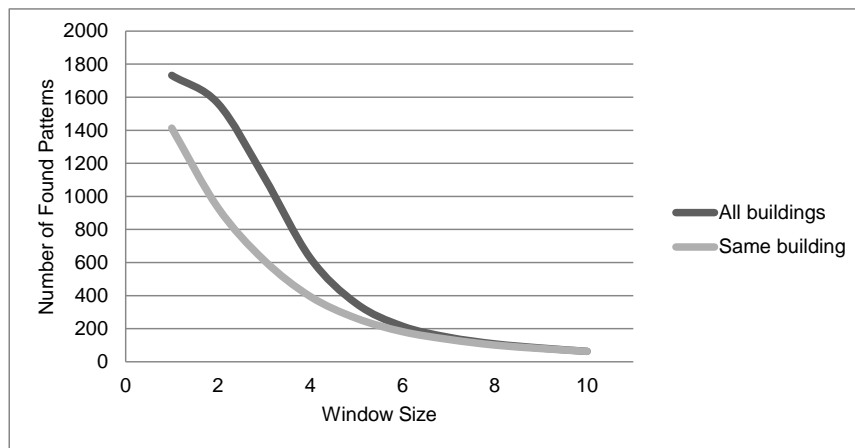


Figure 82: The effect of window size on the number of found matches for TCPM model.

negative effect on the predictive performance. This can be seen in Figure 81 where the MER value slightly increased for a window size of 9. Figure 82 also shows that with a large window size all found patterns will only belong to same building's dataset. The reason behind this phenomenon is the restriction imposed by the 10% similarity criterion as it is difficult to find a large sequence of similar days that achieve this criterion in the dataset of other buildings.

Another important aspect of this evaluation study is the comparison between TCPM and HCPM in terms of their ability to forecast the total day-ahead power consumption. As we clarified before, HCPM is used to forecast the detailed 24-hour day-ahead power consumption. However, it can be used to forecast the total consumption by taking the sum of all hours' predictions. The MER values achieved by each model with respect to different window sizes can be seen in Figure 83. This figure clearly shows that TCPM outperforms HCPM in forecasting the total day-ahead power consumption. The reason behind these results lies in the fact that HCPM in contrast to TCPM follows a clustering approach for determining similarities between different days. Such an approach has a main disadvantage with regard to total day-

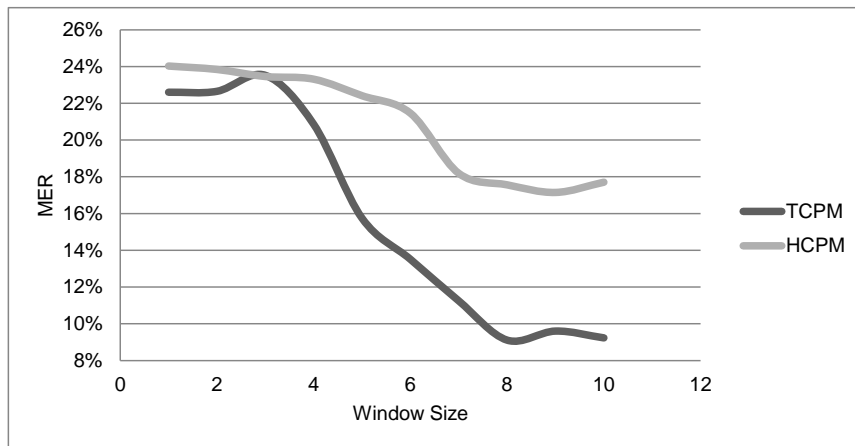


Figure 83: Comparison of TCPM and HCPM in terms of achieved MER values.

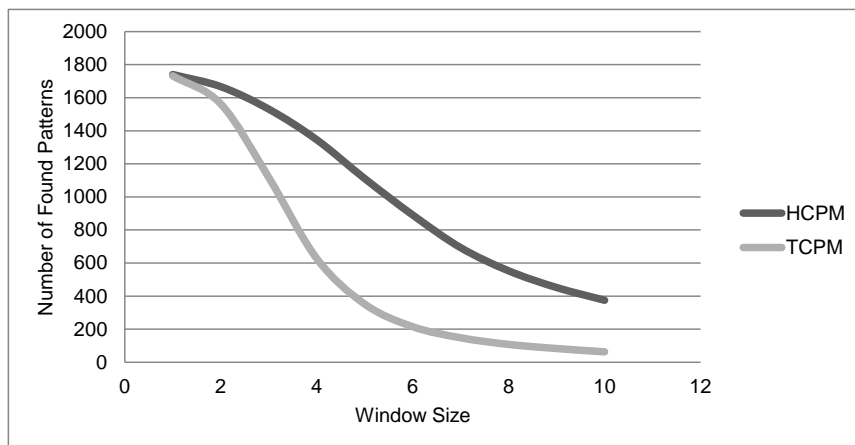


Figure 84: Comparison between the numbers of found matches by HCPM and TCPM

ahead power forecasting. The clustering algorithm relies on calculating the distances to clusters' centroids for determining the cluster to which a day belongs. This may cause two days to be positioned in the same cluster based on their detailed 24-hour power consumption even though they have a significant difference between their total values of power consumption. TCPM avoids this problem by imposing its 10% similarity criterion. This is clear from Figure 84 where we can see that HCPM always finds more matches than TCPM even for large window size.

#### 7.4 SUMMARY

In this chapter, we presented our approaches for forecasting long-term and short-term power consumption of individual buildings as well as of a portfolio of buildings. In Section 7.1, we introduced the CER dataset with which we built and evaluated our forecasting models. We presented our methodology for forecasting long-term power consumption in Section 7.2 where we clarified in details our Building-Performance Multiple Regression model for modeling average monthly power consumption of res-

idential buildings based on their time-independent features. In Section 7.2.2, we introduced and evaluated our hybrid model for forecasting long-term daily power consumption of residential buildings based on time-independent and time-dependent features. Thereafter, we introduced two approaches for forecasting short-term day-ahead power consumption for SME buildings in Section 7.3. We started by explaining the HCPM model for forecasting the detailed 24-hour day-ahead power consumption in Section 7.3.2. We evaluated the predictive performance of HCPM and showed that it outperforms the PSF model in Sections 7.3.2.1 and 7.3.2.2. Then, we explained the TCPM model for forecasting the total day-ahead power consumption in Section 7.3.3. We evaluated the predictive performance of TCPM and compared it to the performance achieved by HCPM in Section 7.3.3.1.



---

## SUMMARY AND OUTLOOK

---

### 8.1 SUMMARY

The recognition and prediction of users' context have always been the focus of researchers in the field of smart environments. The main reason behind this interest lies in the fact that several novel Information and Communications Technology (ICT) services can be realized and implemented given the availability of an accurate recognition and prediction of users' context. Ambient Assisted Living (AAL) represents an example of such services where the researchers develop systems that are able to help elderly people in safely performing their daily activities by recognizing any dangerous situations that may face them and informing the responsible entity [83]. Another example service is power conservation where the researchers develop systems with the main aim of increasing users' awareness of their power consumption so that power wastages can be avoided [42][53]. In this thesis, we presented our four novel contributions for an accurate extraction, recognition and prediction of several aspects of users' context solely based on fine-granular measurements of power consumption as a sensing modality.

The first contribution of this work focused on the recognition of user activities in single-user environments. We developed *SMARTENERGY.KOM*, our hardware/software platform for recognizing user activities in smart environments. We deployed *SMARTENERGY.KOM* in two single-user apartments where we collected information about several environmental parameters, namely motion, temperature and brightness as well as appliance-level fine-granular power consumption data of available electrical appliances. We designed a set of activities which have to be monitored for each apartment. Through a smartphone-based feedback system, the user provided us with his current activity. After obtaining the required dataset, we preprocessed it so that we create the required training and testing datasets for building and evaluating our activity recognition models. We conducted an extensive evaluation study in which we studied the effect of excluding environmental sensors on the predictive performance of *SMARTENERGY.KOM*. The evaluation results showed that *SMARTENERGY.KOM* is able to recognize users' activities with a predictive performance of 86.1% in terms of average f-measure value. Moreover, the results showed that our platform is able to recognize the activities after excluding the environmental sensors from the feature space, however, with a 14% decrease in the predictive performance in terms of f-measure.

The second contribution of this work focused on three main aspects of users' behavior in smart environments. Firstly, we introduced our novel approach for indoor localization of users in smart environments based on their fine-granular appliance-level power consumption data. We evaluated our localization approach using the dataset we collected with *SMARTENERGY.KOM*. The evaluation results showed that our approach is able to recognize the indoor location of users solely based on their power consumption data with a predictive performance of up to 96% in terms of

f-measure values. Secondly, we presented our approach for modeling behavioral patterns of users with the main goal of identifying temporal relations between user's subsequent activities. Using Apriori algorithms, we identified all potential temporal patterns followed by the users. Thereafter, we introduced our approach for incorporating the extracted knowledge into the models of activity recognition. The evaluation results showed that with temporal patterns incorporated into the feature space, our platform is able to recognize users' activities with a predictive performance of 94.5% in terms of average f-measure value solely based on fine-granular power consumption data. Thirdly, we introduced our approach for power profiling in which we mathematically proved the existence of a user-specific power consumption pattern which repeats itself over the days. We conducted an analytical study with which we identified this pattern after excluding the outliers in power consumption.

The third contribution of this work builds upon the first contribution by focusing on activity recognition in multi-user environments. With this contribution, we developed *ML-SMARTENERGY.KOM*, our hardware/software platform for activity recognition in multi-user environments. We deployed *ML-SMARTENERGY.KOM* in a two-user apartment where we installed appliance-level power sensors as well as environmental sensors for sensing motion, brightness and temperature in the environment. The users provided their feedback about their ongoing activities through a smartphone-based feedback system. After a deployment period of three weeks, we got the dataset required for building and evaluating our activity recognition models. As a first step, we built a set of binary relevance models. Each model is trained to recognize one of the activities where the readings provided by power and environmental sensors have been incorporated into the feature space. After evaluating the predictive performance of these models, we incorporated temporal relations between subsequent activities into the feature space as a second step. The evaluation results showed that these temporal relations have an essential effect on enhancing the predictive performance of activity recognition models. As a third step, we excluded all environmental sensors from the feature space and evaluated the predictive performance of the binary relevance models. The evaluation results proved that power sensors are enough for achieving an accurate predictive performance in recognizing activities in multi-user environments. Another essential part of our third contribution is the design of a novel approach for identifying intra- and inter-user dependency relations between concurrent users' activities. Based on the identified dependency relation, we designed the multivariate *ML-SMARTENERGY.KOM* activity recognition model which builds upon these dependency relations for enhancing the predictive performance of activity recognition models. The evaluation results showed that *ML-SMARTENERGY.KOM* model achieved the highest predictive performance in recognizing the activities of both users. Moreover, it outperformed all multi-user activities recognition models presented in Chapter 3 by achieving an average f-measure value of 91%.

The fourth contribution of this work focuses on forecasting long-term and short-term power consumption of individual buildings as well as of a portfolio of buildings. We built and evaluated all of our models using the dataset provided by the Irish commission for energy regulation (CER). We started by developing the Building-Performance Multiple Regression model for the average monthly power consumption of individual buildings. The main goal of this model was to identify the important predictors that affect buildings' power consumption. The initial set of predic-

tors contained demographic features of residents, building characteristics, heating sources and number of available appliances. By utilizing the approach of backward elimination, we were able to identify the important predictors as well as the buildings with abnormal power consumption. Thereafter, we developed our hybrid model for forecasting long-term power consumption of individual buildings and portfolios of buildings. To build this model, we combined two modeling techniques, namely conditional inference trees and multiple linear regression. This model showed a very good predictive performance in forecasting the long-term power consumption for a portfolio of buildings. However, its predictive performance dropped when the power consumption of individual buildings had to be predicted. This can be explained by the fact that the approach of linear regression models the mean power consumption. This causes the model to ignore any consumption differences between individual buildings that are caused by different behavioral patterns of residents. To overcome this problem, we developed two models for forecasting short-term day-ahead power consumption of individual buildings, namely Hourly Consumption Pattern Matching (HCPM) model and Total Consumption Pattern Matching (TCPM) model. HCPM model forecasts the detailed 24-hour day-ahead power consumption of individual buildings while TCPM forecasts their total day-ahead power consumption. The evaluation of HCPM model showed that it achieved a very good predictive performance of 17.1% in terms of Mean Error Relative (MER) and 82.1% in terms of accuracy. Moreover, it outperformed the approach of Pattern Sequence-based Forecasting (PSF) in terms of predictive performance and time efficiency. Moreover, TCPM model showed a very good predictive performance as well by achieving an MER value of 9.1% in predicting the total day-ahead power consumption of individual buildings.

## 8.2 OUTLOOK

In this thesis, we presented four main contributions concerning the accurate identification, recognition and prediction of several aspects of users' context in smart environments. The contributions of this work pave the road for new research ideas in the field of context recognition and prediction. Each of the presented contributions exposes a potential for designing and developing new approaches that can build upon the ideas, systems, results and conclusions presented in this work.

With regard to our platforms for activity recognition in smart environments, namely *SMARTENERGY.KOM* and *ML-SMARTENERGY.KOM*, there is a potential to include and recognize further activities that are not covered in our work. By deploying these platforms in new houses and for new users, the researchers can design new sets of activities to be monitored based on users' daily behavior and their lifestyle. These activities can be chosen based on a specific use case for which the system is deployed such as elderly care, power conservation or increasing the comfort of users where the researchers can study the effect of achieved predictive performance on the use case. It should be noticed here that a new training phase has to be completed before utilizing our platforms for recognizing new sets of activities.

Deploying our two platforms in new houses leads to further potential research work regarding our approach of indoor localization based on power consumption which can be extended to cover richer environments in which the locations of two or

more users have to be recognized. This includes dealing with new types of complexity that arise due to the existence of several persons in one common environment.

Regarding activity recognition in multi-user environments, there is a great potential for deploying and evaluating our approach in environments with more than two users where a wider range of interactions between users' activities can be obtained. As a result, new intra- and inter-user label dependency patterns can be identified using our approach and utilized for achieving an accurate recognition of multiple concurrent activities of several users. The main work

Concerning our models for long-term and short-term forecasting of power consumption, new datasets that cover several years of power consumption can be used to evaluate the predictive performance of all models. With such long time periods, it becomes possible to study and discuss the effects of seasonality on the power consumed by individual buildings as well as portfolios of buildings. Taking seasonal effects into account can lead to an enhanced predictive performance as it allows distinguishing between different patterns of power consumption based on different time periods.

---

## BIBLIOGRAPHY

---

- [1] R. Agrawal and R. Srikant. Fast Algorithms for Mining Association Rules in Large Databases. In *Proceedings of the 20th International Conference on Very Large Data Bases, VLDB '94*, pages 487–499, 1994.
- [2] H. K. Alfares and M. Nazeeruddin. Electric Load Forecasting: Literature Survey and Classification of Methods. *International Journal of Systems Science*, 33(1):23–34, 2002.
- [3] A. Alhamoud, V. Muradi, D. Böhnstedt, and R. Steinmetz. Activity Recognition in Multi-User Environments Using Techniques of Multi-label Classification. In *Proceedings of the 6th International Conference on the Internet of Things, IoT'16*, pages 15–23, New York, NY, USA, 2016. ACM.
- [4] A. Alhamoud, A. A. Nair, C. Gottron, D. Böhnstedt, and R. Steinmetz. Presence Detection, Identification and Tracking in Smart Homes Utilizing Bluetooth Enabled Smartphones. In *Local Computer Networks Workshops (LCN Workshops), 2014 IEEE 39th Conference on*, pages 784–789, Sept 2014.
- [5] A. Alhamoud, F. Ruettiger, A. Reinhardt, F. Englert, D. Burgstahler, D. Böhnstedt, C. Gottron, and R. Steinmetz. SMARTENERGY.KOM: An Intelligent System for Energy Saving in Smart Home. In *Local Computer Networks Workshops (LCN Workshops), 2014 IEEE 39th Conference on*, pages 685–692, Sept 2014.
- [6] A. Alhamoud, P. Xu, F. Englert, A. Reinhardt, P. Scholl, D. Boehnstedt, and R. Steinmetz. Extracting Human Behavior Patterns From Appliance-Level Power Consumption Data. In *European Conference on Wireless Sensor Networks*, pages 52–67. Springer, 2015.
- [7] A. Alhamoud, P. Xu, F. Englert, P. Scholl, T. A. B. Nguyen, D. Böhnstedt, and R. Steinmetz. Evaluation of User Feedback in Smart Home for Situational Context Identification. In *Pervasive Computing and Communication Workshops (PerCom Workshops), 2015 IEEE International Conference on*, pages 20–25, March 2015.
- [8] J. F. Allen and G. Ferguson. Actions and Events in Interval Temporal Logic. *Journal of Logic and Computation*, 4:531–579, 1994.
- [9] S. Aman, M. Frincu, C. Chelmiss, and M. U. Noor. *Empirical Comparison of Prediction Methods for Electricity Consumption Forecasting*. Department of Computer Science, University of Southern California, Los Angeles, CA, 90089, 2012.
- [10] S. Aman, Y. Simmhan, and V. Prasanna. *Improving Energy Use Forecast for Campus Micro-grids Using Indirect Indicators*. Dec 2011.
- [11] A. Anjum and M. U. Ilyas. Activity Recognition Using Smartphone Sensors. In *2013 IEEE 10th Consumer Communications and Networking Conference (CCNC)*, pages 914–919, Jan 2013.

- [12] S. Arlot and A. Celisse. A Survey of Cross-Validation Procedures for Model Selection. *Statist. Surv.*, 4:40–79, 2010.
- [13] A. Benmansour, A. Bouchachia, and M. Feham. Multioccupant Activity Recognition in Pervasive Smart Home Environments. *ACM Comput. Surv.*, 48(3):34:1–34:36, 2015.
- [14] C. M. Bishop. *Pattern Recognition and Machine Learning (Information Science and Statistics)*. Springer-Verlag New York, Inc., Secaucus, NJ, USA, 2006.
- [15] O. Brdiczka, J. L. Crowley, and P. Reignier. Learning Situation Models in a Smart Home. *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, 39(1):56–63, Feb 2009.
- [16] L. Breiman. Random Forests. *Machine Learning Journal*, 45:5–32, 2001.
- [17] P. Brockwell and R. Davis. *Time Series: Theory and Methods*. Springer Series in Statistics. Springer, 2009.
- [18] K. C. Budka, J. G. Deshpande, and M. Thottan. *Communication Networks for Smart Grids: Making Smart Grid Real*. Springer Publishing Company, Incorporated, 2014.
- [19] F. Büsching, M. Bottazzi, and L. C. Wolf. The GAL Monitoring Concept for Distributed AAL Platforms. In *2012 IEEE 14th International Conference on e-Health Networking, Applications and Services (Healthcom) (IEEE Healthcom 2012)*, Beijing, P.R. China, 2012.
- [20] J. Byun and S. Park. Development of a Self-adapting Intelligent System for Building Energy Saving and Context-aware Smart Services. *IEEE Transactions on Consumer Electronics*, 57(1):90–98, February 2011.
- [21] L. Chen, J. Hoey, C. D. Nugent, D. J. Cook, and Z. Yu. Sensor-Based Activity Recognition. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, 42(6):790–808, Nov 2012.
- [22] R. Chen and Y. Tong. A two-stage method for solving multi-resident activity recognition in smart environments. *Entropy*, 16:2184–2203, 2014.
- [23] S. Chen and Y. Huang. Recognizing Human Activities from Multi-Modal Sensors. In *Intelligence and Security Informatics, 2009. ISI '09. IEEE International Conference on*, pages 220–222, June 2009.
- [24] A. Clare and R. D. King. Knowledge Discovery in Multi-label Phenotype Data. In *Proceedings of the 5th European Conference on Principles of Data Mining and Knowledge Discovery, PKDD '01*, pages 42–53, London, UK, UK, 2001. Springer-Verlag.
- [25] Commission for Energy Regulation (CER). *Electricity Smart Metering Technology Trials Findings Report*. ESB Networks, Belgard Square North, Tallaght, Dublin 24., 2011.
- [26] D. J. Cook, A. S. Crandall, B. L. Thomas, and N. C. Krishnan. CASAS: A Smart Home in a Box. *Computer*, 46(7):62–69, July 2013.

- [27] K. Coughlin, M. A. Piette, C. Goldman, and S. Kiliccote. *Estimating Demand Response Load Impacts: Evaluation of Baseline Load Models for Nonresidential Buildings in California*. Lawrence Berkeley National Lab, Tech. Rep. LBNL-63728, 2008.
- [28] K. Coughlin, M. A. Piette, C. Goldman, and S. Kiliccote. *Emergency Demand Response Program Manual, Sec 5.2: Calculation of Customer Baseline Load (CBL)*. New York Independent System Operator, Southern California Edison, Tech. Rep., 2010.
- [29] K. Coughlin, M. A. Piette, C. Goldman, and S. Kiliccote. *10-Day Average Baseline and Day-Of Adjustment*. Southern California Edison, Tech. Rep., 2011.
- [30] K. Crammer and Y. Singer. A Family of Additive Online Algorithms for Category Ranking. *J. Mach. Learn. Res.*, 3:1025–1058, Mar. 2003.
- [31] A. Crandall and D. J. Cook. Learning Activity Models for Multiple Agents in a Smart Space. In H. Nakashima, H. Aghajan, and J. Augusto, editors, *Handbook of Ambient Intelligence and Smart Environments*, pages 751–769. Springer US, 2010.
- [32] N. Cristianini and J. Shawe-Taylor. *An Introduction to Support Vector Machines: And Other Kernel-based Learning Methods*. Cambridge University Press, New York, NY, USA, 2000.
- [33] K. Dembczynski, W. Waegeman, W. Cheng, and E. Hüllermeier. On Label Dependence and Loss Minimization in Multi-Label Classification. *Machine Learning*, 88(1-2):5–45, 2012.
- [34] K. Dembczynski, W. Waegeman, W. Cheng, and E. Hüllermeier. On Label Dependence in Multilabel Classification. In *LastCFP: ICML Workshop on Learning from Multi-label data*. Ghent University, KERMIT, Department of Applied Mathematics, Biometrics and Process Control, 2010.
- [35] A. K. Dey. Understanding and Using Context. *Personal Ubiquitous Comput.*, 5(1):4–7, Jan. 2001.
- [36] A. Doryab and J. Togelius. Activity Recognition in Collaborative Environments. In *The 2012 International Joint Conference on Neural Networks (IJCNN)*, pages 1–8, 2012.
- [37] M. Eichelberg, A. Hein, F. Büsching, and L. Wolf. The GAL Middleware Platform for AAL. In *Proceedings of the First International Workshop on AAL Service Platforms (WASP 2010)*, Lyon, France, July 2010.
- [38] A. Elisseeff and J. Weston. A Kernel Method for Multi-Labelled Classification. In T. G. Dietterich, S. Becker, and Z. Ghahramani, editors, *Advances in Neural Information Processing Systems 14 (NIPS-01)*, pages 681–687, 2002.
- [39] F. Englert, T. Schmitt, S. Koessler, A. Reinhardt, and R. Steinmetz. How to Auto-configure Your Smart Home?: High-resolution Power Measurements to the Rescue. In *Proceedings of the Fourth International Conference on Future Energy Systems, e-Energy '13*, pages 215–224, New York, NY, USA, 2013. ACM.

- [40] S. Fan and R. J. Hyndman. Short-Term Load Forecasting Based on a Semi-Parametric Additive Model. *IEEE Transactions on Power Systems*, 27(1):134–141, Feb 2012.
- [41] J. Fogarty, C. Au, and S. E. Hudson. Sensing from the Basement: A Feasibility Study of Unobtrusive and Low-cost Home Activity Recognition. In *Proceedings of the 19th Annual ACM Symposium on User Interface Software and Technology*, UIST '06, pages 91–100, New York, NY, USA, 2006. ACM.
- [42] I. Georgievski, V. Degeler, G. A. Pagani, T. A. Nguyen, A. Lazovik, and M. Aiello. Optimizing Energy Costs for Offices Connected to the Smart Grid. *IEEE Transactions on Smart Grid*, 3(4):2273–2285, Dec 2012.
- [43] S. Godbole and S. Sarawagi. Discriminative methods for multi-labeled classification. In *Pacific-Asia Conference on Knowledge Discovery and Data Mining*, pages 22–30. Springer, 2004.
- [44] I. Goodfellow, Y. Bengio, and A. Courville. Deep Learning. Book in preparation for MIT Press, 2016.
- [45] M. Görtz. *Effiziente Echtzeitkommunikationsdienste durch Einbeziehung von Kontexten*. PhD thesis, Technische Universität, Darmstadt, August 2005.
- [46] T. Hastie, R. Tibshirani, and J. Friedman. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction, Second Edition*. Springer Series in Statistics. Springer, 2009.
- [47] S. Haykin. *Neural Networks and Learning Machines*. Number Bd. 10 in Neural networks and learning machines. Prentice Hall, 2009.
- [48] L. Hernandez, C. Baladron, J. M. Aguiar, B. Carro, A. J. Sanchez-Esguevillas, J. Lloret, and J. Massana. A Survey on Electric Power Demand Forecasting: Future Trends in Smart Grids, Microgrids and Smart Buildings. *IEEE Communications Surveys Tutorials*, 16(3):1460–1495, Third 2014.
- [49] H. Hippert, C. Pedreira, and R. Souza. Neural Networks for Short-term Load Forecasting: a Review and Evaluation. *Power Systems, IEEE Transactions on*, 16(1):44–55, Feb 2001.
- [50] T. Hong, M. Gui, M. Baran, and H. Willis. Modeling and Forecasting Hourly Electric Load by Multiple Linear Regression with Interactions. In *Power and Energy Society General Meeting, 2010 IEEE*, pages 1–8, July 2010.
- [51] T. Hothorn, K. Hornik, and A. Zeileis. Unbiased Recursive Partitioning: A Conditional Inference Framework. *Journal of Computational and Graphical Statistics*, 15(3):651–674, 2006.
- [52] E. Hüllermeier, J. Fürnkranz, W. Cheng, and K. Brinker. Label Ranking By Learning Pairwise Preferences. *Artificial Intelligence*, 172:1897–1916, 2008.
- [53] M. Jentsch, M. Jahn, F. Pramudianto, J. Simon, and A. Al-Akkad. An Energy-saving Support System for Office Environments. In *Proceedings of the Second International Conference on Human Behavior Understanding*, HBU'11, pages 83–92, Berlin, Heidelberg, 2011. Springer-Verlag.



- [54] H. Jiang, Y. Lee, and F. Liu. Anomaly Detection, Forecasting and Root Cause Analysis of Energy Consumption for a Portfolio of Buildings Using Multi-Step Statistical Modeling, Nov. 1 2012. US Patent App. 13/098,044.
- [55] H. Jiang, Y. Lee, and F. Liu. Anomaly Detection, Forecasting and Root Cause Analysis of Energy Consumption for a Portfolio of Buildings Using Multi-step statistical Modeling, 2012. US Patent App. 13/098,044.
- [56] Z. C. Johanyák and S. Kovács. Distance Based Similarity Measures of Fuzzy Sets. 2005.
- [57] E. Keogh, K. Chakrabarti, M. Pazzani, and S. Mehrotra. Dimensionality Reduction for Fast Similarity Search in Large Time Series Databases. January 2001.
- [58] E. Keogh and A. C. Ratanamahatana. Exact Indexing of Dynamic Time Warping. *Knowledge and Information Systems*, 7(3):358–386, 2005.
- [59] A. Khotanzad, R. Afkhami-Rohani, T.-L. Lu, A. Abaye, M. Davis, and D. J. Maratukulam. Annstlf-a Neural-network-based Electric Load Forecasting System. *IEEE Transactions on Neural Networks*, 8(4):835–846, Jul 1997.
- [60] S.-C. Kim, Y.-S. Jeong, and S.-O. Park. RFID-based Indoor Location Tracking to Ensure the Safety of the Elderly in Smart Home Environments. *Personal Ubiquitous Comput.*, 17(8):1699–1707, Dec. 2013.
- [61] M. Kuhn. *caret: Classification and Regression Training*, 2016. R package version 6.0-64.
- [62] U. Kulau, F. Büsching, and L. Wolf. IdealVolting: Reliable Undervolting on Wireless Sensor Nodes. *ACM Trans. Sen. Netw.*, 12(2):11:1–11:38, Apr. 2016.
- [63] J. R. Kwapisz, G. M. Weiss, and S. A. Moore. Activity Recognition Using Cell Phone Accelerometers. *SIGKDD Explor. Newsl.*, 12(2):74–82, Mar. 2011.
- [64] J. Lin, E. Keogh, S. Lonardi, and B. Chiu. A Symbolic Representation of Time Series, with Implications for Streaming Algorithms. In *Proceedings of the 8th ACM SIGMOD Workshop on Research Issues in Data Mining and Knowledge Discovery*, DMKD '03, pages 2–11, New York, NY, USA, 2003. ACM.
- [65] B. Liu, Y. Ma, and C.-K. Wong. Classification using association rules: weaknesses and enhancements. In *Data mining for scientific and engineering applications*, pages 591–605. Springer, 2001.
- [66] D. Lymberopoulos, A. Bamis, and A. Savvides. Extracting Spatiotemporal Human Activity Patterns in Assisted Living Using a Home Sensor Network. *Universal Access in the Information Society*, 10(2):125–138, 2011.
- [67] L. Mainetti, L. Patrono, and I. Sergi. A Survey on Indoor Positioning Systems. In *Software, Telecommunications and Computer Networks (SoftCOM)*, 2014 22nd International Conference on, pages 111–120, Sept 2014.

- [68] J. Mantyjarvi, J. Himberg, and T. Seppanen. Recognizing Human Motion With Multiple Acceleration Sensors. In *Systems, Man, and Cybernetics, 2001 IEEE International Conference on*, volume 2, pages 747–752 vol.2, 2001.
- [69] F. Martinez Alvarez, A. Troncoso, J. Riquelme, and J. Aguilar Ruiz. Energy Time Series Forecasting Based on Pattern Sequence Similarity. *Knowledge and Data Engineering, IEEE Transactions on*, 23(8):1230–1243, Aug 2011.
- [70] U. Maurer, A. Smalagic, D. P. Siewiorek, and M. Deisher. Activity Recognition and Monitoring Using Multiple Sensors on Different Body Positions. In *International Workshop on Wearable and Implantable Body Sensor Networks (BSN'06)*, pages 4 pp.–116, April 2006.
- [71] Merriam-Webster.com. Activity, 2016. [Online; accessed 9-March-2016].
- [72] K. Metaxiotis, A. Kagiannas, D. Askounis, and J. Psarras. *Artificial Intelligence in Short-term Electric Load Forecasting: a State-of-the-art Survey for the Researcher, Energy Conversion and Management*. 2003.
- [73] T. M. Mitchell. *Machine Learning*. McGraw-Hill, Boston, 1997.
- [74] H. Mori and A. Takahashi. Hybrid Intelligent Method of Relevant Vector Machine and Regression Tree for Probabilistic Load Forecasting. In *Innovative Smart Grid Technologies (ISGT Europe), 2011 2nd IEEE PES International Conference and Exhibition on*, pages 1–8, Dec 2011.
- [75] M. M. Moya and D. R. Hush. Network Constraints and Multi-objective Optimization for One-class Classification. *Neural Netw.*, 9(3):463–474, 1996.
- [76] K. P. Murphy. *Machine Learning: A Probabilistic Perspective*. The MIT Press, 2012.
- [77] N. Murty. *Pattern Recognition : an Algorithmic Approach*. Undergraduate topics in computer science. Springer, New York, 2011.
- [78] W. Niu, J. Long, D. Han, and Y.-F. Wang. Human Activity Detection and Recognition for Video Surveillance. In *Multimedia and Expo, 2004. ICME '04. 2004 IEEE International Conference on*, volume 1, pages 719–722 Vol.1, June 2004.
- [79] S. N. Patel, J. A. Kientz, G. R. Hayes, S. Bhat, and G. D. Abowd. Farther Than You May Think: An Empirical Investigation of the Proximity of Users to Their Mobile Phones. In *Proceedings of the 8th International Conference on Ubiquitous Computing, UbiComp'06*, pages 123–140, Berlin, Heidelberg, 2006. Springer-Verlag.
- [80] D. M. W. Powers. *Evaluation: From Precision, Recall and F-Factor to ROC, Informedness, Markedness and Correlation*, 2007.
- [81] S. Preece, J. Goulermas, L. Kenney, D. Howard, K. Meijer, and R. Crompton. Activity Identification Using Body-Mounted Sensors — A Review of Classification Techniques. *Physiological Measurement*, 30:R1–R33, 2009.
- [82] R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2014.

- [83] P. Rashidi and A. Mihailidis. A Survey on Ambient-Assisted Living Tools for Older Adults. *IEEE Journal of Biomedical and Health Informatics*, 17(3):579–590, May 2013.
- [84] J. Read. *Scalable Multi-Label Classification*. PhD thesis, University of Waikato, 2010.
- [85] J. Read, B. Pfahringer, G. Holmes, and E. Frank. Classifier Chains for Multi-label Classification. In *Proceedings of the European Conference on Machine Learning and Knowledge Discovery in Databases: Part II, ECML PKDD '09*, pages 254–269, Berlin, Heidelberg, 2009. Springer-Verlag.
- [86] A. Reinhardt. *Designing Sensor Networks for Smart Spaces - Unified Interfacing and Energy-Efficient Communication between Wireless Sensor and Actuator Nodes*. PhD thesis, Technische Universität, Darmstadt, 2011.
- [87] A. Reinhardt, P. Baumann, D. Burgstahler, M. Hollick, H. Chonov, M. Werner, and R. Steinmetz. On the Accuracy of Appliance Identification Based on Distributed Load Metering Data. In *Sustainable Internet and ICT for Sustainability, 2012*, pages 1–9, Oct 2012.
- [88] A. Reinhardt, D. Burkhardt, M. Zaheer, and R. Steinmetz. Electric Appliance Classification Based on Distributed High Resolution Current Sensing. In *Local Computer Networks Workshops (LCN Workshops), 2012 IEEE 37th Conference on*, pages 999–1005, Oct 2012.
- [89] L. Rokach and O. Maimon. *Data Mining with Decision Trees - Theory and Applications*. 2<sup>nd</sup> Edition, volume 81 of *Series in Machine Perception and Artificial Intelligence*. WorldScientific, 2014.
- [90] A. Rovai, J. Baker, and M. Ponton. *Social Science Research Design and Statistics: A Practitioner's Guide to Research Methods and IBM SPSS*. Watertree Press, 2013.
- [91] Y. Rui and A. El-Keib. A Review of Ann-based Short-term Load Forecasting Models. In *System Theory, 1995., Proceedings of the Twenty-Seventh Southeastern Symposium on*, pages 78–82, Mar 1995.
- [92] S. J. Russell and P. Norvig. *Artificial Intelligence: A Modern Approach*. Pearson Education, 2 edition, 2003.
- [93] S. Saadaoui and L. Wolf. Architecture Concept of a Wireless Body Area Sensor Network for Health Monitoring of Elderly People. In *Proceedings of IEEE Consumer Communications and Networking Conference (CCNC'07)*, Las Vegas, USA, 2007.
- [94] S. Saadaoui and L. Wolf. Scalability Issues in Wireless Body Area Sensor Networks for Remote Monitoring of Elderly People. In *Proceedings of the 1st International Workshop on Wireless Sensor Networks for Health Care (WSNHC 2007), in conjunction with INSSo7, the 4th International Conference on Networked Sensing Systems*, Braunschweig, Germany, June 2007.

- [95] S. Saguna, A. Zaslavsky, and D. Chakraborty. Complex activity recognition using context driven activity theory in home environments. In *Smart Spaces and Next Generation Wired/Wireless Networking*, pages 38–50. Springer, 2011.
- [96] R. E. Schapire and Y. Singer. Boostexter: A boosting-based system for text categorization. *Machine Learning*, 39(2):135–168.
- [97] J. Schmitt. *Anpassungsfähige Kontextbestimmung zur Unterstützung von Kommunikationsdiensten*. PhD thesis, Technische Universität, Darmstadt, December 2009.
- [98] Y. Schröder, G. von Zengen, S. Rottmann, F. Büsching, and L. Wolf. InPhase: An Indoor Localization System based on Phase Difference Measurements. In *Proceedings of the 1st KuVS Expert Talk on Localization*, Lübeck, Germany, April 2015.
- [99] W. Shen, V. Babushkin, Z. Aung, and W. L. Woon. An Ensemble Model for Day-ahead Electricity Demand Time Series Forecasting. In *Proceedings of the Fourth International Conference on Future Energy Systems, e-Energy '13*, pages 51–62, New York, NY, USA, 2013. ACM.
- [100] R. Silipo and P. Winters. Big Data, Smart Energy, and Predictive Analytics Time Series Prediction of Smart Energy Data. Technical report, 2013.
- [101] M. Sioutis and Y. Tan. User indoor location system with passive infrared motion sensors and space subdivision. In *International Conference on Distributed, Ambient, and Pervasive Interactions*, pages 486–497. Springer, 2014.
- [102] J. R. Smith, K. P. Fishkin, B. Jiang, A. Mamishev, M. Philipose, A. D. Rea, S. Roy, and K. Sundara-Rajan. RFID-based Techniques for Human-activity Detection. *Commun. ACM*, 48(9):39–44, Sept. 2005.
- [103] M. Sokolova and G. Lapalme. A Systematic Analysis of Performance Measures for Classification Tasks. *Elsevier Information Processing & Management*, 45(4):427–437, 2009.
- [104] E. Spyromitros-Xioufis, G. Tsoumakas, and I. Vlahavas. An Empirical Study of Lazy Multilabel Classification Algorithms. In *Proceedings of the 5th Hellenic Conference on Artificial Intelligence: Theories, Models and Applications, SETN '08*, pages 401–406, Berlin, Heidelberg, 2008. Springer-Verlag.
- [105] E. M. Tapia, S. S. Intille, and K. Larson. Activity recognition in the home using simple and ubiquitous sensors. In *International Conference on Pervasive Computing*, pages 158–175. Springer, 2004.
- [106] M. E. Tipping. Sparse Bayesian Learning and the Relevance Vector Machine. *J. Mach. Learn. Res.*, 1:211–244, Sept. 2001.
- [107] G. Tsoumakas and I. Katakis. Multi Label Classification: An Overview. *International Journal of Data Warehouse and Mining*, 3(3):1–13, 2007.
- [108] G. Tsoumakas, I. Katakis, and I. Vlahavas. Mining multi-label data. In *Data mining and knowledge discovery handbook*, pages 667–685. Springer, 2009.

- [109] G. Tsoumakas and I. Vlahavas. Random k-Labelsets: An Ensemble Method for Multilabel Classification. In *Proceedings of the 18th European Conference on Machine Learning, ECML '07*, pages 406–417, Berlin, Heidelberg, 2007. Springer-Verlag.
- [110] T. van Kasteren, A. Noulas, G. Englebienne, and B. Kröse. Accurate Activity Recognition in a Home Setting. In *Proceedings of the 10th International Conference on Ubiquitous Computing, UbiComp '08*, pages 1–9, New York, NY, USA, 2008. ACM.
- [111] V. N. Vapnik. *Statistical Learning Theory*. Adaptive and learning systems for signal processing, communications, and control. Wiley, New York, 1998.
- [112] A. Viera and J. Garrett. Understanding interobserver agreement: The kappa statistic. *Family Medicine*, 37(5):360–363, 5 2005.
- [113] G. von Zengen, Y. Schröder, and L. Wolf. InPhase: A Low Cost Indoor Localization System for IoT Devices. In *Microsoft Indoor Localization Competition, Workshop at IPSN 2015*, Seattle, WA ,USA, April 2015.
- [114] S. Wan and X. H. Yu. Facility Power Usage Modeling and Short-term Prediction with Artificial Neural Networks. In *Advances in Neural Networks - ISNN 2010*, pages 548–555, Mar 2010.
- [115] G. M. Weiss, J. L. Timko, C. M. Gallagher, K. Yoneda, and A. J. Schreiber. Smartwatch-based Activity Recognition: A Machine Learning Approach. In *2016 IEEE-EMBS International Conference on Biomedical and Health Informatics (BHI)*, pages 426–429, Feb 2016.
- [116] R. Weron. *Modeling and Forecasting Electricity Loads and Prices: A Statistical Approach*. The Wiley Finance Series. Wiley, 2007.
- [117] Wikipedia. Activities of Daily Living — Wikipedia, The Free Encyclopedia, 2016. [Online; accessed 9-March-2016].
- [118] I. H. Witten and E. Frank. *Data Mining: Practical Machine Learning Tools and Techniques, Second Edition (Morgan Kaufmann Series in Data Management Systems)*. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 2005.
- [119] X. Yang. Comparison of the Ls-svm Based Load Forecasting Models. In *Electronic and Mechanical Engineering and Information Technology (EMEIT), 2011 International Conference on*, volume 6, pages 2942–2945, Aug 2011.
- [120] Y. Yang. An Evaluation of Statistical Approaches to Text Categorization. *Inf. Retr.*, 1(1-2):69–90, May 1999.
- [121] J. Ye, G. Stevenson, and S. Dobson. KCAR: A Knowledge-driven Approach for Concurrent Activity Recognition. *Pervasive and Mobile Computing*, 19:47–70, 2015.
- [122] Z. Younes, F. Abdallah, and T. D. ux. Multi-label Classification Algorithm Derived From K-Nearest Neighbor Rule With Label Dependencies. In *Signal Processing Conference, 2008 16th European*, pages 1–5, August 2008.

- [123] C. R. Yu, C. L. Wu, C. H. Lu, and L. C. Fu. Human Localization via Multi-Cameras and Floor Sensors in Smart Home. In *2006 IEEE International Conference on Systems, Man and Cybernetics*, volume 5, pages 3822–3827, Oct 2006.
- [124] T. yu Wu, C. chun Lian, and J. Y. jen Hsu. Joint Recognition of Multiple Concurrent Activities using Factorial Conditional Random Fields. In *2007 AAAI Workshop on Plan, Activity, and Intent Recognition, Technical Report WS-07-09. The AAAI Press, Menlo Park, 2007.*
- [125] M.-L. Zhang and Z.-H. Zhou. A k-Nearest Neighbor Based Algorithm For Multi-Label Classification. In *IEEE International Conference on Granular Computing*, volume 2, pages 718–721, July 2005.

---

## AUTHOR'S PUBLICATIONS

---

### A.1 MAIN PUBLICATIONS

1. A. Alhamoud, S. Bergsträsser, D. Burgstahler, D. Böhnstedt, F. Englert, A. Reinhardt, J. Schmitt, and R. Steinmetz. CPaaS: Context Processing as a Service. Software Demonstration, NetSys Communication Software Award, Mar 2013.
2. A. Alhamoud, M. Kreger, H. Afifi, C. Gottron, D. Burgstahler, F. Englert, D. Böhnstedt, and R. Steinmetz. Empirical Investigation of the Effect of the Door's State on Received Signal Strength in Indoor Environments at 2.4 GHz. In *Local Computer Networks Workshops (LCN Workshops), 2014 IEEE 39th Conference on*, pages 652–657, Sept 2014.
3. A. Alhamoud, V. Muradi, D. Böhnstedt, and R. Steinmetz. Activity Recognition in Multi-User Environments Using Techniques of Multi-label Classification. In *Proceedings of the 6th International Conference on the Internet of Things, IoT'16*, pages 15–23, New York, NY, USA, 2016. ACM.
4. A. Alhamoud, A. A. Nair, C. Gottron, D. Böhnstedt, and R. Steinmetz. Presence Detection, Identification and Tracking in Smart Homes Utilizing Bluetooth Enabled Smartphones. In *Local Computer Networks Workshops (LCN Workshops), 2014 IEEE 39th Conference on*, pages 784–789, Sept 2014.
5. A. Alhamoud, F. Ruettiger, A. Reinhardt, F. Englert, D. Burgstahler, D. Böhnstedt, C. Gottron, and R. Steinmetz. SMARTENERGY.KOM: An Intelligent System for Energy Saving in Smart Home. In *Local Computer Networks Workshops (LCN Workshops), 2014 IEEE 39th Conference on*, pages 685–692, Sept 2014.
6. A. Alhamoud, P. Xu, F. Englert, P. Scholl, T. A. B. Nguyen, D. Böhnstedt, and R. Steinmetz. Evaluation of User Feedback in Smart Home for Situational Context Identification. In *Pervasive Computing and Communication Workshops (PerCom Workshops), 2015 IEEE International Conference on*, pages 20–25, March 2015.
7. A. Alhamoud, P. Xu, A. Reinhardt, F. Englert, P. Scholl, D. Böhnstedt, and R. Steinmetz. Extracting Human Behavior Patterns from Appliance-level Power Consumption Data. In *Wireless Sensor Networks.12th European Conference EWSN 2015*, pages 52–67. Springer, February 2015.

### A.2 CO-AUTHORED PUBLICATIONS

8. D. Burgstahler, S. Bergsträsser, D. Böhnstedt, A. Reinhardt, A. Alhamoud, F. Englert, J. Schmitt, and R. Steinmetz. CPaaS: Context Processing as a Service. *PIK - Praxis der Informationsverarbeitung und Kommunikation*, 36:47, Feb 2013.

9. F. Englert, I. Diaconita, A. Reinhardt, A. Alhamoud, R. Meister, L. Backert, and R. Steinmetz. Reduce the Number of Sensors: Sensing Acoustic Emissions to Estimate Appliance Energy Usage. In *Proceedings of the 5th ACM Workshop on Embedded Systems For Energy-Efficient Buildings, BuildSys'13*, pages 12:1–12:8, New York, NY, USA, 2013. ACM.
10. F. Englert, A. El'Hindi, D. Burgstahler, A. Alhamoud, and R. Steinmetz. Reducing the Electricity Consumption of Large outdoor LED Advertising Screens. In *Proceedings of the 5th International Conference on Future Energy Systems, e-Energy '14*, pages 209–210, New York, NY, USA, 2014. ACM.
11. F. Englert, P. Lieser, A. Alhamoud, D. Böhnstedt, and R. Steinmetz. Electricity-Metering in a Connected World: Virtual Sensors for Estimating the Electricity Consumption of IoT Appliances. In *Future Internet of Things and Cloud (FiCloud), 2015 3rd International Conference on*, pages 317–324, Aug 2015.
12. F. Englert, P. Lieser, M. Rettberg-Päplow, A. Alhamoud, D. Böhnstedt, and R. Steinmetz. Lightmon: Apportioning the Effect of Light Switching Events on the Electricity Consumption of Buildings. In *EWSN '16 Proceedings of the 2016 International Conference on Embedded Wireless Systems and Networks*, pages 77–88, Feb 2016.
13. F. Englert, M. Rettberg-Päplow, S. Kössler, A. Alhamoud, T. A. B. Nguyen, D. Böhnstedt, and R. Steinmetz. Enhancing User Privacy by Data Driven Selection Mechanisms for Finding Transmission-Relevant Data Samples in Energy Recommender Systems. In *Networked Systems (NetSys), 2015 International Conference and Workshops on*, pages 1–6, March 2015.
14. P. Lieser, F. Englert, A. Alhamoud, D. Burgstahler, and D. Böhnstedt. Towards Virtual Sensors for Estimating the Electricity Consumption of Networked Appliances. In *Proceedings of the 2015 ACM Sixth International Conference on Future Energy Systems, e-Energy '15*, pages 205–206, New York, NY, USA, 2015. ACM.



---

**CURRICULUM VITÆ**

---

**PERSONAL**

Name	Alaa Alhamoud, M.Sc.
Date of Birth	29.06.1985
Place of Birth	Homs, Syria
Nationality	Syrian

**EDUCATION**

Since 11/2012	Technische Universität Darmstadt, Doctoral candidate at the Department of Electrical Engineering and Information Technology
10/2009 - 03/2012	RWTH Aachen University, Master studies in Media Informatics, Degree: Master of Science
09/2003 - 08/2008	Al-Baath University, Bachelor studies in Information Technology Engineering, Degree: Bachelor of Science
09/2000 - 06/2003	Al-Faraby High School, Homs, Syria, Degree: Bacculaureate

**WORK EXPERIENCE**

Since 11/2015	Technische Universität Darmstadt, Research assistant at the researcher group Distributed Sensing Systems at the Multimedia Communications Lab (KOM)
06/2010 - 04/2011	RWTH Aachen University, Student research assistant at Embedded Software Labora- tory

**TEACHING ACTIVITY**

2013 - 2016	Teaching Assistant, Lecture and Exercise, "Communication Networks I", SS 2013, SS 2014, SS 2015, SS 2016
-------------	--

- 2014 - 2015 Teaching Assistant, Labs and Projects,  
"Multimedia Communications Lab I/II",  
WS 2015/16, SS 2015, WS 2014/15
- SS 2015 Teaching Assistant, Seminar,  
"Current Topics in Web Applications, Information Manage-  
ment and Semantics"
- WS 2013/14 Teaching Assistant, Seminar,  
"Advanced Topics in Future Internet Research"

## SCIENTIFIC ACTIVITIES

- 2014 Reviewer for IEEE 78th Vehicular Technology Conference:  
VTC-Fall, 2013
- 2014 Reviewer for IEEE 77th Vehicular Technology Conference:  
VTC-Spring, 2013

---

## SUPERVISED STUDENTS

---

1. Chadi Izzou Rhaibani. *Occupancy Detection and Prediction based on Fine-granular Power Consumption Profiles*, Master thesis
2. Hatem Khalloof. *An Extensive Analytical Study of Dataport Electricity Dataset for Accurately Predicting Short- and Long-Term Power Consumption and Generation Trends*, Master thesis
3. Shadi Shahoud. *Enhancing Predictive Performance of Multi-Label Classification by Extracting, Representing and Exploiting Label Dependency in Multi-Label Datasets*, Master thesis
4. Hosam Nima. *Situation Detection based on Activity Recognition in Disaster Scenarios*, Master thesis
5. Vaidehi Muradi. *Activity Recognition in Multi-User Environments Using Multi-Label Classification Techniques*, Master thesis
6. Wael Alkhatib. *Hybrid Large Scale Power Consumption Forecasting*, Master thesis
7. Tomislav Hasan. *Scalable Distributed Smart Home Monitoring and Auto-configuration*, Master thesis
8. Dan Le. *Activity Detection in Multi-User Environments*, Master thesis
9. Pei Xu. *An Extensive and In-depth Analytical Study of Tracebase 2.0 Dataset*, Master thesis
10. Haitham Afifi. *Evaluation of Link Quality in Indoor Environments for Wi-Fi Signal*, Bachelor thesis
11. Felix Ruettiger. *Design and Implementation of a Framework for Energy Efficiency in Smart Living Spaces based on Wireless Sensor Networks and Activity Recognition*, Master thesis



---

ERKLÄRUNG LAUT §9 DER PROMOTIONSORDNUNG

---

Ich versichere hiermit, dass ich die vorliegende Dissertation allein und nur unter Verwendung der angegebenen Literatur verfasst habe. Die Arbeit hat bisher noch nicht zu Prüfungszwecken gedient.

*Darmstadt, 14. September 2016*

---

Alaa Alhamoud