Tampere University

Anna Virtanen

PREDICTING THE WORKLOAD OF HOME CARE PERSONNEL USING TIME SERIES FORECASTING

Master of Science Thesis
Faculty of Information Technology and Communication Sciences
Examiners: Tarmo Lipping, Jari Turunen
February 2023

ABSTRACT

Anna Virtanen: Predicting the Workload of Home Care Personnel Using Time Series Forecasting
Master of Science Thesis
Tampere University
Electrical Engineering
February 2023

The substantial increase in the number of elderly leads to increased demand for home care and institutional care services. Forecasting the workload of social and health care personnel is difficult, and the allocation of resources is currently based on the experience accumulated by experts using their "gut" feeling.

Workload prediction can be facilitated with artificial intelligence and machine learning based solutions. In this Master of Science Thesis, the workload of the home care personnel of The Joint Municipal Authority for Social and Healthcare in Central Uusimaa (Keusote) is predicted four months ahead. The forecast is produced for all of Keusote's 27 home care sites using two different methods: a traditional ARIMA model and a modern LSTM neural network.

The models were compared by computing the root-mean-squared error, mean absolute error, mean absolute per cent error and goodness of fit measures for the forecasts produced for each home care unit. The key indicators gave slightly better results for the ARIMA model, but when the forecasts were plotted side by side with the actual workload, it was noticed that the ARIMA-based forecasts often did not have the correct DC level, they diverged from the actual workload, started to oscillate or were constant throughout the forecasting horizon. Therefore the LSTM model, which gave almost equally good results, was selected for production.

In the future, forecasting will probably be based on the Resident Assessment Instrument (RAI) measurements of home care clients. The next step would be to generate a forecast utilising a Markov model, which is a probability-based state transition model.

Keywords: time series, forecasting, demand forecasting, home care, ARIMA, LSTM

TIIVISTELMÄ

Anna Virtanen: Kotihoidon henkilöstön työmäärän ennustaminen aikasarjamenetelmällä
Diplomityö
Tampereen yliopisto
Sähkötekniikka
Helmikuu 2023

Ikääntyneiden määrän voimakas kasvu johtaa kotihoidon ja laitoshoidon palveluiden kysynnän lisääntymiseen. Sosiaali- ja terveydenhoidon henkilöstön työmäärän ennustaminen on vaikeaa, ja ennustaminen perustuu tällä hetkellä asiantuntijoille muodostuneeseen mututuntumaan.

Työmäärän ennustamista voidaan helpottaa tekoälyyn ja koneoppimiseen perustuvilla ratkaisuilla. Tässä diplomityössä ennustetaan Keski-Uudenmaan sote -kuntayhtymän (Keusote) kotihoidon henkilöstön työmäärää neljä kuukautta eteenpäin. Ennuste tuotetaan kaikille Keusoten 27:lle kotihoidon yksikölle kahdella eri metodilla: perinteisellä ARIMA-mallilla ja modernilla LSTM-neuroverkolla.

Malleja vertailtiin keskenään laskemalla kullekkin kotihoidon yksikölle tuotetulle ennusteelle tunnuslukuina keskineliövirheen neliöjuuri, keskipoikkeama, keskipoikkeaman prosenttivirhe ja yhteensopivuuden aste. Tunnusluvut antoivat lievästi parempia tuloksia ARIMA-mallille, mutta kun ennusteet piirrettiin vierekkäin todellisen työmäärän kanssa, huomattiin, että ARIMA-pohjaisissa ennusteissa DC-taso ei usein ollut oikea, ennusteet poikkesivat todellisesta työmäärästä, alkoivat oskilloida tai olivat vakioita koko ennustehorisontin ajan. Siksi tuotantoon valittiin lähes yhtä hyviä tuloksia antanut LSTM-malli.

Tulevaisuudessa ennustaminen perustuu todennäköisesti kotihoidon asiakkaiden Resident Assessment Instrument (RAI) -mittauksiin. Seuraava askel olisi ennusteen tuottaminen Markovin mallilla, joka on todennäköisyyslaskentaan perustuva tilasiirtymämalli.

Avainsanat: aikasarja, ennustaminen, kysynnan ennustaminen, kotihoito, ARIMA, LSTM

PREFACE

This Master of Science thesis marks the culmination of nearly eight years of university education and constitutes a significant milestone in my academic journey. I would like to thank my employer and Keusote for providing me with this interesting thesis topic.

I am particularly grateful to my thesis supervisors, Tarmo Lipping and Jari Turunen, for their guidance, support and constructive feedback throughout this project. I would also like to express my sincere thanks to my colleagues at work for taking the time to review this thesis and provide their insightful comments.

In addition to the support from others, I would like to extend a special acknowledgement to myself for my dedication and hard work. This thesis is the product of countless hours of research, programming, writing and reflection, and I am proud of what I have accomplished.

Finally, I would like to acknowledge the unwavering support and encouragement provided by my friends and family. Their belief in my abilities and constant encouragement has been a source of motivation and inspiration throughout this challenging yet rewarding journey.

Tampere, 9th February 2023

Anna Virtanen

CONTENTS

GLOSSARY

AIC	Akaike Information Criterion
ANN	Artificial Neural Networks
AR	Autoregression
ARIMA	Autoregressive Integrated Moving Average
ARMA	Autoregressive Moving Average
BIC	Bayesian Information Criterion
CAP	Client Assessment Protocol
CPS	Cognitive Performance Scale
HQIC	Hannan-Quinn Information Criterion
Keusote	The Joint Municipal Authority for Social and Healthcare in Central Uusimaa
LSTM	Long Short-Term Memory
MA	Moving Average
MAE	Mean Absolute Error
MAPE	Mean Absolute Percentage Error
Q-Q	Quantile-Quantile
RAI	Resident Assessment Instrument
ReLU	Rectified Linear Unit
RMSE	Root-Mean-Squared Error
RNN	Recurrent Neural Network
RUG	Resource Utilization Group
SQL	Structured Query Language
tanh	Hyperbolic tangent
THL	Finnish Institute for Health and Welfare

1. INTRODUCTION

According to Statistics Finland's latest population projection, the number of elderly will grow strongly in the next 50 years. The increase in the number of elderly is caused by the low birth rate in Finland in the last few years. The effects of the low birth rate in Finland are reflected in the demographic dependency ratio, that is, the number of people under the age of 15 and over the age of 65 per 100 people of working age. At the end of the year 2019, there were 871,036 people under the age of 15 and 3,422,982 between the ages of 15 and 64 in Finland. 1,231,274 people were over the age of 65 at the end of 2019. The demographic dependency ratio was 61.4 in the year 2019 which is the highest since 1922. [1] [2]

Due to large increases in the number of elderly, the need for elderly care services will grow drastically. According to the statistics of the Ministry of Economic Affairs and Employment of Finland, Finland is in need of 20 000 new employees in the social and health care service sector [3]. However, care work is ill-payed, demanding and mentally and physically burdensome. In fact, many care workers intend to leave their employment due to poor working conditions, insufficient living wages and, at the same time, large groups of aged workers will retire. In addition to difficulties with finding care personnel, many regional authorities and local municipalities struggle with diminishing resources and an increased need for care. With these trends ongoing, Finland needs to restructure current service models and produce an increasing amount of care services for the ageing population with more efficiency than before. [4]

A significant problem in home care work management is the difficulty in estimating the workload of the care personnel. Many customers have changes in service needs, there are different services, care personnel's visits have different duration and the nature of performed services is different. In addition, there are changes in service areas, strategic decisions and operational problems in service housing. Currently, the proper allocation of resources is based on the experience accumulated by experts using their "gut" feeling.

Due to difficulties in estimating the workload in home care, many social and health care service providers are interested in machine learning and artificial intelligence based help in decision-making and resource management processes. Machine learning and artificial intelligence have been utilized to an increasing extent in the social and health care sector, especially in demand forecasting. Machine learning produced forecasts can be used to balance the workload for those offices or tasks where the need for resources is the greatest. Additionally, low-priority routine check-ups could be planned to more suitable time windows.

This Master of Science Thesis studies the use of two machine learning models to forecast the workload of home care personnel for four months ahead. The study was conducted for the Joint Municipal Authority for Social and Healthcare in Central Uusimaa (Keusote) for 27 different home care facilities. Two machine learning models were trained for the problem: a traditional ARIMA model and a modern LSTM neural network. At the end of the experiment, the better-performing model was selected for production.

This Thesis begins by giving an introduction to the home care services for the elderly in Finland in chapter 2. Chapter 3 introduces traditional time series forecasting methodologies as well as a modern approach with an LSTM neural network. The execution of the study is presented in chapter 4, and the results are given in chapter 5. Chapter 6 concludes the Thesis with discussion and conclusion.

2. HOME CARE SERVICES

In the Nordic countries, social and health care services are largely provided as a public service. Services are tax-funded to ensure that uniform, high-quality and low-cost care is available to all citizens, instead of relying on person's financial status. The honest principle is that all people use the same services and are treated the same way in similar care situations.

However, in Finland the social and health care service sector is facing challenges in the forthcoming years. According to Statistics Finland, the number of elderly people will grow strongly in the following decades. The increase in the number of elderly is caused by the low birth rate in Finland in the last years (see Figure 2.1), and its effects are reflected in the demographic dependency ratio. [1] [2]

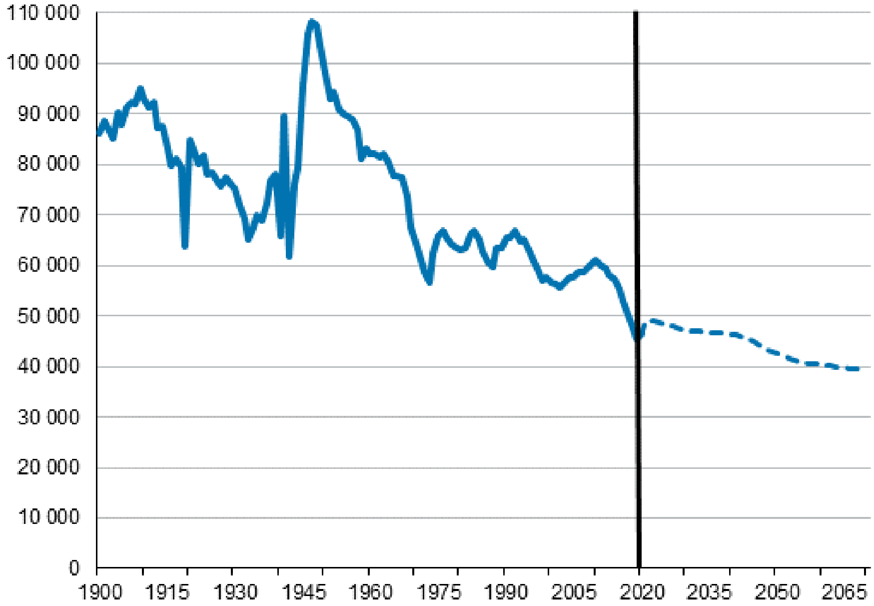


Figure 2.1. *Number of live births in Finland from 1900 to 2020 and a projection until 2070 [1].*

Figure 2.2 highlights more the vast increase in the elderly and the problems with the demographic dependency ratio. It shows those aged 15 or under and those aged 65 or over per 100 working-age persons from 1900 to 2020 and a projection until 2070.

Before the 1990s, home care and residential care were largely provided by regional au-

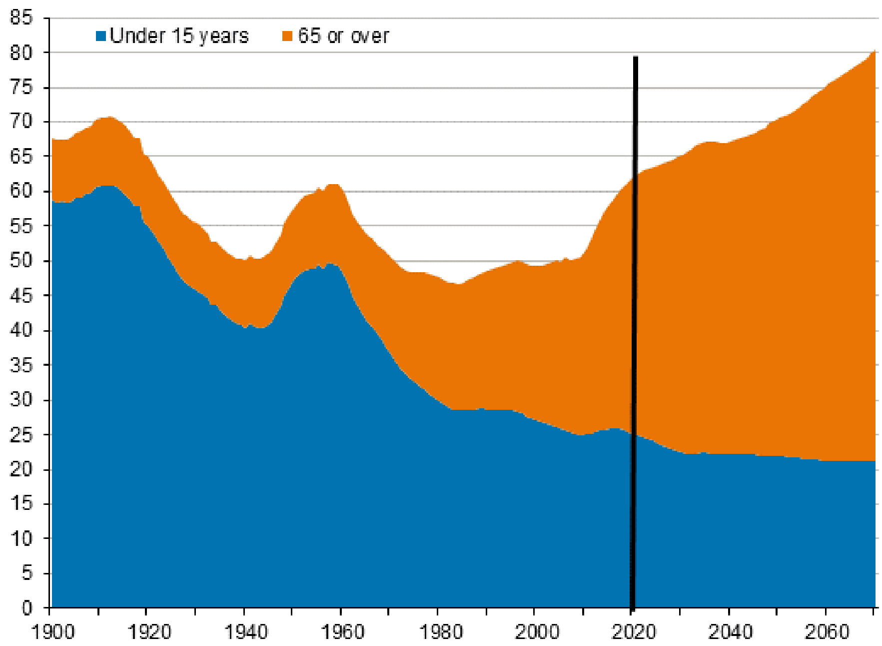


Figure 2.2. *Demographic dependency ratio from 1900 to 2020 and a projection until 2070 [1].*

thorities and local municipalities. Additionally, many non-profit organisations played an important role in providing both home care and residential care services. Currently, however, home care and residential care services are increasingly provided by for-profit service providers and companies. The marketisation of public services has been further developed by the economic acts many governments have to have adopted to cope with the long-lasting crisis in the public financing of social and health care services. An important after-effect of home-care marketisation is that the non-profit service providers have to match up to the for-profit companies to be successful in competitive bids. [4]

2.1 Home Care Workload Estimation in Keusote

The Joint Municipal Authority for Social and Healthcare in Central Uusimaa (Keusote) provides social and health care services in the Järvenpää, Hyvinkää, Nurmijärvi, Tuusula, Mäntsälä and Pornainen regions. Services are organized regionally with the goal of effective and cost-efficient services. The basic agreement of the municipal association was concluded between the municipalities in 2017, and the responsibility for organizing services was transferred from the municipalities to the municipal association on January 1, 2019.

With the establishment of the municipal association, preparations were also made for the social security reform. According to social security legislation that came into use in 2021, the responsibility for organizing social and health care services will be transferred to the well-being services county of Central Uusimaa. The transition will take place in the same geographical area where the municipal corporation already operates. The new Central Uusimaa welfare area has started organizing services on January 1, 2023.

Keusote's housing services for the elderly include home care and residential care, such as assisted living facilities, enhanced service housing and institutional care. Housing services can be organized by Keusote or a private service provider. 24-hour housing services are aimed at customers who need regular care services, whereas home care is mainly intended for elderly people who need assistance to support their everyday life at home. [5]

The Resident Assessment Instrument (RAI) evaluation system is an international evaluation and monitoring system used to determine an aged person's need for services, functional capacity, health status and reserves of energy and strength. It was developed by the international research organization InterRAI [6] and it is nationally used in elderly care, for example, in the United States, most Canadian states, Iceland, Belgium and New Zealand. Additionally, the system is currently being evaluated through research and testing in various other countries worldwide. In 2018, about 35 % of home care clients and about 40 % of 24-hour care clients were assessed with the RAI evaluation system in Finland [7].

In practice, the RAI evaluation system employs a structured query template, which comprises hundreds of questions. Based on the client's responses, the system generates different internationally validated measures. Based on these measures, the client's treatment, rehabilitation and service plan is made and the daily nursing work is planned out.

Based on the results of multiple interviews conducted with Keusote's home care management professionals, the RAI measures that most indicate a high risk of the need for home care services are presented in Table 2.1

Table 2.1. *RAI measures that most indicate a high risk of home care service needs gathered from interviews with Keusote's home care management professionals.*

RAI measure	meaning	scale (low to high)
CAP	Risk of Institutional Care	-
CHESS	Changes in Health, End-stage disease and Symptoms and Signs	0–5
CPS	Cognitive Performance Scale	6-0
CPS0	Cognitive Performance Scale	1-0
MAPLe5	Method for Assigning Priority Levels	1-5

Table 2.1 shows five RAI measures. Client Assessment Protocols or Collaborative Action Points (CAP) are RAI-based triggers used to provide information about the client's

resources and risk factors. They help to create an overall picture of the client's situation and identify the client's strengths as well as problems that require action. [8]

The Changes in Health, End-stage disease and Symptoms and Signs (CHESS) describe the stability of the client's health status. It includes questions about the swelling of lower limbs, shortness of breath, weight loss, reduced fluid intake or food consumption, as well as, changes in decision-making abilities and/or daily activities compared to the previous evaluation. Additionally, the CHESS measure takes into consideration if the life expectancy of the client is less than 6 months. [8]

The Cognitive Performance Scale (CPS) measure describes an aged person's cognition in terms of 1) short-term memory, 2) being understood, 3) the ability to make decisions, 4) the ability to eat independently and 5) the general level of consciousness. The CPS measure helps to identify cognitive decline and potential memory disorder and its severity. The weaker the client's cognition, the higher the value of the CPS meter is. [8]

Method for Assigning Priority Levels (MAPLe5) measure describes factors that have an effect on the client's ability to live at home. Additionally, it also describes the type of resources the client has. [8]

RAI information is also used to support decision-making and management in home care services. For example, the cost-effectiveness and quality of care work can be evaluated with the RAI tool. Resource Utilization Groups (RUG) categorization is a RAI-based measure that is used to determine the number of care personnel and their expertise at a home care site. The RUG categorization uses information about the client's state of health, physical and psychological functioning and special treatments and needs for care. The client is placed in a certain RUG category based on how much human resources are needed to carry out the care and services according to their needs. Personnel resources include for example the number of personnel, their expertise and the time spent on the client's care. RUG categorization defines the resource needs for care at the home care unit level, and it is not suitable for describing the needs of individual clients. The seven RUG clinical main groups, also affecting the need for services, are described in Table 2.2 [8]

An average home care client in the Keusote area has three visits in a week and an average visit takes 21 minutes. Most clients are female and belong to the age group 85-89. The average number of visits and the time spent with the client within a week grouped by age are gathered in Tables 2.3 and 2.4, respectively. The Tables show that the number of visits and the time home care spends with the client increase with age.

Figure 2.3 demonstrates the number of persons and the number of home care visits within a week grouped by age. The age group 85-89 years includes the most people and the majority of them have 1-2 daily visits from home care on average. The number of home

Table 2.2. *The seven RUG clinical main groups.*

RUG clinical main groups
Multidisciplinary rehabilitation
Very high demanding care
Special care
Clinically diverse
Weakening of cognition
Behavioral disorders
Weakening of physical performance

Table 2.3. *The average and the median number of visits grouped by age in the Keusote area from 5.5.2021 to 1.9.2022.*

age group	average	median
65-69	11	7
70-74	11	8
75-79	12	12
80-84	13	14
85-89	13	14
90-94	14	14
95-99	14	14

care visits increases with age, and only a small minority have home care visits more than four times a day on average. This may be due to clients shifting to institutional care with the decline in health or due to natural demise.

Figure 2.4 illustrates the number and the duration of home care visits per age group. It shows that the number of visits is the highest among patients in the age group 85-89. This is most likely caused by this age group including the majority of home care clients. The number of visits grows by the increase of medium-length 10-29 minute visits until the age group 85-89, and at 90 years and over, the number of visits decreases, again most likely due to clients moving to institutional care or due to natural demise.

Home care work management is faced with a significant challenge in accurately assessing the workload of care personnel. This is due to a variety of factors, such as the diversity of clients and their service needs, the diversity of services offered, changes in service areas, strategic decisions, and operational issues in service housing. Currently, the allocation of resources is based primarily on the accumulated experience of experts and their intuition.

Given these difficulties in estimating workload in home care, there is a growing interest among social and health care service providers in utilizing machine learning and artificial

Table 2.4. *The average and median time spent at a client's house within a week grouped by age in the Keusote area from 5.5.2021 to 1.9.2022.*

age group	average (h)	median (h)
65-69	3	2
70-74	4	3
75-79	4	3
80-84	4	4
85-89	4	4
90-94	5	4
95-99	5	4

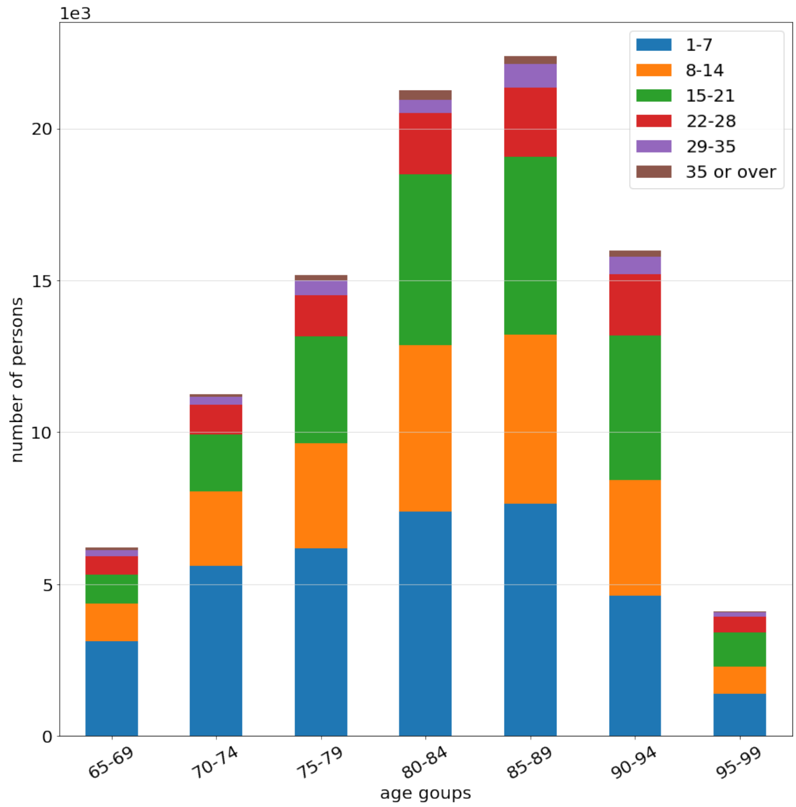


Figure 2.3. *Number of persons and home care visits per week grouped by age in Keusote area from 5.5.2021 to 1.9.2022.*

intelligence to aid in decision-making and resource management processes. Workload forecasting is the process of estimating the amount of work or demand that will be placed on a system or organization in the future. This can include forecasting the number of customers, orders, or projects that will need to be handled, as well as the resources (such as personnel or equipment) that will be required to meet that demand. The goal of workload forecasting is to help organizations plan and prepare for future demand, and to ensure that they have the necessary resources to meet future requirements.

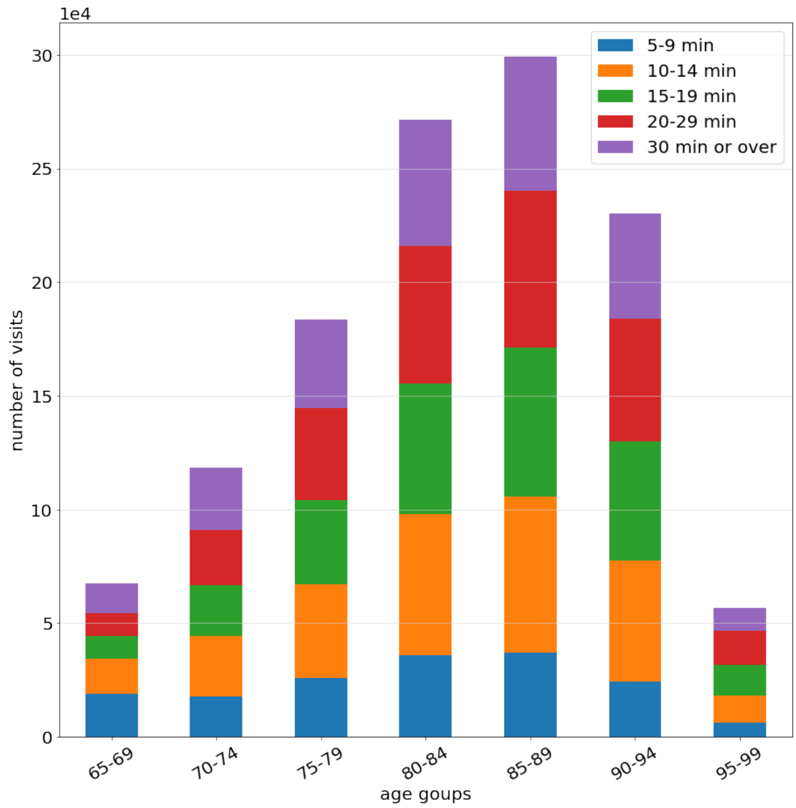


Figure 2.4. *Number and duration of home care visits grouped by age in Keusote area from 5.5.2021 to 1.9.2022.*

2.2 Previous Studies

Workload forecasting has been utilized in many sectors. For instance, an arrival count model for call centres based on a mixed Poisson approach has been introduced in literature [9]. It involves forecasting both arrival counts and average serving times. In the context of home care, this could mean forecasting how many clients will need to be visited, and how long each visit will take.

Forecasting the visits to emergency departments has been featured in many sources, such as [10], which introduces a one-step-ahead forecasting model with the Facebook Prophet machine learning library, and [11] which compares a Long Short-Term Memory (LSTM) network to a Temporal Fusion Transformer (TFT) architecture. In [12] Padthe et al. introduce a way of forecasting with Generalized Linear Model with Poisson Regression (GLM) with different regularization variants. Padthe et al. also experiment with forecasting with a linear Gradient Boosting Machine (GBM). These results indicate that machine learning can be effectively utilized in workload forecasting and further research in this area is needed.

In addition to demand forecasting, the classification of patients according to patient load as well as direct patient load forecasting has also been studied. In [13], Floa et al. review internationally published literature to identify the patient classification systems used to classify nursing intensity and the assessment of staffing resources currently used in home health care. They highlight the validity and reliability of staffing resources and staff allocation. Floa et al. found out that the included papers (13) were considered to be appropriate for the measurement of patients' needs as well as for assessing workload and proper allocation of staff.

Mtonga et al. propose in their paper [14] a deep convolutional neural network (CNN) model for patient load forecasting, that is capable of predicting the future patient load given the current and historical patient load data. Furthermore, Mtonga et al. present a framework for an Internet of Things (IoT) based smart bus transport system, allowing patients stable enough to transfer to query bus location and time-related information and to travel from one health facility to another to avoid overgrowing.

Overall, effective workload forecasting is crucial for maintaining the quality and accessibility of social and health care services, and for improving the overall efficiency of health care organizations. Workload forecasting can be utilized, for example, to determine weekly staffing levels. A mismatch between service demand and staffing ratios leads to overcrowding, excessive waiting times, incomplete preventive service delivery and irritated staff.

3. TIME SERIES FORECASTING

Time series forecasting is an essential part of businesses nowadays. It involves using historical data through statistical analyses and modelling to make assumptions about the future. At the time of computing the forecast, the actual outcome is often not available. A traditional example of time series forecasting is an extrapolation of past weeks' mean temperature values in predicting next week's mean temperature of the day. Forecasting with machine learning and deep learning methods has gained a lot of attention in the past few years. However, traditional forecasting methods are still widely used in many applications. [15]

3.1 Time Series Data

Time series data is a collection of recorded sequential observations over a particular time interval. This ordered set of data points indicates, how things change over time. Time series data is an important source of information in many fields of industry, such as marketing, finance, meteorology, telecommunications, robotics and healthcare. [16]

Forecasting models are characterized by univariate and multivariate types in nature. Univariate models are often more suitable for simpler analyses, whereas advanced multivariate models are used more and more often in solving industry problems. [17]

3.2 Forecasting Methodologies

Traditionally, forecasting applications and online tutorials focus on one-step forecasting, where a forecasting model is trained to forecast one time step ahead. Industrial applications often require more complicated models: the forecasting horizon is typically at least a few weeks, months or even a year ahead. [17]

A multi-step forecasting model can be built in a direct or recursive manner. Direct methods refer to a strategy where each forecasting horizon is computed independently from the other forecasting horizons.

Let us denote a time series of n observations by $[y_{t-1}, ..., y_{t-n}]$ and $h \geq 0$ as the forecasting horizon. To forecast h time steps $[y_t, \ldots, y_{t+h}]$ ahead with the direct method, $h + 1$ independent forecasting models f_{h+1} are learned, and each independent model forecasts a single time step in the forecasting horizon:

$$
\begin{aligned}
\hat{y}_t &= f_1(y_{t-1}, y_{t-2}, ..., y_{t-n}) \\
\hat{y}_{t+1} &= f_2(y_{t-1}, y_{t-2}, ..., y_{t-n}) \\
&. \\
&. \\
&. \\
\hat{y}_{t+h} &= f_{h+1}(y_{t-1}, y_{t-2}, ..., y_{t-n}),
\end{aligned}
$$

where $\hat{y}_t$ is the forecasted time step at time t.

The direct method does not use any predicted values to compute the forecasts and therefore, possible prediction errors will not accumulate in the process. However, training one model for each time step of the forecasting horizon is computationally burdensome, especially when the forecasting horizon is far in the future. Additionally, using separate models means that there is no possibility to model the dependencies between the predictions. [17]

The recursive strategy is the oldest and most intuitive strategy. In this approach, first a single model f is trained to forecast one time step y_t ahead. Then, to forecast the following time step y_{t+1}, the value previously forecasted in the first step is used as part of the input steps and the forecast is computed using the same one-step ahead model. By continuing this manner for h time steps, the forecast for the entire horizon can be produced:

$$
\begin{aligned}
\hat{y}_t &= f(y_{t-1}, y_{t-2}, ..., y_{t-n}) \\
\hat{y}_{t+1} &= f(\hat{y}_t, y_{t-1}, ..., y_{t-n+1}) \\
\hat{y}_{t+5} &= f(\hat{y}_{t+4}, ..., \hat{y}_{t+1}, \hat{y}_t, ..., y_{t-n+5}) \\
&. \\
&. \\
&. \\
\hat{y}_{t+h} &= f(\hat{y}_{t+h-1}, ..., \hat{y}_{t+1}, \hat{y}_t, ..., y_{t-n+h}), n > h.
\end{aligned}
$$

As opposed to the direct strategy, the recursive method is able to learn the dependencies in the time series data and is much faster to train. However, the recursive strategy uses approximations of previously computed forecasts and therefore makes it prone to accumulation of errors. [17]

3.3 Conventional Linear Methods

The most basic building blocks of univariate time series forecasting models are the autoregression (AR) and the moving average (MA) models. Autoregression assumes that future observations are related to prior observations through a linear relationship. Autoregressive model of order p, AR(p), can be written as

$$y_t = c + \phi_1 y_{t-1} + \phi_2 y_{t-2} + ... + \phi_p y_{t-p} + \varepsilon_t, \tag{3.1}$$

where c is a constant, $\phi_1, \phi_2, .., \phi_p$ are the parameters of the model, $y_{t-1}, y_{t-2}, .., y_{t-p}$ are the lag variables included in the model and ε_t is the white noise. [18]

The moving average model focuses on the model error of past predictions and uses them as an impulse for forecasted values. Moving average model of order q, MA(q), can be described by

$$y_t = c + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + ... + \theta_q \varepsilon_{t-q}, \tag{3.2}$$

where $\theta_1, \theta_2, .., \theta_q$ are the parameters of the model and $\varepsilon_t, \varepsilon_{t-1}, .., \varepsilon_{t-q}$ are the white noise error terms. [18]

The autoregressive moving average (ARMA) model is a combination of the AR and MA models. The main principle of combining these models is simply that two models are more powerful than one model. The ARMA model is defined by the following equation:

$$y_t = c + \phi_1 y_{t-1} + ... + \phi_p y_{t-p} + \theta_1 \varepsilon_{t-1} + ... + \theta_q \varepsilon_{t-q} + \varepsilon_t. \tag{3.3}$$

The autoregressive integrated moving average (ARIMA) model assumes that future observations can be represented as a linear function of the differenced observations and residual errors at prior time steps. It is generally considered as a development of the simpler ARMA model. Furthermore, ARIMA includes the idea of integration. [15]

In this connection, integrating (the letter I in ARIMA) stands for a mathematical synonym for differencing a non-stationary time series. A stationary time series does not show a long-term trend, and time series can be made stationary by applying differencing: all actual values are replaced by the difference between the present and the previous value. The order of the difference is often denoted as d. ARIMA(p, d, q) model is described with the following equation:

$$y'_t = c + \phi_1 y'_{t-1} + ... + \phi_p y'_{t-p} + \theta_1 \varepsilon_{t-1} + ... + \theta_q \varepsilon_{t-q} + \varepsilon_t, \tag{3.4}$$

where y'_t is the differenced time series. Sometimes taking the difference more than once may be necessary to achieve a stationary series. However, it is uncommon to go beyond second-order differences anywhere in practice. [18]

Equation 3.4 is often written in the backshift notation as

$$(1 - \phi_1 B - ... - \phi_p B^p)(1 - B)^d \hat{y}_t = c + (1 + \theta_1 B + ... + \theta_q B^q)\varepsilon_t, \tag{3.5}$$

where $\hat{y}'_t = (1 - B)^d \hat{y}_t$. The first part in equation 3.5 represents the AR(p) part, the second differences (d) and the third MA(q) part. [18]

In summary, the parameters of the ARIMA(p, d, q) model are specified as follows: p stands for the number of lag variables included in the ARIMA model, often called the lag order. Variable d denotes the difference between consecutive observations and is generally referred to as the degree of differencing. Letter q stands for the magnitude of the moving average window, and it is appointed as the order of the moving average.

3.4 Artificial Neural Networks

Artificial neural networks (ANN) are advanced machine learning methods inspired by the biology of the brain. They allow complex nonlinear relationships between the variable and its predictors and are nowadays state-of-the-art methods for complex forecasting tasks and thus featured in many sources [15] [16] [18].

A neural network can be considered as a network of neurons, often referred to as nodes, which are organized in layers. The input nodes form the bottom layer and the output nodes form the top layer. The intermediate layers are called hidden layers. The forecasts are obtained by a linear combination of the inputs and passing the result through an activation function. Activation functions are used to regulate values flowing through the network. The most common choices of activation functions are hyperbolic tangent, rectified linear unit (ReLU) and sigmoid function, all featured in Figure 3.1. [15]

Before making any forecasts, the weights of the network need to be adjusted to best fit the data. This is called training and it is done with the backward propagation of errors using the backpropagation algorithm. The algorithm is designed to work back from output nodes to input nodes and to compute the gradient of the loss function with respect to the weights of the network. The data is passed to the network iteratively batch by batch and after each iteration, the weights of the network are updated to minimize loss. One of the most common optimization algorithms to find the set of weights to minimize loss

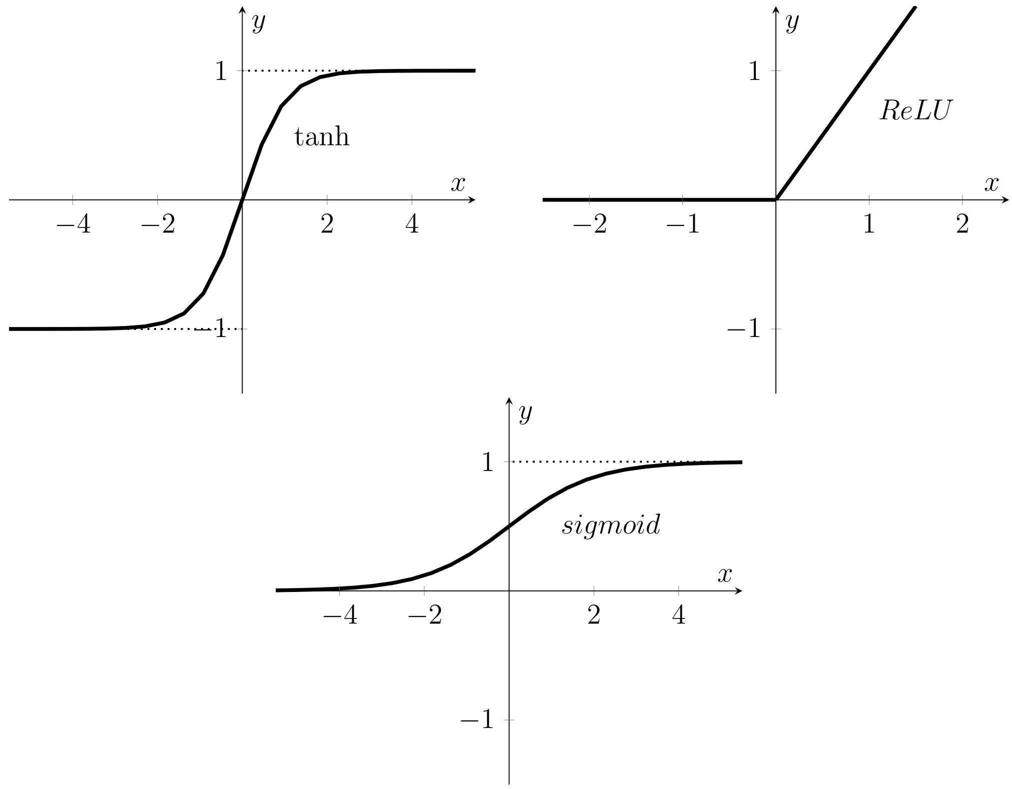


Figure 3.1. *ANN activation functions: hyperbolic tangent, rectified linear unit and sigmoid functions.*

is gradient descent, where the weights are updated along the direction of the steepest descent. [15]

A network architecture where every node in the previous layer is connected to each node in the following layer is called a fully connected network. There are also other shapes of network architectures such as convolutional neural networks (CNNs), where some of the hidden layers perform mathematical operations called convolutions, and recurrent networks, where information can pass through nodes connected in a cycle, allowing the output of some nodes to affect the succeeding input to the same nodes. [15]

3.4.1 Recurrent Neural Networks

Recurrent neural networks (RNNs) are a class of artificial neural networks. The difference between a traditional, dense ANN and RNN is a feedback loop. The difference is illustrated in more detail in Figure 3.2.

The feedback loop enables the inputs of an RNN to have a feedback relation with each other: RNNs use the feedback loop, often referred to as the hidden state, to be able to remember what it has seen before. When an RNN cell receives an input, it also processes the content of the memory and combines the information before making a prediction. This characteristic feature makes RNNs great for modelling sequence types of data, such as

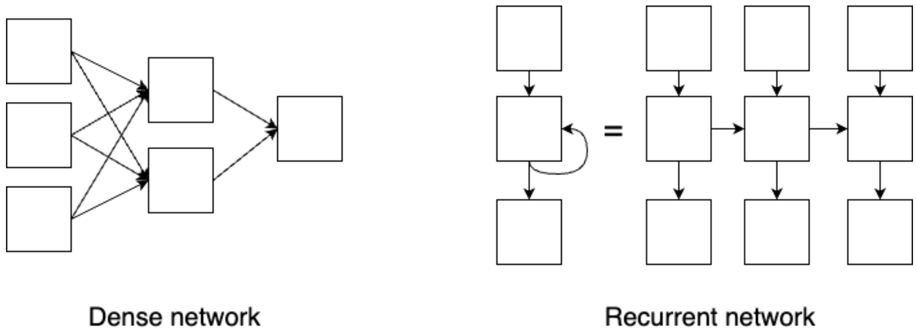


Figure 3.2. *A traditional, dense ANN vs RNN.*

time series data, as it is capable of learning long-term processes. [15]

As described before, a single RNN cell receives not one but two inputs. This is described in more detail in Figure 3.3: X is the input at time t, Y is the prediction at time t and a represents the weights that pass from left to right in Figure 3.2 and enable the model fit as a sequence through time.

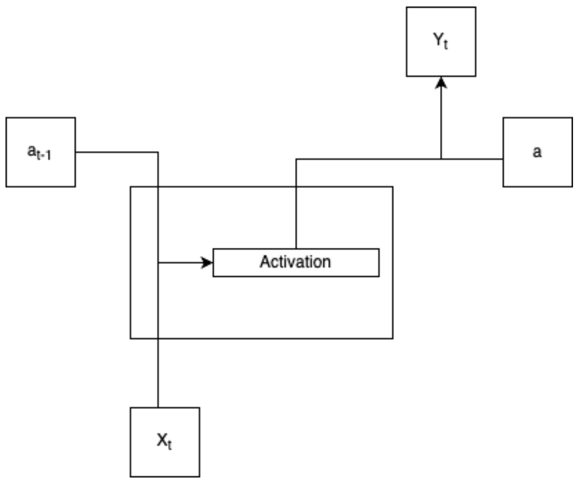


Figure 3.3. *A schematic overview of an RNN cell, adapted from [18].*

Hyperbolic tangent is the standard activation function in RNNs. ANNs commonly use rectified linear units (ReLU) activation layers, but with long sequences, such as time series data, the repeated multiplications with weights act like exponentiation and make the system explode. Tanh activation layer forces the values to stay between -1 and 1 and therefore suits better for time series forecasting purposes.

RNNs' limitations are their short-term memory. If an input sequence is long enough, RNNs are not able to transfer information from earlier time steps to later time steps, and they sometimes leave out important information from the beginning of the time series sequence during the backpropagation process. In RNN backpropagation processes, the gradient decreases as it is backpropagated through time. If the value of the gradient becomes too small, it is not able to contribute to the learning process. This issue is called the vanishing gradient problem, and it is solved in more advanced RNN architectures, such as long short-term memory (LSTM). [16]

3.4.2 Long Short-Term Memory (LSTM)

LSTM is a subclass of RNNs with a special type of cell structure. They were already introduced in 1997 [19], and are now one of the state-of-the-art models for time series forecasting. As mentioned before, LSTMs help preserve the error propagated through time and layers without any risk of losing important information. This is done with a special type of architecture that includes cell states and gates that can regulate the flow of information. The cell learns to make decisions about how much, what and when to release and store information by an iterative process of making guesses, backpropagating the error and adjusting the weights with gradient descent. A schematic model of a single LSTM cell is shown in Figure 3.4.

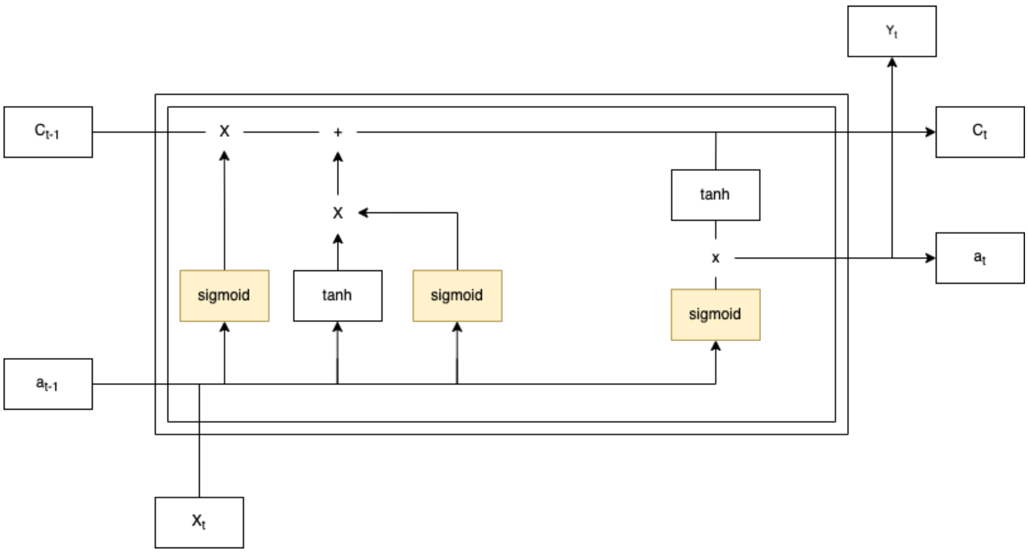


Figure 3.4. *A structure of a single LSTM cell, adapted from [18].*

In Figure 3.4, the horizontal line at the top represents the cell state. C_{t-1} is the cell state at time $t-1$ and a_{t-1} is the hidden state coming from the previous LSTM cell at time $t-1$. C_t is the new cell state and a_t is the outgoing hidden state. X_t is the input coming to the LSTM cell and Y_t is the output of the cell.

The cell state is regulated with gates. Gates are the way of letting information through the cell and are marked with yellow in Figure 3.4. The three gates are regulated with sigmoid neural net operations in the cell. The first gate is called the forget gate, the second the input gate and the final the output gate. Each of these three gates passes the input X_t and information from the previous hidden state a_{t-1} through a sigmoid function in order to regulate the input values between 0 and 1. The closer the 0 means to forget and the closer to the 1 means to remember.

The forget gate is the first sigmoid-regulated gate on the left side of the cell. Its job is to determine what information should be forgotten. The function of the forget gate is

$$f_t = \sigma(W_f \cdot [a_{t-1}, x_t] + b_f), \tag{3.6}$$

where σ is the sigmoid operation, W_f are the weights for the forget gate neurons and b_f is the bias for the forget gate.

The second gate in the cell is the input gate. The input gate will decide what new information should be stored in the cell state. The formula for the input gate is:

$$i_t = \sigma(W_i \cdot [a_{t-1}, x_t] + b_i), \tag{3.7}$$

where W_i are the weights for the input gate neurons and b_i is the bias for the input gate. After the input gate has decided what values should be updated, a tanh layer creates a vector of new candidate values $\tilde{C}_t$ that could be added to the state:

$$\tilde{C}_t = tanh(W_c \cdot [a_{t-1}, x_t] + b_c), \tag{3.8}$$

where tanh is the hyperbolic tangent operation, W_c are the weights and b_c is the bias for the vector of new candidate values. The values passed through the input gate and the values from the new candidate vector are then multiplied to create an update to the state. The new cell state is then updated by forgetting the things the forget gate decided should be forgotten and updating the things the input gate determined should be updated, scaled by how much each state value should be updated. Or in other words, written with a formula:

$$C_t = f_t * C_{t-1} + i_t * \tilde{C}_t. \tag{3.9}$$

The third gate in the cell is the output gate. The output gate decides what the next hidden state should be. Firstly, the current input and the previous hidden state are passed through a sigmoid function:

$$o_t = \sigma(W_o \cdot [a_{t-1}, x_t] + b_o), \tag{3.10}$$

where W_o are the weights for the output gate neurons and b_o is the bias for the output gate. Secondly, the newly modified cell state is passed through a hyperbolic tangent function. The output from tanh function is then multiplied with the output from the sigmoid function to decide, what information the hidden state should keep:

$$h_t = o_t * tanh(C_t), \tag{3.11}$$

which is the hidden state at time t. [16]

3.5 Model Evaluation Metrics

A machine learning model should not be trained with the whole dataset. The universal principle is to leave some data for testing the model and evaluate its performance with a few evaluation metrics. Some of the most well-known metrics in time series forecasting are the root-mean-squared error (RMSE), mean absolute error (MAE) and mean absolute percentage error (MAPE). [15]

The RMSE is the square root of the mean squared error (MSE):

$$RMSE = \sqrt{\frac{1}{h}\sum_{i=0}^{h-1}(y_i - \hat{y}_i)^2}, \tag{3.12}$$

where h is the forecasting horizon, y_i stands for the actual values and $\hat{y}_i$ for the forecasted values.

The RMSE uses squared errors rather than original errors since original errors can involve positive or negative values and after summation, could cancel each other out. Taking a square root of the average of the squared errors makes the scale of the error the same as the original variables and consequently, making the error easier to interpret.

The MAE is computed by taking the row-wise absolute differences between the predicted and the actual values:

$$MAE = \frac{1}{h}\sum_{i=0}^{h-1}\left|y_i - \hat{y}_i\right|. \tag{3.13}$$

Because the MAE is computed by taking the absolute values of the errors before averaging them, summing the errors will not make them cancel each other out. Like the RMSE, the MAE also yields a score in the same range as the original variables.

Generally, the RMSE is favoured over the MAE in error calculation, since RMSE uses squares rather than absolute values and therefore, is easier to use in mathematical computations that require taking derivatives. The derivatives of squared errors are easier to compute than the derivatives of absolute errors, and as derivatives are widely used in function optimization and minimization, it is a significant criterion. [15]

The MAPE is computed by taking the error for each prediction and dividing it by the actual value. This will make the error measure a percentage and therefore standardized:

$$MAPE = \frac{1}{h} \sum_{i=0}^{h-1} \left| \frac{y_i - \hat{y}_i}{y_i} \right|. \tag{3.14}$$

Standardization of the error between zero and one makes it easy to communicate of the performance results. The MAPE can be easily converted to a goodness of fit measure by computing 1 - MAPE. It is often easier to communicate model performance in terms of a positive result compared to a negative one.

The MAPE has one significant drawback: when the actual value is zero, the formula will divide the error by the actual value. In other words, this scenario leads to a division by zero, which is problematic.

4. HOME CARE WORKLOAD PREDICTION IN KEUSOTE

The determination of workload for home care personnel represents a significant challenge in contemporary home care management. In addition to the anticipated strong growth in the number of elderly individuals in the coming decades, the frequent changes in customer need for home care services, the diversity of services offered and the varying nature of home care visits, as well as changes in service areas, strategic decisions and operational issues related to service provision play a significant role in the complexity of determining appropriate staffing levels in the home care service sector. The appropriate staffing levels need to meet customer needs for service, ensure customer satisfaction and at the same time, minimize service costs.

The aim of this study was to create a machine learning model that is able to predict the workload of home care personnel in the Keusote operating area. Two models were designed to forecast the daily workload of home care personnel in each Keusote's home care site four months ahead: a traditional ARIMA model and a modern LSTM neural network. The better-performed model was selected for production to weekly forecast the workload for each home care unit at the Keusote operating area. The forecasted values are sent to a third-party interface through which the forecast is available to home care personnel, for example, in resource management.

4.1 Data Storage

Keusote's patient information data is stored in a data warehouse. The warehouse architecture is based on the Virta initiative [20]. The Virta initiative is a joint project by DigiFinland, the Ministry of Social Affairs and Health, the well-being services counties, the Finnish Institute for Health and Welfare (THL), the National Supervisory Authority for Welfare and Health (Valvira) and the Social Insurance Institution of Finland (Kela). The goal of the project is 1) to facilitate the management of welfare areas with reliable and up-to-date information, 2) to support the regions in increasing their capacity for information management and 3) to develop a data warehouse reference architecture that especially takes into consideration of processing of personal data. Virta reference architecture is illustrated in Figure 4.1.

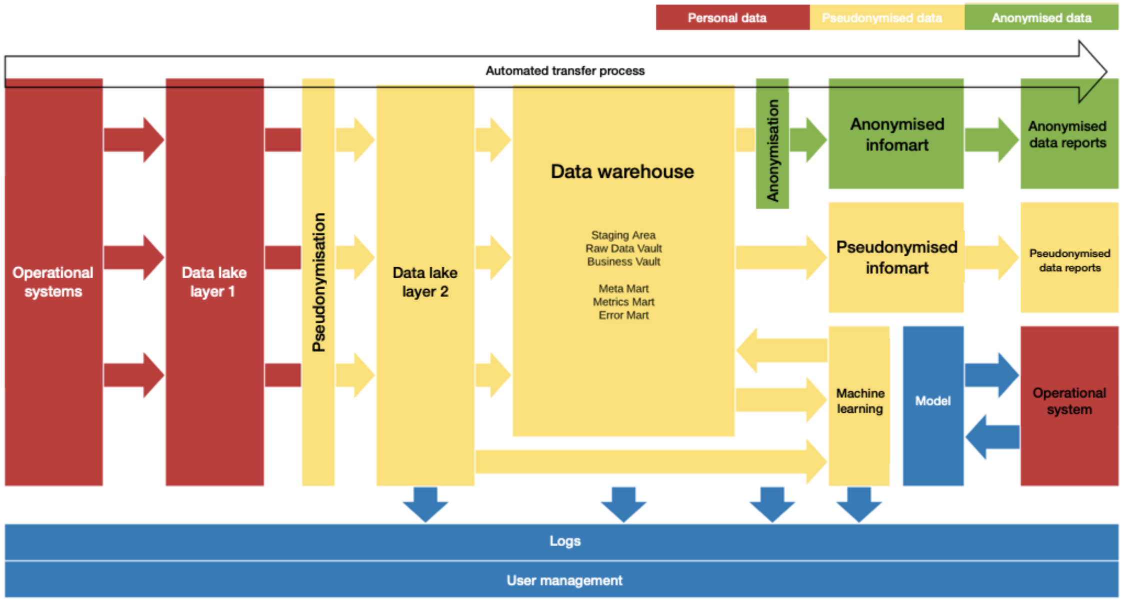


Figure 4.1. *Virta reference architecture [20].*

The warehouse architecture is based on Azure Cloud services, and a more detailed architecture and technology description can be seen in Figure 4.2. Data is first gathered from various internal sources, such as patient information, finance and human resource systems and dumped into an Azure data lake. Personal data, subject to European General Data Protection Regulation (GDPR) [21], is pseudonymised at the data lake. Pseudonymisation means processing personal data in such a way that makes it impossible to identify individuals from the data without the use of additional information. To be more precise, personally identifiable information fields within the data are replaced by an artificial identifier, a pseudonym. The pseudonym serves to identify the data within the warehouse, but the real-world identity of the individual remains unknown.

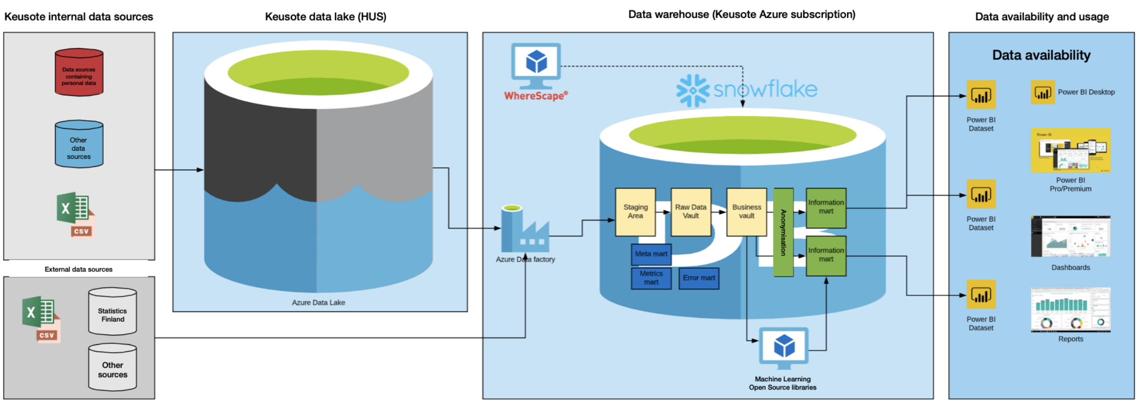


Figure 4.2. *Keusote's data warehouse architecture.*

Next, the pseudonymised data is transported through Azure Data Factory processes to Snowflake database service staging area. Azure Data factory processes also fetch data from external data sources and feed it to the staging area. The Snowflake data warehouse is built according to the Data Vault 2.0 principle [22]. Data Vault 2.0 is a largely used system of business intelligence in data warehousing and information delivery. It is designed to provide long-term storage of data without any use of re-engineering. The data warehouse architecture consists of three main architectural layers of Data Vault 2.0: the raw data warehouse (RDW), the business data vault (BDV) and the information delivery layer, also referred to as the publish layer. In the raw data warehouse, data is divided into easily expandable data models: hubs, links and satellites using business keys [22]. At the business data vault layer, business rules are applied and larger data constructions are created, such as "visit" or "diagnose". The publish layer is used to deliver information to the end-user, for example, in the form of reports.

When data is moved from the business vault layer to the publish layer, it is anonymised. Anonymisation means processing the data in such a way that makes it impossible to identify individuals. In addition to preventing identification by a third party, identification must be prevented also by the data controller itself. Contrary to pseudonymisation, data identifiability is reduced by changing the identifier with each query. Additionally, the precision of the data is reduced, for example, by replacing age fields with age group fields.

4.2 Derived Variables

When a nurse visits a home hospice care client, the visit is entered into a patient information system, and each visit has a certain duration. In this study, to forecast the amount of daily work of home care personnel four months ahead, daily totals of the duration of the visits have been selected as the dependent variable. The study used pseudonymised data in order to connect information for example about patient's age, visits, visit duration, and RAI assessment results.

As a result of several discussions with Keusote's home care professionals, three information sources were identified to have an effect on home care personnel's workload and that could be utilized in making the forecast. The independent variables were 1) daily visit duration sums, 2) patient-specific treatment and service plan information and 3) patient-specific values of certain RAI parameters.

Eventually, patient-specific treatment and service plans and RAI parameters could not be utilized in forecasting. Since the primary objective of the model is to forecast the workload of home care personnel, it is imperative that each independent variable is directly related to the workload (= the sum of daily visit durations). While the data warehouse did contain information on patient-specific treatment and service plans, it did not provide insight into the amount of work and effort required by care personnel for a specific service. Furthermore, treatment and service plans are subject to frequent updates in response to changes in the customer's needs for service. Unfortunately, the data warehouse did not contain information on the original plans prior to any updates and, therefore, prevented the use of treatment and service plans in training the forecasting algorithms.

Problems with RAI parameters were similar than with patient-specific treatment and service plans. Since RAI measurement is a fairly new method to evaluate patients' need for care services, there was not enough RAI data available in the data warehouse. In fact, not a single patient had any change in RAI parameters in the data warehouse and therefore, comparing changes in RAI parameters with increasing or decreasing workload was not possible. Later during the study, it was found that some RAI data might be located in a different data system and not yet accessible in the data warehouse. However, as there was no certain information about the existence of new RAI data, the study was decided to be carried out with the data that was available in the data warehouse at the time of the study.

To sum it all up, after these findings the original plan to develop a multivariate model had to be changed to a univariate model. Eventually, the forecasting model was chosen to be trained with daily visit duration sums while still keeping the aim of forecasting the workload for four months ahead.

4.3 Forecasting

After the information sources to best estimate the workload were selected, the data had to be reshaped to suit a univariate forecasting problem. This was done with multiple SQL queries in the data warehouse. The dataset was limited to include home care visits in the Keusote area from 5.5.2021 to 1.9.2022 so that the first 365 days were used for training and the last 120 days (= four months) were used for testing. Home care visits were extracted from other visits by the following criteria: 1) the form of service being home care, home nursing and home help service, 2) the type of visit being a professional's visit to the customer's home and 3) the truth value of being a visit being 1, meaning True. Additionally, visits having a longer duration than the number of minutes in a 24-hour period were filtered out as errors. An example of the final univariate dataset is shown in Figure 4.3.

The data was reindexed to catch possible missing dates and missing data. If a missing date was found, the data for that missing row was fixed as the median value for that care site.

The forecasts were programmed in the Azure Databricks environment. Azure Databricks is a data analytics platform, and it is optimized for Microsoft Azure cloud services. It provides a web-based platform for working with Spark, which offers automated cluster management and IPython-style notebooks.

In total, there were 27 different care sites in the Keusote area, for which a four-month forecast had to be produced. Two models were chosen to be trained for the forecasting

Row	DATE	CARE_SITE_1	CARE_SITE_2	CARE_SITE_3	CARE_SITE_4
1	2022-05-01	1203	1267	1501	1476
2	2022-04-30	1328	1399	1654	1413
3	2022-04-29	2660	1566	2076	1553
4	2022-04-28	1839	1517	1969	1879
5	2022-04-27	1582	1360	2076	1782
6	2022-04-26	1536	1697	1818	1300
7	2022-04-25	1878	1534	1959	2050
8	2022-04-24	1448	1485	1483	1228
9	2022-04-23	1192	1352	1408	1255
10	2022-04-22	1694	1563	1849	1591

Figure 4.3. *An example of the forecasting dataset.*

problem: a traditional ARIMA model and a modern LSTM neural network. One year of historical data was selected for training, and the forecasts for each 27 care sites were computed independently without taking any influence from each other.

After making the forecasts for each 27 home care units, the forecasted values were evaluated with RMSE, MAE and MAPE error metrics. Additionally, the forecasted workload and the actual workload were plotted in the same figure to evaluate the goodness of the forecast.

4.3.1 Forecasting with Auto-ARIMA

The ARIMA forecast was programmed with the auto-ARIMA process using Python's pmdarima library [23]. It is used to identify the most optimal parameters for an ARIMA model: the number of lag variables p, order of differencing d and the magnitude of the moving average window q. Basically, auto-ARIMA takes the data and fits many models with different ARIMA parameters to the data. Each model characteristic is compared and the best fit is chosen according to the selection criteria.

As part of the auto-ARIMA process, the order of differencing d is determined with a series of differencing tests, such as Kwiatkowski–Phillips–Schmidt–Shin, Augmented Dickey-Fuller and Phillips–Perron tests. The optimal values for p and q are decided by fitting many models with different ranges of defined start_p, start_q, max_p and max_q, and selecting the model with the best information criterion based on Akaike Information Criterion (AIC), Corrected Akaike Information Criterion, Bayesian Information Criterion (BIC), Hannan-Quinn Information Criterion (HQIC), or "out of bag" validation scoring. [23]

For example, the auto-ARIMA process suggests ARIMA(8,1,7) model (Figure 4.4) for care site 5. The output summary uses the same template for all autoregressive models and therefore the header displays SARIMAX (seasonal autoregressive integrated moving average with exogenous factors) results even though the model is actually an ARIMA model. The output summary shows basic information such as the name of the variable attempted to predict, date and time information and the number of observations. The Log-Likelihood, AIC, BIC and HQIC help compare models with different parameters for the best fit. The log-likelihood recognises the distribution that best fits the dataset. AIC, BIC and HQIC punish the model for complexity as complex models may result in overfitting, are hard to interpret and the overall computational efficiency is not optimal.

```
                               SARIMAX Results
==============================================================================
Dep. Variable:                      y   No. Observations:                  365
Model:               SARIMAX(8, 1, 7)   Log Likelihood               -2455.545
Date:                Wed, 02 Nov 2022   AIC                           4945.090
Time:                        13:52:32   BIC                           5011.342
Sample:                    05-05-2021   HQIC                          4971.422
                         - 05-04-2022
Covariance Type:                  opg
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
intercept     -7.6905      9.168     -0.839      0.402     -25.660      10.279
ar.L1         -0.9458      0.109     -8.640      0.000      -1.160      -0.731
ar.L2         -0.6453      0.154     -4.193      0.000      -0.947      -0.344
ar.L3         -0.7472      0.157     -4.762      0.000      -1.055      -0.440
ar.L4         -0.6266      0.150     -4.170      0.000      -0.921      -0.332
ar.L5         -0.7578      0.154     -4.916      0.000      -1.060      -0.456
ar.L6         -0.6373      0.152     -4.195      0.000      -0.935      -0.340
ar.L7          0.2833      0.152      1.869      0.062      -0.014       0.580
ar.L8          0.2317      0.095      2.438      0.015       0.045       0.418
ma.L1          0.0821      0.078      1.049      0.294      -0.071       0.235
ma.L2         -0.1855      0.074     -2.516      0.012      -0.330      -0.041
ma.L3          0.2672      0.060      4.484      0.000       0.150       0.384
ma.L4         -0.0961      0.065     -1.472      0.141      -0.224       0.032
ma.L5          0.2123      0.054      3.961      0.000       0.107       0.317
ma.L6         -0.0067      0.072     -0.094      0.925      -0.147       0.134
ma.L7         -0.7642      0.053    -14.426      0.000      -0.868      -0.660
sigma2      4.558e+04   2572.582     17.717      0.000    4.05e+04    5.06e+04
===================================================================================
Ljung-Box (L1) (Q):                   0.00   Jarque-Bera (JB):               852.85
Prob(Q):                              0.97   Prob(JB):                         0.00
Heteroskedasticity (H):               0.87   Skew:                             1.70
Prob(H) (two-sided):                  0.46   Kurtosis:                         9.68
===================================================================================

Warnings:
[1] Covariance matrix calculated using the outer product of gradients (complex-step).
```

Figure 4.4. *Auto-ARIMA's automatically produced output summary of the model selection process.*

The output summary in Figure 4.4 additionally shows the error term sigma2 and the lag variables from L1 to L8. Statistically insignificant variables at 5% risk level are the intercept, ar.L7, ma.L1, ma.L4, and ma.L6, as their p-values are above the 0.05 threshold. Ljung-Box test states that the null hypothesis suggesting the errors are white noise, cannot be rejected as the probability is above 0.05. Heteroscedasticity tests whether the error residuals have the same variance. Since the probability is again above the threshold of 0.05, the null hypothesis cannot be rejected. Jarque-Bera tests the null hypothesis that the data is normally distributed. The probability of this test is zero which states that the data is not normally distributed. Additionally, the Jarque-Bera test indicates that the data is right-skewed and it has a large kurtosis.

The auto-ARIMA process outputs diagnostics graphs for the standardized residuals. The residual plots can be seen in Figure 4.5 for care site 5.

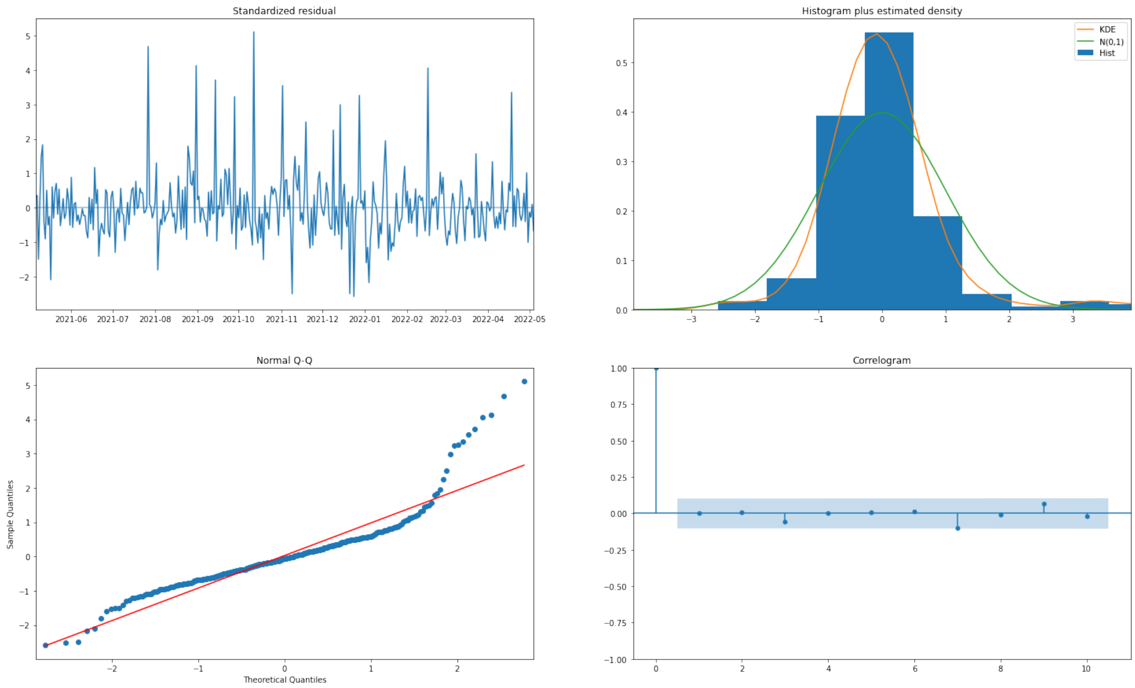


Figure 4.5. *Diagnostic plots for standardized residuals for auto-ARIMA based ARIMA(8, 1, 7) model for care site 5.*

At the top left corner in Figure 4.5, standardized residuals fluctuate around a mean of zero. The density plot at the top right corner shows a histogram and estimated density function of standardized residuals along with a normal(0,1) density function plotted for reference. It shows a slightly right-skewed distribution with a mean of zero. The normal quantile-quantile (Q-Q) plot at the bottom left corner shows that most data points are located more or less in line with the red line. The deviation at the right side suggests that the distribution is right skewed. The correlogram at the bottom right corner shows no autocorrelation in the data.

4.3.2 Forecasting with LSTM

The application of an artificial neural network (ANN) based Long Short-Term Memory (LSTM) model for forecasting required certain modifications to the dataset prior to feeding the dataset values to the model. Specifically, in order to align the dataset values with the hyperbolic tangent activation function of the LSTM units, the dataset values had to be rescaled to a range between -1 and 1. This was accomplished by utilizing the MinMaxScaler functionality of the scikit-learn library [24].

Subsequently, the dataset had to be modified to look like a supervised learning problem. A supervised learning problem is a type of machine learning problem in which the model is trained on a labeled dataset, that is, a dataset in which the outcome or target variable is already known. The goal of a supervised learning algorithm is to learn a mapping from the input variables to the output variable and make predictions about new data based on the mapping.

In practice, to frame the dataset as a supervised learning problem, it was rephrased such that the input data consisted of seven previous time steps, while the next time step was designated as the output:

input	output
1, 2, 3, 4, 5, 6, 7	8
2, 3, 4, 5, 6, 7, 8	9
3, 4, 5, 6, 7, 8, 9	10.

The LSTM model was created with the Sequential class from Keras [25] to linearly stack layers into the model. Different numbers and combinations of layers and cells were tested until the performance of the model did not significantly improve. Eventually, the network consisted of three LSTM layers and one dense layer at the end, with each LSTM layer having 64 cells. The dense layer consisted of one cell to match the desired output size. The model loss was computed with the mean-squared error, which is similar to the RMSE introduced in chapter 3 in equation 3.12 but without the square root. The optimization algorithm was selected to be Adam, and the activation function a hyperbolic tangent to regulate values between -1 and 1. The LSTM network structure in total is presented in Program 4.1.

```
model = Sequential()
model.add(LSTM(n_neurons, return_sequences=True,
          batch_input_shape=(n_batch, X.shape[1], X.shape[2]),
          stateful=True))
model.add(LSTM(n_neurons, return_sequences=True))
model.add(LSTM(n_neurons))
model.add(Dense(y.shape[1]))
model.compile(loss='mean_squared_error', optimizer='adam')
```

Program 4.1. *LSTM network structure.*

The LSTM model was programmed to produce the four-month forecast with a recursive technique: first, seven days of historical data were used to produce a one-step-ahead forecast. Next, the newly predicted day was transferred to the seven-day historical data set and the first day was dropped from the beginning of the period making it seven days long again. Then a new one-step-ahead forecast was again computed, and this process was iterated until the target length of four months was reached. The recursive forecasting process is illustrated in more detail in Figure 4.6.

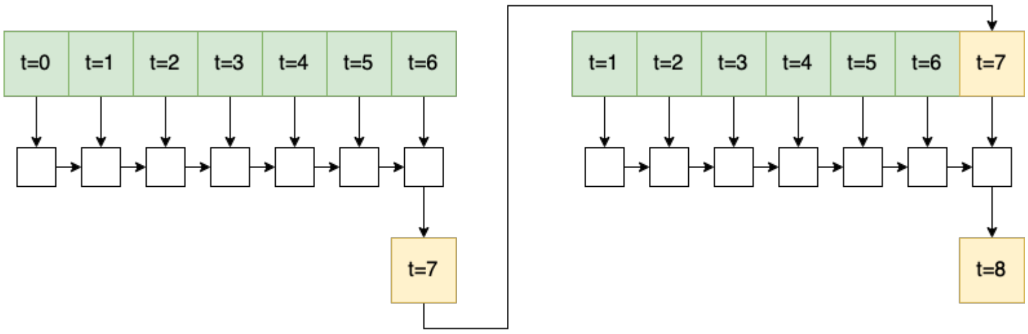


Figure 4.6. *A drawing of the recursive forecasting strategy of the LSTM network.*

After computing the forecasts, a reverse transformation was applied using the MinMaxScaler to facilitate the comparison of the forecasts with the observed workload values. Furthermore, statistical metrics Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE), and a measure of goodness of fit were computed for each home care site in order to facilitate a more comprehensive comparison between the forecasts and the actual observations.

5. RESULTS

The forecasting results for all care sites can be seen in Appendix A in Tables A.1 and A.2 for the ARIMA and the LSTM models, respectively. Figure 5.1 shows a visual summary of these tables as a boxplot presentation.

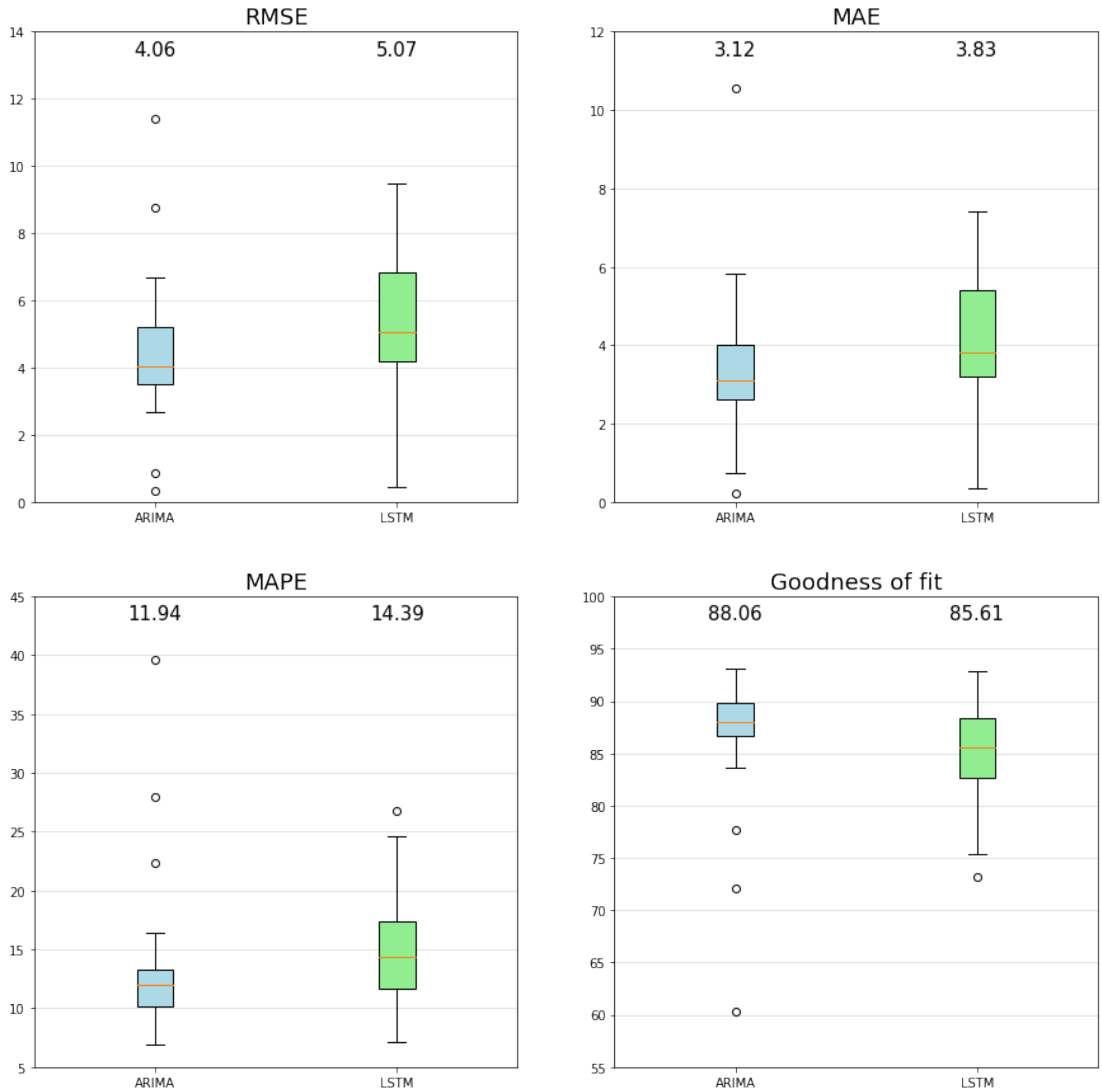


__Figure 5.1.__ A visual summary of Tables A.1 and A.2 presented as boxplots. The median value of the boxplot is shown above the plot.

The figure shows four subfigures, one for each model evaluation metric described in section 3.5. Each subfigure displays two boxplots, one for the ARIMA model-based forecasts and one for the LSTM model-based forecasts. The comparison of the two reveals that the boxplots computed from the ARIMA model errors are comparatively shorter in comparison to those generated from the LSTM model errors. Moreover, the boxplots derived from the LSTM model-based forecasts have longer whiskers compared to those generated from the ARIMA model-based forecasts. These observations demonstrate that the results from LSTM-based forecasts exhibit greater variability as compared to the forecasts produced by the ARIMA model. This can be interpreted as the LSTM-based forecast errors having a higher variance. Furthermore, the box plots constructed from the ARIMA model errors possess more outliers in comparison to those generated from the LSTM model errors. Interestingly, the box plots constructed from the RMSE and MAE error metrics for the LSTM model do not feature any outliers.

The median values for each boxplot in each subfigure in Figure 5.1 are relatively close. RMSE median values around four and five hours and MAE median values around three and four hours indicate promising results for both the ARIMA and LSTM model-based forecasts. The mean value of around four hours of forecasting error per day is deemed acceptable. The median values for MAPE around twelve and fourteen per cent correspond to a goodness of fit median values of 88.06 % for the ARIMA model and 85.61 % for the LSTM model.

In general, the error metric values RMSE, MAE and MAPE appear to be slightly better for the forecasts computed with the ARIMA model than with the LSTM model, leading to a slightly better overall goodness of fit with the ARIMA model. As a result, by only looking at the error metrics and goodness of fit measures for the forecasts, the ARIMA model seems like a better fitting model for this type of forecasting problem. However, upon further examination, it becomes apparent that this is not the case and a more thorough analysis of the results will provide an explanation for the situation.

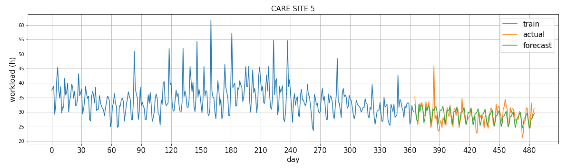
Table 5.1 shows three care sites with the best forecasting results obtained with the ARIMA model. LSTM model results for the corresponding care sites are shown for reference.

Table 5.1. *ARIMA model results for the best three forecasts according to the goodness of fit. LSTM model results for the same care sites are shown for reference.*

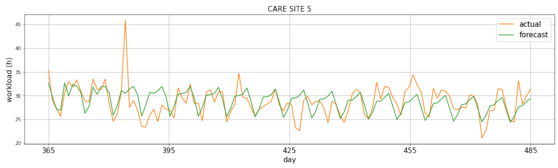
	ARIMA				LSTM			
care site id	RMSE (h)	MAE (h)	MAPE (%)	Goodness of fit (%)	RMSE (h)	MAE (h)	MAPE (%)	Goodness of fit (%)
care site 5	2.75	1.99	6.96	93.04	5.00	3.88	13.90	86.10
care site 25	3.25	2.65	9.05	90.95	5.09	4.22	14.39	85.61
care site 19	3.79	3.12	9.27	90.73	6.80	5.89	16.87	83.13

The table highlights that the ARIMA model achieved a goodness of fit of approximately 90% for these best three forecasts. Conversely, the goodness of fit for the LSTM model is slightly inferior, with values ranging from 83% to 86%. The error metrics RMSE and MAE are comparatively small for these care sites, around 3 hours for the ARIMA model and about 4 hours for the LSTM model, with the exception of care site 19 where the RMSE value with the LSTM model is 6.80 hours. The model evaluation metrics RMSE, MAE and MAPE are all smaller for the ARIMA model results as compared to the LSTM model results.

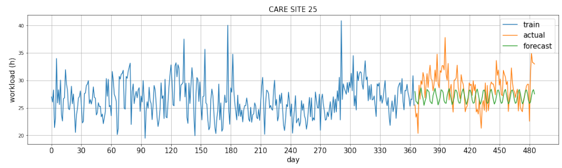
Nevertheless, an inspection of the plotted forecasts in Figures 5.2 and 5.3 offers a deeper understanding of the characteristics of the forecasts. Figure 5.2 displays the historical workload of one year prior to the start of the forecast, the forecasted workload and the actual workload of the four-month forecasting horizon for the best three forecasts computed with the ARIMA model. Figure 5.3 provides a more detailed close-up of the forecasted workload and the actual workload for these three forecasts.



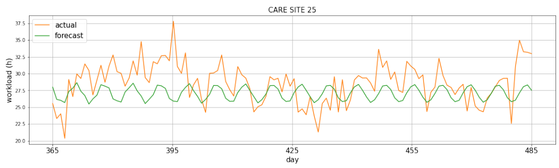
(a) *Care site 5 with ARIMA.*



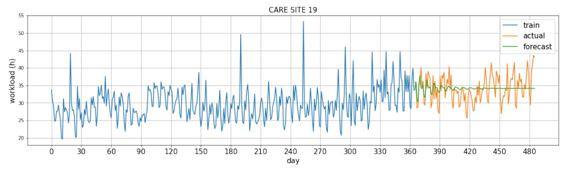
(b) *Care site 5 with ARIMA.*



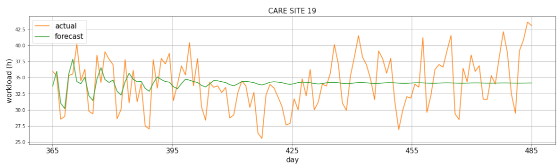
(c) *Care site 25 with ARIMA.*



(d) *Care site 25 with ARIMA.*



(e) *Care site 19 with ARIMA.*

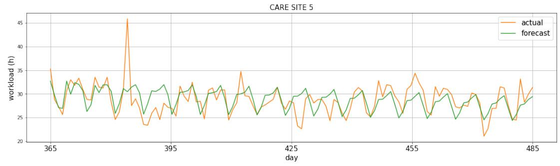


(f) *Care site 19 with ARIMA.*

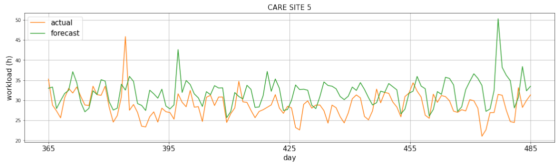
Figure 5.2. *Left-hand side shows the historical workload of one year before the start of the forecast, the forecasted workload and the actual workload for the top 3 best forecasts with the ARIMA model. Right-hand side shows a close-up of the forecasted workload and the actual workload.*

The ARIMA model demonstrates an excellent fit for care site 5 in Figure 5.3a. However, Figure 5.3c reveals that the ARIMA forecast exhibits oscillatory behaviour and an incorrect DC level. Furthermore, Figure 5.3e highlights a quite common phenomenon observed in multiple ARIMA-based forecasts within the study: after a short period of time, the model sticks to forecast the median value of the workload instead of attempting to forecast any peaks or sudden increases or decreases in workload.

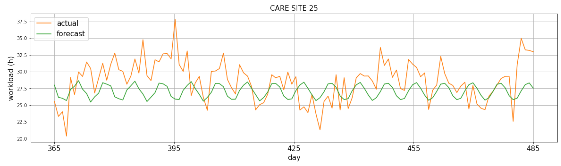
This is due to the residual errors being calculated as the difference between the observed value and the estimated workload value; forecasting the mean value is a safe way to min-



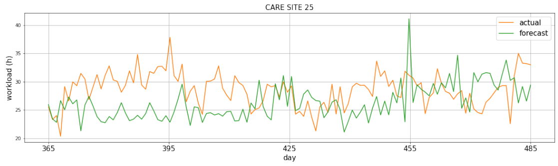
(a) *Care site 5 with ARIMA.*



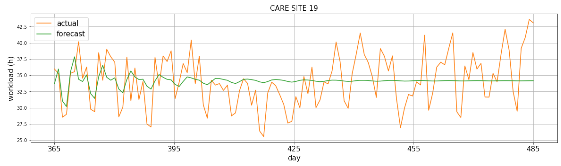
(b) *Care site 5 with LSTM.*



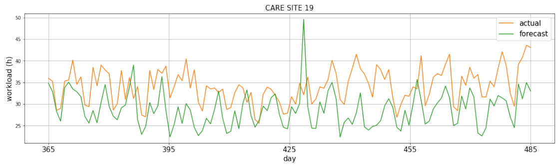
(c) *Care site 25 with ARIMA.*



(d) *Care site 25 with LSTM.*



(e) *Care site 19 with ARIMA.*



(f) *Care site 19 with LSTM.*

Figure 5.3. *A close-up of the ARIMA model results plotted for the top 3 best forecasts. LSTM model results for the same care sites are plotted for reference.*

imize the error and thus, resulting in slightly improved RMSE, MAE, MAPE and goodness of fit values. This phenomenon was observed in multiple ARIMA-based forecasts, in care sites 1, 4, 7, 12, 13 and 19.

Conversely, Figures 5.3b, 5.3d, 5.3f demonstrate that the LSTM model has a more robust approach on how the workload develops throughout the forecasting horizon. The DC level is accurate, and the model also attempted to predict peaks in workload, such as around day 455 of the forecasting horizon in Figure 5.3d.

Additionally, the LSTM model never stuck to predict a constant workload value, whereas the ARIMA model did for some home care sites. Based on the interviews with Keusote's home care management professionals, one of the main requirements for the forecast was to provide a comprehensive assessment of the expected workload throughout the forecasting horizon. This includes identifying times of increased and decreased demand, as well as, predicting instances of peak demand where additional workforce may be needed. Therefore, the ability to accurately predict peaks and avoid forecasting constant workload values throughout the forecasting horizon is a crucial characteristic of the model and one of the main requirements of the forecasting task.

The three poorest forecasting results with the ARIMA model are shown in Table 5.2. LSTM model results for the same care sites are shown for reference.

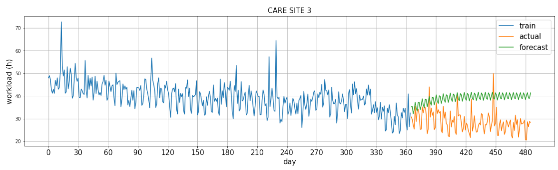
Table 5.2 illustrates the lowest three forecasting outcomes with the ARIMA model. The worst three results are given for care sites 3, 10, and 22. The performance of the model, as indicated by the RMSE, MAE, and MAPE values, is relatively poor for care site 3,

Table 5.2. *ARIMA model results for the worst three forecasts according to the goodness of fit. LSTM model results for the same care sites are shown for reference.*

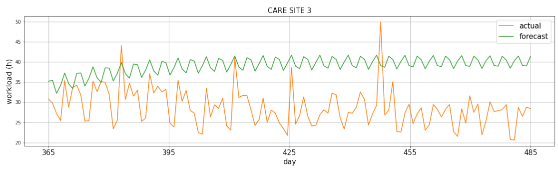
care site id	ARIMA RMSE (h)	ARIMA MAE (h)	ARIMA MAPE (%)	ARIMA Goodness of fit (%)	LSTM RMSE (h)	LSTM MAE (h)	LSTM MAPE (%)	LSTM Goodness of fit (%)
care site 3	11.49	10.63	39.99	60.01	8.66	7.39	26.76	73.24
care site 10	5.72	4.55	27.95	72.05	9.47	7.40	45.95	54.05
care site 22	0.87	0.73	22.30	77.70	1.02	0.82	23.69	76.31

resulting in a goodness of fit of 60%. However, the results for care sites 10 and 22 exhibit a more satisfactory performance. The table also reveals that while the LSTM model performs better in terms of forecasting for care site 3, it does not necessarily demonstrate a superior performance for care sites 10 and 22.

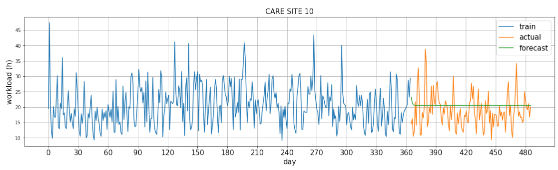
Again, a more in-depth examination of the forecast plots presented in Figure 5.4 provides a clearer understanding of the nature of the forecasts. The examination reveals that the forecast for care site 3 diverges significantly from the actual workload and starts to oscillate, leading to poor performance as indicated by the error metric calculations. Although the error measures for care sites 10 and 22 do not exhibit significant issues as portrayed in Table 5.2, the relatively high goodness of fit of around 70% can largely be attributed to the ARIMA model consistently forecasting a nearly constant value throughout the forecast horizon.



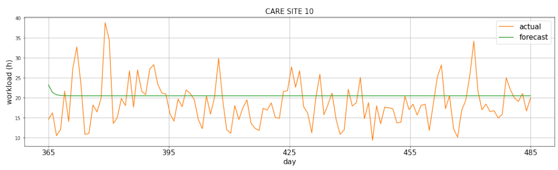
(a) *Care site 3 with ARIMA.*



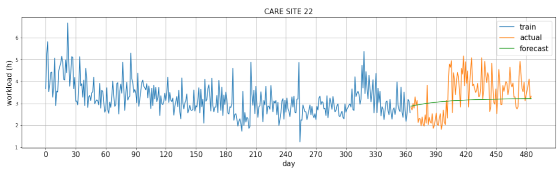
(b) *Care site 3 with ARIMA.*



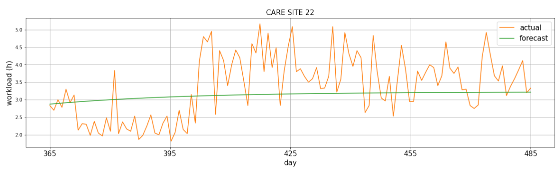
(c) *Care site 10 with ARIMA.*



(d) *Care site 10 with ARIMA.*



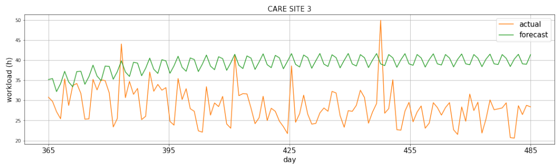
(e) *Care site 22 with ARIMA.*



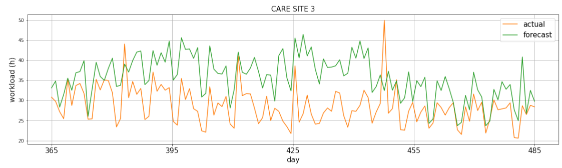
(f) *Care site 22 with ARIMA.*

Figure 5.4. *Left-hand side shows the historical workload of one year before the start of the forecast, the forecasted workload and the actual workload for the top 3 worst forecasts with the ARIMA model. Right-hand side shows a close-up of the forecasted workload and the actual workload.*

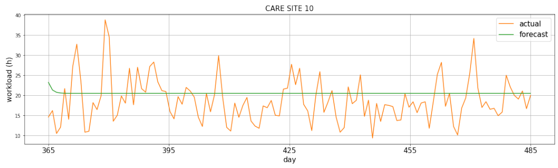
Figure 5.5 further clarifies the differences between the nature of ARIMA-based and LSTM-based forecasts. The results demonstrate that while the ARIMA model was unable to generate a satisfactory forecast, the LSTM model was successful in doing so. Specifically, the starting and ending points of the waves in the forecast by the LSTM model for care site 3, as shown in Figure 5.5b, align well with the actual workload, although the forecast predicts a generally higher workload compared to the actual workload.



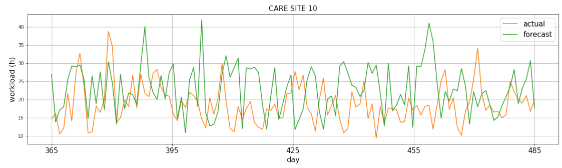
***(a)** Care site 3 with ARIMA.*



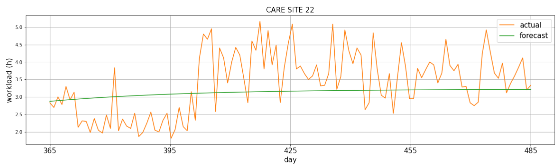
***(b)** Care site 3 with LSTM.*



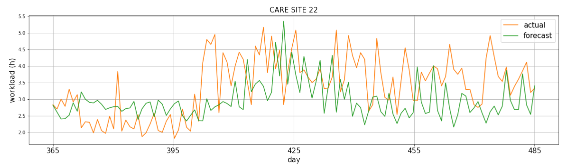
***(c)** Care site 10 with ARIMA.*



***(d)** Care site 10 with LSTM.*



***(e)** Care site 22 with ARIMA.*



***(f)** Care site 22 with LSTM.*

***Figure 5.5.** A close-up of the ARIMA model results plotted for the worst three forecasts. LSTM model results for the same care sites are plotted for reference.*

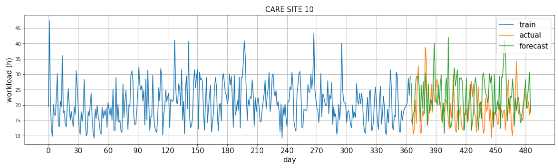
Table 5.3 presents the lowest three forecasts produced by the LSTM model. ARIMA model results for the corresponding care sites are shown for reference.

***Table 5.3.** LSTM model results for the worst three forecasts according to the goodness of fit. ARIMA model results for the same care sites are shown for reference.*

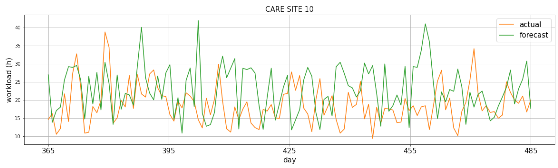
care site id	LSTM				ARIMA			
	RMSE (h)	MAE (h)	MAPE (%)	Goodness of fit (%)	RMSE (h)	MAE (h)	MAPE (%)	Goodness of fit (%)
care site 3	8.66	7.39	26.76	73.24	11.49	10.63	39.99	60.01
care site 9	6.77	5.44	24.64	75.36	3.63	3.11	13.12	86.88
care site 10	9.47	7.40	45.95	54.05	5.72	4.55	27.95	72.05

Similar to the results obtained from the ARIMA model, the goodness of fit for the worst three forecasts is approximately 70% with the exception of care site 10, which exhibits a goodness of fit of 54.05%. It is noteworthy that care sites 3 and 10 presented challenges for both the ARIMA and LSTM models in terms of forecasting workload accurately.

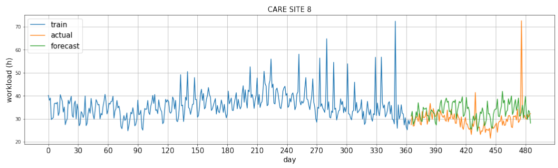
Table 5.3 suggests that the ARIMA model outperforms the LSTM model in terms of forecasting accuracy. However, a visual comparison of the plotted forecasts for the two models in Figures 5.6 and 5.7 reveals that the LSTM-based forecasts are superior, as the ARIMA model's forecasts are relatively constant throughout the forecast horizon. Ultimately, even the worst three forecasts computed with the LSTM model are more reliable than the ARIMA model's forecasts for the given care sites, despite the numerical results suggesting otherwise.



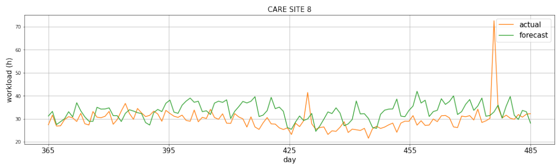
***(a)** Care site 10 with LSTM.*



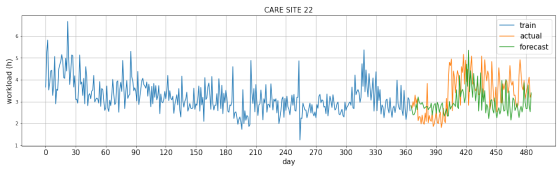
***(b)** Care site 10 with LSTM.*



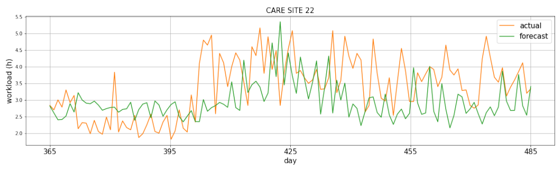
***(c)** Care site 8 with LSTM.*



***(d)** Care site 8 with LSTM.*

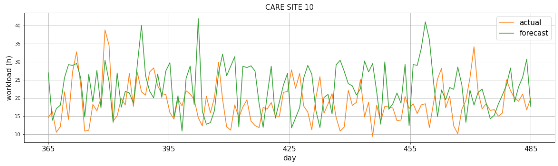


***(e)** Care site 22 with LSTM.*

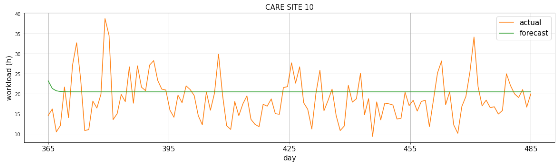


***(f)** Care site 22 with LSTM.*

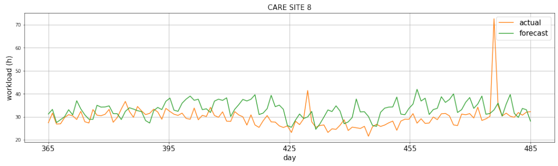
***Figure 5.6.** Left-hand side shows the historical workload of one year before the start of the forecast, the forecasted workload and the actual workload for the worst three forecasts with the LSTM model. Right-hand side shows a close-up of the forecasted workload and the actual workload.*



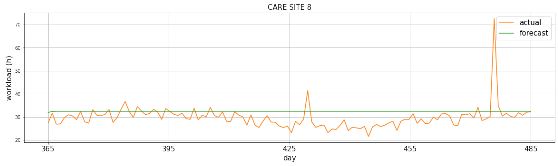
(a) Care site 10 with LSTM.



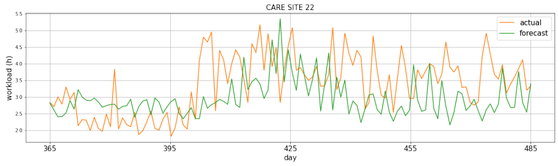
(b) Care site 10 with ARIMA.



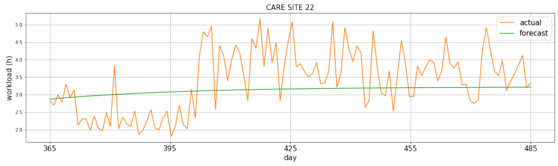
(c) Care site 8 with LSTM.



(d) Care site 8 with ARIMA.



(e) Care site 22 with LSTM.



(f) Care site 22 with ARIMA.

Figure 5.7. *LSTM model results plotted for the worst three forecasts. ARIMA model results are plotted for reference.*

In summary, the comparison between the ARIMA and LSTM models for forecasting workload reveals that the ARIMA model provides slightly superior numerical results in terms of the evaluation metrics RMSE, MAE, MAPE and goodness of fit. Nonetheless, a visual inspection of the forecast plots highlights that the ARIMA model had a tendency to predict a constant workload value, while the LSTM model aimed to identify peak demand instances. While the evaluation metrics favour the ARIMA model, the ability to accurately predict peaks and to steer away from forecasting constant workload values throughout the forecasting horizon are critical aspects of the model and a key requirement of the forecasting task.

6. DISCUSSION AND CONCLUSION

Determining the appropriate workload of home care personnel is a prevalent challenge in the field of home care management. Specifically, the identification of appropriate staffing levels on a weekly basis is essential for meeting the future demands of customers, ensuring customer satisfaction, and minimizing service costs.

The goal of this study was to develop a machine learning model to forecast the daily workload of home care personnel in home care units operating in the Keusote area. The forecasting horizon was medium-long, spanning four months ahead. A total of 27 home care units were included in the study, for which the four-month forecast was produced.

Two distinct forecasting models were designed to tackle the forecasting problem: a traditional ARIMA model and a modern LSTM neural network. Initially, three information sources were identified as potentially impacting the workload: patient-specific treatment and service plans, patient-specific RAI measures, and historical workload. However, after a thorough examination of these sources, it was determined that patient-specific treatment and service plans, as well as patient-specific RAI measures, had to be excluded from the study. As a result, the original plan to develop a multivariate model had to be changed to a univariate model, which resulted in a significant reduction of the model's forecasting power.

The ARIMA model utilized an auto-ARIMA forecasting process to define the model parameters (p, d, and q), while the LSTM model employed a recursive forecasting technique in which future time steps were forecasted based on previously forecasted values.

The results of the study were encouraging, particularly in light of the initial starting point where two out of three information sources had to be excluded from the study. The median value of the MAPE for the ARIMA model was 11.94, resulting in a goodness-of-fit score of 88.06%, and the median value of MAPE for the LSTM model was slightly higher at 14.39, resulting in a goodness-of-fit value of 85.61%. However, a visual examination of the forecasts revealed that the LSTM model had a better ability to produce reliable forecasts over the four-month forecasting horizon and demonstrated an aptitude for forecasting peaks in workload. As a result, the LSTM model was selected for production.

It is worth noting that the results of the LSTM model may have been influenced by the recursive forecasting technique, which can lead to the accumulation of errors. The LSTM model may have produced better results than the ARIMA model if more information sources, such as patient-specific treatment and service plans and patient-specific RAI parameters, could have been included in the training. The use of exogenous variables, such as holidays, could also be exploited in future research, as the workload during the holidays is often greater than on normal weekdays.

The analysis of the data revealed some limitations in terms of data quality. There were instances of biases in the records, as the majority of visit durations were divisible by five due to the use of a phone application by nurses to record the visit duration and the prompt's availability of such durations as speed selector options. Additionally, some errors were found in the records, as some of the recorded home care visit durations exceeded 24 hours. These records were removed from the training data. Furthermore, some records were missing for some days for certain home care units. These missing values were substituted with the median workload (=sum of daily visits durations) of the respective home care unit. The errors and biases present in the data can potentially compromise the accuracy and reliability of the forecasting models. Therefore, a thorough examination of data quality is crucial in detecting and correcting data deficiencies in order to obtain robust and trustworthy forecasts.

Additionally, the results of both the LSTM model and the ARIMA model may have been affected by the temporal context in which the data was collected, as the data was gathered during the COVID-19 pandemic. During the pandemic, many resources were transferred to the tasks of pandemic prevention and control and the treatment of corona patients. For example in elderly care services, the prolonged pandemic and the growing shortage of home care staff delayed service need assessments and limited the availability of home care services. Additionally, the backlog in treatment and services has increased in areas such as oral health care, non-urgent specialised medical care and, in particular, basic public services for vulnerable clients. On the other hand, the use of remote services has increased during the pandemic. The use of remote digital services could also be taken into account in the future development of the forecast.

The study also raises the issue of appropriate accuracy metrics. As the residual errors are calculated by comparing the observed workload value to the estimated value, forecasting a constant value throughout the forecasting horizon serves as a reliable method for minimizing error and subsequently producing slightly superior values for the RMSE, MAE, MAPE and goodness of fit. This ultimately provides a preference for the ARIMA model over the LSTM model, even though a visual inspection of the forecasts suggested otherwise. However, most prediction performance metrics used in regression are based on the difference between the observed and predicted values and, therefore, studying alternative ways to measure prediction performance may prove worthwhile. In addition, an interesting area for future research could be to examine the forecasting error with a few different forecasting horizons and to analyse, how different forecasting horizons would affect the RMSE, MAE and MAPE accuracy metrics.

Furthermore, in future research, it would be advantageous to explore other ways of forecasting. One potential model to consider is the Markov model, which is a probability-based model of state transitions. A state-based model would allow taking into account home care clients in a specific state and the probabilities of clients transitioning from one state to another. In this context, each state would correspond to a particular duration and intensity of home care visits. The state assignment of home care clients could be computed by utilizing the Resident Assessment Instrument (RAI) measurements of home care clients, as they provide insight into the needs for service. This approach would enable the computation of the workload based on the number of clients in different states and the probabilities of state transitions. The Markov model would enable a more detailed understanding of the nature of clients in a certain state and help anticipate a large number of clients moving from a less labour-intensive state to a more labour-intensive one.

REFERENCES

[1] Official Statistics of Finland (OSF). *Population Projection [e-publication]*. 2021. URL: https://www.stat.fi/til/vaenn/2021/vaenn_2021_2021-09-30_tie_001_en.html (visited on 12/22/2022).

[2] Official Statistics of Finland (OSF). *Population Structure [e-publication]*. 2021. URL: https://www.stat.fi/til/vaerak/2019/vaerak_2019_2020-03-24_tie_001_en.html (visited on 12/22/2022).

[3] Eija-Leena Koponen, Ulla-Maija Laiho, and Mika Tuomaala. "Mistä tekijät sosiaali- ja terveysalalle — työvoimatarpeen ja -tarjonnan kehitys vuoteen 2025 [Where to Find Workers for Social and Health Care — Analysis of Need and Offer of Workforce Until 2025]". In: (2012).

[4] Flavia Martinelli, Anneli Anttonen, and Margitta Mätzke. *Social Services Disrupted: Changes, Challenges and Policy Implications for Europe in Times of Austerity*. New Horizons in Social Policy series. Cheltenham, Gloucestershire: Edward Elgar Publishing, 2017, pp. 219–238.

[5] Keusote - Keski-Uudenmaan sote. *Kotihoito ikääntyneille, säännöllisen kotihoidon palvelu [Home Care for the Elderly, Regular Home Care Service]*. 2022. URL: https://www.keusote.fi/palveluhakemisto/saannollinen-ja-tilapainen-kotihoito/ (visited on 09/20/2022).

[6] InterRAI. *InterRAI*. URL: https://interrai.org/ (visited on 01/25/2023).

[7] Parliament of Finland. *Hallituksen esitys eduskunnalle laiksi ikääntyneen väestön toimintakyvyn tukemisesta sekä iäkkäiden sosiaali- ja terveyspalveluista annetun lain muuttamisesta [The Government's Proposal to the Parliament as a Law on Supporting the Functional Capacity of the Elderly and on Amending the Act on Social and Health Services for the Elderly]*. 2020. URL: https://www.eduskunta.fi/FI/vaski/HallituksenEsitys/Sivut/HE_4+2020.aspx.

[8] Finnish Institute for Health and Welfare (THL). *Information on the RAI Assessment System*. 2022. URL: https://thl.fi/en/web/ageing/assessment-of-service-needs-with-the-rai-system (visited on 09/26/2022).

[9] Sivan Aldor-Noiman, Paul D. Feigin, and Avishai Mandelbaum. "Workload Forecasting for a Call Center: Methodology and a Case Study". In: *The Annals of Applied Statistics* 3.4 (2009), pp. 1403–1447.

[10] Antti-Jussi Mäkipää. "Forecasting Emergency Department Arrivals with Facebook Prophet library". Bachelor's thesis. Tampere University, Faculty of Information Technology and Communication Sciences, 2021, p. 20.

[11] Eetu Pulkkinen. “Forecasting Emergency Department Arrivals with Neural Networks”. Bachelor’s thesis. Tampere University, Faculty of Information Technology and Communication Sciences, 2020, p. 21.

[12] Karthik K Padthe, Vikas Kumar, Carly M Eckert, Nicholas M Mark, Anam Zahid, Muhammad Aurangzeb Ahmad, and Ankur Teredesai. “Emergency Department Optimization and Load Prediction in Hospitals”. In: *arXiv.org* (2021).

[13] Jill Flo, Bjørg Landmark, Siri Tønnessen, and Lisbeth Fagerström. “Patient Classification Systems Used to Classify Nursing Intensity and Assess Nursing Staffing Resources in Home Health Care: A Scoping Review”. In: *International Journal of Nursing Studies* 99 (2019).

[14] Kambombo Mtonga, Santhi Kumaran, Chomora Mikeka, Kayalvizhi Jayavel, and Jimmy Nsenga. “Machine Learning-Based Patient Load Prediction and IoT Integrated Intelligent Patient Transfer Systems”. In: *Future Internet* 11.11 (2019), pp. 236–.

[15] Joos. Korstanje. *Advanced Forecasting with Python: With State-Of-the-Art-Models Including LSTMs, Facebook’s Prophet, and Amazon’s DeepAR*. Berkeley, CA: Apress, 2021.

[16] Francesca Lazzeri. *Machine Learning for Time Series Forecasting with Python*. Indianapolis, Indiana: John Wiley & Sons, 2021.

[17] Souhaib Ben Taieb, Gianluca Bontempi, Amir F Atiya, and Antti Sorjamaa. “A Review and Comparison of Strategies for Multi-Step Ahead Time Series Forecasting Based on the NN5 Forecasting Competition”. In: *Expert Systems with Applications* 39.8 (2012), pp. 7067–7083.

[18] Rob J. Hyndman and George Athanasopoulos. *Forecasting: Principles and Practice*. 2st edition. Amsterdam, Netherlands: OTEXTS, 2018.

[19] Sepp Hochreiter and Jürgen Schmidhuber. “Long Short-Term Memory”. In: *Neural Computation* 9.8 (1997), pp. 1735–1780.

[20] DigiFinland. *Virta Project*. 2018. URL: https://digifinland.fi/en/our-operations/virta-project/ (visited on 08/24/2022).

[21] *Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the Protection of Natural Persons with Regard to the Processing of Personal Data and on the Free Movement of such Data, and Repealing Directive 95/46/EC (General Data Protection Regulation)*. 2016. URL: https://eur-lex.europa.eu/eli/reg/2016/679.

[22] Daniel Linstedt and Michael Olschimke. *Building a Scalable Data Warehouse with Data Vault 2.0*. 1st edition. Amsterdam, Netherlands: Morgan Kaufmann, 2016.

[23] *Python’s auto-ARIMA Documentation*. URL: https://alkaline-ml.com/pmdarima/modules/generated/pmdarima.arima.auto_arima.html#pmdarima-arima-auto-arima.

[24] *Scikit-Learn Documentation*. URL: https://scikit-learn.org/stable/index.html#.

[25] *The Sequential Class Documentation*. URL: https://keras.io/api/models/sequential/.

APPENDIX A: RESULT TABLES

Table A.1. *Auto ARIMA model results for all care sites. The best fit and the worst fit are highlighted with green and red colour.*

care site id	RMSE (h)	MAE (h)	MAPE (%)	Goodness of fit (%)
care site 1	6.68	4.33	14.67	85.33
care site 2	3.73	2.42	10.19	89.81
care site 3	11.49	10.63	39.99	60.01
care site 4	3.52	2.87	12.21	87.79
care site 5	2.75	1.99	6.96	93.04
care site 6	4.33	3.55	10.20	89.80
care site 7	6.45	5.49	16.38	83.62
care site 8	5.70	3.91	13.55	86.45
care site 9	3.63	3.11	13.12	86.88
care site 10	5.72	4.55	27.95	72.05
care site 11	4.76	2.97	12.51	87.49
care site 12	8.76	5.83	9.84	90.16
care site 13	3.61	2.88	9.05	90.95
care site 14	4.80	3.68	10.75	89.25
care site 15	4.70	3.97	9.37	90.63
care site 16	3.72	3.07	11.82	88.18
care site 17	3.18	2.48	11.58	88.42
care site 18	3.49	2.69	10.46	89.54
care site 19	3.79	3.12	9.27	90.73
care site 20	2.67	2.16	12.38	87.62
care site 21	0.32	0.22	11.19	88.81
care site 22	0.87	0.73	22.30	77.70
care site 23	4.07	3.02	12.09	87.91
care site 24	4.67	3.30	11.94	88.06
care site 25	3.25	2.65	9.05	90.95
care site 26	5.52	4.37	13.16	86.84
care site 27	4.91	4.09	13.19	86.81

Table A.2. *LSTM model results for all care sites. The best fit and the worst fit are highlighted with green and red colour.*

care site id	RMSE (h)	MAE (h)	MAPE (%)	Goodness of fit (%)
care site 1	7.66	5.21	16.18	83.82
care site 2	5.70	3.83	16.27	83.73
care site 3	8.66	7.39	26.76	73.24
care site 4	4.12	3.12	13.22	86.78
care site 5	5.00	3.88	13.90	86.10
care site 6	4.42	3.52	9.72	90.28
care site 7	9.13	6.27	17.78	82.22
care site 8	6.86	5.00	16.97	83.03
care site 9	6.77	5.44	24.64	75.36
care site 10	9.47	7.40	45.95	54.05
care site 11	5.07	3.16	14.04	85.96
care site 12	9.38	6.20	9.73	90.27
care site 13	4.65	3.56	11.53	88.47
care site 14	6.85	5.36	15.21	84.79
care site 15	4.14	3.25	7.19	92.81
care site 16	3.45	2.74	11.39	88.61
care site 17	3.67	3.03	13.07	86.93
care site 18	5.25	4.35	15.74	84.26
care site 19	6.80	5.89	16.87	83.13
care site 20	2.99	2.36	13.61	86.39
care site 21	0.44	0.35	18.45	81.55
care site 22	1.02	0.82	23.69	76.31
care site 23	4.26	3.24	11.81	88.19
care site 24	7.31	6.16	23.94	76.06
care site 25	5.09	4.22	14.39	85.61
care site 26	4.53	3.68	10.19	89.81
care site 27	4.54	3.61	10.69	89.31