

**UNIVERSITY OF VAASA**

**SCHOOL OF TECHNOLOGY AND INNOVATIONS**

**COMMUNICATION AND SYSTEMS ENGINEERING**

RAJA VARDHAN REDDY MARTHALA

**WIFI BASED INDOOR POSITIONING – A MACHINE LEARNING  
APPROACH**

Master's thesis for the degree of Master of Science in Technology submitted for  
assessment.

Vaasa, May 22, 2020.

Supervisor

Professor Heidi Kuusniemi

Instructors

Professor Mohammed S. Elmusrati

Associate Professor Petri Välisuo

## *ACKNOWLEDGEMENTS*

First and foremost, I would like to express my gratitude to my thesis supervisor Heidi Kuusniemi for offering a research assistant position in Digital Economy. I appreciate her continuous guidance, knowledge sharing and feedback. Her hardworking nature, immense knowledge and attitude towards team members, are my inspiration.

I am very thankful to my thesis instructor and our master's programme director, Professor Mohammed S. Elmusrati. I appreciate his efforts, deep knowledge and support, throughout my master's programme. I got interest towards Artificial Intelligence from his machine learning course.

I am also thankful to my other instructor, Associate Professor Petri Välisuo. I appreciate his support and discussions throughout my thesis. I gained new skills in machine learning from him.

Special thanks to the programmer, Jyri Nieminen for this contribution in creating indoor app and my teacher, Tobias Glocker his support throughout the programme. Thanks to Kannan Selvan, my friends and to everyone who supported in completion of my thesis.

I would like to express my love towards my parents, Sudarshan Reddy Marthala and Sumalatha Marthala for their unconditional love. I have no words to express their support from my pre-school to still now. I express my gratitude to my uncle, Surendra Marthala for encouraging me to apply for master's in Finland. I thank my sibling, cousins, grandparents and extended family for being my strength. I express my devotion towards the god Lord Venkateshwara swamy for giving me blessings, happiness and a making me a part of Marthala family.

I take this opportunity to express my gratitude towards University of Vaasa for offering quality education. I express my deepest gratitude to the beautiful country, Finland for providing me free education and quality life. In simple "Minä rakastan Suomea".

Vaasa, 22.05.2020.

Raja Vardhan Reddy Marthala

## TABLE OF CONTENTS

LIST OF FIGURES .....	5
LIST OF TABLES.....	6
ABBREVIATIONS .....	7
ABSTRACT .....	10
1. INTRODUCTION .....	12
1.1. Overview of Indoor Positioning Systems .....	12
1.2. Motivation behind Indoor Positioning System (IPS) with Wi-Fi .....	13
1.3. Aim .....	14
1.4. Thesis Structure .....	14
2. EXISTING INDOOR POSITIONING TECHNOLOGIES .....	16
2.1. IPS based on RF Technologies .....	16
2.1.1. Wi-Fi .....	16
2.1.2. Bluetooth .....	18
2.1.3. Ultra Wide Band (UWB).....	19
2.1.4. ZigBee .....	21
2.2. IPS based on Non-RF Technologies .....	21
2.2.1. Visual Navigation.....	21
2.2.2. Geomagnetic Field.....	22
2.2.3. Inertial Navigation System.....	23
2.2.4. Infrared Beacon .....	24
2.2.5. Visible Light Communication .....	24
2.3. Techniques for Indoor Positioning.....	26
2.3.1. Fingerprinting .....	26
2.3.2. Trilateration .....	31
2.3.3. Triangulation .....	33
2.3.4. Pedestrian Dead Reckoning.....	34
3. DEVELOPMENT OF IPS ANDROID APP BASED ON HERE SDK.....	35
3.1. Introduction.....	35
3.2. HERE SDK .....	36
3.3. HERE Indoor Radio Mapper.....	36
3.4. Android Studio .....	37
3.5. Technobothnia Demo App.....	38

4. MACHINE LEARNING IN GENERAL .....	40
4.1. Introduction .....	40
4.2. Machine Learning Algorithms .....	41
4.2.1. Support Vector Regressor (SVR).....	42
4.2.2. Decision Tree Regressor (DTR).....	45
4.2.3. Random Forest Regressor (RFR) .....	47
4.2.4. Extremely Randomized Trees Regressor (ETR).....	50
4.3. Applying Machine Learning in Wi-Fi based IPS .....	50
4.3.1. Model Motivation .....	50
5. EXPERIMENT AND RESULTS.....	52
5.1. Introduction to the Experiment Setup .....	52
5.2. Implementation of the Machine Learning Models .....	53
5.3. Results .....	55
6. CONCLUSIONS AND FUTURE WORK .....	59
6.1. Conclusions .....	59
6.2. Future Work .....	60
REFERENCES .....	61

## LIST OF FIGURES

<b>Figure 1.</b> Existing Wi-Fi Infrastructure in the Technobothnia Technology Center Layout at the University of Vaasa Campus.....	17
<b>Figure 2.</b> A suggested Bluetooth Technology Infrastructure Suitable for an IPS to the Technobothnia Technology Center Layout at the University of Vaasa Campus. ....	19
<b>Figure 3.</b> Visual Light Communication with LED Lights.....	25
<b>Figure 4.</b> RSS1, RSS2, RSS3 are the Received Signal Strengths from Access Points AP1, AP2, AP3 Respectively.....	26
<b>Figure 5.</b> Online & Offline Phases of the Fingerprinting Method. ....	27
<b>Figure 6.</b> Illustrates Weighted Centroid Localization. ....	29
<b>Figure 7.</b> An Illustration of Trilateration. ....	32
<b>Figure 8.</b> An Illustration of Triangulation. ....	33
<b>Figure 9.</b> Basic Understanding of Pedestrian Dead Reckoning. ....	34
<b>Figure 10.</b> Showing Current Position & Initial Position on Technobothnia Floor Plan in the HERE Indoor Radio Mapper App.....	36
<b>Figure 11.</b> Showing RSSI Values in the Radio Map Admin Tool. ....	37
<b>Figure 12.</b> Showing the Android App Developing Environment in Android Studio. ....	38
<b>Figure 13.</b> The First Page & the Venue Map on the Technobothnia Demo App.....	39
<b>Figure 14.</b> Communication between the App & the HERE Environment. ....	39
<b>Figure 15.</b> Regression, Classification & Clustering in Machine Learning. ....	42
<b>Figure 16.</b> Schematic Diagram of Support Vector Regression.....	44
<b>Figure 17.</b> Simple Explanation of Decision Tree.....	45
<b>Figure 18.</b> High-Level Diagram of Random Forest Regressor Algorithm. ....	49
<b>Figure 19.</b> Showing the Spyder IDE for the IPS.....	52
<b>Figure 20.</b> High-level explanation of the implemented ML models. ....	54
<b>Figure 21.</b> A 3D Plot Showing True vs Predicted Coordinates of the SVR in a Metric Local Reference Frame. ....	55
<b>Figure 22.</b> Violin Plots Showing the 2D Error Distribution & the Median (in meters) of the Six IPS Algorithms Being Compared. ....	57
<b>Figure 23.</b> Violin Plots Showing the 3D Error Distribution & the Median (in meters) of the Six IPS Algorithms Being Compared. ....	57

## LIST OF TABLES

<b>Table 1.</b> Tabular column showing performance comparisons between the different algorithms implemented. ....	56
---	----

## ABBREVIATIONS

2D	2 Dimensional
3D	3 Dimensional
AGV	Automated Guided Vehicle
AI	Artificial Intelligence
AOA	Angle of Arrival
AP	Access Point
BLE	Bluetooth Low Energy
SIG	Special Interest Group
DT	Decision Tree
DTR	Decision Tree Regressor
ECC	European Communications Committee
ETR	Extremely Randomized Trees Regressor
FCC	Federal Communication Commission
GHz	Giga Hertz
GNSS	Global Navigation Satellite System
GPS	Global Positioning System
IDE	Integrated Development Environment
IEEE	Institute of Electrical and Electronics Engineers
IMU	Inertial Measurement Unit
IoT	Internet of things
IPS	Indoor Positioning System

IR	Infrared
K-NN	K-Nearest Neighbor
LED	Light Emitting Diode
LOS	Line of Sight
MAC	Media Access Control
ML	Machine Learning
NLOS	Non-line of Sight
NN	Neural Network
OOB	Out of Bag
PCA	Principle Component Analysis
PD	Photo diode
PDOA	Phase Difference of Arrival
PDR	Pedestrian Dead Reckoning
PF	Particle Filtering
RF	Radio Frequency
RFR	Random Forest Regressor
RNN	Recurrent Neural Network
RSS	Received Signal Strength
RSSI	Received Signal Strength Indicator
SDK	Software Development Kit
SVM	Support Vector Machines
SVR	Support Vector Regressor
TDOA	Time Difference of Arrival



TOA	Time of Arrival
TV	Television
TWR	Two Way Ranging
TWTF	Two-Way Time of Flight
USA	United States of America
UWB	Ultra Wide Band
VLC	Visual Light Communication
WLAN	Wireless Local Area Network

---

**Faculty of Technology**

<b>Author:</b>	Raja Vardhan Reddy Marthala
<b>Topic of the Thesis:</b>	Wi-Fi Based Indoor Positioning – A Machine Learning Approach
<b>Supervisor:</b>	Professor Heidi Kuusniemi
<b>Instructors:</b>	Professor Mohammed S. Elmusrati Associate Professor Petri Välisuo
<b>Degree:</b>	Master of Science in Technology
<b>Major of Subject:</b>	Communication and Systems Engineering
<b>Year of Entering the University:</b>	2016
<b>Year of Completing the Thesis:</b>	2020

**Pages: 68**

---

**ABSTRACT**

Navigation has become much easier these days mainly due to advancement in satellite technology. The current navigation systems provide better positioning accuracy but are limited to outdoors. When it comes to the indoor spaces such as airports, shopping malls, hospitals or office buildings, to name a few, it will be challenging to get good positioning accuracy with satellite signals due to thick walls and roofs as obstacles. This gap led to a whole new area of research in the field of indoor positioning. Many researches have been conducting experiments on different technologies and successful outcomes have been seen. Each technology providing indoor positioning capability has its own limitations.

In this thesis, different radio frequency (RF) and non-radio frequency (Non-RF) technologies are discussed but focus is set on Wi-Fi for indoor positioning. A demo indoor positioning app is developed for the Technobothnia building at the University of Vaasa premises. This building is already equipped with Wi-Fi infrastructure. A floor plan of the building, radio maps and a fingerprinting database with Wi-Fi signal strength measurements is created with help of tools from HERE technology. The app provides real-time positioning and routing as a future visitor tool.

With the exceeding amounts of available data, one of the highly popular fields is applying Machine Learning (ML) to data. It can be applied in many disciplines from medicine to space. In ML, algorithms learn from the data and make predictions. Due to the significant growth in various sensor technologies and computational power, large amounts of data can be stored and processed. Here, the ML approach is also taken to the indoor positioning challenge. An open-source Wi-Fi fingerprinting dataset is obtained from Tampere University and ML algorithms are applied on it for performing indoor positioning. Algorithms are trained with received signal strength (RSS) values with their respective reference coordinates and the user location can be predicted. The thesis provides a performance analysis of different algorithms suitable for future mobile implementations.

---

**KEYWORDS:** Indoor Positioning, Wi-Fi, HERE, Radio Signals, Machine Learning, Android App, Python, Scikitlearn, Wireless Communications.

# 1. INTRODUCTION

## 1.1. Overview of Indoor Positioning Systems

Navigation has been in existence for many thousands of years, during ancient times people used stars as a reference to navigate. As years passed technology advanced. Especially the advancement in space technology has led to the invention of satellites. In 1973, the US government launched the GPS (Global Positioning System) project, a groundbreaking system to provide outdoor localization, and it came into practice in 1980 for civilian use. Yao & Ma (2018: 5375) stated that GPS can provide 5-10 m meter level of accuracy in outdoors and it is reliable. Its application areas are wide such as car navigation, logistics & assets tracking, defense, surveying, precise timing etc.

Statistics show that "80%- 90% of people's lifetime are indoors, while 70% of mobile phones and 80% cellular data are from indoors" (Gan, Yu, Huang & Li 2017). This percentage shows that there is a requirement of accurate and reliable indoor navigation technology to address the application needs GPS addresses outdoors. Application areas of indoor positioning technology include hospitals, logistics, shopping mall, tunnels, automated guided vehicles (AGV), mining, smart buildings and warehouses. Accuracy is one of the important and challenging factor in indoor localization or tracking an asset. With GPS that provides accurate and reliable localization, those services are however limited to outdoors. GPS, as well as other Global Navigation Satellite Systems (GNSS), works with signals received from the satellites to the receiver. As the signal has to travel tens of thousands of kilometers from satellites, by the time it reaches the receiver it becomes weak but that is still good enough for positioning outdoors. When it comes to an indoor scenario there are a lot of factors that affect GNSS signals. (Gan, Yu, Huang & Li 2017.) Puricer & Kovar (2017) states the materials in constructions of building affects also GNSS signal reception indoors. GPS signals cannot penetrate through thick walls. Apart from the lack of line of sight (LOS) signal reception, low received signal power GNSS signals are not accurate and reliable for indoor navigation. This leads to a need for major research and development efforts in indoor navigation technologies to

achieve better accuracy, reliability, easy installation and low-cost devices.

## 1.2. Motivation behind Indoor Positioning System (IPS) with Wi-Fi

Radio signals are primarily meant for communication purposes and their key feature wireless transmission provides a possibility for the implementation of an IPS without any extra infrastructure once implemented. As we know, GPS/GNSSs are excellent in providing navigation services but limited to outdoors only. Navigation inside multi-floor buildings, structures with thick concrete roofs, tunnels and mines cannot thus rely on satellite signals. A solution to close this gap is to make use of existing radio signals and infrastructures as well as their inherent capabilities to provide IPS capability. Different radio frequency technologies such as Wi-Fi, UWB (ultra wide-band), Bluetooth and ZigBee are able to provide an IPS. The technology is selected depending upon the application requirements such as costs, accuracy and range. UWB has less range when compared with Bluetooth and Wi-Fi, and multiple sensors are needed for a smaller area. In this case, Wi-Fi has more advantages over other technologies, due to that Wi-Fi can be used for internet communication and as an infrastructure for an IPS at the same time and no extra device infrastructure is in principal need.

In this thesis project, the main objective is to provide an IPS utilizing an existing infrastructure. So, no need of new hardware which cuts down extra costs. Along with that the user does not require extra hardware to be installed for navigation indoors but also a mobile phone suffices as the user platform. The application area in this work is a building containing a technology center at the university premises which has existing Wi-Fi infrastructure in place. In this building, there are different kinds of technical labs with lots of visitors and students roaming around the premises and the idea behind the work is to create an IPS and routing system for the technology center users. With the help of software tools and a dedicated Android apps, Wi-Fi signals strengths are captured in the premises with the existing infrastructure and analyzed. Furthermore, an Android app is developed as a visitor application with a HERE SDK platform that provides an IPS routing service with the existing Wi-Fi infrastructure.

### 1.3. Aim

The aim of this project is to implement an IPS in the building in focus (the Technobothnia technology center) as an Android mobile application built with an HERE software development kit (SDK) for users needing routing guidance in the premises. This app is designed in a way that it gives real time positioning and routing within the building. No extra hardware is included other than the existing Wi-Fi infrastructure at hand. In addition to the IPS implemented as a user app, received signal strength (RSS) measurements from Wi-Fi base stations are collected in the premises along with manual true coordinates marked in the building i.e. the reference points. This collected data constitutes the training phase for indoor Wi-Fi positioning and is needed for algorithm comparison purposes. The collected signals are preprocessed and fed to machine learning (ML) localization algorithms being compared.

In summary, the main objectives of this thesis are as follows

1. Creating an IPS demo to premises having existing Wi-Fi infrastructure
2. Developing an Android mobile app on the HERE SDK platform as a practical user routing example
3. Creating a fingerprinting database along with manually marked true coordinates as IPS training data
4. Comparing different ML localization models with the received signal strength data in terms of accuracy and performance

### 1.4. Thesis Structure

This thesis is structured as follows:

Firstly, in chapter 2, basic understanding of various indoor radio navigation systems are given in theory. Explanations are given about capabilities and characteristics of different radio-frequency (RF) and also non-RF technologies that can be used for IPS. Under RF technologies, Wi-Fi, Bluetooth, UWB and ZigBee are described. From these, the main

focus of this thesis will be on achieving an IPS demo with especially the Wi-Fi technology. Along with that, theoretical background of non-RF technologies such as visual navigation, geo-magnetic field, inertial navigation systems and visual light communication are discussed. Improving the achievable IPS accuracy with machine learning techniques are also explained for some of these.

Secondly, high level explanation is given about positioning techniques such as fingerprinting, trilateration, triangulation and pedestrian dead reckoning.

Chapter 3 focuses on tools and the SDK that are required to develop an android mobile application for IPS. The basic idea of the associated mobile development environment, radio map creation, the SDK and the architecture of the app are given. High level explanation about the HERE radio mapper, the Android studio and the HERE SDK are discussed.

Chapter 4 introduces the trending topic “Machine Learning”. The primary concepts of machine learning algorithms are explained in particular. Applying ML in Wi-Fi based IPS.

In chapter 5 Experiment is carried out. Results are explained with figures and comparisons are made for better understanding and evaluation of the different methods.

Chapter 6 discusses about the conclusions of thesis and gives an outlook on further developments.

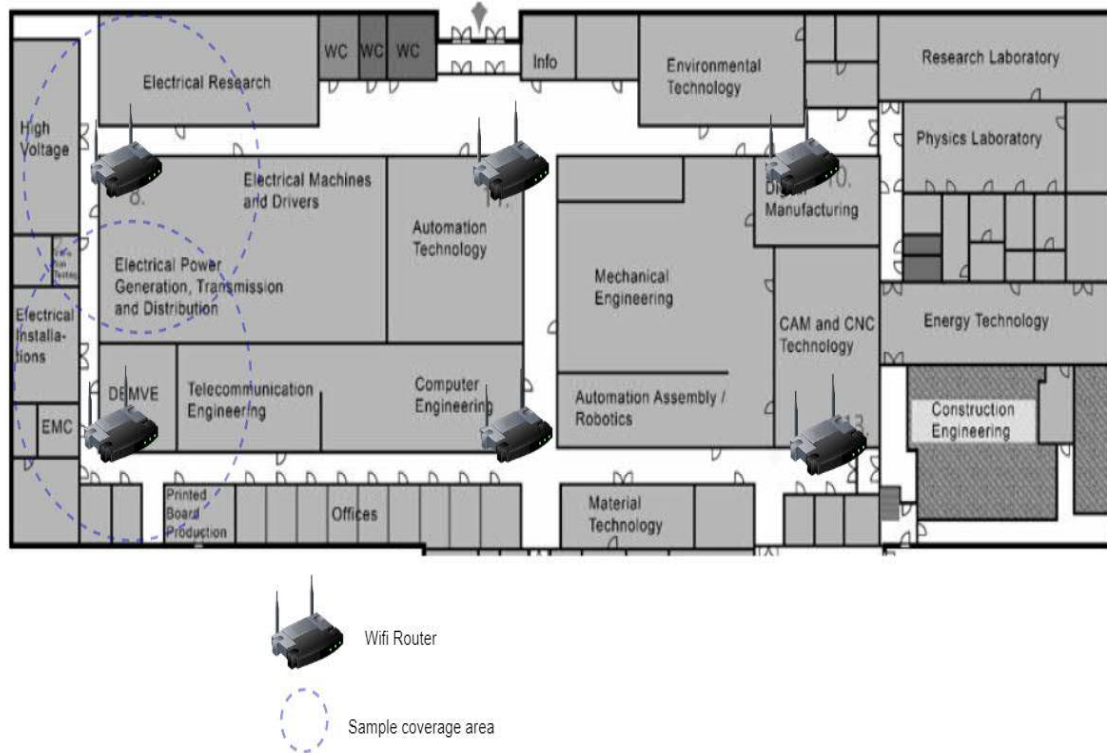
## 2. EXISTING INDOOR POSITIONING TECHNOLOGIES

### 2.1. IPS based on RF Technologies

#### 2.1.1. Wi-Fi

Many technologies such as WLAN/Wi-Fi, Bluetooth, UWB, Geomagnetism, RF, IR ID, Ultrasonic, cellular networks, computer vision and ZigBee are used in indoor localization (Rusli, Ali, Jamil, & Din 2016). Wi-Fi is used widely in IPS implementations because of its cost-effectiveness, existing infrastructure and deployment that is not complex. Currently, Wi-Fi devices that communicate over 5 GHz is used over 2.5 GHz. Less interference, more stability, better speed and less noise are the advantages with the 5 GHz connection. There are different techniques used in implementing an IPS with WLAN. One of the most commonly used techniques is the Wi-Fi fingerprinting method. It works with received signal strengths from the access points being the measurements and with which the position is being estimated. (Doiphode, Bakal, & Gedam 2016.) There are two modes namely in the localization process of a fingerprinting method: an offline mode and an online mode to perform the indoor positioning function. In the offline mode, a fingerprint database is created by collecting received signal strength measurements from Wi-Fi routers. Along with these received signal strength indicators (RSSI), location information of the measurement reference points need to be collected too. Received signal strengths are nothing but the amount of power received from the transmitter to the receiver. The best practice to identify a fingerprint is to take an average from the readings that were collected. A single reading cannot be identified as a fingerprint due to the presence of noise in the environment. Thereafter the online mode is used for positioning. Fingerprinting is described in more detail in subchapter 2.3. Another location determination method that uses geometry is the trilateration method. Location is given at the point of intersection of three circles formed by distance measures to Wi-Fi access points from the user device. In this method, distances are calculated with various signal measurement techniques such as Time Difference of Arrival (TDoA), Time of Arrival (ToA), Angle of Arrival (AoA) and Received Signal Strength and more.





**Figure 1.** Existing Wi-Fi Infrastructure in the Technobothnia Technology Center Layout at the University of Vaasa Campus.

As described in (Rashid, Chowdhury & Nawal 2017), in the ToA technique, the distance is computed by measuring the time taken by the signal to reach the receiver from the access point. Geometric triangle relationship is used to determine the position coordinates. Diffraction and reflection of the radio frequency signal from the floor and walls cause multipath effects which results in poor ranging accuracy. Temporal absorption techniques can be implemented to reduce the multipath effect.

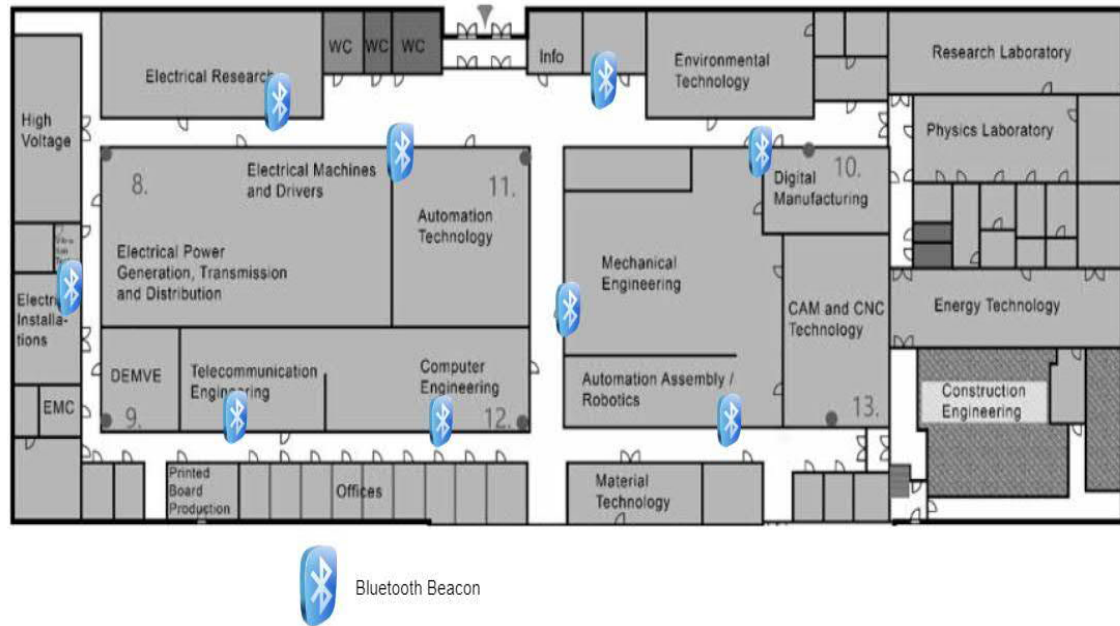
In an AoA method, location is computed by analyzing the angle of arrival of the signal. The challenging part is the implementation of this method in larger areas. As only two access points are used at any instance, cumulative measurement error of the AoA increases significantly. Both the ToA and the AoA are described in more detail in subchapter 2.3.

Indoor positioning with Wi-Fi is generally very cost-effective and easily deployable to existing infrastructures, as the one shown in Fig. 2.1, but it has its limitations too. As discussed in (Rashid et al. 2017), accuracy obtained through Wi-Fi-based positioning will be around 2.5 meters which is quite impressive but some applications demand even millimeter-level accuracy. For example, robots working on high-precision tasks can require even millimeter level of accuracy to spot and pick up assets.

### 2.1.2. Bluetooth

“Bluetooth indoor positioning is one of the hotspots in LBS since Bluetooth 1.0 published in 1999” (Jianyong, zili, haiyong & zhaohui 2014). Nowadays almost every smartphone has Bluetooth. Main advantages of using Bluetooth for also positioning is cost effective and easy deployment. Jeon, Kong, Nam & Yim (2015) has implemented an IPS with Bluetooth RSSIs, accelerometers and a barometer fused together. They carried out an experiment with Bluetooth RSSI and without Bluetooth RSSI. The outcome was that the system with Bluetooth RSSI gave better accuracy than without it. According to the experiment results, the obtained error rate was 12% and achieved accuracy 88%. By this, we can conclude that using Bluetooth signals with a related infrastructure is a suitable choice for implementing an IPS from the range of technology possibilities.

Bluetooth Low Energy (BLE) is a new standard for wireless personal area network technology developed by Bluetooth Special Interest Group (SIG) (Ji, Kim, Jeon & Cho 2015). The operating frequency of Bluetooth and BLE is 2.4 GHz. The throughput of BLE is smaller than traditional Bluetooth which is not supportive of voice transmission. When comparing power consumption, BLE consumes much less power than the classical one. In addition, the advantage of Bluetooth over Wi-Fi is that Wi-Fi typically needs electrical connections, but Bluetooth can be operated with a battery. In particular, it can be operated with a coin battery up to a few years. The range of transmission is directly proportional to the power consumed. In terms of BLE device size, the beacons are very small fitting almost in all places.



**Figure 2.** A suggested Bluetooth Technology Infrastructure Suitable for an IPS to the Technobothnia Technology Center Layout at the University of Vaasa Campus.

Rida, Liu, Jadi, Algawhari & Askourih (2015) carried out an experiment with Bluetooth signals and the trilateration method. The setup was made with Bluetooth 4.0 and beacons were programmed to broadcast signals at every 400 ms and then go to a sleep mode. As the trilateration algorithm is not so complex it can be easily implemented in hardware. Results of these experiments are quite impressive with given on average 0.5-1 meters of error.

Looking more closely at the limitations of Bluetooth technology, the downside is that it cannot cover a wide operational range, due to that the low consumption of power leads to weak signal strength transmission.

### 2.1.3. Ultra Wide Band (UWB)

Ultra Wide Band (UWB) is one of the technologies that suits well for an Indoor Positioning System. Dabove, Pietra, Piras, Jabbar & Kazim (2018) has presented an indoor positioning solution with two-way time of flight (TWTF) system for range measurements. In experimentation,  $100 \pm 25$  mm of average 3D accuracy was obtained

in the tests. The system is further tested in a harsh environment, which gave a decent horizontal accuracy of 87.4 mm. This technology suits well for an environment where the multipath effect is high. In all, Two Way ranging (TWR), ToA and TDoA techniques can be used to measure high-resolution time in UWB systems. As stated by Dabove et al. (2018) about UWB limitations, "Federal Communications Commission (FCC) in the USA has limited the unlicensed use of UWB to an equivalent isotropic radiated power density of -41.3 dBm/MHz and restricted the frequency band to 3.1 GHz - 10.6 GHz (respectively 6.0 GHz - 8.5 GHz in accordance to the European Communications Committee (ECC)) in order to avoid interference". In their results, it is further concluded that by implementing two-way time of flight (TWTF) and a multilateration method for estimating Trans-Receiver tag's position, better accuracy, high data rate and low power consumption would have been obtained. The advantages of UWB systems include that the distance estimated is directly correlated to the bandwidth of the signal. Limitations include the fact that such a system requires a network infrastructure for positioning to be constructed that ultimately may lead to high cost.

L. Yao, Wu, Yao, & Liao (2017) described an indoor positioning UWB solution with Inertial Measurement Unit (IMU) assistance. Interest behind these two technologies combined is they are complementary to each other. For data fusion, a Kalman filter was used. Their experiment proved that there is a further improvement in accuracy in the technology combination. This gives a clear picture that adding more data/data fusion leads to better accuracy. Along with this, power consumption and processing time should be kept in mind.

Due to objects and walls in typical environments for IPS implementation, signals get reflected that creates a severe multipath effect, and blockages that lead to non-line of sight (NLOS) situations in signal reception. Through UWB, these problems can largely be overcome commercially available UWB positioning devices nowadays include e.g. the DWM 1000 from DecaWave and Pozxy from Kickstarter.

#### 2.1.4. ZigBee

ZigBee technology is especially suitable to be implemented for isolated locations and harsh radio environments. Some of the applications existing with ZigBee are electric meter readings, lighting control and smoke detection. In general, there is a maximum number of devices of 65535 that can be connected to any ZigBee network. ZigBee technology based on the IEEE 802.15.4 standard is one of the mostly applied. It has the capability of providing received signal strength (RSS) measurements that is suitable to provide indoor positioning functionality. At known physical locations are multiple reference sensors and the sensors without known location are called blind nodes or target nodes. RSS measurements received by the reference sensors from the target node are used for positioning. Advantages include the high-level communication protocol used, low cost and low power consumption. (Liu, Chen, Kao, Hong, & Yang 2017.)

### 2.2. IPS based on Non-RF Technologies

#### 2.2.1. Visual Navigation

Visual-based navigation is infrastructure free, which means that there is no need for preinstalled devices in the environment such as in Wi-Fi and Bluetooth positioning. Only self-contained sensors are included. The advantage is the infrastructure-free environment where there is no need for prior knowledge about floor maps and the base stations in the indoor environment. In visual navigation, a video camera can serve as a visual gyroscope and an odometer (Ruotsalainen, 2013). Turn rate and translation from the consecutive images can be extracted. Positioning can be also done with known visual markers. In this method, visual markers are placed in the venue with location information included, which naturally is not anymore infrastructure-free. Latitude, longitude and altitude can be included in the marker information. Through camera, visual angle is measured to the marker and the location is computed. In other related methods, a database can be built with snapshots of the venue. When a camera is moving along the venue, current snapshots are taken to interpolate with the snapshots already present in the database and the positioning

is computed. Rantanen, Mäkelä, Ruotsalainen & Kirkko-Jaakkola (2018) experimented by fusing inertial navigation and visual navigation. The machine learning algorithm Random Forest classifier was used to classify different movements of the user such as climbing, walking etc. Tests carried out in co-operation with the Finnish Defense Forces showed excellent accuracy of 92% and 94% for the first and second dataset experimented, respectively. Overall a data fusion is a good approach to handle tactical and rescue applications.

### 2.2.2. Geomagnetic Field

As presented by Cao & Kang (2017), sea turtles use the principle of geomagnetic navigation for long-distance migration, and the same can be applied in electrical systems. Advantages of geomagnetic navigation are typically their low power consumption, being weather proof, no radiation emission and feasibility of implementation. These features make geomagnetic navigation well suitable for IPS. Finnish engineers at IndoorAtlas Ltd ([www.indooratlas.com](http://www.indooratlas.com)) developed an indoor navigation technology based on geomagnetic field and achieved precision up to 0.1-2.0 meters. Their system has been evaluated with experiments to estimate pedestrian position with a real-time geomagnetic field. A database is built offline with magnetic sensor information. In particular, attitude and heading of a smartphone was measured by Xsens MTi-300 equipment ([www.xsens.com](http://www.xsens.com)). Inertial sensors were used to estimate the user direction. All the data collected during offline was preprocessed and errors caused due to concrete walls were removed. Next, the processed data was fed to an algorithm for positioning. Results showed that positioning error around 1.2 meters were obtainable.

As discussed by Li, Gallagher, Dempster & Rizos (2012), advantages of geomagnetic navigation are that it does typically not require any infrastructure or network. Looking on to the disadvantages, the magnetic field might be affected by electrical devices, pipes and steel in the buildings. In other technologies, generally fingerprint data consists of more information which furthermore leads to better positioning performance. Unfortunately, only three component (X, Y, Z) intensities are existing in magnetic field data which gives

less information when compared to other positioning technologies. In this case true north is unknown, with help of accelerometer only vertical intensity and horizontal intensity are extracted.

### 2.2.3. Inertial Navigation System

Accelerometers and gyroscopes are inertial sensors that can be used for navigation. Inertial navigation has gained attention because the sensors of low cost can be found that provide good accuracy. Main advantages are that no prior knowledge of the environment is needed apart from the initialization, no extra infrastructure is needed and external factors such as constructions cannot effect these sensors. This feature makes inertial navigation capable of being used in tactical situations. These sensors provide speed and direction of the object and by knowing the initial position, it is possible to estimate the future position known as Dead Reckoning. (Batistić & Tomić 2018.) The main drawback is that a small error in direction estimation can lead to a large error in position estimation. Particle filtering (PF) was introduced to reduce errors. Obstacles (like walls, furniture) were taken into consideration by the particle filter to eliminate positions that are practically impossible. PF is a numerical approximation to a Bayesian filter. Bayesian filter uses probability distribution for estimating location and the PF represents the probability by using a set of weighted samples. (Batistić et al. 2018.) So, in practice the PF algorithm is used to reduce error accumulation. An experiment carried out by Han & Zhao (2017) showed that inertial navigation combined with other technologies was improved in accuracy.

Pedestrian Dead Reckoning (PDR) is one of the promising indoor positioning technique. Step detection and step length estimation are important in this technique. Based on the characteristics of human walking, the current position can be updated from the previous position with incremental locomotion. Pedestrians can be tracked with this technique with readily available sensors in smartphones. Errors in estimating step length and heading direction produce cumulative errors that affects the PDR technique. A machine learning algorithm decision tree (J48 Classifier) has been implemented to classify the basic holding styles of a smartphone. Wi-Fi fingerprinting and map matching has been further combined with PDR which resulted in 51.02% reduced positioning error. (Ahmed, Diaz, Angel, &

Minguez 2017.)

#### 2.2.4. Infrared Beacon

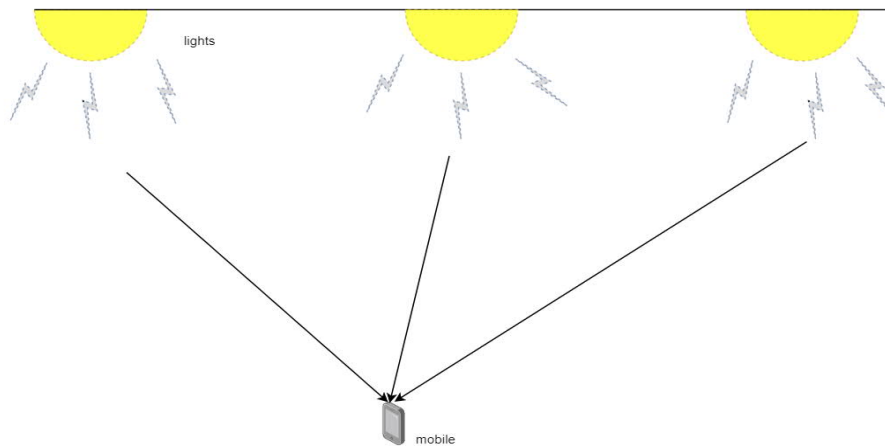
Working principle of infrared beacon is emitting a detecting the light that has wavelength ranging from 1 nm to 10 mm. Beacon takes turns in transmitting and receiving so they never misjudge their own signals. Every day example for IR is TV remote.

Yao & Ma (2018) proposed a method that consists of fixed infrared beacons and a K real-time grouping algorithm for indoor positioning. They stated that inaccuracy in different positioning methods is due to self-stability and drift in practical applications. In their procedure, reference points are fixed infrared beacons, a pan-tilt camera is fixed above the user for tagging measurements and the height is computed in real-time depending upon the angle difference between the vertical position and the horizontal position after the rotation and fixed beacon's height. This is effective in dealing with drift in positioning data according to simulation results. In the hardware configuration processor module, angle sensor module and camera module are included. The rotation angle of the platform is the key to measuring the positioning information. Once the data is acquired it is fed into the K-time group algorithm and position is estimated. When compared with wireless signals this method is more stable and there is no need to worry about the signal attenuation problems. The outcome of the experiment was that by using K-time grouping, algorithm positioning accuracy had improved.

#### 2.2.5. Visible Light Communication

Visible Light Communication (VLC) is one of the technologies that can also be used for indoor positioning. In this, the light emitting diode (LED) and the photodiode (PD) are the key components. PD is used as a receiver that senses the intensity of the light whereas LED acts as a transmitter. In this environment, LED's are placed at know points and through receivers, the intensity of light is calculated. The distance between the receiver and transmitter are estimated with positioning techniques and the current position of the receiver will be localized.





**Figure 3.** Visual Light Communication with LED Lights.

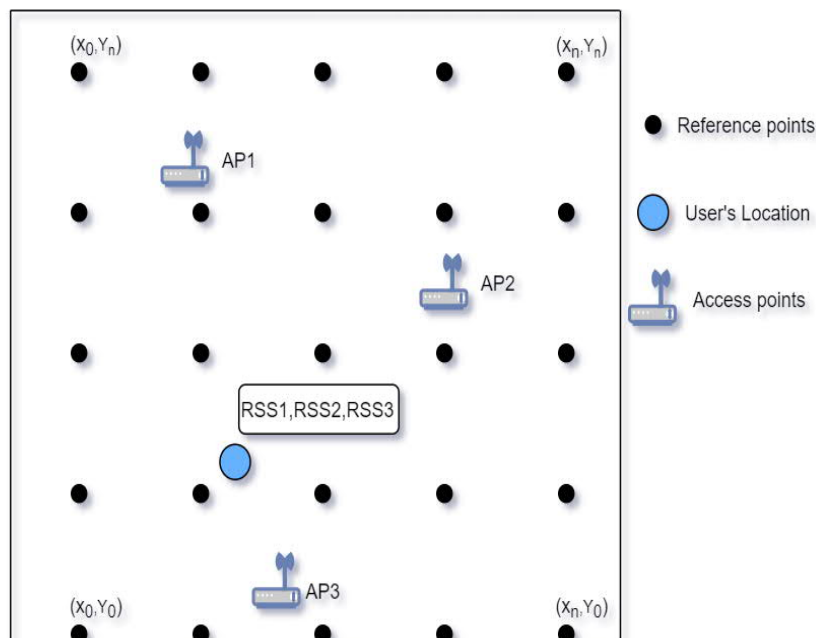
Main advantages of VLC positioning over other technologies are low energy consumption, robustness, extended lifetime, can be used in RF sensitive areas such as hospitals, provides high precision, does not suffer from interferences such as with RF systems, no extra infrastructure other than the existing LED lamps needed, and is non-sensitive to electromagnetic fields and other environmental interferences.

Three positioning algorithms typically used for VLC are Receiver Signal Strength (RSS), TDOA and Phase Difference of Arrival (PDOA). Synchronization is challenging when implementing TDOA and PDOA methods. The processor has to detect and differentiate the phase difference between two or more received signals which makes the system complex. RSS method is preferred over these methods. From Li, Cao & Chen (2018), an experiment was conducted that provided 0.79 cm as the average error between predicted coordinates and actual Coordinates. Neural Network (NN), a machine learning based algorithm was used in the VLC positioning method. Attribute feature consisted of the time difference between the location and LED. Fixed point coordinates were physical locations of tags. The Neural Network learns the relationship between attributes and tags. Data undergoes a training phase and a validation phase. Hyper parameters in the NN are tuned for accuracy. From simulation analysis, 1.62 cm was the obtained positioning error. (Li et al. 2018.)

## 2.3. Techniques for Indoor Positioning

### 2.3.1. Fingerprinting

Fingerprinting is one of the most commonly used positioning estimation techniques. Storing the signal strengths in a database along with other parameters is the key. This method can be used with different technologies such as Wi-Fi and the geomagnetic field. (Karppinen 2018.) A map with radio properties is created with different signals along with other information that is stored in the database. These properties at different locations can be used as reference points. A new device's location can be determined by matching the signal properties of the device with reference points' signal properties stored in the database. According to Engström & Helander (2015), the main advantage of this method is that even without knowing the physical location of the access points (AP) positioning can be estimated.



**Figure 4.** RSS1, RSS2, RSS3 are the Received Signal Strengths from Access Points AP1, AP2, AP3 Respectively.

Fingerprinting is done in two phases, offline and online

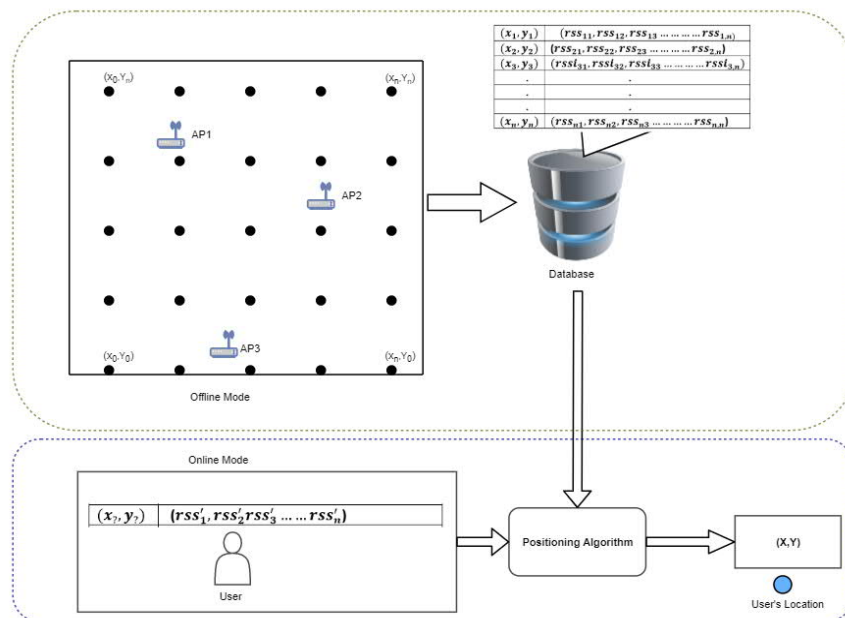
### Offline phase:

Firstly, an indoor area is marked with reference points. A coordinate system is used to refer to each of the reference points. Once this is done, radio data is collected at the points. A specific radio map will be created with the data collected and stored in a database. This is also known as the calibration phase

In the radio map, each reference point has a unique identity and RSS values. Radio signals from AP's undergo multipath and signal reflections due to the indoor environment at hand. The database will be stored with the unique identities along with a vector of RSSs and the corresponding coordinates.

### Online phase:

In this mode, user's device collects the RSS information from the available AP's and sends them to the database for fingerprint matching. In return, user receives a location depending on the best possible match of fingerprinting.



**Figure 5.** Online & Offline Phases of the Fingerprinting Method.

### 2.3.1.1. Weighted Centroid Algorithm

In this algorithm, RSSI and distances from the reference nodes are two important elements for positioning. Firstly, Weights are calculated based on RSSI between reference nodes and unknown node, without considering environmental effects on RSSI values. (Fan, He, Tao & Xu 2013.) The following is the mathematical explanation

Let's assume

$A_i$  Be the know reference node at position  $(P_i, Q_i)$  ( $1 \leq i \leq n$ ) and the estimated position of unknown node  $M$  be  $(P_z, Q_z)$ .

Number of beacon nodes be  $n$  and their coordinates

are  $A_1(P_1, Q_1), A_2(P_2, Q_2), A_3(P_3, Q_3) \dots A_n(P_n, Q_n)$  respectively

$T_1, T_2, T_3, \dots, T_n$  are the distance from reference node to unknown node respectively.

Consider two nodes  $A_1$  and  $A_3$  and their centroid position is  $A_{13}(P_{13}, Q_{13})$ .  $D_1$  and  $D_3$  Are distance to  $A_{13}$  from  $A_1$  and  $A_3$  respectively.

From centroid principle

$$\frac{(P_{13}-P_1)}{T_1} = \frac{(P_{13}-P_3)}{T_3} \quad (1)$$

$$\frac{(Q_{13} - Q_1)}{T_1} = \frac{(Q_{13} - Q_3)}{T_3} \quad (2)$$

Above equation can be written as

$$P_{13} = \frac{\frac{P_1}{T_1} + \frac{P_3}{T_3}}{\frac{1}{T_1} + \frac{1}{T_3}}, Q_{13} = \frac{\frac{Q_1}{T_1} + \frac{Q_3}{T_3}}{\frac{1}{T_1} + \frac{1}{T_3}} \quad (3)$$

For the coordinates P and Q are

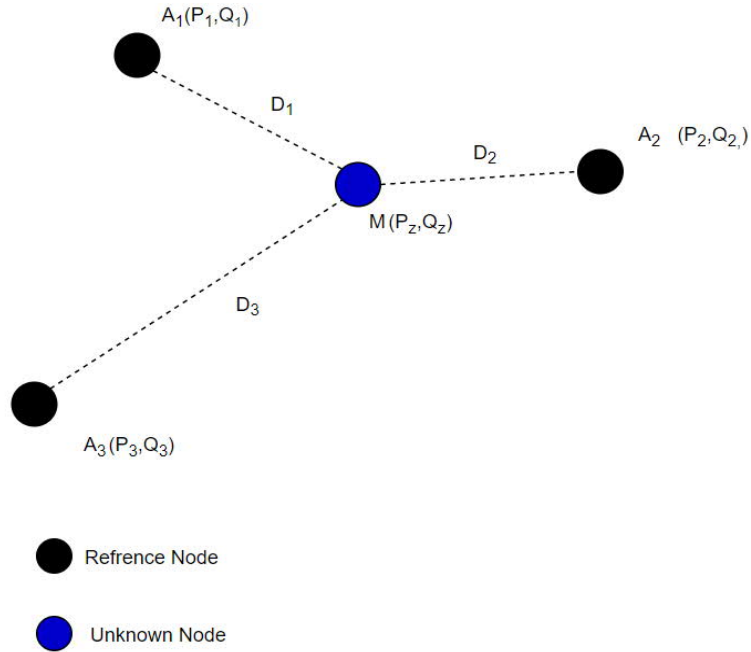
$$P_{13} = \frac{\frac{P_1}{T_1} + \frac{P_3}{T_3} + \dots + \frac{P_n}{T_n}}{\frac{1}{T_1} + \frac{1}{T_2} + \dots + \frac{1}{T_n}}, Q_{13} = \frac{\frac{Q_1}{T_1} + \frac{Q_3}{T_3} + \dots + \frac{Q_n}{T_n}}{\frac{1}{T_1} + \frac{1}{T_3} + \dots + \frac{1}{T_n}} \quad (4)$$

From the above equations it is clear that, weight value of a node is directly proportional to distance.  $W_i$  is the weight of each reference node. If communication between unknown node and known node is not successful, then  $W_i = 0$ .

Now, the weighted centroid formula can be written as

$$P_z = \frac{\sum_{i=1}^n W_i \times P_i}{\sum_{i=1}^n W_i}, Q_z = \frac{\sum_{i=1}^n W_i \times Q_i}{\sum_{i=1}^n W_i} \quad (5)$$

(Fan et al. 2013.)



**Figure 6.** Illustrates Weighted Centroid Localization.

### 2.3.1.2. Log Gauss Likelihood Algorithm

In Log Gauss likelihood estimation algorithm, number grid points are calculated from the training samples. Then training matrix and test matrix are generated from the training and test samples respectively. Each point is given a probability corresponding to the likelihood that user's/current RSS vector corresponds to that point. (Lohan, Torres-Sospedra, Richter, Leppäkoski, Huerta & Cramariuc 2017.) The following steps gives the mathematical explanation.

Steps involved:

- Training data will be collected form the fingerprint data base and the current RSS value is collected from the device
- Likelihood is estimated by Gaussian similarity
- Replace zero's with a very small value, let say  $1 \times 10^{-6}$
- For increased stability and efficiency apply  $\log$
- Calculate cost function that is add all individual probabilities
- Lastly find N points with highest probability to estimate the position

Let  $x_i$  be RSS training samples  $i = 1, 2 \dots N$

RSS current are the samples from the user

Let's assume estimating parameters of exponential function in Gaussian

$$x[n] = r^n + w[n], n = 0, 1 \dots N - 1 \quad (6)$$

Where  $w[n] \sim (0, \sigma^2)$

Here we estimate is  $r > 0$

Function to represent likelihood is

$$p(x; r) = \frac{1}{(2\pi\sigma^2)^{\frac{N}{2}}} \exp \left[ -\frac{1}{2\pi\sigma^2} \sum_{n=0}^{N-1} (x(n) - r^n)^2 \right] \quad (7)$$

For increased stability and efficiency apply log to the above function

$$\log(p(x; r)) = \log \left[ \frac{1}{(2\pi\sigma^2)^{\frac{N}{2}}} \exp \left[ -\frac{1}{2\pi\sigma^2} \sum_{n=0}^{N-1} (x(n) - r^n)^2 \right] \right] \quad (8)$$

$$\log(p(x; r)) = -\frac{N}{2} \log(2\pi\sigma^2) - \frac{1}{2\pi\sigma^2} \sum_{n=0}^{N-1} (x(n) - r^n)^2 \quad (9)$$

$$\mu = \frac{\sum_n^N x[n]}{N} \quad (10)$$

$$\sigma^2 = \frac{\sum_n^N (x[n] - \mu)^2}{N} \quad (11)$$

Then cost function  $J(x)$  is calculated by adding all individual probabilities

$$J(x) = \sum_n^{N-1} (\log(p(x; r))) \quad (12)$$

Next cost function is sorted in descending order along with their respective  $id$ 's from 1 to M. Here M will be number of points with highest probability.

Then position will be the mean of training data 1 to M.

$$Position = \frac{\sum_{i=1}^M x_i(n)}{M} \quad (13)$$

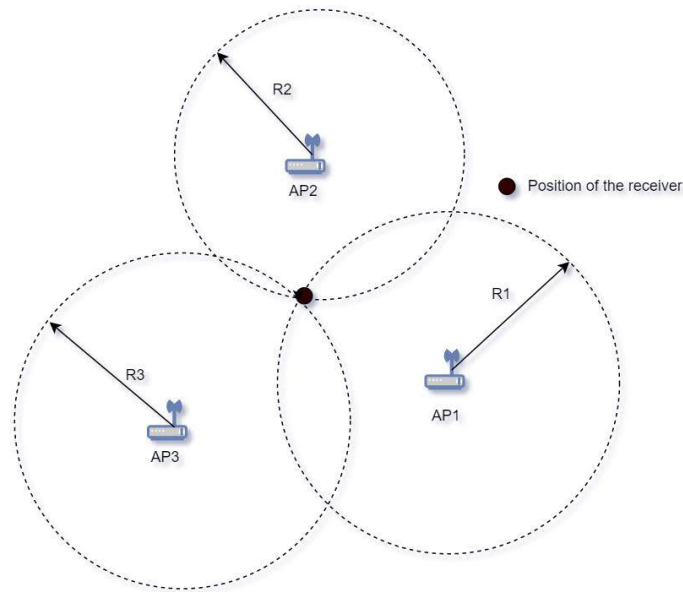
(Huttunen 2013.)

### 2.3.2. Trilateration

In GPS navigation and surveying, the trilateration method is widely used. As its name states, tri, three known APs are used to estimate the position of the user. Device's position is located by calculating the radius of the AP's signal range which is determined by RSSI

between the AP and the device, or propagation time. Each AP forms one sphere. Intersection of three sphere's radius will be the device location. So, a minimum of three AP's are needed to determine the location of the device. As well known, RSS and propagation time can be affected by obstacles in the indoor environment. So, the line of sight signal is preferred to avoid poor accuracy due to weak and delayed signals.

Trilateration indoor positioning shows good accuracy with UWB technology even with surrounding indoor obstacles. But due to its small range, more nodes are needed to cover an area. Wi-Fi is also one of the technologies where trilateration method can be applied to. (Engström et al. 2015.)



**Figure 7.** An Illustration of Trilateration.

Figure 2.7 illustrates trilateration, with  $(X_1, Y_1)$  &  $R_1$ ,  $(X_2, Y_2)$  &  $R_2$ ,  $(X_3, Y_3)$  &  $R_3$  are the coordinates and distances to the unknown coordinates of the user receiver from AP1, AP2 and AP3 respectively.  $(X_0, Y_0)$  are the coordinates of user receiver. User's coordinates can be solved with the equations in a two-dimensional case

$$(X_1 - X_0)^2 + (Y_1 - Y_0)^2 = R_1^2 \quad (14)$$



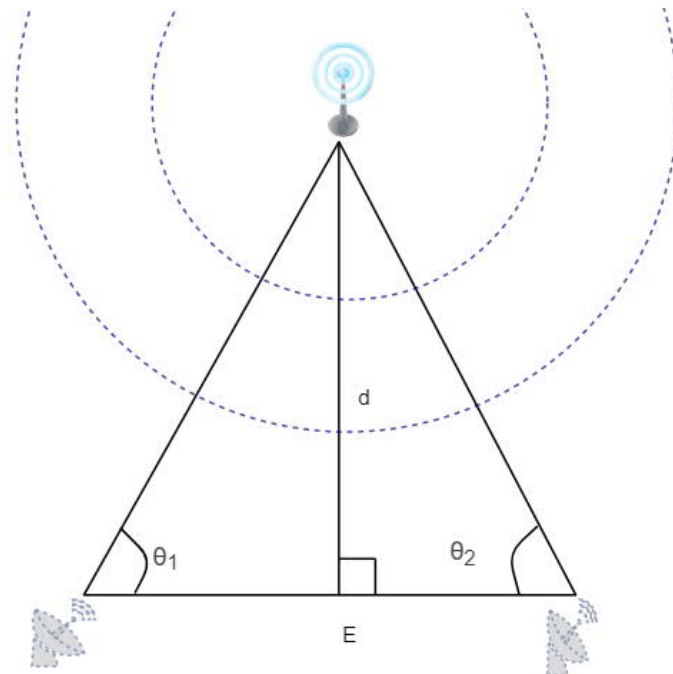
$$(X_2 - X_0)^2 + (Y_2 - Y_0)^2 = R_2^2 \quad (15)$$

$$(X_1 - X_0)^2 + (Y_3 - Y_0)^2 = R_3^2 \quad (16)$$

To calculate the distance between the device and AP, there are two different lateration techniques typically used: Time of Arrival (TOA) and Time difference of Arrival (TDOA). There, RSSI or signal travel times are calculated and radio signal velocities are multiplied to obtain the distances. (Karppinen 2018.)

### 2.3.3. Triangulation

Triangles are used for geometric calculations in the triangulation method. This requires two receivers, namely the transmitted signal need to be received by the two receivers. They key is calculating the angle of the arriving signal at each receiver.



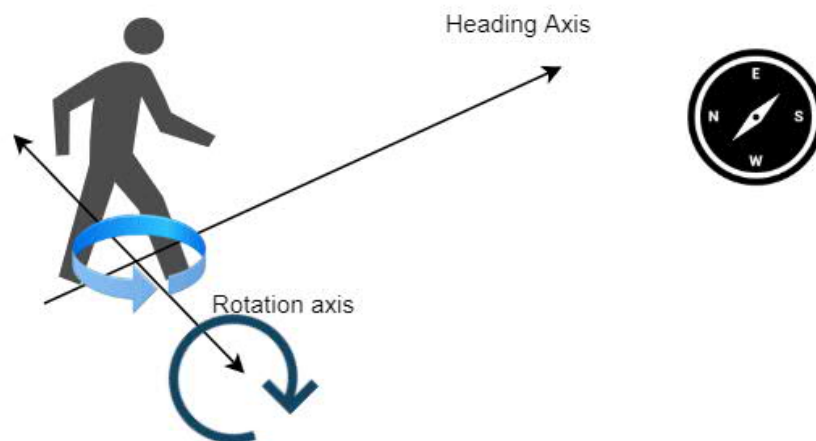
**Figure 8.** An Illustration of Triangulation.

$$d = E \frac{\sin(\theta_1) + \sin(\theta_2)}{\sin(\theta_1 + \theta_2)} \quad (17)$$

After obtaining the angles  $\theta_1$  and  $\theta_2$  the distance  $d$  is determined by the equation.  $E$  is the known distance between those receivers. (Engström et al. 2015.)

#### 2.3.4. Pedestrian Dead Reckoning

Inertial sensor modules consist of accelerometers, magnetometers and gyroscopes. These sensors provide information about user's movement. Accelerometers can be used to count the number of steps and the direction is determined by the gyroscope or magnetometer. User's location is obtained by adjusting the average step length for the walking speed. As there will be no reference to the true location there will be successive increase in the error. (Kang & Han 2015.)



**Figure 9.** Basic Understanding of Pedestrian Dead Reckoning.

### 3. DEVELOPMENT OF IPS ANDROID APP BASED ON HERE SDK

#### 3.1. Introduction

As mentioned in the above chapter, IPS can be achieved with multiple technologies. The reason for choosing Wi-Fi technology is that the infrastructure is already existing and no other extra cost is typically included. In this case, IPS is used by humans with the employment of Wi-Fi technology that provides decent accuracy for indoor navigation. HERE Technologies, a mapping and navigation service provider for both indoors and outdoors, provide both a software development kit (SDK) and Radio mapper tool to be used for indoor navigation functionality. The idea utilizing these software tools is to build an Android mobile app, which gives indoor navigation and routing to the Technobothnia technology center building as a visitor guidance tool. The SDK takes care of the RSSI measurements and positioning. The radio mapper is a tool used to create radio maps on the site of the venue at hand. In general, HERE provides maps for both public and private venues. As the Technobothnia building comes under private venues, HERE developed a 3D indoor map specifically for this building in order to be able to provide the center visitors with 3D routing.

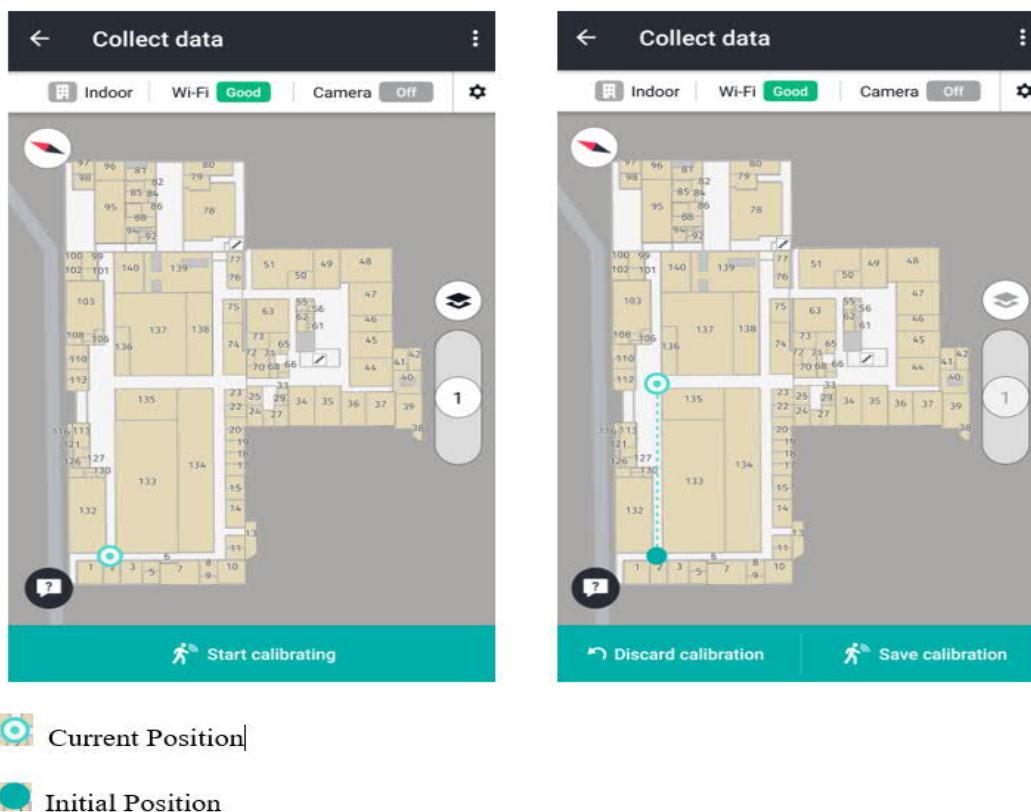
An Android app is developed on this SDK for positioning and routing. This app can be installed on both mobiles and tablets as well. Once the app is installed and initiated, it loads the 3D indoor map of the Technobothnia building next to the University of Vaasa campus and shows the user's current position based on the data that is collected during the radio mapping phase. It is possible to select the destinations within in the environment for navigation purposes. User's real time location can be seen as a blue dot on the map and it follows the user's movement and thus provides tracking.

### 3.2. HERE SDK

HERE maps provides an SDK tool (Software Development Kit) for establishing an IPS. After login to the HERE account ([www.developer.here.com](http://www.developer.here.com)) with user credentials, the Here SDK for Android can be chosen. This SDK is a system, which can only be viewed in terms of its inputs and outputs, and modifications cannot be done inside the SDK.

### 3.3. HERE Indoor Radio Mapper

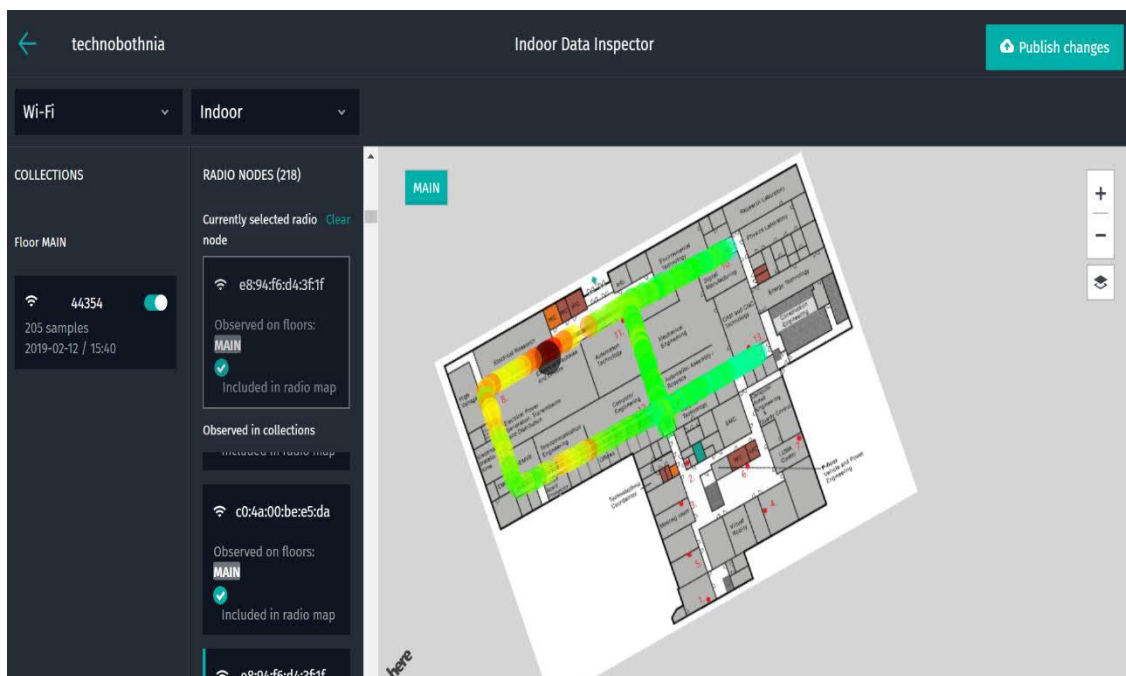
The radio mapper is an app provided by HERE. After login, a venue can be added and given a name to. A floor plan with .png format is uploaded. Then carefully the floor map can be placed on the outdoor map provided by HERE and saved.



**Figure 10.** Showing Current Position & Initial Position on Technobothnia Floor Plan in the HERE Indoor Radio Mapper App.

Next, the radio mapping functionality starts with the floor map being displayed and the user's position (blue dot) can be marked on the map. After calibration, and saving the settings, user's position can be marked again on the map. Now, the user can start walking in the path where radio data will be collected and samples can be seen being collected. Once the path is done, the track can be saved. Radio data are published by clicking on the publish button and changes can be seen in the Radio Map Admin Tool.

The created radio maps can be visualized in the Radio Map Admin Tool. In this, the floor map is seen along with RSSI distributions at different access points (MAC addresses).

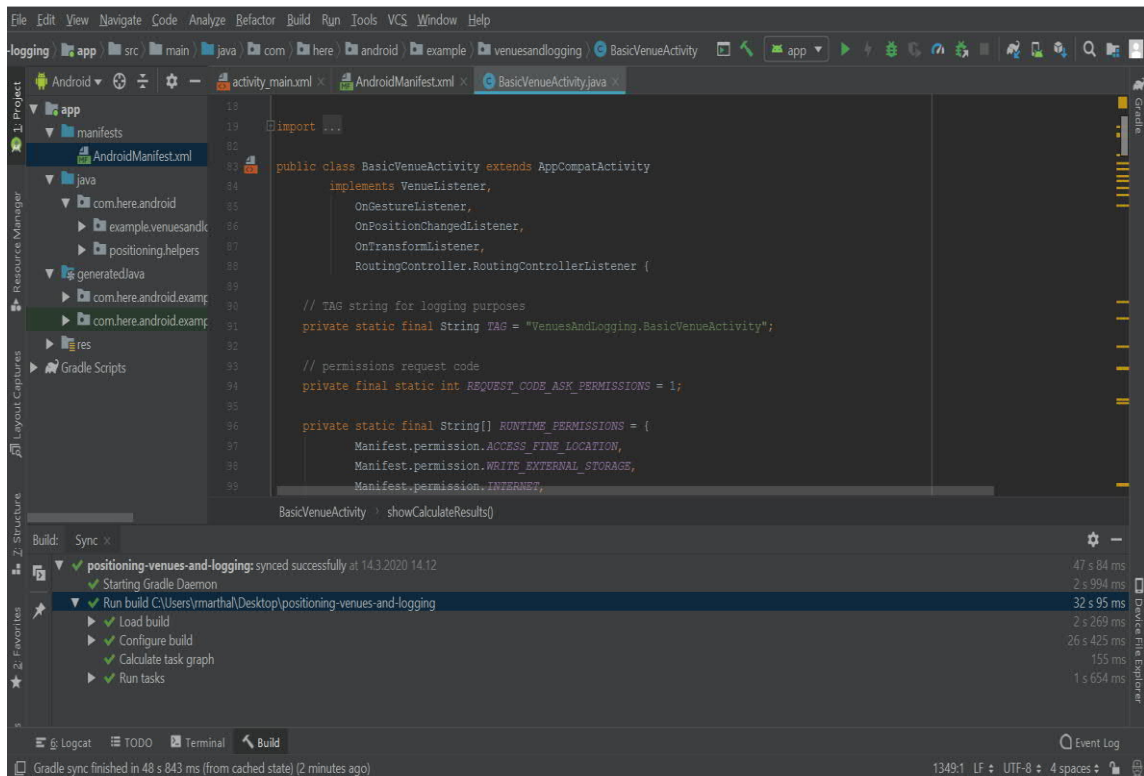


**Figure 11.** Showing RSSI Values in the Radio Map Admin Tool.

### 3.4. Android Studio

An IDE (Integrated Development Environment) to develop Android apps was next utilized. This official IDE from Google and can be downloaded from (<https://developer.android.com/studio>). The IPS Android app is developed using the Android studio with the HERE SDK integrated. The user interface (UI) is designed in a

manifest file and the activity is the main file of development. Android emulators are built in this IDE and it is therefore possible to run the app before it is being installed on the device.

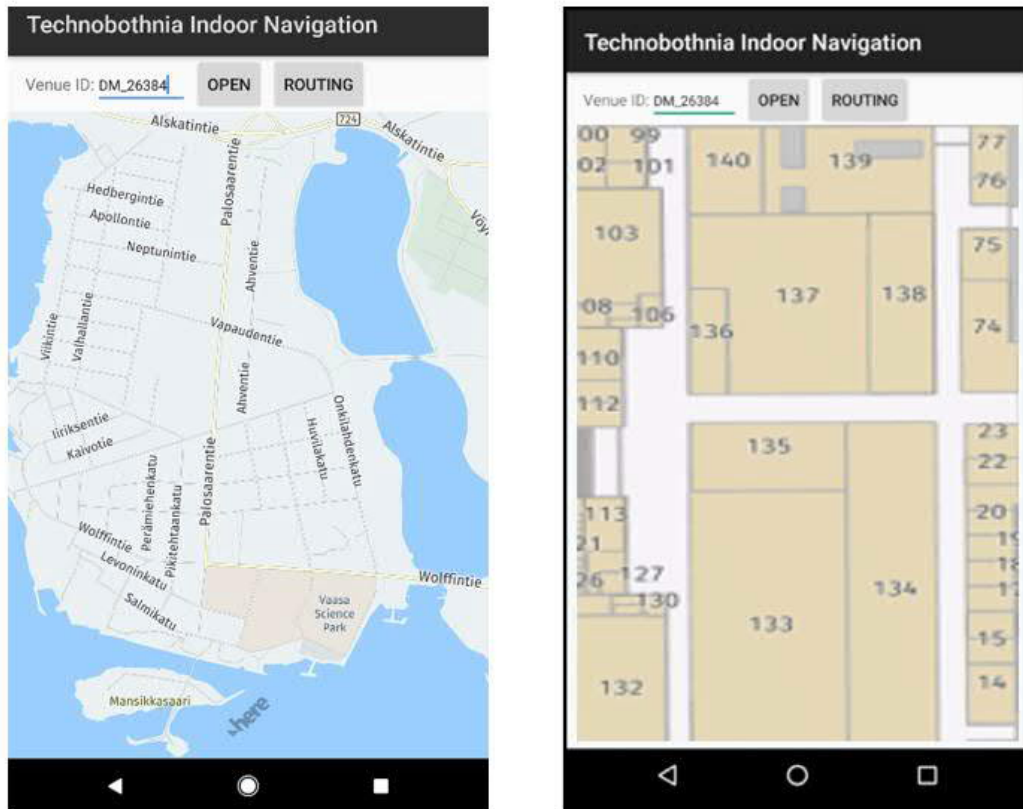


**Figure 12.** Showing the Android App Developing Environment in Android Studio.

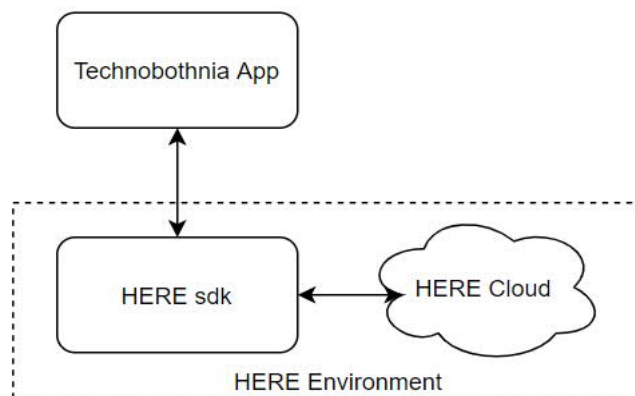
### 3.5. Technobothnia Demo App

A demo app for indoor positioning in the Technobothnia technology center in Vaasa, Finland is developed with Android Studio and built on the HERE SDK. The app can be downloaded through the link <https://sites.univaasa.fi/indoornavigation/technobothnia-indoor-navigation/>. After the download, the app can be opened and when entering the venue ID *DM\_26384*, the Technobothnia map will be shown on the screen. The SDK communicates with the HERE cloud to synchronize the radio data. Once data is synchronized, it is possible to use the app offline. Routing is possible within the venue by

selecting a starting point and an ending point. There are options to choose the fastest or the shortest route.



**Figure 13.** The First Page & the Venue Map on the Technobothnia Demo App.



**Figure 14.** Communication between the App & the HERE Environment.

## 4. MACHINE LEARNING IN GENERAL

### 4.1. Introduction

Machine learning (ML), a subset of AI, has grown significantly in recent years. Professor Andrew Ng stated that "AI is new electricity". This means that the invention of electricity led to many other applications that changed human life forever in the same way AI can transform our current lives into being much more comfortable and secure. (Wharton 2017.)

Machine learning provides the ability for systems to automatically learn and improve from experience without being explicitly programmed. Machine learning is most often divided into supervised learning and unsupervised learning. The type of learning in which we have historical data, feed to an algorithm and make predictions for the information is called supervised learning. In unsupervised learning, systems find the hidden patterns in the data. Nowadays, due to vast growth in sensor technologies, IoT (Internet of Things), processing power and storage capacity, lots of data is being generated and stored. These generated data will be useful for analyzing the behavior of the system and finding the hidden patterns will help to extract useful information. So, the data is fed to the machine learning algorithms for prediction or classification depending on the application.

ML as a tool for positioning has advantages over other positioning algorithms. In IPS, machine learning algorithms are implemented to improve accuracy, with typically less computational power. Many papers have been published with the implementation of machine learning algorithms in indoor navigation, and positioning showed improvement in accuracy. Jedari, Wu, Rashidzadeh & Saif (2015) published a paper on indoor positioning with RSSI and machine learning algorithms. K nearest neighbor (k-NN), a rules-based classifier (JRip) and random forest algorithms are experimented with to predict indoor positioning. An RN-131-EK Wi-Fi board was used for RSSI fingerprinting. The data collected were fed into those algorithms and their performance was analyzed. From results, random forest classifiers were ahead with 91% of accuracy which was quite good. Similarly, 77.40% improvement with k-NN.

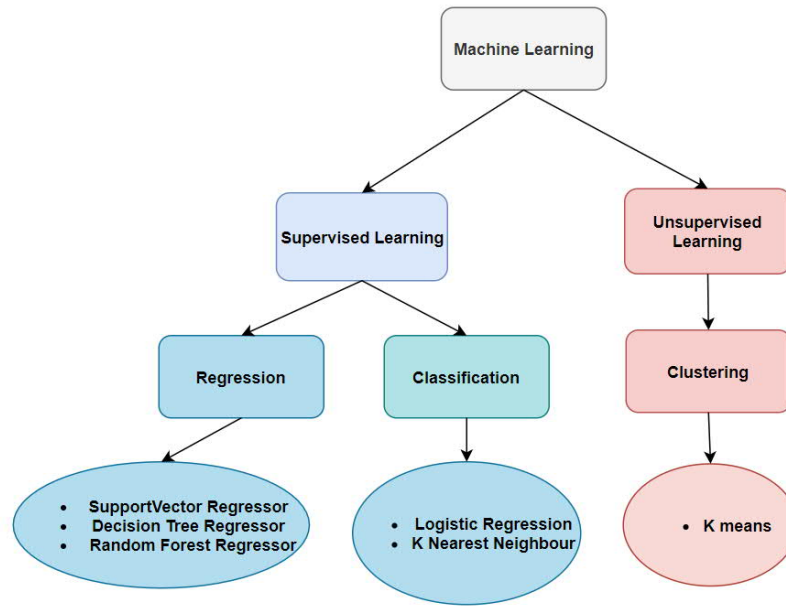


A paper published by Salamah, Tamazin, Sharkas & Khedr (2016) states that enhanced accuracy, less computation time and cost can be achieved through machine learning. RSSI fingerprint data is collected and processed with Principle Compact Analysis (PCA) to extract information from a collected radio map. Apart from that, without losing the information, a multivariate data matrix can be reduced. In this method, k-NN, random forest classifier, Support Vector Machine (SVM) classifier and decision tree algorithms are fed with RSSI data. Results show that in static mode random forest classifier computation time was reduced 70% and in a dynamic case 33% by the k-NN. Lukito & Chrismanto (2017) has done a similar experiment with RSSI fingerprints from Wi-Fi nodes but used Recurrent Neural networks (RNN). The forward connection might not always be present in RNN. In comparison with J48, SVM, Naïve Bayes, the Multi-layer perceptron RNN exhibited better accuracy.

#### 4.2. Machine Learning Algorithms

Machine learning algorithms are selected depending upon whether it is a regression or classification problem. In regression, values are continuous and output will be a numerical value. Algorithm learns from the numerical data and predicts in numerical. For classification problem, algorithm learns from both numerical and categorical data. In this output will be a binary (0 or 1) which are classes or categories. Both regression and classification comes under supervised learning. As there will be no prior information about the data in unsupervised learning, algorithm makes clusters from the data by learning the hidden patterns which gives information.

In this thesis, ML algorithms are used to predict the coordinates of the user's location from the RSSI values which are numeric. It is clearly seen as a regression problem and the following algorithms are chosen.



**Figure 15.** Regression, Classification & Clustering in Machine Learning.

#### 4.2.1. Support Vector Regressor (SVR)

A Support Vector Machine (SVM) is one of the machine learning algorithm for multi-class classification. In this technique, the exploitation of support vectors from the learning dataset makes it possible to realize a discriminant function. The idea can be straight forwardly applied to Support Vector Regressor (SVR). (Kikuchi, Matsuyama, Sano & Tsuji 2006.)

Support Vector Regressor (SVR) algorithm is used to predict numerical values. This algorithm comes under supervised learning that means predictions are made from the historical data. Training of SVR includes solving quadratic programming problem, which has inequality constraints. (Nagatani & Abe 2007.) Considering ‘rbf’ kernel, experiments are conducted with other parameters are discussed in chapter 5(Scikit learn 2019.) Mathematical explanation of the algorithm is as follows

Let  $N$  be number of input-output pairs  $(x_i, y_i)(i = 1, 2, 3 \dots N)$ .  $g(x)$  Be the mapping function and the input vector  $x$  is mapped into high dimensional feature space. Then the approximation function  $f(x)$  be

$$f(x) = W^T g(x) + b \quad (18)$$

$W$  is the weight vector

$b$  is the bias term

Loss function is defined as

$$L(y, f(x)) = \begin{cases} 0 & \text{if } |y - f(x)| \leq \varepsilon, \\ |y - f(x)| & \text{otherwise,} \end{cases} \quad (19)$$

$\varepsilon$  is user-defined threshold.

Solving the regression problem

Minimize

$$\frac{1}{2} W^T W + C \sum_{i=1}^N (\varepsilon_i + \varepsilon_i^*) \quad (20)$$

$$\text{Subject to } \begin{cases} y_i - W^T g(x_i) - b \leq \varepsilon + \varepsilon_i \\ W^T g(x_i) + b - y_i \leq \varepsilon + \varepsilon_i \\ \varepsilon_i \geq 0, \varepsilon_i^* \geq 0 \text{ for } (i = 1, 2 \dots N) \end{cases}$$

$C$  is the margin parameter

$\varepsilon_i, \varepsilon_i^*$  are the slack variables of  $x_i$

Maximize

$$-\frac{1}{2} \sum_{i,j=1}^N (\alpha_i - \alpha_i^*)(\alpha_j - \alpha_j^*) g^T(x_i) g(x_j) - \varepsilon \sum_{i=1}^N (\alpha_i + \alpha_i^*) + \sum_{i=1}^N y_i (\alpha_i - \alpha_i^*) \quad (21)$$

$$\text{Subject to } \begin{cases} \sum_{i=1}^N (\alpha_i - \alpha_i^*) = 0, \\ 0 \leq \alpha_i, \alpha_i^* \leq C \end{cases}$$

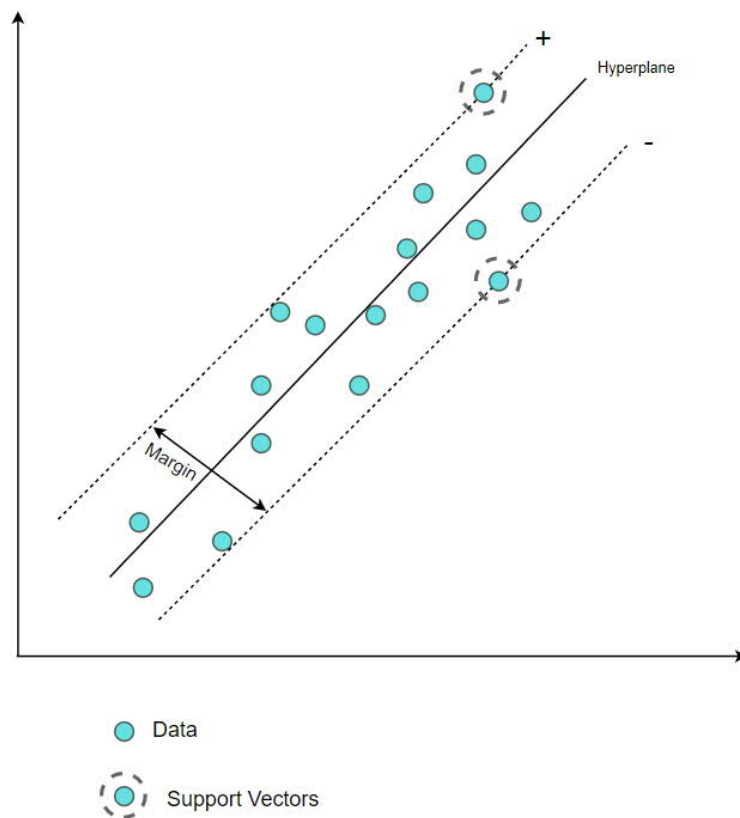
$\alpha_i, \alpha_i^*$  are the Lagrange multipliers associated with  $x_i$

$H(x_i, x_j) = g^T(x_i)g(x_j)$  is a kernel.

Final approximation equation will be

$$f(x) = \sum_{i=1}^N (\alpha_i - \alpha_i^*) H(x_i, x_j) + b \quad (22)$$

(Nagatani et al. 2007.)

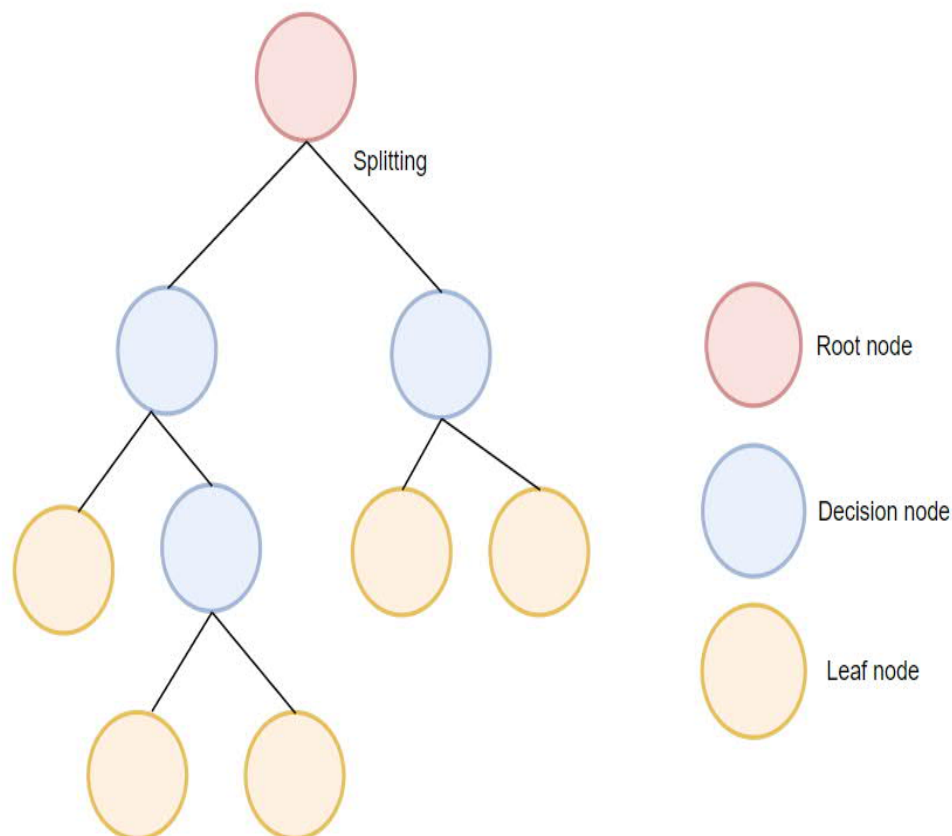


**Figure 16.** Schematic Diagram of Support Vector Regression.

#### 4.2.2. Decision Tree Regressor (DTR)

Decision Trees (DT) is used for classification and regression, which is non-parametric supervised learning. This algorithm can handle both categorical and numerical data. As the name indicates tree like structures are formed in which data is made into small subsets. Trees are mainly divided in to three nodes namely Root node, Decision node and Leaf node. Root node is the topmost node of the tree, Decision node is the node that has two or more branches depending upon the data and in Leaf node the numerical target is represented. (Sayad 2020.)

The main advantage of this algorithm is simple to understand and visualization of trees are possible.



**Figure 17.** Simple Explanation of Decision Tree.

DTs are mathematically explained as follows

Let,  $x_i \in R^n, i = 1, \dots, l$  be the training vector,

$y \in R^l$  be label vector,

$Q$  is the data at node  $m$

For each candidate split  $\theta = (j, t_m)$ ; where  $j, t_m$  are features and threshold respectively

$Q_{left}(\theta), Q_{right}(\theta)$  are subsets after data partition

$$Q_{left}(\theta) = (x, y) | x_j \leq t_m$$

$$Q_{right} = Q / Q_{left}(\theta)$$

From impurity function  $H()$ , Impurity of  $m$  is calculated as follows

$$G(Q, \theta) = \frac{n_{left}}{N_m} H(Q_{left}(\theta)) + \frac{n_{right}}{N_m} H(Q_{right}(\theta)) \quad (23)$$

Parameters that minimizes impurity is  $\theta^* = \operatorname{argmin}_{\theta} G(Q, \theta)$

Next repeat for  $Q_{left}(\theta^*)$  and  $Q_{right}(\theta^*)$  subsets until it reaches maximum allowable depth

$$N_m < \min_{samples} \text{ or } N_m = 1$$

In regression target will be continuous

For node  $m$ ,

$R_m$  is the representing region

$N_m$  is the observations

$X_m$  is train data at node  $m$

Mean Squared Error is calculated at terminal nodes by mean values that reduces  $L2$  error

$$\bar{y}_m = \frac{1}{N_m} \sum_{i \in N_m} y_i$$

$$H(X) = \frac{1}{N_m} \sum_{i \in N_m} (y_i - \bar{y}_m)^2 \quad (24)$$

Mean Absolute Error is calculated at terminal nodes by median values that reduces  $L1$  error

$$\bar{y}_m = \frac{1}{N_m} \sum_{i \in N_m} y_i$$

$$H(X) = \frac{1}{N_m} \sum_{i \in N_m} |y_i - \bar{y}_m| \quad (25)$$

(Scikit learn 2019.)

#### 4.2.3. Random Forest Regressor (RFR)

Random forest method comes under supervised learning, can be used to solve classification and regressions. This method was first introduced in 2001 by Leo Brieman. For small quantity of data with more number of features this method suits well (Zhang, Wang, Jiang, Fan & Dan 2015). "A random forest is a meta estimator that fits a number of classifying decision trees on various sub-samples of the dataset and uses averaging to improve the predictive accuracy and control over-fitting" (Scikit learn 2019).

Random forest regression makes predictions by combining decision from a sequence of base models. Each base model represents a decision tree and their cumulative output is random forest. This model is very good at learning non-linear interaction between features

and target. (Acharya, Armaan & Anthony 2019.) The following will be mathematical explanation of random forest regression algorithm.

Basically, a collection of tree structure classifiers makes a random forest classifier.

$$h(X; \theta_k), k = 1, 2 \dots K$$

$\theta_k$  are independent and identical distributed random vectors

$(X, Y)$  is an original data set where  $X$  is input variable and  $Y$  is output variable. Primarily with bootstrap extract  $K$  samples from dataset  $(X, Y)$ , having each sample size as of original dataset. Next step is getting  $K$  prediction results from the  $K$  regression model built for each sample. The final results will be the average of  $K$  results. The regression model sequence after  $K$  rounds will be  $\{h(X; \theta_1), h(X; \theta_2) \dots h(X; \theta_k)\}$ . The final prediction model will be as follows.

$$\bar{h}(X) = \frac{1}{K} \sum_{k=1}^K h(X; \theta_k) \quad (26)$$

Step-by-step process in the random forest regression algorithm

- Let  $n$  be the total number of samples in a dataset and by using bootstrap a small set of samples  $b$  is created from  $n$ . Then  $b$  regression trees are grown. In this recursive process, samples that are not chosen for  $b_{out-of-bag}$  are used as test samples

At each node of regression tree  $m_{try}$  variables are specified as alternate branch variables. The optimum rule for selecting the best branch is

$$m_{try} \ll \frac{P}{3}, \quad \text{where } m_{try} \ll P$$

Where  $P$  is the variables in dataset

$$m_{try} \ll P$$



- From top to bottom each tree starts growing branches. For the trees to stop growing further, the node is set to least size as condition.
- Finally random forest regression with *b regression* is developed. Outcome is computed with residual mean square, predicted by out-of-bag data (OOB). Equation for predictions is as follows

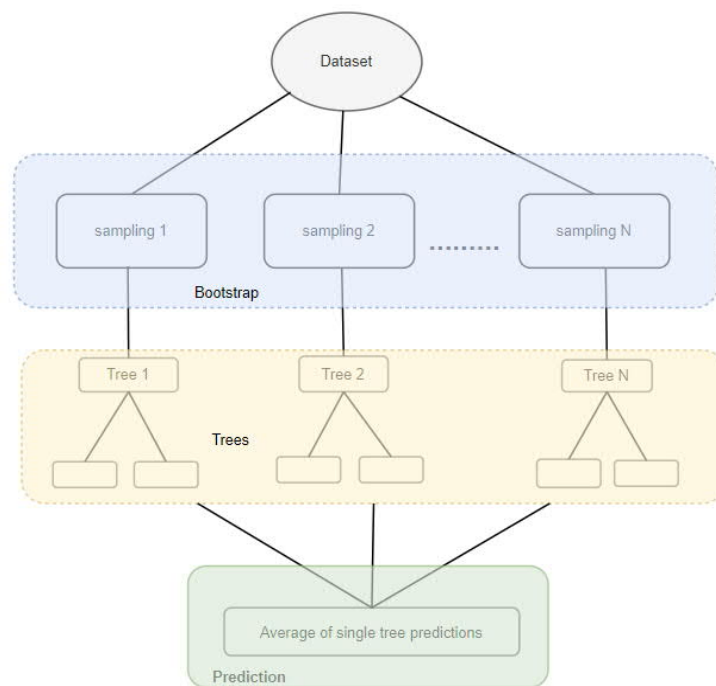
$$\hat{y}_i^{OOB} = \frac{1}{n} \sum_{i=1}^n y_i \quad (27)$$

$$MSE_{OOB} = \frac{1}{n} \sum_{i=1}^n \{y_i - \hat{y}_i^{OOB}\}^2 \quad (28)$$

The outcome can be calculated by averaging the outputs of all trees and the prediction accuracy assessed for by MSE.

$y_i$  is the actual value in OOB

$\hat{y}_i$  is the predicted value in OOB. (Zhang et al. 2015.)



**Figure 18.** High-Level Diagram of Random Forest Regressor Algorithm.

#### 4.2.4. Extremely Randomized Trees Regressor (ETR)

Extremely Randomized Trees (Extra Trees) Regressor share similar properties with random forest and comes under averaging methods which is a part of ensemble methods. The other part under the ensemble method is boosting method. From scikit learn "The goal of ensemble methods is to combine the predictions of several base estimators built with a given learning algorithm in order to improve generalizability / robustness over a single estimator". Multiple estimators are built independently and then predictions' average is taken in an averaging method. Because of reduced variance, combined estimator is better than individual single base estimator. In boosting method base estimators are built sequentially and concentrates on lowering the bias of combined estimator. Here the main aim is to create a powerful ensemble by combining several weak models. (Scikit Learn 2019).

Extra Trees algorithm develops an ensemble of unpruned randomized decision trees. In comparison with random forest algorithm, the main differences are that the split value is chosen fully at random and instead of bootstrap, entire set of training samples are used in growing one tree. Main advantages of this algorithm are due to parallelization - it is more efficient computational wise, high dimensional feature vectors can be easily handled and due to randomization there will be reduction in variance. Complexity of training reduces as nodes are split at random. Only limitation is that convergence speed will be slower, but performance can be comparable or better than random forest when compared. (Pinto et al. 2015 & Xingfang, 2012.)

### 4.3. Applying Machine Learning in Wi-Fi based IPS

#### 4.3.1. Model Motivation

Machine learning algorithms are selected depending upon the data available and the type of output needed. Two main problems are regression in which output will be numerical and the classification has binary output. So, it is necessary to have enough knowledge on

the data that will be fed to the algorithms. Data used in thesis is a crowdsourced Wi-Fi database of the 4-floor building in Tampere University, Tampere, Finland. (<https://zenodo.org/record/1001662#.XsQxaWgzY2z>). (Lohan et al. 2017.)

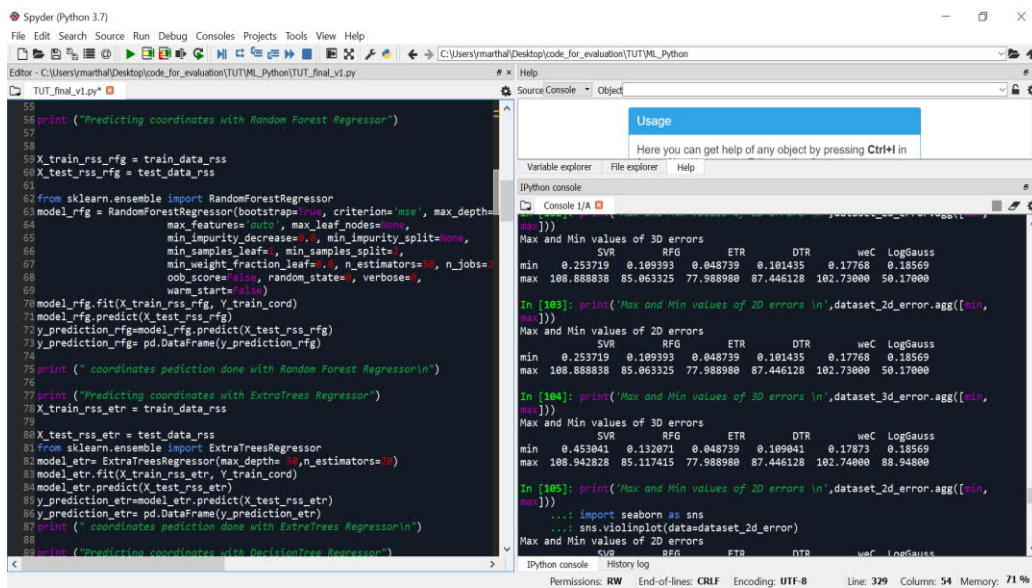
The data consists of Wi-Fi fingerprinting observations, more precisely RSS values in dB, Mac Addresses of base stations and coordinates. The main objective is predicting the user's location from the given RSS values from Wi-Fi base stations. User's location is defined with coordinates. In brief, coordinates are predicted from the measured RSS values which constitutes clearly a regression problem. The model is being trained along with three coordinates XYZ of the user location with their respective RSS values and it is important to predict those three coordinates at once. Obtaining this three coordinates at a time is the main motivation in choosing the Extra Tree regressor, Decision tree Regressor (DTR), Support vector regressor (SVR) and Random forest regressor to be the ones to be compared. "Multioutput regression support can be added to any regressor". (Scikit Learn 2019.) In case of the SVR and the DTR, three coordinates are predicted at a time by adding multioutput regression to them.

## 5. EXPERIMENT AND RESULTS

This chapter presents the experiment setup from Tampere University's open source Wi-Fi RSS database for fingerprinting positioning, the programming environment implemented, the libraries behavior with the machine learning models and finally the results of applying machine learning to a Wi-Fi IPS. The intention is to give a high level of understanding on how the data is fed to the algorithms and an overall performance comparison.

### 5.1. Introduction to the Experiment Setup

The machine learning part of the experiment is carried out in Anaconda environment (Anaconda Inc, Texas, USA) (<https://www.anaconda.com/>) and algorithms are developed with the Spyder IDE (integrated development environment). Spyder is a scientific Python development environment (<https://www.spyder-ide.org/>). The programming language used is Python with version 3.7.3. Along with that numpy, pandas, matplotlib, seaborn and scikit libraries are used for supporting various purposes.



```

55
56 print ("Predicting coordinates with Random Forest Regressor")
57
58
59 X_train_rss_rfg = train_data_rss
60 X_test_rss_rfg = test_data_rss
61
62 from sklearn.ensemble import RandomForestRegressor
63 model_rfg = RandomForestRegressor(bootstrap=True, criterion='mse', max_depth=
64     max_features='auto', max_leaf_nodes=None,
65     min_impurity_decrease=0.0, min_impurity_split=None,
66     min_samples_leaf=1, min_samples_split=10,
67     min_weight_fraction_leaf=0.0, n_estimators=10, n_jobs=
68     oob_score=True, random_state=None, verbose=0,
69     warm_start=False)
70 model_rfg.fit(X_train_rss_rfg, Y_train_cord)
71 model_rfg.predict(X_test_rss_rfg)
72 y_prediction_rfg=model_rfg.predict(X_test_rss_rfg)
73 y_prediction_rfg= pd.DataFrame(y_prediction_rfg)
74
75 print (" coordinates prediction done with Random Forest Regressor\n")
76
77 print ("Predicting coordinates with ExtraTrees Regressor")
78 X_train_rss_etr = train_data_rss
79
80 X_test_rss_etr = test_data_rss
81 from sklearn.ensemble import ExtraTreesRegressor
82 model_etr= ExtraTreesRegressor(max_depth= 10, n_estimators=10)
83 model_etr.fit(X_train_rss_etr, Y_train_cord)
84 model_etr.predict(X_test_rss_etr)
85 y_prediction_etr=model_etr.predict(X_test_rss_etr)
86 y_prediction_etr= pd.DataFrame(y_prediction_etr)
87 print (" coordinates prediction done with ExtraTrees Regressor\n")
88
89 print ("Predicting coordinates with DecisionTree Regressor")

```

```

Usage
-----
Here you can get help of any object by pressing Ctrl+H in
-----
Variable explorer  File explorer  Help
Python console
Console I/A
Max and Min values of 3D errors
SVR    RFG    ETR    DTR    weC    LogGauss
min    0.253719  0.109393  0.048739  0.101435  0.17768  0.18569
max    108.888838  85.063325  77.988980  87.446128  102.73000  50.17000
In [103]: print('Max and Min values of 2D errors \n',dataset_2d_error.agg(['min',
Max and Min values of 2D errors
SVR    RFG    ETR    DTR    weC    LogGauss
min    0.253719  0.109393  0.048739  0.101435  0.17768  0.18569
max    108.888838  85.063325  77.988980  87.446128  102.73000  50.17000
In [104]: print('Max and Min values of 3D errors \n',dataset_3d_error.agg(['min',
Max and Min values of 3D errors
SVR    RFG    ETR    DTR    weC    LogGauss
min    0.453041  0.132071  0.048739  0.109041  0.17873  0.18569
max    108.942828  85.117415  77.988980  87.446128  102.74000  88.94800
In [105]: print('Max and Min values of 2D errors \n',dataset_2d_error.agg(['min',
Max and Min values of 2D errors
SVR    RFG    ETR    DTR    weC    LogGauss

```

Figure 19. Showing the Spyder IDE for the IPS.

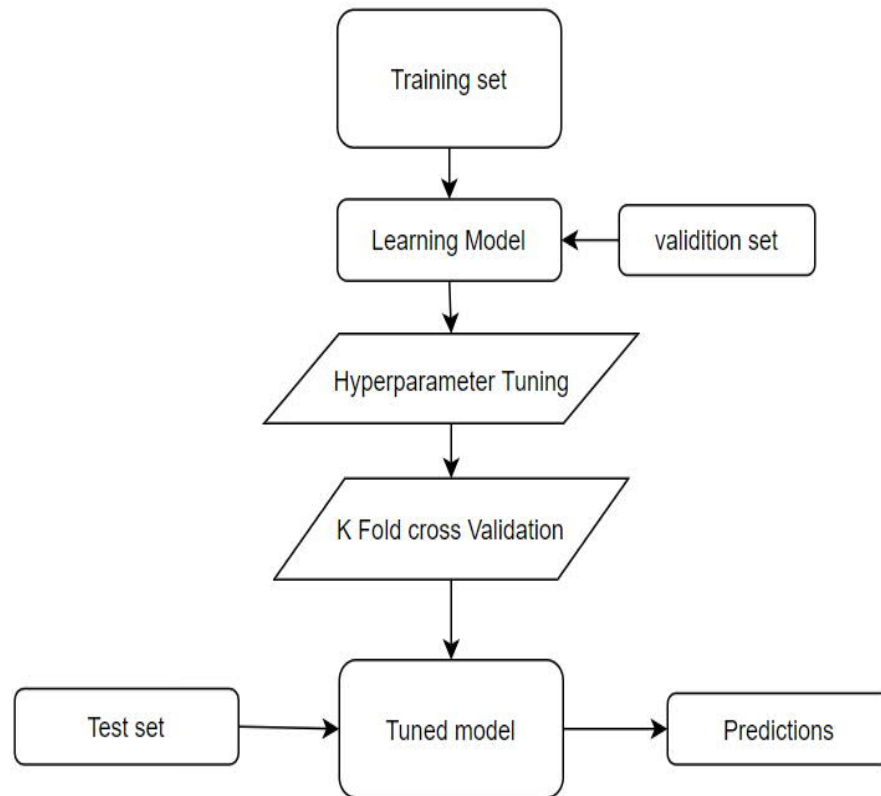
The open source data is divided into two sets, training and test sets, as presented in the data source description in Zenodo(<https://zenodo.org/record/1161525#.XsSJ4mgzBIU>) by the Tampere University (formerly Tampere University of Technology) research team. This radio map consists of 446 reference points and 489 access points. Here the aim is to predict the user's coordinates from the measured RSS values given in the data set and compare with the true reference coordinates. Features are RSS values and labels are coordinates. By feeding features and labels to the model along with required parameters, the system starts to learn about the data. Cross validation is done with the reference tracks to check how well the model is able to predict. In general, 80% of data is used for training and 20% is used for testing. About 10-20% of training data is used for cross validation. In this experiment, K-fold cross validation is carried out and hyper parameter tuning is done with Grid Search and Randomized Search. Once the model is fine-tuned, then test data is fed to the train model to get predictions.

## 5.2. Implementation of the Machine Learning Models

Here a high-level explanation of the machine learning model is given step by step

- Dataset is spilt in to training and test sets with ratio of 80/20.
- Assuming the training set,  $X_{train\ RSS}$  and  $Y_{train\ coordinates}$ . Similarly test set  $X_{test\ RSS}$  and  $Y_{test\ coordinates}$ .
- 20% of training set is kept aside as validation set.
- $X_{train\ RSS}$  and  $Y_{train\ coordinates}$  are fed to the algorithm with parameters, then models start to learn.
- Validation set is used to check the model's accuracy
- Mostly hyper parameter tuning will be required for better accuracy and this is carried out with either grid search or randomized search.
- After tuning, new parameters are obtained and model is updated with those parameters and proceeds to K-fold cross validation.
- After tuning models show better accuracy than before
- Now,  $X_{test\ RSS}$  are fed to the model to predict  $y_{predicted\ coordinates}$

- The close values of  $Y_{test\ coordinates}$  and  $y_{predicted\ coordinates}$  are the best models is predicting



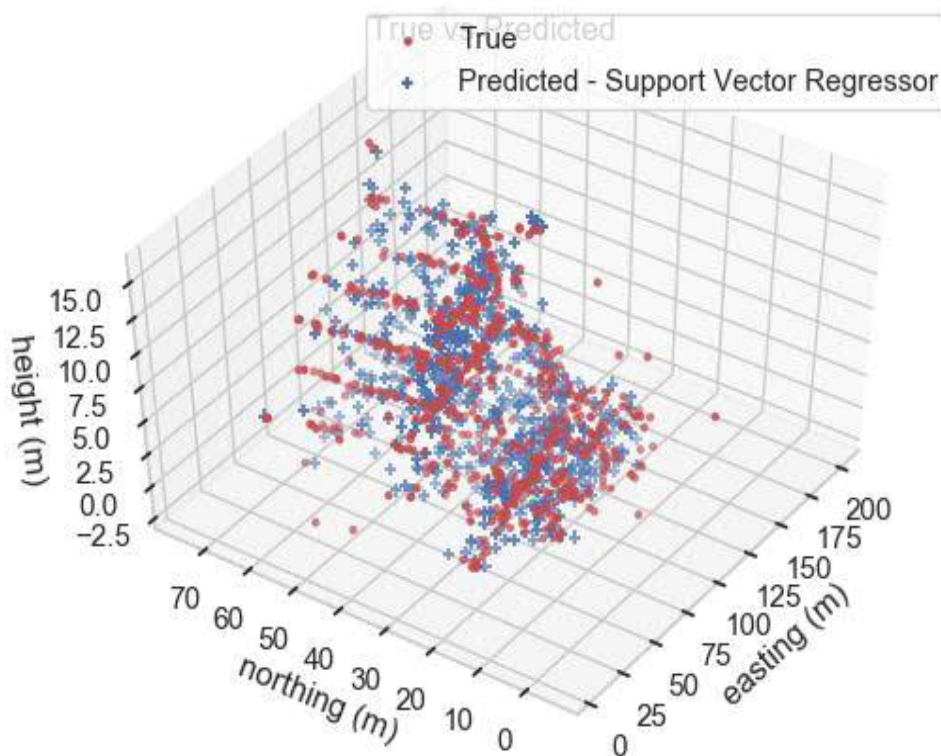
**Figure 20.** High-level explanation of the implemented ML models.

Six different algorithms are being compared in this performance comparison with the dataset: RandomForest (RFR), DecisionTree (DTR), Support Vector Machine (SVR), EXTRA Trees (ETR), Weighted Centroid – (WeC) and a Log Gaussian – (LogGauss). The WeC and the LogGauss are benchmarking implementations from Tampere University accompanying the open source Wi-Fi RSS dataset with reference coordinates. These are being compared with the four ML implementations (RFR, DTR, SVR, and ETR). Explaining SVR from chapter 4.2.1, rbf kernel is used in the algorithm and parameters are  $C=100$ ,  $cache\_size=200$ ,  $coef=0.1$ ,  $degree=1$ ,  $epsilon=0.5$ ,  $gamma='auto'$ ,  $kernel='rbf'$ ,  $max\_iter=-1$ ,  $shrinking=True$  and  $tol=0.001$ .

### 5.3. Results

In this section, overall results are presented and analyzed. Results are displayed in the form of a 3D plot of true vs predicted coordinates for the algorithms being compared, a tabular column stating 2D and 3D error and violin plots to visualize the error distributions obtained.

In the following 3D plot, true vs predicted values are plotted for the support vector regressor algorithm in a metric local reference frame. Multioutput regression is added to this SVR to predict all three XYZ coordinates at a time. This plot gives a clear picture about the distribution of the predicted coordinates when compared with true coordinates that were provided with the fingerprinting dataset.



**Figure 21.** A 3D Plot Showing True vs Predicted Coordinates of the SVR in a Metric Local Reference Frame.

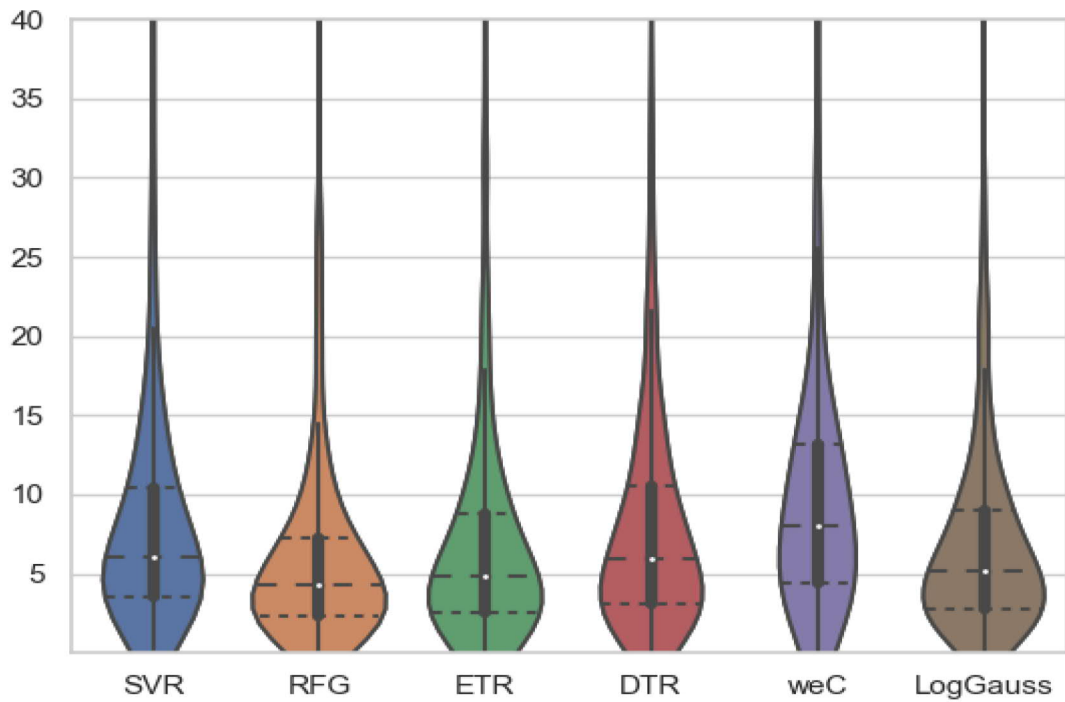
The table below presents a result comparison for all the six algorithms being compared. From the tabular column below it is clear that the RFR has the least errors when compared with the remaining algorithms and its number of outliers are in a single digit. Outliers are considered end result values which are above 40 meters in error. The weighted centroid has the highest amount of errors both in 3D and 2D, and has 21 outliers. Support vector and decision tree implementations are almost equal in comparison.

**Table 1.** Tabular column showing performance comparisons between the different algorithms implemented.

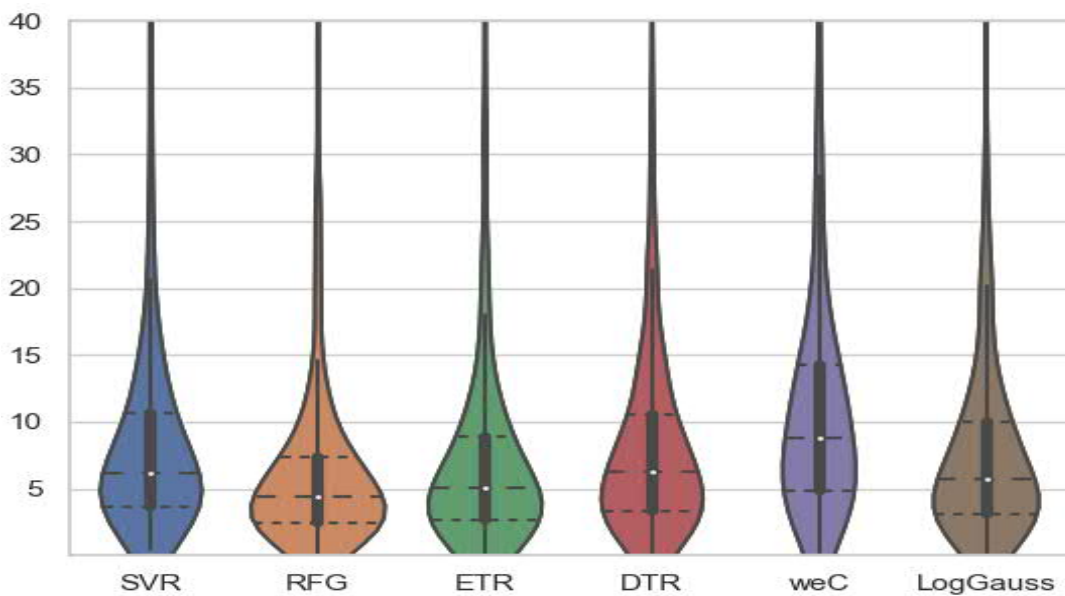
Algorithm	Mean Positioning Error		Median Positioning Error		Min error 3D	Max error 3D	Min error 2D	Max error 2D	No.of outliers (>40 m)
	2D error in meters	3D error in meters	2D error in meters	3D error in meters					
RandomForest (RFR)	6.306	6.422	4.299	4.452	0.132	85.117	0.109	85.063	4
DecisionTree (DTR)	8.789	8.983	6.171	6.373	0.109	87.446	0.101	87.446	11
Support Vector Machine (SVR)	8.828	8.955	6.041	6.185	0.453	108.942	0.253	108.888	12
EXTRA Trees (ETR)	7.31	7.421	4.759	4.842	0.048	77.988	0.048	77.988	8
Weighted Centroid – (WeC)	10.974	11.930	8.023	8.776	0.178	102.740	0.177	102.730	21
Log Gaussian – (LogGauss)	7.061	8.174	5.190	5.719	0.185	88.948	0.185	50.170	3

Below, violin plots provide information about the 2D and 3D error distributions in more detail. Here it is limited to 40 meters because the points above that might not have a useful information and left out for better visualization.





**Figure 22.** Violin Plots Showing the 2D Error Distribution & the Median (in meters) of the Six IPS Algorithms Being Compared.



**Figure 23.** Violin Plots Showing the 3D Error Distribution & the Median (in meters) of the Six IPS Algorithms Being Compared.

Key findings from the performance comparison experiment:

- Handling the missing values and sparse data is challenging.
- Good knowledge about data is necessary when selecting the ML algorithm.
- Decision tree and support vectors can only give a single output limiting the prediction to only one type of a coordinate at once. This can be overcome by adding multi-regression to the algorithms, then it is possible to predict all three coordinates at a time.

## 6. CONCLUSIONS AND FUTURE WORK

This chapter presents the overall conclusions about the Wi-Fi based IPS analysis with ML that was discussed in the thesis and possible future works and developments.

### 6.1. Conclusions

This thesis mainly focuses on Wi-Fi based indoor positioning. The basic functionality of Wi-Fi is to provide internet wirelessly but making use of those radio signals and their geospatially distributed strengths to achieve positioning in parallel is an innovative engineering invention. Apart from Wi-Fi, different technologies were discussed through which indoor positioning can be determined. When compared with costs, infrastructure implementation and accuracy, Wi-Fi has typically more advantages. One part of thesis concentrates on developing an indoor positioning Android app as a visitor guidance tool on the HERE SDK for the Technobothnia technology center building on the campus of the University of Vaasa, Vaasa, Finland. Technobothnia is already equipped with Wi-Fi routers/access points. The app utilizes RSS observations from the routers and the SDK provides positioning and routing inside the building as background service. Radio maps are created and saved in the database of HERE. Satisfactory level of positioning is provided by the app, with room level performance enough for visitors to use as guidance. As the radio signals are affected by the surrounding environment, sometimes the positioning solution can occasionally vary and drift. The implemented visitor guidance app is a first draft with plenty of room for upgrading, which is more closely discussed under future work.

Other part of thesis concentrates on machine learning which one of the fastest growing fields is these days. ML techniques for indoor positioning are performed on the open source Wi-Fi radio data from Tampere University (formerly Tampere University of Technology), Tampere, Finland, which has been obtained via 21 different devices and different users and made available for the scientific community. As discussed earlier, the dataset consists of RSS values and their respective reference coordinates in one of their

multi-floor office buildings. Utilizing this dataset and accompanying benchmarking software, positioning algorithms are implemented and compared. Here the idea has been to experiment on the same dataset with various ML algorithms in particular. After preprocessing the dataset, four ML algorithms were implemented to predict the user coordinates and these were compared against the two benchmarking implementations. The interesting part about the ML models is that no radio signal properties are taken into consideration. Violin plots give clear information about error distributions of the results of the six positioning methods being compared. Random Forest, EXTRA Trees and the Log Gaussian implementations provide the best overall performance. It is clear that the more data, the better the results are. Results could be further improved by feeding relative data additionally into the process, for example, a magnetometer that gives information about direction.

## 6.2. Future Work

From the conclusions, it is clear that there is a need for improving the implemented Android app for Technobothnia visitor guidance, which in this thesis has been a first demonstration of a functioning Wi-Fi based IPS for a venue with existing radio infrastructure. Additionally, this research on ML implementations for an IPS can be taken further by analyzing also other radio signals such as UWB or making use of cellular networks. Fusing various radio signals with satellite navigation signals is one additional way for further research.

There is lots of advancement in machine learning on the horizon. Models and libraries are starting to have so called lite versions, which can be implemented in the mobiles and do not consume excessive amounts of power. A mobile app implementation that integrates an ML algorithm utilizing HERE SDK's output and additional information followed by implementation of a user interface is part of future work for the Technobothnia indoor positioning functionality.

## REFERENCES

- Acharya, Mohan S, Asfia Armaan & Aneeta S Antony (2019). *A Comparison of Regression Models for Prediction of Graduate Admissions* [Online]. International Conference on Computational Intelligence in Data Science [Cited 28 Dec. 2018]. Available from Internet: <URL: <https://ieeexplore-ieee-org.proxy.uwasa.fi/stamp/stamp.jsp?tp=&arnumber=8862140>>
- Ahmed, Dina Bousdar, Estefania Munoz Diaz, Jose Angel & Conejo Minguez (2017). *Exploiting wearable devices for the calibration of inertial navigation systems* [Online]. International Conference on Indoor Positioning and Indoor Navigation (IPIN) [Cited 28 Dec. 2018]. Available from Internet: <URL: <https://ieeexplore-ieee-org.proxy.uwasa.fi/stamp/stamp.jsp?tp=&arnumber=8115931>>
- Batistić, Luca, Mladen Tomić (2018). *Overview of Indoor Positioning System Technologies* [Online]. International Conference on Indoor Positioning and Indoor Navigation (IPIN) [Cited 28 Dec. 2018]. Available from Internet <URL: <https://ieeexplore-ieee-org.proxy.uwasa.fi/stamp/stamp.jsp?tp=&arnumber=8400090>>
- Cao, Limeng & Ruiqing Kang (2017). *Smartphone Indoor Positioning System Based on Geomagnetic Field* [Online]. *Proceedings of International Joint Conference on Neural Networks*, [Cited 28 Dec. 2018]. Available from Internet: <URL: <https://ieeexplore-ieee-org.proxy.uwasa.fi/stamp/stamp.jsp?tp=&arnumber=8243065>>
- Dabove, Paolo, Vincenzo Di Pietra, Marco Piras, Ansar Abdul Jabbar & Syed Ali Kazim (2018). *Indoor positioning using Ultra-Wide Band (UWB) technologies: positioning accuracies and sensors' performances* [Online]. [Cited 28 Dec. 2018]. Available from Internet: <URL: <https://ieeexplore-ieee-org.proxy.uwasa.fi/stamp/stamp.jsp?tp=&arnumber=8373379>>

- Doiphode, Siddhesh R., Dr. J. W. Bakal & Prof. Madhuri Gedam (2016), *A Hybrid Indoor Positioning System based on Wi-Fi Hotspot and Wi-Fi fixed nodes* [Online]. 2nd IEEE International Conference on Engineering and Technology (ICETECH) [Cited 28 Dec. 2018]. Available from Internet: <URL: <https://ieeexplore-ieee.org.proxy.uwasa.fi/stamp/stamp.jsp?tp=&arnumber=7569191>>
- Engström, Hampus & Fredrik Helander (2015). *Evaluation and testing of techniques for indoor positioning* [Online]. [Cited 28 Dec. 2018]. Available from Internet: <URL: <https://www.eit.lth.se/sprapport.php?uid=901>>
- Fan, Hongbo, Guanglin He, Siqian Tao & Hongyu Xu (2013). *Weighted Centroid Localization Algorithm Based on Improved RSSI Ranging* [Online]. International Conference on Mechatronic Sciences, Electric Engineering and Computer (MEC) [Cited 28 Dec. 2018] Available from Internet: <URL: <https://ieeexplore-ieee.org.proxy.uwasa.fi/stamp/stamp.jsp?tp=&arnumber=6885124>>
- Gan, Xingli, Baoguo Yu, Lu Huang & Yaning Li (2017). *Deep Learning for Weights Training and Indoor Positioning Using Multi-sensor Fingerprint* [Online]. International Conference on Indoor Positioning and Indoor Navigation (IPIN) [Cited 28 Dec. 2018]. Available from Internet: <URL: <https://ieeexplore-ieee.org.proxy.uwasa.fi/stamp/stamp.jsp?tp=&arnumber=8115923>>
- Han, Boxiong & Long Zhao (2017). *An indoor positioning algorithm based on Wi-Fi fingerprint and inertial navigation system* [Online]. Proceedings of the 36th Chinese Control Conference [Cited 28 Dec. 2018]. Available from Internet: <URL: <https://ieeexplore-ieee.org.proxy.uwasa.fi/stamp/stamp.jsp?tp=&arnumber=8400090>>
- Huttunen, Heikki (2013). *Statistical Signal Processing* [Online]. [Cited 28 Dec. 2018]. Available from Internet: <URL: <http://www.cs.tut.fi/~hehu/SSP/lecture7.pdf>>

- Jedari, Esrafil, Zheng Wu, Rashid Rashidzadeh & Mehrdad Saif (2015). *Wi-Fi Based Indoor Location Positioning Employing Random Forest Classifier* [Online]. International Conference on Indoor Positioning and Indoor Navigation (IPIN) [Cited 28 Dec. 2018]. Available from Internet: <URL: <https://ieeexplore-ieee-org.proxy.uwasa.fi/stamp/stamp.jsp?tp=&arnumber=7346754>>
- Jeon, Je-sung, Youngsun Kong, Yunyoung Nam & Kangbin Yim (2015). *An Indoor Positioning System using Bluetooth RSSI with an Accelerometer and a Barometer on a Smartphone* [Online]. 10th International Conference on Broadband and Wireless Computing, Communication and Applications [Cited 28 Dec. 2018]. Available from Internet: <URL: <https://ieeexplore-ieee-org.proxy.uwasa.fi/stamp/stamp.jsp?tp=&arnumber=7424881>>
- Ji, Myungin, Jooyoung Kim, Juil Jeon, Youngsu Cho (2015). *Analysis of Positioning Accuracy corresponding to the number of BLE beacons in Indoor Positioning System* [Online]. [Cited 28 Dec. 2018]. Available from Internet: <URL: <https://ieeexplore-ieee-org.proxy.uwasa.fi/stamp/stamp.jsp?tp=&arnumber=7224764>>
- Jianyong, Zhu, Chen zili, Luo haiyong & Li zhaohui (2014). *RSSI Based Bluetooth Low Energy Indoor Positioning* [Online]. International Conference on Indoor Positioning and Indoor Navigation [Cited 28 Dec. 2018]. Available from Internet: <URL: <https://ieeexplore-ieee-org.proxy.uwasa.fi/stamp/stamp.jsp?tp=&arnumber=7275525>>
- Kang, Wonho & Youngnam Han (2015). *SmartPDR: Smartphone-Based Pedestrian Dead Reckoning for Indoor Localization* [Online]. [Cited 28 Dec. 2018]. Available from Internet: <URL: <https://ieeexplore-ieee-org.proxy.uwasa.fi/stamp/stamp.jsp?tp=&arnumber=6987239>>
- Karppinen, Annukka (2018). *Indoor Positioning: Technologies and Use Cases in Retail Context* [Online]. [Cited 28 Dec. 2018]. Available from Internet: <URL:

[https://aalto.doc.aalto.fi/bitstream/handle/123456789/35535/master\\_Karppinen\\_Annukka\\_2018.pdf?sequence=2&isAllowed=y](https://aalto.doc.aalto.fi/bitstream/handle/123456789/35535/master_Karppinen_Annukka_2018.pdf?sequence=2&isAllowed=y)

Kikuchi, Shohei, Shinya Matsuyama, Akira Sano & Hiroyuki Tsuji (2006). *Subspace-based Indoor Mobile Positioning Using Support Vector Regressor* [Online]. [Cited 28 Dec. 2018]. Available from Internet: <URL:<https://ieeexplore-ieee.org.proxy.uwasa.fi/stamp/stamp.jsp?tp=&arnumber=4109818>>

Li, Binghao, Thomas Gallagher, Andrew G Dempster & Chris Rizos (2012). *How feasible is the use of magnetic field alone for indoor positioning?* [Online]. International Conference on Indoor Positioning and Indoor Navigation (IPIN) [Cited 28 Dec. 2018]. Available from Internet: <URL: <https://ieeexplore-ieee.org.proxy.uwasa.fi/stamp/stamp.jsp?tp=&arnumber=6418880>>

Li, Xiaoji, Yanping Cao & Chen Chen (2018). *Machine Learning Based High Accuracy Indoor visible Light Localization Algorithm* [Online]. IEEE International Conference on Smart Internet of Things [Cited 28 Dec. 2018].

Liu, Fang-Tsung, Chiung-Hsing Chen, Yi-Chun Kao, Chih-Ming Hong & Chia-Ying Yang (2017). *Improved ZigBee Module Based on Fuzzy Model for Indoor Positioning System* [Online]. Proceedings of the 2017 IEEE International Conference on Applied System Innovation IEEE-ICASI 2017 [Cited 28 Dec. 2018]. Available from Internet: <URL: <https://ieeexplore-ieee.org.proxy.uwasa.fi/stamp/stamp.jsp?tp=&arnumber=7988150>>

Lohan, Elena Simona, Joaquin Torres-Sospedra, Philipp Richter, Helena Leppäkoski, Joaquin Huerta & Andrei Cramariuc (2017). *Crowdsourced WiFi-fingerprinting database and benchmark software for indoor positioning* [Online]. Zenodo repository [Cited 28 Dec. 2018]. Available from Internet: <URL: <https://zenodo.org/record/1001662#.XsaqQGgzY2z>>



- Lukito, Yuan & Antonius Rachmat Chrismanto (2017). *Recurrent Neural Networks Model for WiFi-based Indoor Positioning System* [Online]. International Conference on Smart Cities, Automation & Intelligent Computing Systems [Cited 28 Dec. 2018]. Available from Internet: <URL: <https://ieeexplore-ieee-org.proxy.uwasa.fi/stamp/stamp.jsp?tp=&arnumber=8267833>>
- Nagatani, Takashi & Shigeo Abe (2007). *Backward Variable Selection of Support Vector Regressors by Block Deletion* [Online]. Proceedings of International Joint Conference on Neural Networks [Cited 28 Dec. 2018]. Available from Internet: <URL: <https://ieeexplore-ieee-org.proxy.uwasa.fi/stamp/stamp.jsp?tp=&arnumber=4371285>>
- Pinto, Adriano, Sergio Pereira, Higino Correia, J. Oliveira, Deolinda M. L. D. Rasteiro & Carlos A. Silva (2015). *Brain Tumour Segmentation based on Extremely Randomized Forest with High-Level Features* [Online]. [Cited 28 Dec. 2018]. Available from Internet: <URL: <https://ieeexplore-ieee-org.proxy.uwasa.fi/stamp/stamp.jsp?tp=&arnumber=7319032>>
- Puricer, Pavel, & Pavel Kovar (2007). *Technical Limitations of GNSS Receivers in Indoor Positioning* [Online]. [Cited 28 Dec. 2018]. Available from Internet: <URL: <https://ieeexplore-ieee-org.proxy.uwasa.fi/stamp/stamp.jsp?tp=&arnumber=4234236>>
- Rantanen, Jesperi, Maija Mäkelä, Laura Ruotsalainen & Martti Kirkko-Jaakkola (2018). *Motion Context Adaptive Fusion of Inertial and Visual Pedestrian Navigation* [Online]. International Conference on Indoor Positioning and Indoor Navigation (IPIN). [Cited 28 Dec. 2018]. Available from Internet: <URL: <https://ieeexplore-ieee-org.proxy.uwasa.fi/stamp/stamp.jsp?tp=&arnumber=8533872>>
- Rashid, Md Tahmid, Md Shadman Sakib Chowdhury & Mehbas Fairuz Nawal (2016). *Composite Indoor Localization and Positioning System Based on Wi-Fi Repeater Technology* [Online]. BRAC University Dhaka, Bangladesh [Cited 28 Dec. 2018].

Available from Internet: <URL: <https://ieeexplore-ieee-org.proxy.uwasa.fi/stamp/stamp.jsp?tp=&arnumber=7848402>>

Rida, Mohamed Er, Fuqiang Liu, Yassine Jadi, Amgad Ali Abdullah Algawhari & Ahmed Askourih (2015). *Indoor location position based on Bluetooth Signal Strength* [Online]. Tongji University Shanghai, China, 2nd International Conference on Information Science and Control Engineering [Cited 28 Dec. 2018]. Available from Internet: <URL: <https://ieeexplore-ieee-org.proxy.uwasa.fi/stamp/stamp.jsp?tp=&arnumber=7120717>>

Ruotsalainen, L. (2013). Vision-Aided Pedestrian Navigation for Challenging GNSS Environments. Suomengeodeettisen laitoksen julkaisuja - Publications of the Finnish Geodetic Institute; 151. Suomen geodeettinen laitos. [Online] [Cited 28 Dec. 2018]. Available from Internet: <URL: [https://tutcris.tut.fi/portal/en/publications/visionaided-pedestrian-navigation-for-challenging-gnss-environments\(6aac7979-7679-465d-9949-6ba6cc8ef79c\).html](https://tutcris.tut.fi/portal/en/publications/visionaided-pedestrian-navigation-for-challenging-gnss-environments(6aac7979-7679-465d-9949-6ba6cc8ef79c).html)>

Rusli, Mohd Ezanee, Mohammad Ali, Norziana Jamil & Marina Md Din (2016). *An Improved Indoor Positioning Algorithm Based on RSSI-Trilateration technique for Internet of Things (IOT)* [Online]. Universiti Tenaga Nasional, Malaysia. International Conference on Computer & Communication Engineering [Cited 28 Dec. 2018]. Available from Internet: <URL: <https://ieeexplore-ieee-org.proxy.uwasa.fi/stamp/stamp.jsp?tp=&arnumber=7808286>>

Salamah, Ahmed H., Mohamed Tamazin, Maha A. Sharkas & Mohamed Khedr (2016). *An Enhanced WiFi Indoor Localization System Based on Machine Learning* [Online]. International Conference on Indoor Positioning and Indoor Navigation (IPIN) [Cited 28 Dec. 2018]. Available from Internet: <URL: <https://ieeexplore-ieee-org.proxy.uwasa.fi/stamp/stamp.jsp?tp=&arnumber=7743586>>

Sayad, Saed (2020). *Decision Tree – Classification* [Online]. [Cited 28 Dec. 2018]. Available from Internet: <URL: [https://www.saedsayad.com/decision\\_tree.htm](https://www.saedsayad.com/decision_tree.htm)>

Scikit learn (2019). *Sklearn.svm.SVR*[Online]. [Cited 25 Feb. 2018]. Available from Internet: <<https://scikitlearn.org/stable/modules/generated/sklearn.svm.SVR.html>>

Scikit learn (2019). *1.10 Decision Trees* [Online]. [Cited 25 Feb. 2018]. Available from Internet: <<https://scikit-learn.org/stable/modules/tree.html#mathematical-formulation>>

Scikit learn (2019). *1.11.Ensemble Methods* [Online]. [Cited 25 Feb. 2018]. Available from Internet: <<https://scikit-learn.org/stable/modules/ensemble.html#forest>>

Scikit learn (2019). *1.12.Multiclass and Multilabel algorithms* [Online]. [Cited 25 Feb. 2018]. Available from Internet: <<https://scikit-learn.org/stable/modules/multiclass.html#multioutput-regression>>

Scikit learn (2019). *3.2.4.3.2.sklearn.ensemble.RandomForestRegressor* [Online]. [Cited 25 Feb. 2018]. Available from Internet: <<https://scikit-learn.org/stable/modules/generated/sklearn.ensemble.RandomForestRegressor.html>>

Wharton (2017). *Why AI is new electricity?* [Online]. University of Pennsylvania [Cited 25 Feb. 2018]. Available from Internet: <URL: <https://knowledge.wharton.upenn.edu/article/ai-new-electricity/>>

Xingfang, GU, MAO Yaobin & KONG Jianshou (2012). *Ensemble Tracking Based on Randomized Trees* [Online]. Proceedings of the 31st Chinese Control Conference [Cited 28 Dec. 2018]. Available from Internet: <URL: <https://ieeexplore-ieee.org.proxy.uwasa.fi/stamp/stamp.jsp?tp=&arnumber=6390591>>

Yao, Leehter, Yeong-Wei Andy Wu, Lei Yao & Zhe Zheng Liao (2017). *An Integrated IMU and UWB Sensor Based Indoor Positioning System* [Online]. International Conference on Indoor Positioning and Indoor Navigation (IPIN) [Cited 28 Dec. 2018]. Available from Internet: <URL: <https://ieeexplore-ieee.org.proxy.uwasa.fi/stamp/stamp.jsp?tp=&arnumber=8115911>>

- Yao, Wanye & Liqi Ma (2017). *Research and Application of Indoor Positioning Method Based on Fixed Infrared Beacon* [Online]. Proceedings of International Joint Conference on Neural Networks [Cited 28 Dec. 2018]. Available from Internet: <URL: <https://ieeexplore-ieee-org.proxy.uwasa.fi/stamp/stamp.jsp?tp=&arnumber=8482658>>
- Yao, Wanye, & Liqi Ma (2018). *Research and Application of Indoor Positioning Method Based on Fixed Infrared Beacon* [Online]. Proceedings of the 37th Chinese Control Conference [Cited 28 Dec. 2018]. Available from Internet: <URL: <https://ieeexplore-ieee-org.proxy.uwasa.fi/stamp/stamp.jsp?tp=&arnumber=8482658>>
- Zhang, Ling, Pukai Wang, Tianyuan Jiang, Gehua Fan & Caihong Dan (2015). *Prediction of Torpedo Initial Velocity Based on Random Forests Regression* [Online]. International Conference on Intelligent Human-Machine Systems and Cybernetics [Cited 28 Dec. 2018]. Available from Internet: <URL: <https://ieeexplore-ieee-org.proxy.uwasa.fi/stamp/stamp.jsp?tp=&arnumber=7334717>>