Transylvania University

Masters Thesis

Health Monitoring for the Insurance Industry

Author:
Adriana Micu

Supervisor:
Silviu Dumitrescu,
Phd. Associate professor

23rd June 2015

Contents

List of Figures

Chapter 1

Introduction: new technology for your health

1.1 Health: What does it stand for?

The World Health Organisation defines health as ?a state of complete physical, mental, and social well-being and not merely the absence of disease or infirmity.? That is why, when a person says he or she is healthy, it can mean a lot of different things and one person alone cannot determine his or her state of health. This is when the doctors come to help.

Most people go to the doctor if they have an ache of some kind to see what is happening, if there is something wrong with their physical health. But sometimes our health parameters could not be at their optimum, and still not showing any signs in this matter. This is the reason why it is important that a person does regular checkups of his overall health.

But is this enough? In the modern society, sometimes people don?t have time to make the regular checkup, or they forget. And how about the daily routine which helps a person be healthier, like doing some sport every, or almost every day, eating healthy, having enough quality sleep during the night, trying to be less stressed out? These factors, if not taken into consideration, can affect our health on a longer period of time and we don?t notice it until it gets worse. Nowadays we don?t have time to always think about these elements which are actually such small things that can help us improve or maintain our health parameters at optimum levels.

1.2 To help a persons health: What does each of us do?

The duty of taking care of our health is in the hands of each of us and there are also the health insurance companies which help us preventing and treating the health problems of which we cannot take care ourselves.

Everyone has a health insurance at a statutory or at a private insurance company. This health insurance is an extremely important part of a healthy life because it is meant to promote wellness and health also by helping you get the checkups and treatments you need for a minimal cost. Every insurance company has different health plans which may or may not suit one persons needs, depending on your health or the special areas of need.

Each of us is responsible for having a health insurance, but also there are the other things which every person can do, to maintain or improve their health every day, because a doctor is not there 24/7, but we have to be responsible for our health. Everyone can to their best to prevent injuries and help their bodies function at their best.

There may be things in our genetics that are perceived as diseases. This is something that we cannot control, but we have control over how we treat our illness. For example, if a person has genetically inherited a heart disease, he or she can control its evolution by having regular checkups and monitor his or her disease state.

The lifestyle of a person is an element of physical and psychical health which we have the most control over. This includes our diet, our emotional health, our level of physical activity, our sleep habits and our behaviours. For example, tobacco use is a problem for many people, but the health care plan can also help by offering a cessation program. Another good example would be the low level of physical activity when it comes to a lot of people. These people have to be highly motivated so that they do their exercises and keep their bodies healthy like this. If the person alone cannot do this, the insurance company is the one which could help in this matter also. Another interesting and important element is the diet in every persons life, because it has a big impact on the his or her health. Monitoring the food someone eats would be helpful but also not really easy, because the targeted person would always have to write down somewhere everything he or she eats, so this is a harder to control element. An easier to control element would be the sleep habits of people of which we can take care by sleeping enough hour and avoiding exhaustion our body and mind energy completely.

1.3 And there is more: What makes our work easier?

As I said before, it is not always easy nowadays to keep track of the health status of a person. The health insurance company is helping by trying to provide a good suited health care plan for everyone, but that is not enough for someone to always know that he or she is healthy, or that his or her illness is being treated or kept at the lower risk levels possible or its actually getting worse day by day. This is why people can do everything the doctor says and have healthy daily habits.

But how does someone know he or she is always doing the right thing? And how does the health insurance company and the doctor know that the insured really did everything he or she has been told? How can the insured be more motivated to do everything the doctor says? There is a nice way, which would bring benefits as well for the insurance company, as for the insured person.

In the past few months, there have appeared many different types of fitness and health wearables. These devices gather important data about the person who wears them and this data can be used as well for the benefit of the insurance company and for the benefit of the insured.

1.4 New technology: What are the benefits?

The insured person gets many benefits from the use of this health and fitness wearables. Because it is hard to keep track of our health status, although we know we are healthy persons, it is important to analyse some of our vital health signals and know if anything could come up in the near future. Also, if we know we are not completely healthy persons, and we need to monitor our activity or our heart rate levels for example, it is important to see and know the status of these all the time.

It is good to let the insurance companies get the health data from us, so that they can develop the best health care plan for our needs and analyse the evolution of our symptoms professionally. Another benefit of the insured person is that the health insurance company can offer him or her a bonus at the end of a year or a trimester or so, if that person has taken good care of his or her health parameters.

The health insurance company must also have benefits from this system usage. It will know if the insured person is in healthy parameters and that a standard health care plan fits them or a more advanced personalised one is the best option. Also, the health insurance company will have information about different health problems of the insured person that they don't know about and this way they can develop personalised health care plans. Also, a very important advantage for the insurance company is that it will save money, because a healthy person means less costs.

Another important benefit of the health insurance company is that they can ”monitor” the insured person, and know, with some limitations, if the person does what the doctor says, for example, if the doctor says the person has an irregular heart rate which happens because he or she only sleeps 4 - 5 hour per night, and the doctor tells the person to sleep at least 7 hours per night, and that person does this from then on, the insurance company can see if the irregular heart rates have come to a stabilised rate and he or she can be congratulated and encouraged by getting, for example, a bonus for his or her health care plan.

1.5 Internet of Things: How can this help?

For some years, health insurance companies, for example in the U.S., have been encouraging the insured people to live a healthy life by penalising those who made bad health choices. In the past year, this model has evolved to giving bonuses to those insured people who make the right health choices. This model was also adopted in Europe and now it is getting improved by the new fitness and health wearables which appear everywhere.

1.5.1 Internet of Things: How does it actually work?

IoT got born and evolved from the convergence of the Internet, of course, as it name says, micro-electromagnetical systems and wireless technologies. In 1994 Reza Raji described the concept in IEEE Spectrum as ”moving small packets of data to a large set of nodes, so as to integrate and automate everything from home appliances to entire factories”.

The integration with the IoT implies that the device to be connected utilises and IP address as a unique identifier. Due to the limited address space of IPv4, objects in the IoT will have to use IPv6 so that these adapt to the extremely large address space required. The objects connected do not have to be only devices with sensory capabilities, but also devices which provide capabilities to manifest an action. So this means that if all objects and people would have a unique identifier, computers could manage and inventory them.

An example of the application of the IoT in our daily lives are intelligent shopping systems which monitor the purchasing habits of specific users by tracking their mobile phones. These users can then get information on special offers on their favourite products, or even location of items they need, which were transmitted from their fridge to their smartphone.

According to Gartner, Inc. (a technology research and advisory corporation), there will be nearly 26 billion devices on the Internet of Things by 2020. As per a recent survey and study done by Pew Research Internet Project, a large majority of the technology experts and engaged Internet users who responded?83 percent?agreed with the notion that the Internet/Cloud of Things, embedded and wearable computing (and the corresponding dynamic systems [22]) will have widespread and beneficial effects by 2025.

1.5.2 Wearables and Internet of Things: How can this help our health?

Wearables and the Internet of Things (IoT) may give the impression that it is all about the sensors, hardware, communication middleware, network and data but the real value is in insights. These wearable sensors get a lot of different information from the person who wears it, from the steps that person made during a period of time to his heart rate value or his blood pressure values and so on.

But only the creation and evolution of this sensors alone doesn't really bring that much. This is where the Internet steps in to connect everyThing. The sensors can connect to other sensors and to other types of devices and the important data can be gathered and used for different helpful purposes.

All these wearable sensors can gather vital health signs from a person. IoT devices can be used to enable remote health monitoring or, for example, emergency notification systems. These health monitoring sensors can range from a simple step count or activity monitoring to blood pressure or heart rate monitors or even to advanced devices capable of monitoring specialised implants or older people who need permanent tracking of their health. These kind of devices also encourage healthy living by giving the person who uses them permanent information for example about his activity, his sleep quality, his heart rate and so on.

The IoT is revolutionising healthcare in many ways, having a broad range of healthcare applications. It's potential is already being noticed and benefited from.

For example, clinical care is being improved. Hospitalised patients whose physiological status required close attention can be constantly monitored using the IoT-driven monitoring which is also a noninvasive monitoring. This approach replaces the process of having a health professional come by at regular intervals to check the patient's vital signs, because these are already continuously monitored and analysed and sent to the health professional for further analysis and review.

Another healthcare field which is being improved is the remote monitoring. The result is that, for example patients with chronic diseases may be less likely to develop complications, and other complications could be diagnosed earlier than they would be in another way.

Early intervention/prevention is another field which benefits of the use of the IoT. This area is focused mostly on healthy people which can also benefit from IoT-driven monitoring of their daily activities. Elderly care is one example, because having a monitoring device that can detect a fall or other interruptions in the everyday activity and report it to the emergency room is very helpful.

Chapter 2

Aims and Objectives

This paper presents an original approach to implement a new innovative healthcare system. This solution is focused on the early intervention/prevention healthcare area.

The following chapters present the business details and the technical details along with the way of implementing the healthcare solution.

In this chapter I am explaining the objectives and also the arguments which lead to these particular objectives. I am also tackling the problem of the importance of the system for the insurance companies and also for the insured persons, but nevertheless, for the research purposes which evolve with this type of data from the human body.

The next chapter is all about the new technology and the implementation of a solution prototype. I am describing every technology used and why I chose it and of course, how I implemented it to be the best fit for the project.

In the fourth chapter I am documenting the results achieved thanks to the prototype of the system I implemented and also the further results which can be easily reached with it.

The fifth and last chapter is about the conclusion and about the future research and future directions.

2.1 What does the system bring?

The system I implemented is suited to be used with a large variety of wearable sensors. For this prototype, I only use one type of wearable sensor which measures a persons heart rate, because this is what the use cases of the system are focused on in this phase.

The concept of this system brings together the newest and most interest raising domains in the last period of time. These are: the Internet of Thing, the Cloud based, distributed Systems and Analytics, each of them with their specific tools and new technologies. I will describe more about each technology and how it works in the context of this system in the third chapter.

2.2 Who benefits and how do they benefit from this solution?

As described in the introduction, I was saying that there are three areas of benefit from this system.

First of all, the area of the people who want to have a healthy lifestyle, and which are interested in having different information about the status of their health. A person can see the status of different vital signs like heart rate, activity, blood pressure, sleep information and so on. For the first phase of this system, the heart rate data and the activity data can be seen. Also, besides just seeing his data, a person using this system can also have access to a simple web application which shows how the heart rate and the activity data is correlated, and also, the person can see there some valuable insights on his data, which tell him if his heart rate is normal for his age, gender, height, weight and activity type. These are the benefits that the user of this system gets directly from it, not having to communicate his data to an insurance company. The benefits of sharing the data with the insurance company were presented in the introduction. Of course, there are also different advantages a person gets by gathering his data and sharing it for research purposes.

Next, the area of the insurance companies, can have a lot of benefits directly from this system. Besides the ones already mentioned in the introduction of this work, that having their insured people ”monitored” when it comes to their health status so that they can better fit every type of person, the insurance company could also get valuable information from the research that can be done having the data from their insured clients. On the other hand, there is also the probability that the data which the insurance company thinks it comes from the same insured person, to actually represent a fraud. This can easily happen if the insured person decides to send someone else’s data instead of his to the insurance company. The reason for this, could be for example, if the insurance company has a bonus program which gives the insured person a percent of cost reduction from his insurance policy if he sticks to some different lifestyle habits imposed by his doctor and so on. Maybe the person is not that interested in keeping up 100% with what his doctor ordered and he gives his wearables to another family member for example. There can be thousands if not millions of examples of how and why an insured person would want to fraud this type of system.

This is the reason why I chose to implement a way in which fraud can be detected by looking only at someone’s heart rate data for a short period of time. The methods used for this, and the implementation is presented in the third chapter.

Chapter 3

Technologies and Implementation

This Chapter presents the technologies I used to create the project prototype. Every technology is described shortly and then I give arguments about why I chose to use that particular technology and not another. Next, I am explaining how the technologies are implemented.

This system is using the most recent evolved technologies which are still constantly emerging in a rapid way. New technologies are being developed all the time thanks to today's society which is already different to the one that existed actually not so many years ago. our society is constantly changing. One of the most important and noticeable characteristic of our today's world is represented by the transformation of the information technology. As we can observe, we are living in an information society where the new technologies are playing a very big part. Nowadays, we cannot imagine our society and even our lives functioning without the new modern technologies.

Modern technologies have revolutionised a lot of areas from people's communication to electronic shops, process automation or even intelligent machines. The Internet hosts now an enormous information base and with the help of technology, this information has been able to bring a lot of benefits reaching everywhere in the world.

Another very important area which is benefiting from applying the new technologies is the healthcare area. It is proven that the human health has improved in the past years thanks to the lots of information which brings more and more knowledge about the human body. This way, the body's functions can be improved, new tools to help heal it can be created and like this, life lasts longer. This is not the only benefit. Also, people can live more comfortably and they can recover from wounds and diseases that even some years ago have been deadly.

This project will introduce another type of system which helps keeping us healthy and motivated to have a healthy lifestyle.

3.1 The Concept: How does it begin?

Wearable computing is one of the most exciting trends in gadget technology today, and for good reason. Collecting data about our habits, from what we eat to where we go to lunch, will let us learn about our daily patterns and change behaviorus to improve our lives.

When it comes to the insurance industry, wearables are starting to represent a very important area of interest due to the fact that if the insured persons are willing to use these wearables, the insurance company can get a lot of interesting facts from that data and can use to improve the services provided for every insured user and to reward him for healthy lifestyle habits.

Of course this type of approach certainly brings the fraud subject in discussion. As easy as an insured user can send his data to the insurance company, he can also send some other person's data in his name. This would create a real chaos in the data of the insurance company, and also, the insights about that specific data would actually not be relevant anymore.

Therefore, I chose to implement a way such a disaster can be avoided, by using only one type of data from the insured person. This will be explained further in another subchapter.

Next, I drew a picture of the concept for a better overview of how everything works together. Every part will be presented in detail in the following subchapters.

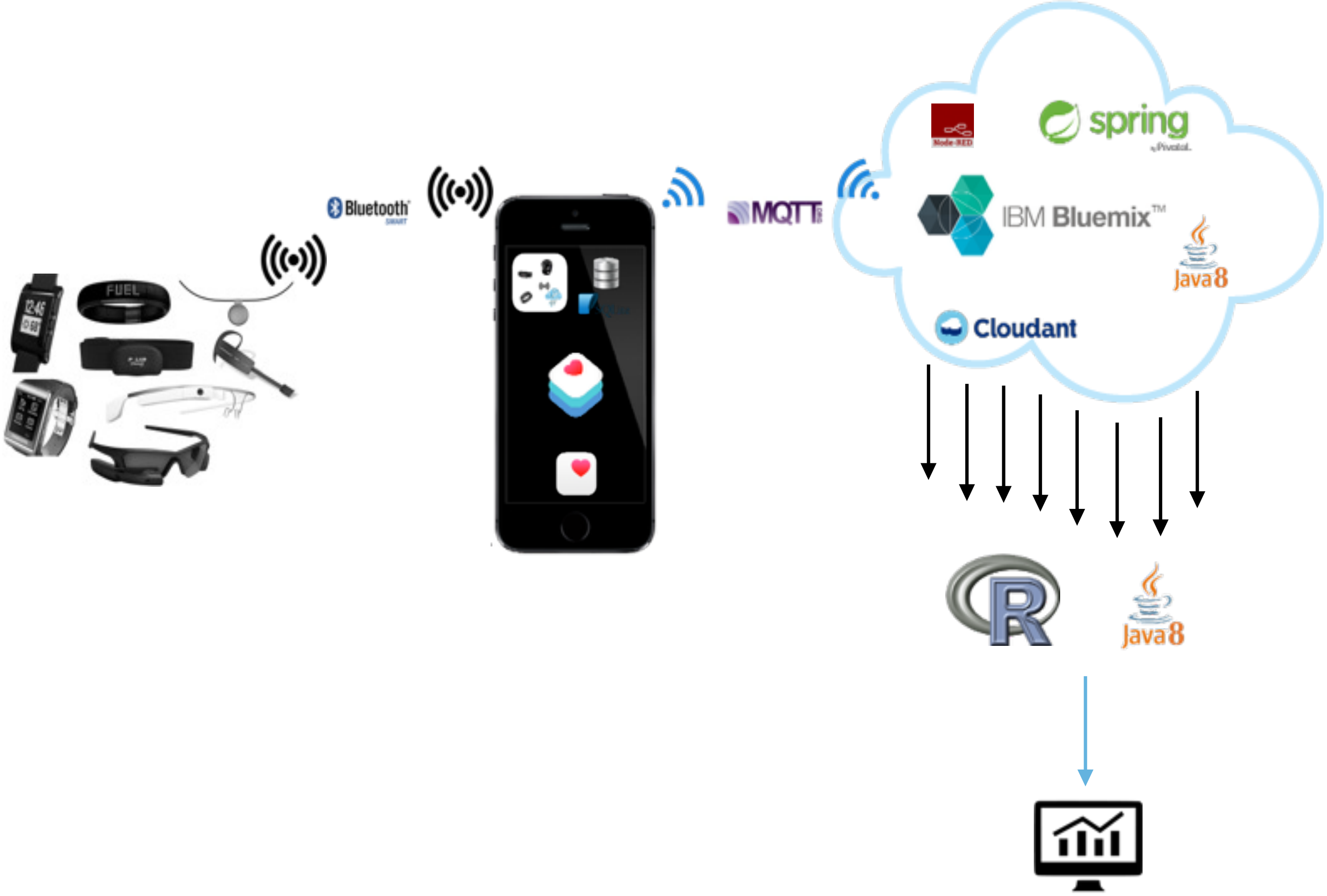


Figure 3.1: The Concept: Overview

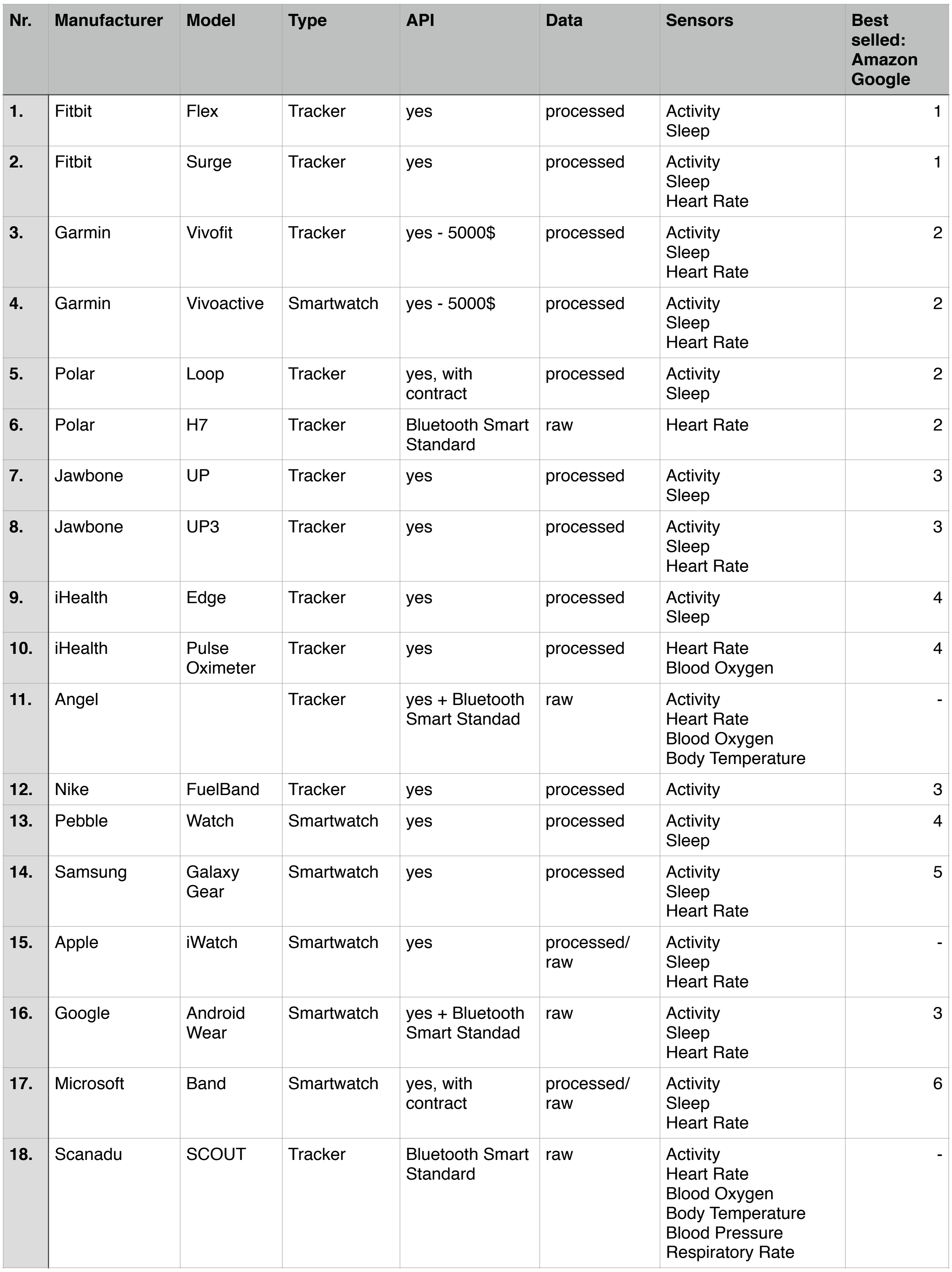
3.2 Sensors: What is the most useful?

As described, the systems takes advantage of the data gathered from lots of different sensors.

The first step in identifying the best suited sensors or smartwatches for the project was to see which are the most used and why are these the most used and what do they bring. I made a list of different sensors and smartwatches and compared them. Next I transposed this list into o table so that I have a better overview of the features and uses of the sensors and the smartwatches I analysed. The table is on the next page.

Analysing the table I put together, I could decide on which sensors would better fit my project. The most important criteria in my decision was, of course to be found in the API column, but also, what type of data the tracker is giving and which sensors it has, were important decision factors. As we can observe, there are some manufacturers of trackers and smartwatches. I took these 13 manufacturers into consideration because of their characteristics like being the best sellers on the market or having the most sensors or giving unprocessed data from the tracker.

In the table, I observed what each manufacturer has brought to the market. For my system, I chose to work with the sensors which use the Bluetooth Smart Standard API, so that the project can make use of different types of devices, not having to implement a different way of retrieving the data from the devices for each one. This is a very important aspect, because having to support a different API for every device could not have been a choice. The Bluetooth Smart Standard is especially created for applications in the healthcare, fitness, security and home entertainment industries. The Bluetooth Special Interest Group predicts that more than 90 percent of Bluetooth-enabled smatrphones will support Bluetooth Smart by 2018. Taking this aspects into consideration I took the decision to use those sensors which use the Bluetooth Smart Standard without any other specific APIs.



Nr.	Manufacturer	Model	Type	API	Data	Sensors	Best selled: Amazon Google
1.	Fitbit	Flex	Tracker	yes	processed	Activity Sleep	1
2.	Fitbit	Surge	Tracker	yes	processed	Activity Sleep Heart Rate	1
3.	Garmin	Vivofit	Tracker	yes - 5000$	processed	Activity Sleep Heart Rate	2
4.	Garmin	Vivoactive	Smartwatch	yes - 5000$	processed	Activity Sleep Heart Rate	2
5.	Polar	Loop	Tracker	yes, with contract	processed	Activity Sleep	2
6.	Polar	H7	Tracker	Bluetooth Smart Standard	raw	Heart Rate	2
7.	Jawbone	UP	Tracker	yes	processed	Activity Sleep	3
8.	Jawbone	UP3	Tracker	yes	processed	Activity Sleep Heart Rate	3
9.	iHealth	Edge	Tracker	yes	processed	Activity Sleep	4
10.	iHealth	Pulse Oximeter	Tracker	yes	processed	Heart Rate Blood Oxygen	4
11.	Angel		Tracker	yes + Bluetooth Smart Standad	raw	Activity Heart Rate Blood Oxygen Body Temperature	-
12.	Nike	FuelBand	Tracker	yes	processed	Activity	3
13.	Pebble	Watch	Smartwatch	yes	processed	Activity Sleep	4
14.	Samsung	Galaxy Gear	Smartwatch	yes	processed	Activity Sleep Heart Rate	5
15.	Apple	iWatch	Smartwatch	yes	processed/ raw	Activity Sleep Heart Rate	-
16.	Google	Android Wear	Smartwatch	yes + Bluetooth Smart Standad	raw	Activity Sleep Heart Rate	3
17.	Microsoft	Band	Smartwatch	yes, with contract	processed/ raw	Activity Sleep Heart Rate	6
18.	Scanadu	SCOUT	Tracker	Bluetooth Smart Standard	raw	Activity Heart Rate Blood Oxygen Body Temperature Blood Pressure Respiratory Rate	-

Figure 3.2: Sensors: The comparison

Another valuable aspect which I took into consideration was the type of data that the device provides. Processed data could lose its accuracy in the phase of processing on which me, as a developer, would not have access to. So the best decision to make, was to look at the devices which provide raw data directly from the sensors so that I could further process it. Having this other fact in mind, I noticed, looking at the information in the table, that raw data mostly comes from the devices which use the Bluetooth Smart Standard. This brought another plus for those devices.

I also took into consideration the sensors every tracker or smartwatch disposes of. The fact that a device has many sensors is good, but it is also good if, for example the device only has one sensor for which it is really specialised for. I looked at the devices which have the Bluetooth Smart Standard and raw data capabilities and noticed that these devices also have a lot of sensors at disposal. Like this, I did not see any reason for which I should not choose these kind of devices.

Looking also at the best sellers on the market until now, I observed that the best sellers were not exactly the sensors I chose, but this market is a new one, and an enormously growing one, so I did not want to make the current selling numbers an important factor in taking my decision upon the chosen devices. Also, the selling market I looked at: Amazon and Google is not a specific one, but the general one, which has not have the main focus on the health and medical market. Judging by the next shown statistics, from Transparency Market Research, the devices I chose to use for my system are very promising.

Taking this statistics into consideration, I could see that there is a noticeable growth in the Health and Medical area, so this helped me sustain the choice I made as the right one for this system.

The wearables I chose to work with for the prototype are: Polar H7, Angel, Android Wear and Scanadu. For the further development of the project, I will also work with the other wearables.

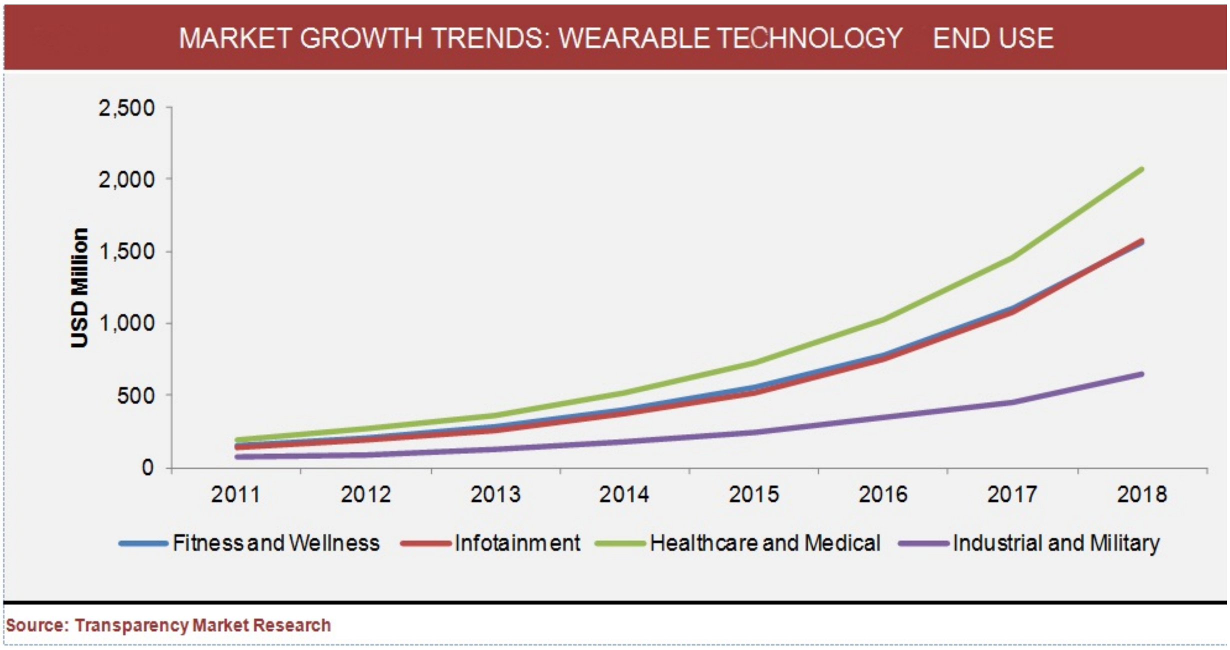


Figure 3.3: Wearbles: The areas

3.3 Bluetooth Low Energy: How do the wearables connect?

First of all, I want to introduce a short definition of the Bluetooth technology: Bluetooth is a wireless technology standard for exchanging data over short distances (using short-wavelength UHF radio waves in the ISM band from 2.4 to 2.485 GHz) from fixed and mobile devices, and building personal area networks (PANs). Invented by telecom vendor Ericsson in 1994, it was originally conceived as a wireless alternative to RS-232 data cables. It can connect several devices, overcoming problems of synchronisation.

Bluetooth is managed by the Bluetooth Special Interest Group (SIG), which has more than 25,000 member companies in the areas of telecommunication, computing, networking, and consumer electronics. [1]

Next, I want to present the Bluetooth Low Energy (BLE) technology [**?** , 3] Bluetooth low energy (Bluetooth LE, BLE, marketed as Bluetooth Smart) is a wireless personal area network technology designed and marketed by the Bluetooth Special Interest Group aimed at novel applications in the healthcare, fitness, beacons, security, and home entertainment industries. Compared to Classic Bluetooth, Bluetooth Smart is intended to provide considerably reduced power consumption and cost while maintaining a similar communication range.

Bluetooth Smart was originally introduced under the name Wibree by Nokia in 2006. It was merged into the main Bluetooth standard in 2010 with the adoption of the Bluetooth Core Specification Version 4.0. [2]

As it can be noticed, Bluetooth and BLE are two different technologies, not just other revisions or something like that. The Classic Bluetooth technology provides a robust wireless connection between devices ranging from headsets and cars to industrial controllers and it is designed for continuous, streaming data applications including voice. So this type of applications would be better suited with the Classic Bluetooth technology. On the other hand, BLE, as it name says, has an extremely low power consumption, unique characteristics and new features enabling new applications that were not practical with Classic Bluetooth. This means that even coin cell battery-operated sensors and actuators in medical, industrial, consumer and fitness applications can now very easily connect to the BLE technology. Nowadays also smartphones are BLE enabled, so it is no problem connecting these to other BLE enabled devices and sensors to gather information from them.

I chose to take advantage of the BLE technology because by analysing its characteristics it is the best fit for the devices I am using in the project.

As I said in the chapter 3.2 about the sensors that I chose, they all use the Bluetooth Low Energy standard specification and like this they can all connect to a smartphone using the same smartphone application implementation.

3.3.1 The Implementation

As the system gathers the data from the sensors of the wearables it somehow has to connect these with another device, which in this case is represented by a smartphone. Following the BLE specifications, there are APIs implemented for iOS smartphones, Android smartphones, Windows smartphones or Blackberry smartphones. In this project, I also used the API which follows the BLE standard specifications so that every sensor which follows this specification will be able to connect to the implemented smartphone application. This is a very important aspect which also represents a great advantage for the implementation of this system.

The BLE enabled wearable device emits data. The BLE enabled smartphone has the application, which will be presented in the next chapters, installed. As the application starts, it requests access to enable Bluetooth visibility, if this is not started. After it is started, the smartphone pairs with the wearable and the application detects this using a specific callback. When this callback is triggered, the type of wearable is identified and if it is a right one, which has the different health or activity sensors enumerated in the chapter before, it will try to read its services and then the characteristics belonging to every discovered service. When a specific characteristic is discovered, the information can be read. This is the way the Bluetooth connection works in a usual use case and in the application I implemented for this system.

To better understand what I presented in the paragraph before, I will start by introducing the concept of central and peripheral devices in Bluetooth LE. The two major roles involved in all Bluetooth Low Energy communication are known as the central and the peripheral. The central has the role if the coordinator. It wants data from a number of its workers in order to accomplish a particular task. A peripheral, on the other side, is the coordinated device, the worker, which gathers and publishes information that is consumed by other devices. In the scenario of this project, a smartphone plays the role of the central and the wearables play the role of the peripherals. The job of the central, the smartphone in this case, is to have the Bluetooth turned on so that it can make its presence known. The central scans for different packets which can contain some data such as the peripheral's name, the peripheral's manufacturer and it can also include some extra information related to what that peripheral connects. For example, the packet can also contain heartbeats per minute data, or steps count data and so on. The job of the central is to scan for this packets and to identify any peripherals it finds relevant and to connect to these wearables. Once the connection takes place, the central needs to chose the data it is interested in. This data is organised into the two concepts I mentioned above: services and characteristics. A service is a collection of data describing a feature of a service. A characteristics describes in more details a service. For example, in the application implemented for this project, one of the services I used was the Heart Rate Service, which has an assigned universally unique identifier (UUID), specified in the

Bluetooth LE documentation. This service contains more characteristics like the Heart Rate Measurement, the Body Sensor Location to the Heart Rate Control Point. Each of this characteristics also have their assigned UUIDs. All of this information is available to the central to manage after the connection to the peripheral was successful.

The following image describes how BLE works in the context of this project.

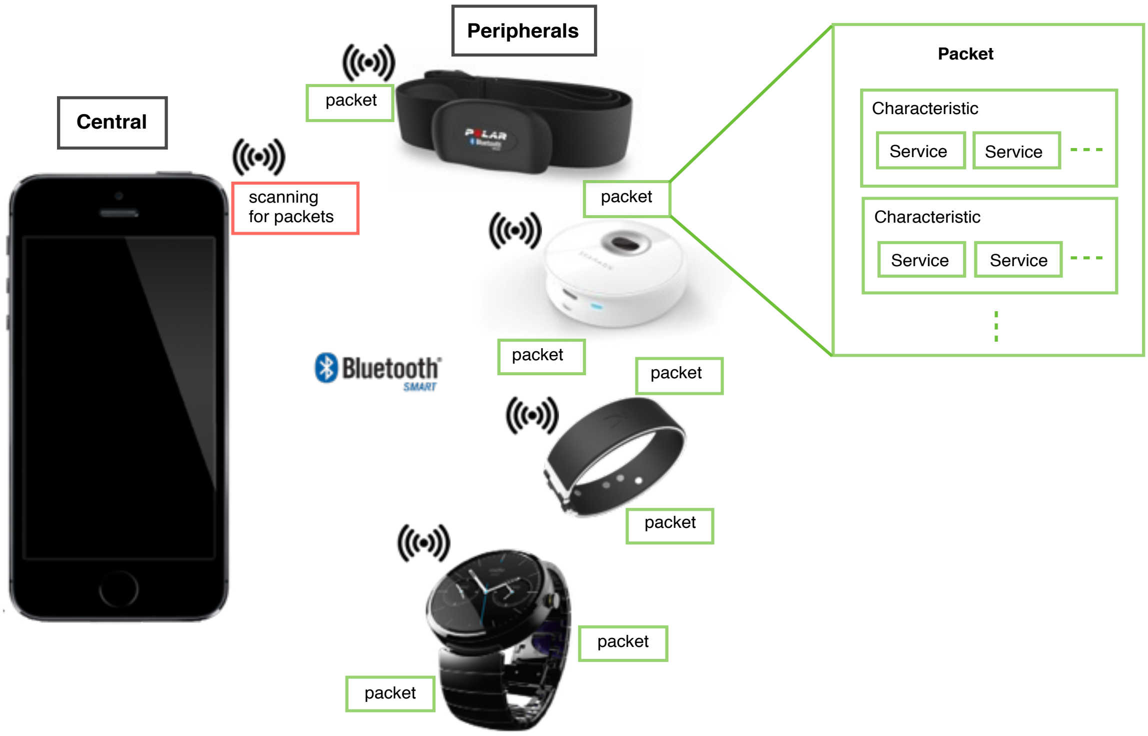


FIGURE 3.4: Bluetooth: The implementation

3.4 MQTT: a new way of communication for IoT applications

MQTT was invented by Dr Andy Stanford-Clark of IBM, and Arlen Nipper of Arcom (now Eurotech), in 1999. After a couple of years it has been made Open Source. It has grown more popular once the mobile devices have taken off and it is in the process of becoming a standard. MQTT is a Client Server publish-subscribe messaging protocol. It has good properties for an IoT application such as: it?s lightweight, simple and easy to implement. It is design to be used on top of the TCP/IP protocol. It is also ideal for mobile applications because it has a low power draw and small network bandwidth consumption.

The messaging pattern - publish-subscribe - requires a message broker. The broker is an adapter that in an intermediary step translates the message from the protocol of the sender to the protocol of the receiver. The messaging broker is responsible for distributing messages to clients based on the topic of the message.

Some of the components as defined in the OASIS standard for the MQTT protocol are: Application message - the message carried by the protocol across a network. It contains specific fields such as Quality of Service or a Topic Name. The QoS flag controls the level of assurance you want when sending a message to the broker. A message with QoS = 0 will be received at most once by subscribed clients. For QoS = 1 a message will be received at least once and for QoS = 2 will be received exactly once by the clients. A topic name is a label that is matched with the subscriptions known by the server. Client - a program or device that uses MQTT. The clients establish the connection with the server. It has the following functions: publishes messages that other clients might be interested, subscribes or unsubscribes to request/remove application messages and disconnects from a server. The client can be implemented in multiple programming languages such as: Java, C++, JavaScript, Objective C or Python. Server - a program or a device that acts as a bridge between clients that send messages and clients that are subscribed. It has the following attributes: accepts network connections, accepts messages from clients, processes subscribe/unsubscribe requests from clients and forwards messages that match clients subscriptions. Session - stateful interaction between client and server. It can last as long as a network connections or it can span on multiple connections.

The simplicity of MQTT can be seen also when we compare it with enterprise brokers such as AMQP. The design is much simpler and more focused. It can be more efficiently used in an embedded system. The design of MQTT's broker is more suited for Internet of Things style applications. A broker can support thousands of concurrent client connections.

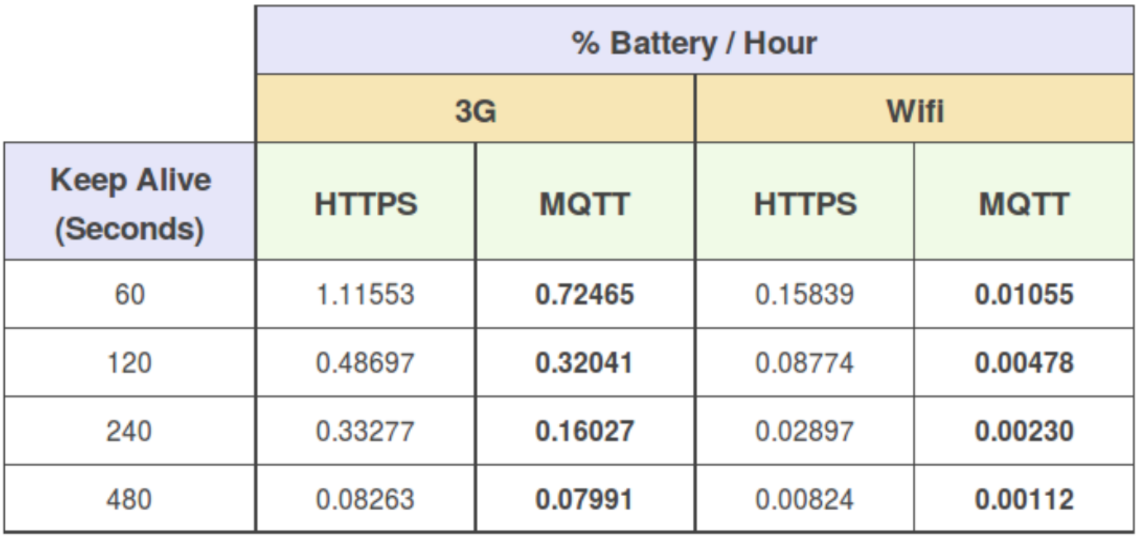
There are some notable projects that use MQTT in real world applications. For example Facebook has used some parts of the protocol in Facebook Messenger. DeltaRails in their IECC Signaling Control System uses MQTT for communication between some of their components.

3.4.1 Why MQTT and not HTTP/REST?

REST is a request-response protocol for client-server applications. In the next paragraphs I am comparing MQTT with the popular RESTful web protocol.

The first very important aspect to compare is how establishing and maintaining a connection works. Compared to HTTP/REST, MQTT is more feature rich regarding this topic. For example the function keep alive can detect if a connection is lost. The client and server send to each other keep alive messages to maintain the connection. This is a faster procedure that the one that HTTP uses ? it waits for a long TCP/IP timeout.

Below we see a table with the cost of maintaining a connection on a period of time for the two protocols:

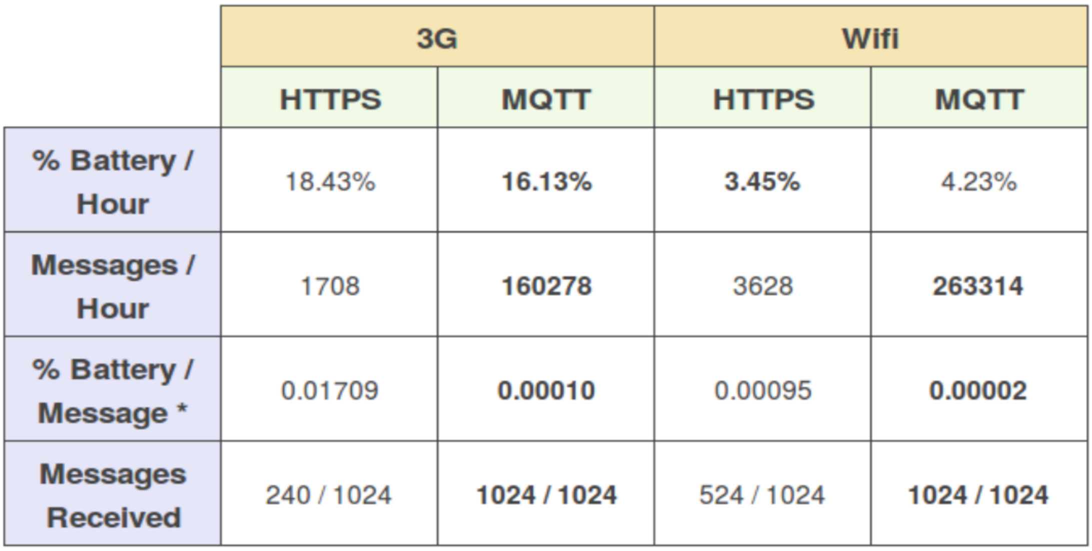


	% Battery / Hour			
	3G		Wifi	
Keep Alive (Seconds)	HTTPS	MQTT	HTTPS	MQTT
60	1.11553	**0.72465**	0.15839	**0.01055**
120	0.48697	**0.32041**	0.08774	**0.00478**
240	0.33277	**0.16027**	0.02897	**0.00230**
480	0.08263	**0.07991**	0.00824	**0.00112**

Figure 3.5: MQTT vs HTTP: Battery consumption

From the table we can notice that MQTT is more efficient. The longer the connection, the 'cheaper' MQTT gets when it comes to the precent of drained battery.

Another comparison between the two protocols is made by the number of messages that can be sent in a given amount of time. A comparison can be seen in the following table. In the table there are four categories of comparison. In each of the categories MQTT wins by quite a margin. This protocol can send many more messages in an hour. Also the battery consumption for each message is smaller compared to HTTP. We can even see that only MQTT manages to send all the messages in a given amount of time.



	3G		Wifi	
	HTTPS	**MQTT**	**HTTPS**	**MQTT**
% Battery / Hour	18.43%	**16.13%**	**3.45%**	4.23%
Messages / Hour	1708	**160278**	3628	**263314**
% Battery / Message *	0.01709	**0.00010**	0.00095	**0.00002**
Messages Received	240 / 1024	**1024 / 1024**	524 / 1024	**1024 / 1024**

Figure 3.6: MQTT vs HTTP: Messages

Given the facts that were presented we can safely say that MQTT is a better choice concerning mobile devices. These are the reasons I chose to use MQTT over REST for this IoT application.

3.4.2 The implementation

In the context of the Mobile Health Monitoring application, the MQTT protocol is used for the communication to the server. The client, the mobile application, sends a connection request to the MQTT broker on the server. This connection request consists of the host of the MQTT broker to which the client application has to connect to, the port of the server and the client id. The connection request has different callbacks for a successful or an unsuccessful connection. If the connection is successful, the client can start publishing messages to a topic or it can subscribe for receiving messages from a topic. In the Mobile Health Monitoring application, after a successful connection, the application starts publishing messages. The publish request consists of the topic the message will be published on, the actual message and the quality of service value.

The messages that have to be sent to the backend are sent in a batch. This batch has a given number of messages. Each message is set to the backend separately. As a message arrives to the backend successfully, it triggers the success callback on the client application. This successful callback enables the next message in the batch to be sent and so on.

The connection to the server is kept all the time, even if the application is in background.

3.5 Smartphones: How do these expand our universe?

Every other year, or even more times a year, a new batch of mobile devices is released that is more powerful than the generation before it. With the help of the smartphones people have been able to automate a lot of at home or even at work tasks without the need of anything else. Smartphones can very good take the place of a computer in a lot of cases. Only a few years ago, the smartphones were not powerful enough to fulfil all of the computing demands. This is not the case any more. Nowadays, a smartphone possesses an amazing processing power, better battery life, improved networking speeds and larger screen sizes.

3.5.1 How do smartphones come in play when it comes to healthcare?

Some new smartphones come with integrated applications that can count steps or that can monitor other activities of a person. Researchers say that the smartphones can be as accurate as a wearable when it comes to monitoring these kind of information. But when it comes to health vital signs like the heart rate or the blood oxygen, only a wearable can have sensors to monitor those vitals. The smartphone still has a role here though. It can connect to the wearable and receive the data from it. This received information can be shown as human readable on the smartphone, or the smartphone can process this information in different ways, something that a wearable cannot do. Also, a smartphone can work like a gateway to open the communication to a powerful server which can analyse the information received from the wearables further with more computing demanding algorithms which are not suited for a mobile device. This is also the case in the context of this system.

The smartphone has the application which I implemented installed. This application scans for the bluetooth advertising packages and receives them. Next, the application takes from the packages exactly the data it wants and then sends it to the backend server. This is the big picture of the way the application is working. Further in this chapter, I will describe in a detailed manner how the application is implemented and how it works.

3.5.2 Mobile Health Monitoring: Application flow

To describe the application flow I am using for the example the Polar H7 Heart Rate Sensor and an iPhone 5S with a Bluetooth connection and an Internet connection.

The application is developed in Objective C so it is meant for iOS devices. As a future target, an application for the Android devices will be next. This system can be used by every person who wants to stay healthy and aware of his body health changes all the time.

Another advantage of building this application for the iOS devices has another advantage: the built in iOS Health application. This application can manage a lot of types of health and activity information from step count data, sleep analysis data, nutrition data to body vitals data like heart rate or blood pressure and others. In the Health app all this data can be seen on separate charts by day, week, month or even year. Also, another very useful thing, this application can be integrated with other health applications by using the iOS HealthKit framework for development. HealthKit allows applications that provide health and fitness services to share their data with the Health app and with each other. A user's health information is stored in a centralised and secure location and the user decides which data should be shared with another application. [4]

The Polar H7 Heart Rate sensor must be placed on the chest in order for it to measure the heart rate. This sensor continuously advertises bluetooth packages containing the heart rate measurement information. On the other side, the iOS device, must have the Bluetooth turned on so that it can scan for the packages advertised by the wearable device.

The application I implemented is installed on the iPhone. At the beginning it requests access to the Accessories so that it can start the Bluetooth if it is not started. Then the iPhone pairs with the wearable and starts receiving advertising packages. Also at the beginning, the application requests access to the iPhone's Health application. It takes advantage of this built in iPhone application to show the received data on different types of charts. For example the following picture shows how the heart rate data sent to the Health app can be seen in the chart in the Health app. Also, in the figure we can observe the Walking and Running Distance measured by the sensor installed in the iPhone hardware.

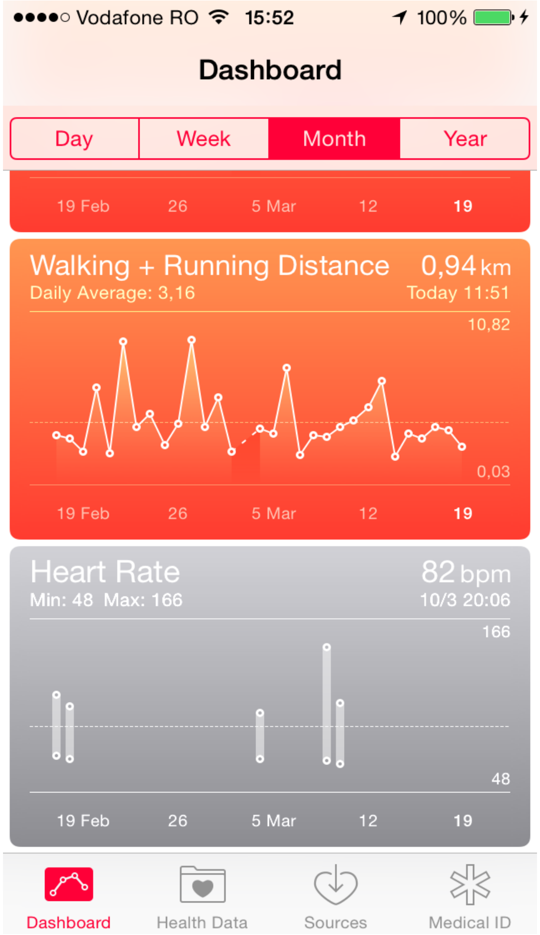


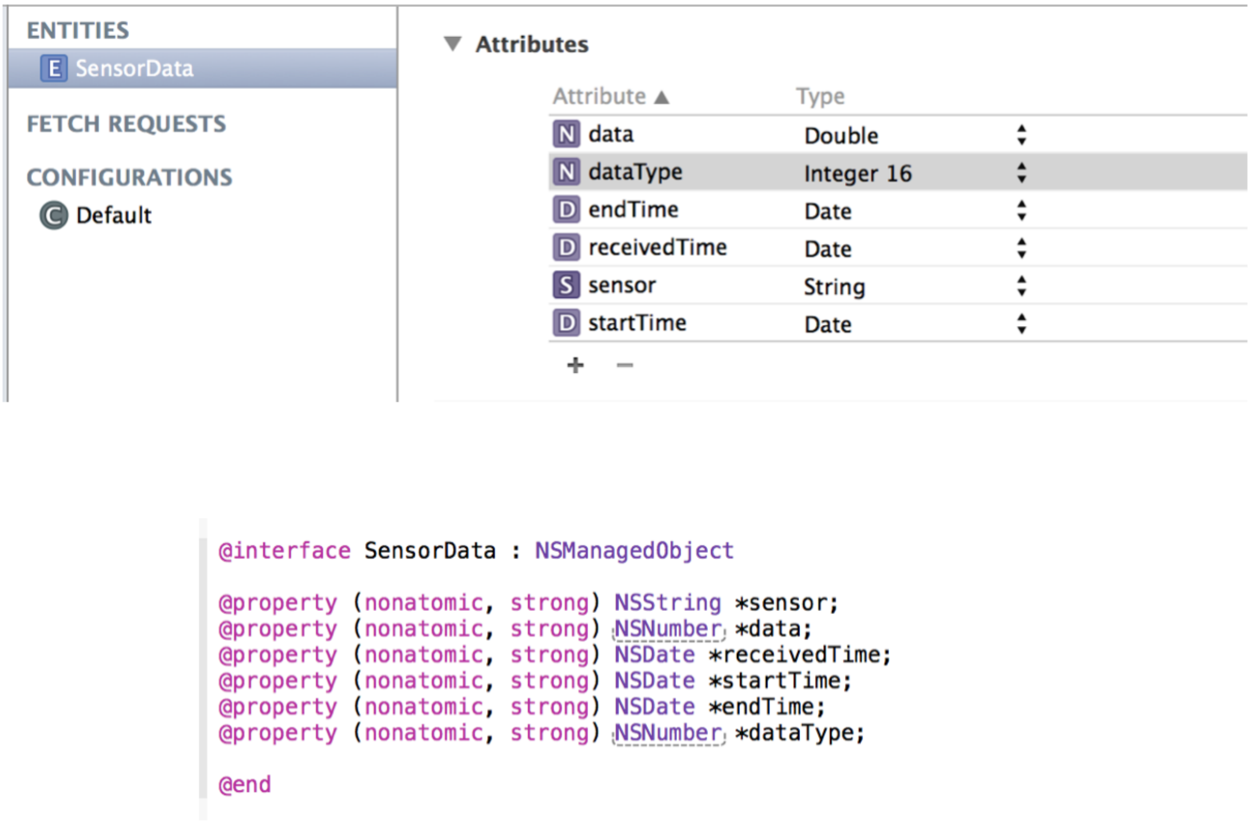
FIGURE 3.7: Healthapp: Heart Rate data over a month

The iOS application implemented for this system takes advantage of the Health app to have different types of data shown over periods of time in an eye catching Graphical User

Interface. Also, on the other hand, the Mobile Health Monitoring application requests, depending on what it needs, different types of data from the Health app: birthdate, sex, weight, blood type, different body measurements, fitness data, nutrition data and so on.

After the application gets the access from the user to write and read data from the Healhapp, every time a heart rate measurement is found in a bluetooth advertising package, it is sent to the Healthapp where it can be seen at every moment.

At the same time, as the data is continuously received from the bluetooth sensor, it is also saved locally in an SQLite database on the iPhone. SQLite is a relational database management system which is embedded into the end program. For this application I only need a table and a mapped entity for the sensor data. The following screenshots present the "Sensor" table and its corresponding entity.



Attribute ▲	Type
N data	Double
N dataType	Integer 16
D endTime	Date
D receivedTime	Date
S sensor	String
D startTime	Date

```
@interface SensorData : NSManagedObject

@property (nonatomic, strong) NSString *sensor;
@property (nonatomic, strong) NSNumber *data;
@property (nonatomic, strong) NSDate *receivedTime;
@property (nonatomic, strong) NSDate *startTime;
@property (nonatomic, strong) NSDate *endTime;
@property (nonatomic, strong) NSNumber *dataType;

@end
```

FIGURE 3.8: Sensor Table and Entity

The "Sensor" entity has six attributes. The "sensor" string is used to save the sensor from which the data arrived. The "data" number holds the actual information for example the number of heartbeats which are received. The "receivedTime" date represents the date at which the information was received from the sensor. The "startTime" and "endTime" dates represent the starting and the ending date for an information received from the sensor, for example, the step count is received in an interval and I save the starting and the ending point of this interval into the database. The "dataType" number represents the type of information which was received from the sensor, for example heart rate is represented by the number 1, step count by the number 2 and so on. These numbers are saved in an enumeration to increase compile-time checking and to avoid errors from passing invalid constants.

The Mobile Health Monitoring application does one more additional task at the beginning: it also connects to the backend system. The successful connection is announced through a callback which triggers a counting task. The counting tasks checks how many records there are in the SQLite "SensorData" table. and if there are a specified number of entries, for example 1500, the app starts sending this batch to the server. The sent data is deleted from the table. After this task is done, the counting tasks starts again and when it reaches the specified maximum amount of data it sends it again to the backend system and so on. The data from the bluetooth wearables is received by the application at a very quick pace. If this continuously incoming streams of data would have been sent to the backend directly, which means also continuously, the application would have needed to use an enormous amour of memory, and after a while it would have crashed because the system would have been out of memory. Another reason for not sending data to the server continuously is that also the Internet connection of the device would have been used completely, and the other applications that need an Internet connection wouldn't have been working properly any more. But thanks to the way I chose to implement the Mobile Health Monitoring application, the app uses a very small amount of the smartphone's memory, and also the Internet connection is not overflown.

The next figure presents the application flow which I have described in words before, in the form of a schema.

Sending the application into background does not affect the flow of the application. The Mobile Health Monitoring application is designed to work as a service application. It

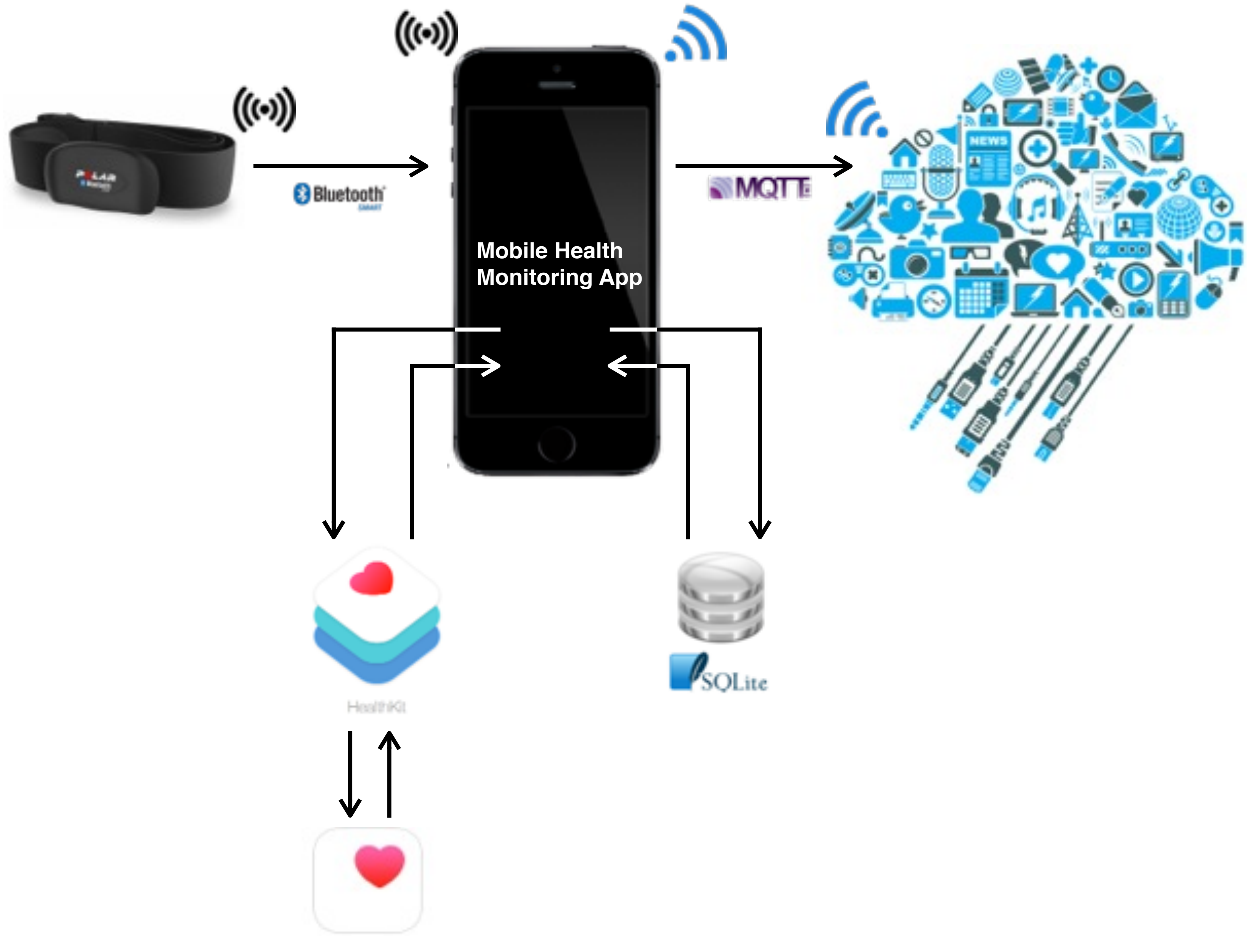


FIGURE 3.9: Application Flow

can also be easily integrated into other applications.

The application would follow this normal flow also with the other sensors. Thanks to using the BLE standard, the others sensors can connect to the same implementation of the application, not having to adapt this for every other wearable.

3.6 The Cloud Backend: Who is taking care of all the data and how?

In the previous subchapters I was talking about a backend application I implemented. Now I will go into more detail about it.

The connection from the application to the backend is made through the MQTT protocol. The connection is made to a Node-Red application which directly sends the received data to the NoSQL database in the BlueMix Cloud.

3.6.1 BlueMix Cloud: What does it mean?

Cloud computing is defined as a type of computing that relies on sharing computing resources rather than having local servers or personal devices to handle applications.

In cloud computing, the word cloud (also phrased as ”the cloud”) is used as a metaphor for ”the Internet,” so the phrase cloud computing means ”a type of Internet-based computing,” where different services - such as servers, storage and applications - are delivered to an organization’s computers and devices through the Internet.

The goal of cloud computing is to apply traditional supercomputing, or high-performance computing power, normally used by military and research facilities, to perform tens of trillions of computations per second, in consumer-oriented applications such as financial portfolios, to deliver personalized information, to provide data storage or to power large. [5]

The definition above is the definition which applies also to the BlueMix Cloud. Its base is is the same as for all other cloud computing systems. I won't go into the details which differentiate BlueMix from the Google Cloud Platform or the Amazon Cloud or others.

3.6.2 Node-Red: Connecting the connections

Node-RED visual coding tools are simplifying the job of wiring up today's world of computers, sensors and online services.

In the system I implemented, I used Node-Red to wire up the incoming streams of data from the sensors through the smartphone with the database in the cloud.

Below is a figure for the visual explanation:

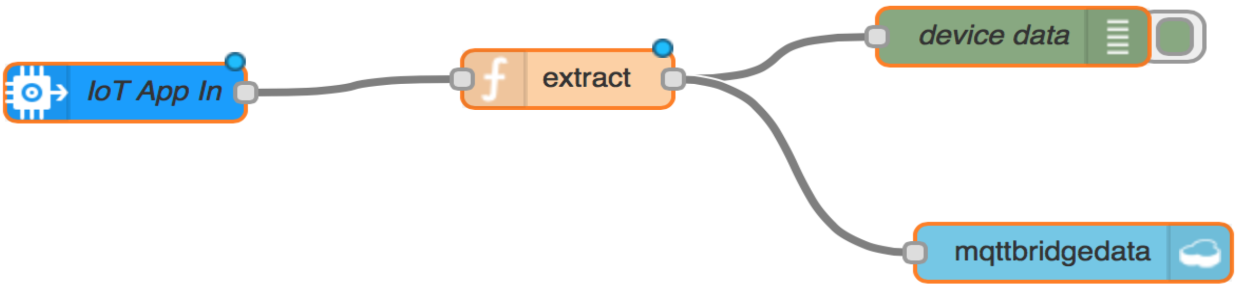


Figure 3.10: Node-Red

In the figure above, there are four nodes. The first node, IoT App In, represents the input node. This node is configured to listen to the event of incoming data from a device. The next node, extract, represents a function which extracts the relevant data from the payload which is received in the previous node. For example here, for the heart rate I only need the to extract the number of beats and the date and time at which the data was registered from the sensor. The third node, device data, does not affect the Node-Red flow in any way as it is only used here for debugging reasons. The last node, mqttbridgedata, is the one which receives the extracted data and sends it to the database in the cloud. This node is configured with the specific URL and credentials for the database to which it has to connect and send the data to.

Now that the data is in the database in the cloud, I can retrieve it from there in another cloud-based application to analyse the data.

3.6.3 The Database: Where all the data is stored

For the implementation of this system I used the CouchDB NoSQL database. This database can be easily replaced with another NoSQL database like MongoDB or DynamoDB, for example, or even with Hadoop or Cassandra if there is a massive amount of data.

NoSQL databases are the best fit for cloud computing systems, because nowadays, in the web world, the need for scalable databases has been increased with the growing data in the Internet world. These needs are being fulfilled by the NoSQL databases with their high scalability and easy programmable model.

The CouchDB NoSQL database stores the data in JSON (JavaScript Object Notation) document format. The access to these documents is made through querying indexes with the web browser, via the HTTP-based REST API.

Traditional relational databases allow running any queries you like as long as the data is structured correctly. In contrast, CouchDB uses predefined map and reduce functions in a style known as MapReduce. These functions provide great flexibility because they can adapt to variations in document structure, and indexes for each document can be computed independently and in parallel. [6]

MapReduce is a programming model for processing large data sets with a distributed algorithm in a cluster. It is composed of a Map() function which does a filtering or sorting on the data and then a Reduce() function which does a summary of the retrieved data. This is how the data is queried in CouchDB and in this project.

3.6.4 The REST Web Service in Java

To retrieve the data from the database, I implemented a REST Web Service in Java. The web service connects to the database using its specific URL and credentials. After it is connected, it makes different requests to the database to retrieve the needed data.

So that the connection to the database is possible, I am using the CouchDBConnector from ektorp. There are other several connectors which can be used, but this one was best suited for my implementation.

The server is located in the BlueMix cloud and the dependency injections are managed by the Spring Boot framework which I am using for this system. Spring Boot is an application framework for the Java platform. Its core features can be used by any Java application, and there are also extensions for building web applications on top of the Java EE platform.

After the connection, the next step of the web service is to get the requested data from the database. When it gets this data, it exports it in .CSV (comma separated values) files so that they can be used for further analysis.

The same web service provides an interface which can be used by a GUI (Graphical User Interface) application to show data and other conclusions drawn from that specific data. The GUI application I implemented is an independent web application. Any type of GUI application can be used with the web service in this project. More about the small GUI application in the following section of this chapter.

3.6.4.1 The GUI Web Application

The GUI web application I implemented for this prototype is a very simple one from the GUI point of view. It just shows the correlation between the heart rate of a user and his corresponding activity status in a chart. Furthermore, it shows a list in which the user can see the total time of the data shown in the chart, the total resting time and the mean heart rate for that resting time and also if there were any, how many stress periods the user had, how much they lasted and what the mean heart rate during the stress periods was. In the same list, there is also shown the total moving time and he mean heart rate for this time. Also, for the moving time the user can see how many very low, low, medium, high and very high intensity training periods he had and how much they lasted and, of course, the mean heart rate he had during each period.

As a conclusion, the user can also see, according to his age, gender, weight and height, if he is in the optimal parameters or not.

All this information is computed in the backend application which just responds to the requests from the GUI with the results.

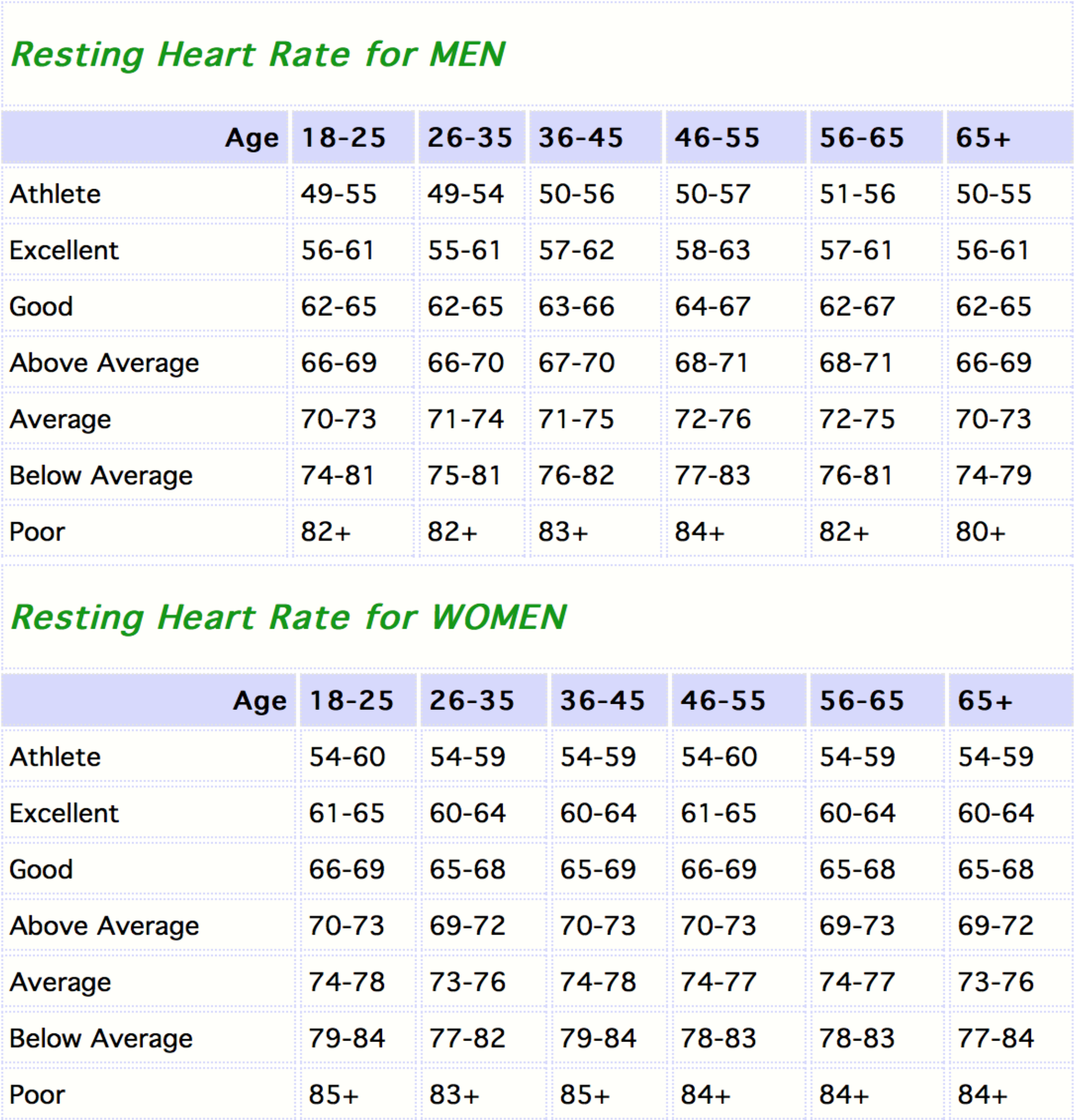
3.6.4.2 Data transformation and analysis

The data for the chart in the GUI application, the heart rate and the activity data, is parsed from the JSON which the web service receives from the database and then another JSON object is made available to the GUI application. Then from this data, which also has some timestamps available, the total time is calculated. The activity data is measured in numbers of steps. This data shows if the person was in a resting period (not moving) or in a training period (moving).Therefore I can calculate the total resting and the total moving time and also the corresponding mean heart rate.

During resting, a person can also have stress periods, and I detect this in the backend by differences of the current heart rate from the mean heart rate. If the current heart rate is very different from the mean heart rate to doesn't have to mean that it denotes a stress period, it could actually be a recovery period which resulted from a training/moving period. Therefore I calculate the stress periods accordingly, by taking into consideration if the period before was a resting or a moving period and looking at how much the recovery period is lasting. The result shows how many stress periods the suer had and how much they lasted and also the mean heart rate for these.

During training, there can also be different periods representing different training intensities. Judging by how many steps a person does in a period of time, I can calculate the type of training intensity: very low, low, medium, high or very high.

The conclusions I draw are according to some tables which show the optimum mean heart rate values. The tables I am using are taken from the http://www.topendsports.com site, "The Sport + Science Resource". The site provides a wide range of quality information about sports, science, fitness and nutrition, and much more. For example, I am showing the tables for the resting mean heart rates for men and women in the following figure:



Resting Heart Rate for MEN

Age	18-25	26-35	36-45	46-55	56-65	65+
Athlete	49-55	49-54	50-56	50-57	51-56	50-55
Excellent	56-61	55-61	57-62	58-63	57-61	56-61
Good	62-65	62-65	63-66	64-67	62-67	62-65
Above Average	66-69	66-70	67-70	68-71	68-71	66-69
Average	70-73	71-74	71-75	72-76	72-75	70-73
Below Average	74-81	75-81	76-82	77-83	76-81	74-79
Poor	82+	82+	83+	84+	82+	80+

Resting Heart Rate for WOMEN

Age	18-25	26-35	36-45	46-55	56-65	65+
Athlete	54-60	54-59	54-59	54-60	54-59	54-59
Excellent	61-65	60-64	60-64	61-65	60-64	60-64
Good	66-69	65-68	65-69	66-69	65-68	65-68
Above Average	70-73	69-72	70-73	70-73	69-73	69-72
Average	74-78	73-76	74-78	74-77	74-77	73-76
Below Average	79-84	77-82	79-84	78-83	78-83	77-84
Poor	85+	83+	85+	84+	84+	84+

Figure 3.11: Resing Heart Rate: Men and Women

After having all the analysis done, I can draw some conclusions about the user's fitness status according to the tables above.

There is also another table with optimal values for every training intensity and every age, and with the help of this I can compare the results I got for a specific user to the optimal values and see if his/her training heart rate shows that he/she is in good/optimal parameters or not.

And the following figures show how everything looks like for the user in the web application. Of course, as I mentioned before, any GUI can be used with the REST web service I implemented, so this is just a prototype phase of a simple GUI to see an example.

The first figure shows the chart which allows the user to see how his heart rate and activity data look and how they correlate.

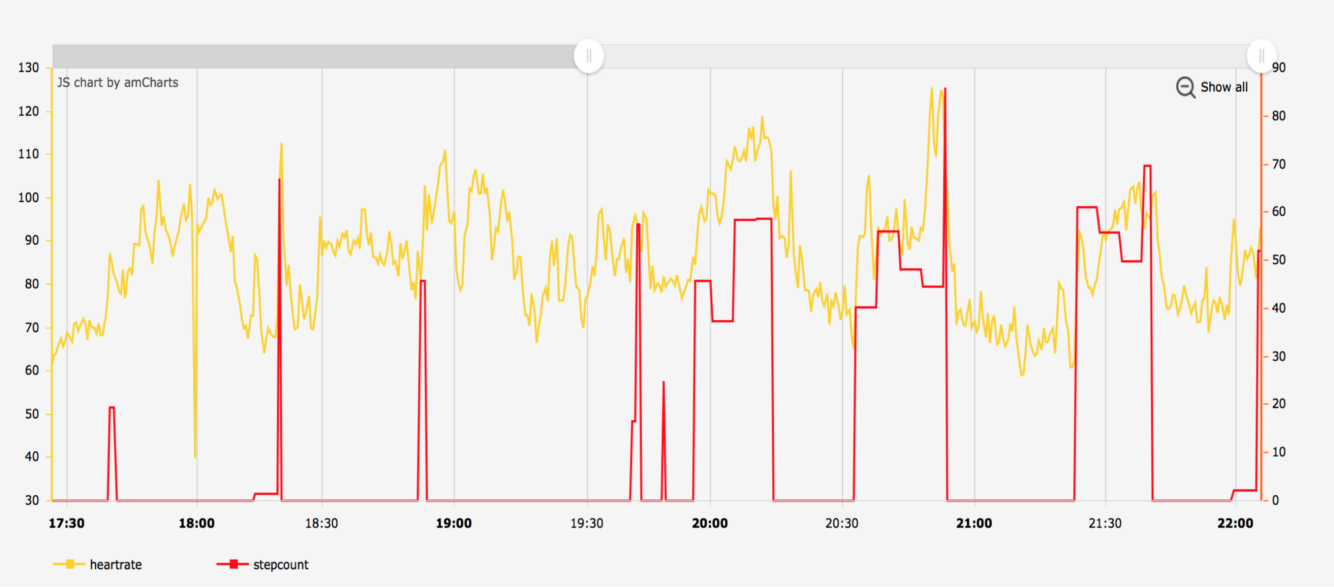
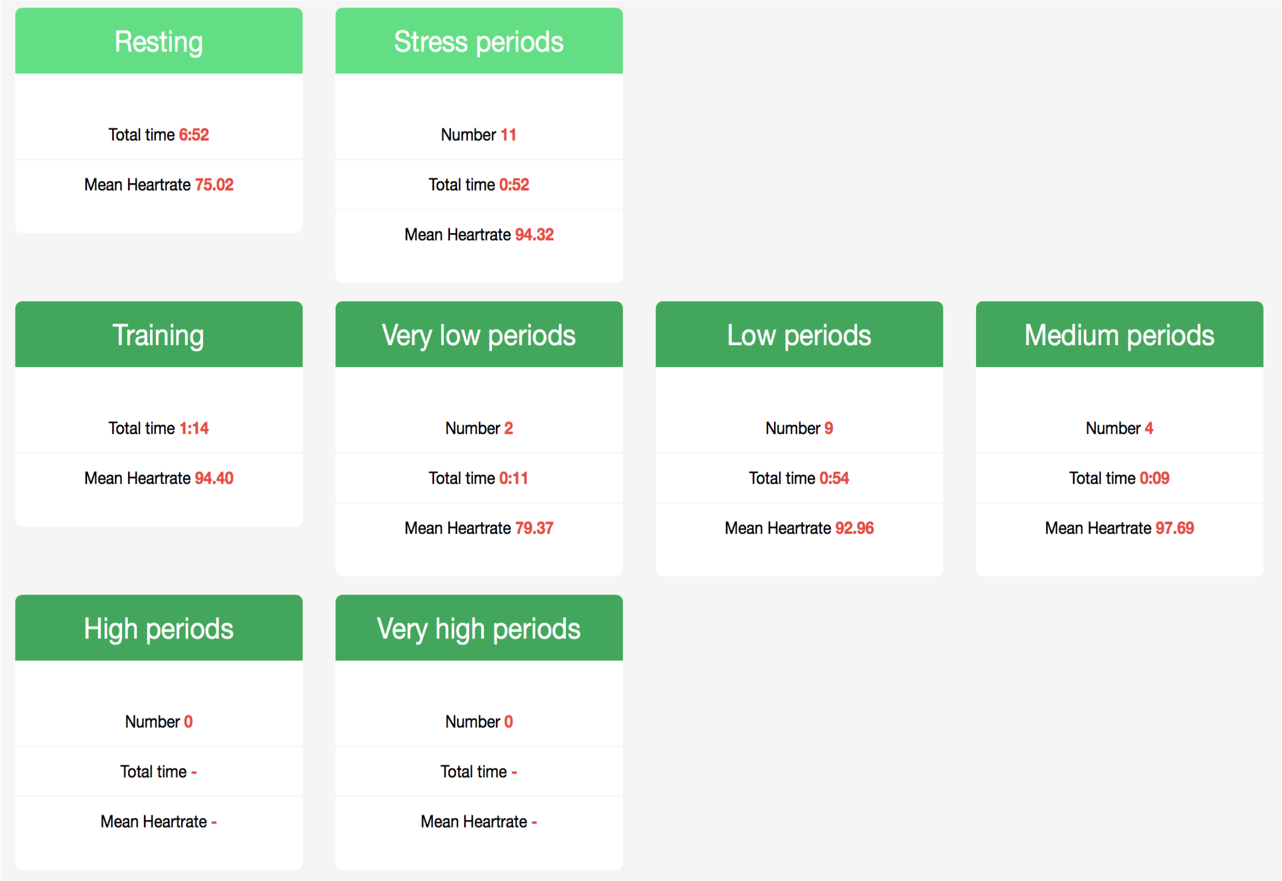


Figure 3.12: Web application: Chart with heart rate and activity

For example for this data I could get the following calculations and results with the corresponding conclusions:



Resting

Total time 6:52

Mean Heartrate 75.02

Stress periods

Number 11

Total time 0:52

Mean Heartrate 94.32

Training

Total time 1:14

Mean Heartrate 94.40

Very low periods

Number 2

Total time 0:11

Mean Heartrate 79.37

Low periods

Number 9

Total time 0:54

Mean Heartrate 92.96

Medium periods

Number 4

Total time 0:09

Mean Heartrate 97.69

High periods

Number 0

Total time -

Mean Heartrate -

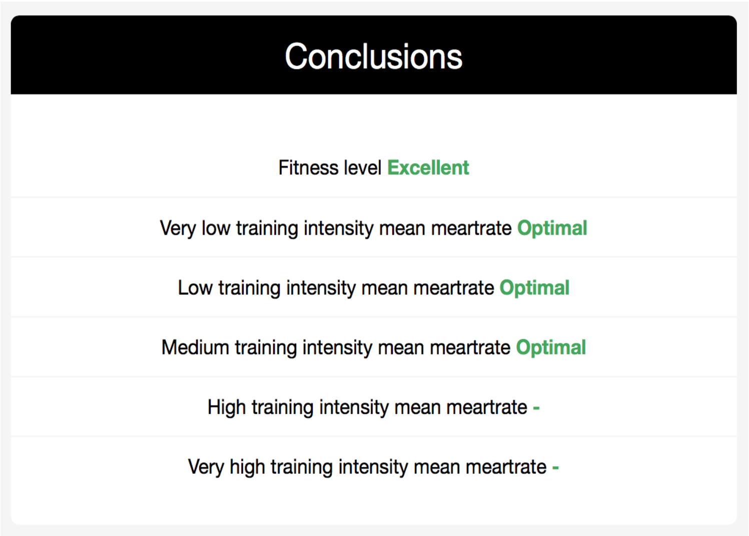
Very high periods

Number 0

Total time -

Mean Heartrate -

FIGURE 3.13: Web application: Results



Conclusions

Fitness level Excellent

Very low training intensity mean meartrate Optimal

Low training intensity mean meartrate Optimal

Medium training intensity mean meartrate Optimal

High training intensity mean meartrate -

Very high training intensity mean meartrate -

FIGURE 3.14: Web application: Conclusions

Chapter 4

Analytics: Person Classification by Heart Rate

How do we get answers?

Analytics is the discovery and communication of meaningful patterns in data. Especially valuable in areas rich with recorded information, analytics relies on the simultaneous application of statistics, computer programming and operations research to quantify performance. Analytics often favors data visualization to communicate insight.

Firms may commonly apply analytics to business data, to describe, predict, and improve business performance. Specifically, areas within analytics include predictive analytics, enterprise decision management, retail analytics, store assortment and stock-keeping unit optimization, marketing optimization and marketing mix modeling, web analytics, sales force sizing and optimization, price and promotion modeling, predictive science, credit risk analysis, and fraud analytics. Since analytics can require extensive computation (see big data), the algorithms and software used for analytics harness the most current methods in computer science, statistics, and mathematics. [7]

In this project I implemented two different approaches for creating a prediction model which can analyse if the data it receives is from the person it has been trained to recognise or not. The business context this system is used gives this analytics task very much importance.

4.1 Person Classification using Heart Rate data

Heart rate, or heart pulse, is the speed of the heartbeat measured by the number of poundings of the heart per unit of time ? typically beats per minute (bpm). The heart rate can vary according to the body's physical needs, including the need to absorb oxygen and excrete carbon dioxide. Activities that can provoke change include physical exercise, sleep, anxiety, stress, illness, ingesting, and drugs.
The normal resting adult human heart rate ranges from 60?100 bpm. [8]

I will start by showing some plots with how the time series data with the heart rate for 2 different people looks like.

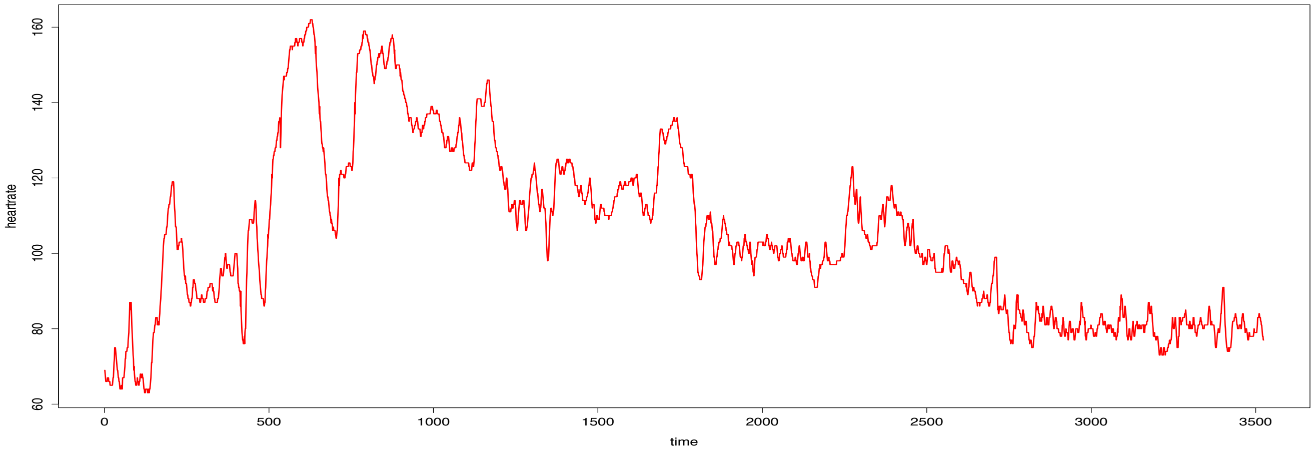


Figure 4.1: Plot: Person 1a

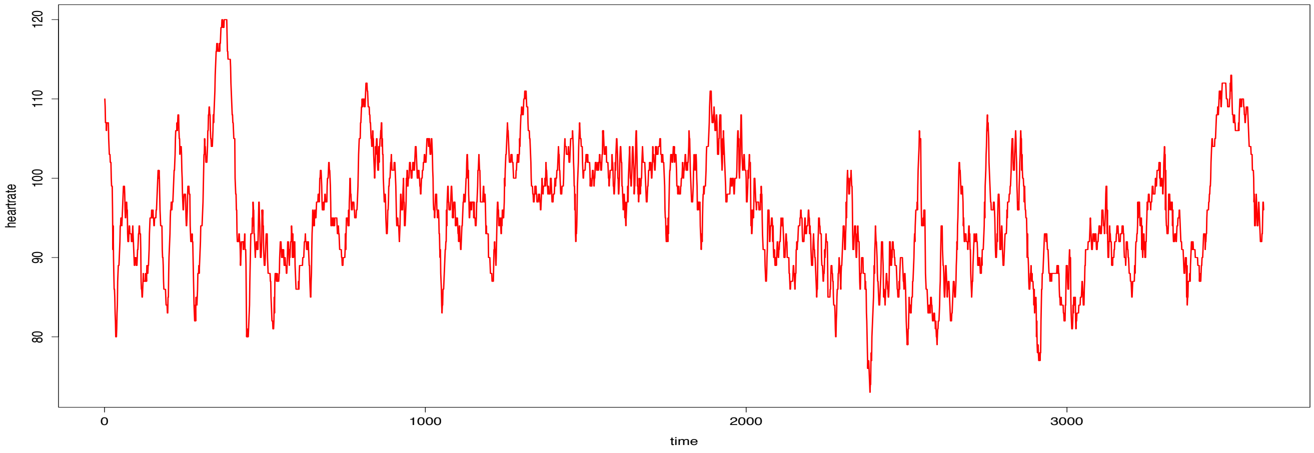


Figure 4.2: Plot: Person 1b

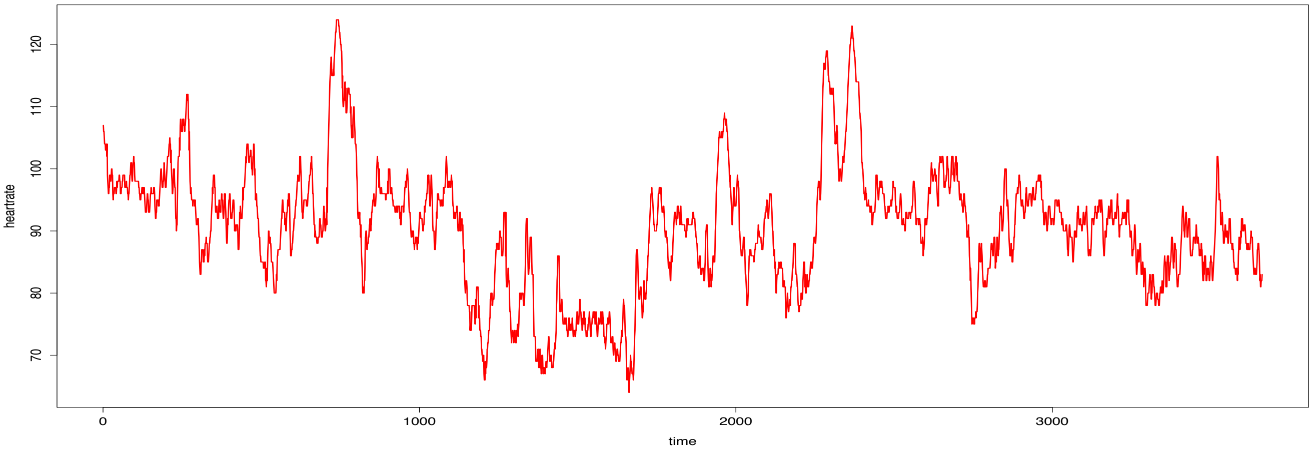


FIGURE 4.3: Plot: Person 2

The charts are plotted in R. R is a programming language and software environment for statistical computing and graphics. R has become one of the most widely used statistical software in the recent years.

So looking at the charts above. Could someone differentiate between person 1 and person 2? Or could someone say that the first 2 plots belong to the same person? The answer is obviously: no. But by analysing the data thoroughly, I could find specific patterns which can actually differentiate those data sets in the charts.

My aim, using the following two approaches is to implement an algorithm in such a way that it can recognise another or the same person from a small amount of testing data. The training data on the other hand should have enough values so that the algorithm can learn enough specific patterns for a person.

4.2 About the Data Set

The initial data set consists a time series. For every registered time, there is a specific heart beat number. The time is measured in milliseconds and the heart rate is represented by an integer value. This is how I sent the data from the mobile application on the smartphone to the database in the cloud. This way I have the accurate data directly from the heart rate sensor.

4.2.1 Heart Rate Variability

Having the heart rate values, and the exact millisecond at which every heart rate value was registered, I calculated the heart rate variability values for each beat.

Heart rate variability (HRV) is the physiological phenomenon of variation in the time interval between heartbeats. It is measured by the variation in the beat-to-beat interval. [9] So this means that the HRV indicates the fluctuations of heart rate around an average heart rate. An average heart rate of 60 beats per minute (bpm) does not mean that the interval between successive heartbeats would be exactly 1.0 sec, instead they may vary from 0.5 sec up to 2.0 sec.

The image below shows some heart rate measures and the intervals, calculated in milliseconds, between every beat which vary from a beat to another.

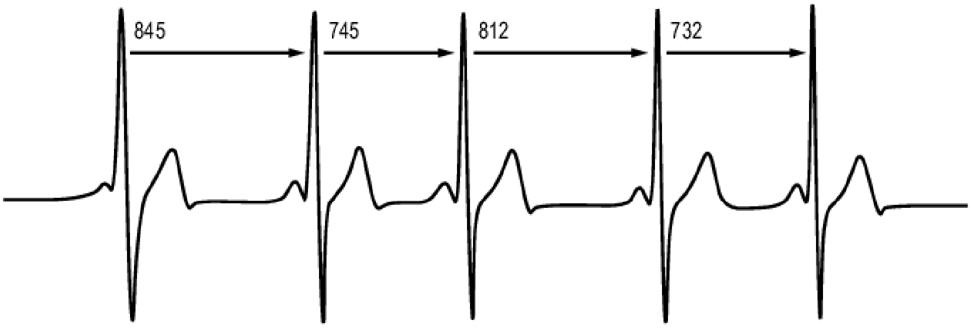


Figure 4.4: HRV: Fluctuations

HRV analysis has gained significant clinical attention as can be seen from the large number of research efforts of the past two decades. Artificial intelligence and machine learning methods constitute a powerful tool in HRV analysis. Hon and Lee found a diminished beat-to-beat variation in the fetal heart rate and treated it as an indicator of distress [14]. Since then, heart rate variability (HRV), a measurement of the beat-to-beat variations in the instantaneous heart rate or R-R intervals, has received a lot of research interest. HRV is regarded as an indicator of the health status of the autonomic nervous system [15]. Irregular HRV is always associated with disorders of internal secretion or the cardiovascular system, such as coronary artery disease (CAD) and congestive heart failure (CHF) [16]. Therefore, HRV plays an important role in the assessment of the cardiovascular system and overall health.

So HRV is a very important measurement of a persons health. Still, the charts for the heart rate variability would be approximately the equivalent of the charts I presented before for the heart rate measurement. This does not really help when it comes to analysing if the data is from the same person or from someone else. But there is something that can help: HRV has a lot of features which can be extracted from it and then analysed. I found a way of using some specific features of the HRV to implement an pattern recognition algorithm which learns how to differentiate people.

4.3 Preprocessing: Data Cleaning

In general, physiological signals need to be pre-processed since noises are introduced at some point along the process of data acquisition. The preprocessing step is essential for removing the noises in the received signal, and for enhancing it to prepare it for the analysis. There are three types of time series pre-processing stages: ectopic beat/interval correction, detrending, and resampling. For this system, I implemented the ectopic beat correction.

Ectopic intervals are abnormal intervals that result from false/missed heart beat readings. These ectopic intervals need to be removed from the time series to avoid errors, as reported by Thuraisingham. [17]

There are many approaches to detecting and removing the ectopic beats reported in the literature. In this project, the time interval difference between two consecutive heart beats, which differs by more than 20% from the previous heart beat interval, is considered an ectopic interval. [18] So this is how the detection of the ectopic beats works. I am looking at the differences between consecutive heart beat intervals and I am selecting the ones which differ more than the set threshold: 20%. Ectopic beats have to be corrected. This means removing the ectopic beat and replacing it with a correct beat, so that the removal does not affect the result of the analysis of the time series data set.

4.3.1 Ectopic beats correction

There are some replacement techniques for the correction of the ectopic beats based on mean/median methods. I wanted to use a more complex and more reliable technique for this project, so I chose the cubic spline interpolation method. [19]

Given a set of $n + 1$ data points (x_i, y_i) where no two x_i are the same and $a = x_0 < x_1 < ... < x_n = b$, the spline $S(x)$ is a function satisfying:

1. $S(x) \,\epsilon\, C^2[a, b]$;
2. On each subinterval $[x_{i-1}, x_i], S(x)$ is a polynomial of degree 3, where $i = 1, ..., n$.
3. $S(x_i) = y_i,$ for all $i = 0, 1, ..., n$.

Then we can assume that:

$$S(x) = \begin{cases} C_1(x), & x_0 \leq x \leq x_1 \\ ... & \\ C_i(x), & x_{i-1} \leq x \leq x_i \\ ... & \\ C_n(x), & x_{n-1} \leq x \leq x_n \end{cases}$$

where each $C_i = a_i + b_i x + c_i x^2 + d_i x^3$ $(d_i \neq 0)$ is a cubic function, $i = 1, ..., n$.
So, to determine the cubic spline $S(x)$, a_i, b_i, c_i, and d_i, have to be determined for each i, so that the first and the second derivates $C'(x_i)$, and $C''(x_i)$ are linear. After these coefficients are determined, the spline $S(x)$ can be used to calculate the corrected value of the ectopic beat x according to the intervals before and after the beat taken into consideration.

I implemented this method in Java. After detecting an ectopic beat, I am replacing it using the above mentioned interpolation function. For the interpolation function, I am taking into consideration all the intervals from a minute before the ectopic beat occurred until its occurrence, and also all the intervals which occur from the ectopic beat until a minute after it. So that I can apply the function accordingly, I have to sort the data in the interval mentioned above and retrieve only the unique values. Next, so that I can create the cubic spline interpolation function, I set my x points as the heart rate measurement values and the y points as the HRV measurement values, each corresponding to the heart rate measurements. This is how I get the new result with the corrected ectopic beat. I am overwriting the ectopic beat value with the corrected one for the specific interval in the time series data set.

After this data cleaning step I can go on analysing the HRV knowing that there are no more inaccurate values in the data set.

4.4 Feature Extraction

Statistical features from HRV time series can be extracted using three different methods: time, frequency, and non-linear domain. For this implementation, I extracted features from the time and non-linear domain.

4.4.1 Time Domain Analysis

In the time domain analysis, statistical measures are obtained directly from the HRV time series data set. The features are calculated for intervals of 60 seconds. The data is segmented in 60 second segments.

Mean Heart Rate and HRV For each segment, the mean heart rate and the mean HRV are calculated. The following formula is used for both features:

$$Mean = \frac{1}{n}\sum_{i=1}^{n} x_i \text{ ,where n = size of the segment of 60 seconds,} \tag{4.1}$$

$$x_i = \text{value of the HR/HRV for the i-th segment point}$$

Standard Deviation for HR and HRV For each segment, the standard deviation for the HR and the standard deviation for the HRV are calculated.

In statistics, the standard deviation (SD) is a measure that is used to quantify the amount of variation or dispersion of a set of data values. A standard deviation close to 0 indicates that the data points tend to be very close to the mean (also called the expected value) of the set, while a high standard deviation indicates that the data points are spread out over a wider range of values. [10]

The standard deviation is equal to the square root of the variance. The variance is the mean of the square of the deviations of each data point from the mean. The following formula is implemented to calculate the standard deviations for each segment:

$$\sigma = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})^2} \text{ ,where n = size of the segment of 60 seconds,} \tag{4.2}$$

$$x_i = \text{value of the HR/HRV for the i-th segment point,}$$

$$\overline{x} = \text{mean HR/HRV value for that segment}$$

Median HRV For each segment, the mean HRV value is calculated using the following formula:

$$Median = \frac{n+1}{2} th\ term \text{ ,where n = size of the segment of 60 seconds,} \tag{4.3}$$

$$\text{n = odd number}$$

$$Median = ((\frac{n}{2})th\ term + (\frac{n}{2}+1)th\ term)/2 \text{ ,where n = size of the segment of 60 seconds,} \tag{4.4}$$

$$\text{n = even number}$$

NN50 and NN20 For each segment, I calculate the number of the successive HRV intervals which differ more than 50 ms, respectively, more than 20 ms.

pNN50 and pNN20 Also, for each segment, I calculate NN50, respectively NN20 divided by the total number of all HRV intervals in the selected segment. These two features show a lot of detail for the selected data segment, thanks to the small differences they are calculated for.

Root Mean Square of Successive Differences (RMSSD) The RMSSD is also calculated for every segment of 60 seconds. It represents the square root of the mean of the squares of the successive differences between adjacent HRV intervals. The following formula is implemented for this features:

$$RMSSD = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n-1}(HRV_{i+1} - HRV_i)^2} \text{ ,where n = size of the segment of 60 seconds} \tag{4.5}$$

Standard Deviation of Successive Differences (SDSD) This time, different than the standard deviation feature mentioned above, instead of calculating the values according to the mean HRV of the segment, I am calculating the values according to the mean standard deviation of that segment. So first, for every segment, I calculate the mean standard deviation, and then the standard deviation of the differences between the successive HRVs depending on the mean standard deviation.

So this means that the following formula describes the SDSD:

$$SDSD = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n-1}(D_i - D_{mean})^2} \text{ ,where n = size of the segment of 60 seconds} \tag{4.6}$$

and

$$D_{mean} = \frac{1}{n-1}\sum_{i=1}^{n-1} D_i^2 \text{ , where n = values in the segment,} \tag{4.7}$$

$$\text{D = standard deviation of an interval}$$

Minimum and Maximum NN For every segment, I retrieve the minimum and the maximum value of the HRV.

HRV triangular index (HRVTI) and Triangular Interpolation of the NN interval histogram (TINN) The HRVTI represents a geometric measure of the HRV.

First of all I needed to build the histogram for each selected interval. The HRVTI considers the major peak of the histogram as a triangle with its baseline width corresponding to the amount of the HRV interval variability, its height corresponds to the most frequently observed duration of the intervals and its area corresponds to the total number of all HRV intervals used to construct it.

The sample density distribution D is expressed as a triangle, which assigns the number of equally long NN intervals to each value of their lengths. The most frequent NN interval length X is established, that is Y=D (X) is the maximum of the sample density distribution D. The HRV triangular index is the value obtained by dividing the area integral of D by the maximum Y. [11] This means that HRV is the total number of all HRV intervals divided by the height of the histogram of all NN intervals measured on a discrete scale with bins of 7.8125 ms (1/128 seconds). This is how the formula is implemented also in this project.

TINN is another geometrical measure which can be retrieved form the HRV values. Baseline width of the minimum square difference triangular interpolation of the highest peak of the histogram of all NN intervals.

The following figure gives a better understanding of the two features mentioned above:

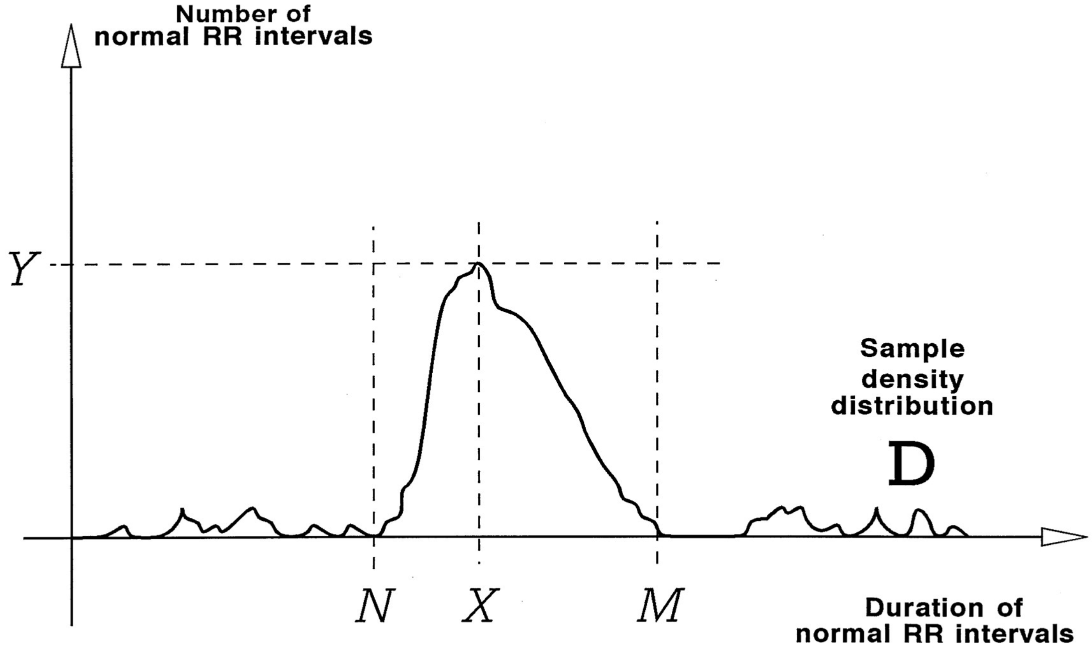

Number of normal RR intervals
Y
Sample density distribution
D
N
X
M
Duration of normal RR intervals

FIGURE 4.5: HRV: Geometric features

According to the figure presented, the two features can be formulated in the following way:

$$HRVTI = nr/Y \text{ , where nr = number of all HRV intervals in a segment}$$

and

$$TINN = M - N$$

4.4.2 Non-Linear Domain Analysis

The non-linear phenomenon of the heart rate variability also has to be explained using some specific methods, because the heart signals show that they can also have complex and irregular behaviour. Due to the complexity of the mechanism of the heart rate, the HRV analysis based on non-linear dynamics can provide valuable information. Non-linear behaviour can assume chaotic behaviour, but a lot of testing has shown that heart rate variability cannot be described as a chaotic process. [20]

Short term variability (SD1) Short term variability represents the standard deviation of the instantaneous beat-to-beat variability. For each segment, the SD1 value is calculated depending on the SDSD variables, with the following formula:

$$SD1_i = \sqrt{\frac{1}{2}(SDSD_i^2)} \text{ , where i = index of the current interval} \tag{4.8}$$

Long term variability (SD2) Long term variability represents the standard deviation of continuous variability. For each segment, the SD2 value is calculated depending on the values for the standard deviation of the HRV and the SDSD variables, with the following formula:

$$SD2_i = \sqrt{(2 * SDHRV_i^2) - \frac{1}{2}SDSD_i^2} \text{ , where i = index of the current interval} \tag{4.9}$$

Spectral Entropy (SpEn) The spectral entropy shows the complexity of the input HRV time series segment. Large values of this feature show that the data has a high irregularity and smaller values indicate more regular time series. This shows us how much information that segment contains. In this project, the Shannon entropy is used with the following formula:

$$SpEn = \sum_{i=1}^{n} p_i \log \frac{1}{p_i} \text{ , where n = size of the segment of 60 seconds,} \qquad (4.10)$$

$$\text{p = probabilities of the values in the current segment}$$

Having all these features calculated, I export them in a .CSV file so that I can load them wherever I need them for the next task: pattern recognition and classification

4.5 Pattern Recognition and Classification

Pattern recognition is a branch of machine learning that focuses on the recognition of patterns and regularities in data, although it is in some cases considered to be nearly synonymous with machine learning. Pattern recognition systems are in many cases trained from labeled ”training” data (supervised learning), but when no labeled data are available other algorithms can be used to discover previously unknown patterns (unsupervised learning). [12]

In they project, pattern recognition and classification techniques are used to find out if a data set belongs to a specific person or not. So this means that I have a one-class pattern matching and classification problem. In the next section I will go on and demonstrate how one-class recognition of heart rate data can be performed at the 90%+ level of accuracy, thanks to an appropriate choice of features and thanks to the pattern recognition algorithm I implemented.

4.5.1 Feature selection

To discover different persons by analysing their heart rate data, I had to find the features which differ the most. That is why, I started by analysing the extracted features mentioned above from four different different subjects.

The mean heart rate is not analysed for differences because I treaded this feature as a reference feature.

The results were the following:

1. Mean Heart Rate Variability: A persons heart rate variability can vary a lot, but it also has some specific values. These specific values are mostly not the same in every person, which is why the differences also proved that this feature will play an important role in the classification.

2. Standard Deviation for Heart Rate: If we look back at the charts in the Figures 4.1 - 4.3, it's easy to notice, even with no calculations, that the heart rate of two persons cannot really be very different. During a day everyone can have normal resting heart rates from 50 to 100 beats per minute (BPM), so no matter the length of the segment for the calculation of the standard deviation for the heart rate, I would not get very different values for different people. This means that this feature will not help me get a better classification result. This is why I chose to leave it out.

3. Standard Deviation for Heart Rate Variability: This feature on the other hand holds a lot of information. For the same reasons as the mean heart rate variability feature, this features is taken into consideration for the further analysis.

4. Median Heart Rate Variability: Like I said, the heart rate variability has a lot of information to give to the analysis algorithm, so this feature, which is directly linked to the HRV is also taken into consideration.

5. NN50 and NN20: These two features, although they provide direct information from the HRV data, do not show any potential for getting a better classification, because they are just a counting value which is not relevant in the classification for this project.

6. pNN50 and pNN20: These two features also provide essential information about the HRV, so they will be used in the following classification approaches.

7. RMSSD: This feature shows that together with other features taken into consideration, it can also provide important information in the further processing, so it will be taken into consideration.

8. SDSD: As this feature is directly linked to the standard deviation for the heart rate variability, there is no doubt that it should be kept for the further investigation.

9. Minimum and Maximum NN: As the heart rate variability can have specific values for a person, it is important to know the minimum and the maximum segment values.

10. HRVTI and TINN: These two geometrical features could be very important, if I would want to classify some other types of classes, like cardiac arrhythmias, or other heart rate symptoms and diseases, but because they are mostly related to the histogram distribution of the heart rate variability, they are not relevant for the classification I am implementing in this project.

11. SD1 and SD2: The features for the short and long term variabilities are very useful as they are directly linked to the standard deviation of successive differences feature and to the standard deviation of the heart rate variability feature.

12. SpEn: The spectral entropy, although it shows interesting details, about the amount of information every segment of heart rate variability segments holds, it is not relevant for determining if this information belongs to one or the other person, because for every person, the amount of information the data holds can have the same variability.

So this means that out of the 17 features extracted, I am taking into consideration 12 as principal features after the analysis described above.

4.5.2 Approach 1: One-Class Support Vector Machines (OC-SVM)

4.5.2.1 Support Vector Machines (SVM)

The Support Vector Machine classifier has become very popular in the past years for solving problems in classification, regression, and novelty detection. As the classification problem I am tackling is a novelty detection problem, the SVM algorithm seems to be a good fit. An important property of the SVMs is that the determination of the model parameters corresponds to a convex optimisation problem, which means that any local solution is also a global optimum.

For a better understanding I will present the SVM model in a general manner, but I will not get into any details, because this work does not make use of the classic SVM model, but the core properties are also the same. This is why I want to present them.

The linear SVM is a model which receives a data set with named classes and then it tries to separate them using a decision boundary, called the separating hyperplane. In the following figure we consider that x's represent one class and o's represent another class. This is how the SVM classification model would look like:

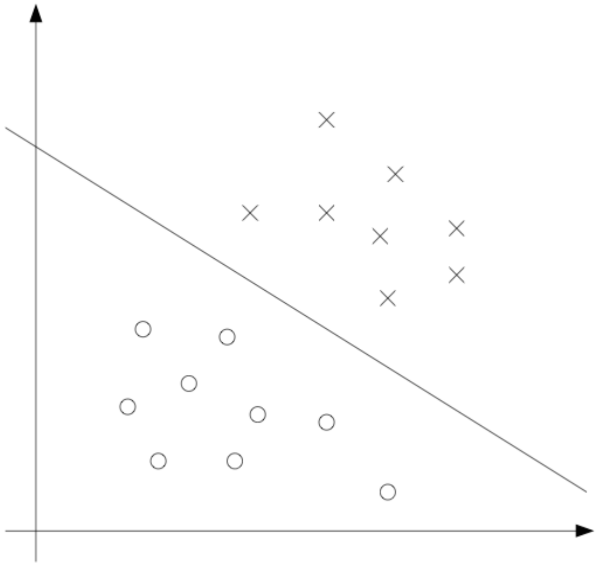


FIGURE 4.6: SVM: Linear

Starting from the point in the presented figure, the SVM creates an optimal margin between the classes.

In addition to performing linear classification, SVMs can perform non linear classification using the "kernel trick" and mapping their inputs into high-dimensional feature spaces. Linear SVMs are using the dot product as the kernel function $K_{lin}(x_i, x_j) = x_i \cdot x_j$. The kernel function for a nonlinear SVM which is better suited when there are a lot of features which have a classification in a multidimensional space, is a nonlinear function:

1. ploynomial: $K_{poly}(x_i, x_j) = (x_i \cdot x_j + 1)^p$
2. radial-basis: $K_{rbf}(x_i, x_j) = exp(\frac{-(x_i - x_j)^2}{s^2})$
3. sigmoid: $K_{sig}(x_i, x_j) = tanh(s(x_i \cdot x_j) + c)$

4.5.2.2 OC-SVM: The technique

The rules and properties mentioned in the section above apply also of this special type of SVM.

This time, the support vectors characterising the object's class are obtained only from the positive training examples. These support vectors actually don't define a separating hyperplane, but a separating hypersphere which encloses all possible representations of the training object. All observation which do not fit into this hypersphere are considered outliers/novelty. This is exactly what I needed to implement to try to detect if data from a different person that the expected one arrives.

Schölkopf et al. (1999) proposed a method of adapting the SVM model to the one-class classification problem. The main characteristic is that after transforming the feature via a chosen kernel, the origin is treated as the only member of the one class from the origin, and then, the techniques of the standard two-class SVM are used. This is what the Schölkopf Methodology assumes: Suppose that a dataset has a probability distribution P in the feature space. Find a "simple" subset S of the feature space such that the probability that a test point from P lies outside S is bounded by some a priori specified value. Supposing that there is a dataset drawn from an underlying probability distribution P, one needs to estimate a "simple" subset S of the input space such that the probability that a test point from P lies outside of S is bounded by some a prior specified $v \epsilon (0, 1)$. The solution for this problem is obtained by estimating a function f which is positive on S and negative on the complement S. In other words, Schölkopf et al., developed an algorithm which returns a function f that takes the value +1 in a "small" region capturing most of the data vectors, and -1 elsewhere. [13] This means that the algorithm can be summarised as mapping the data into a feature space using an appropriate kernel function and then, trying to separate the mapped vectors from the origin with maximum margin:

$$f(x) = \begin{cases} +1 \text{ if } x \,\epsilon\, S \\ -1 \text{ if } x \,\epsilon\, \overline{S} \end{cases} \tag{4.11}$$

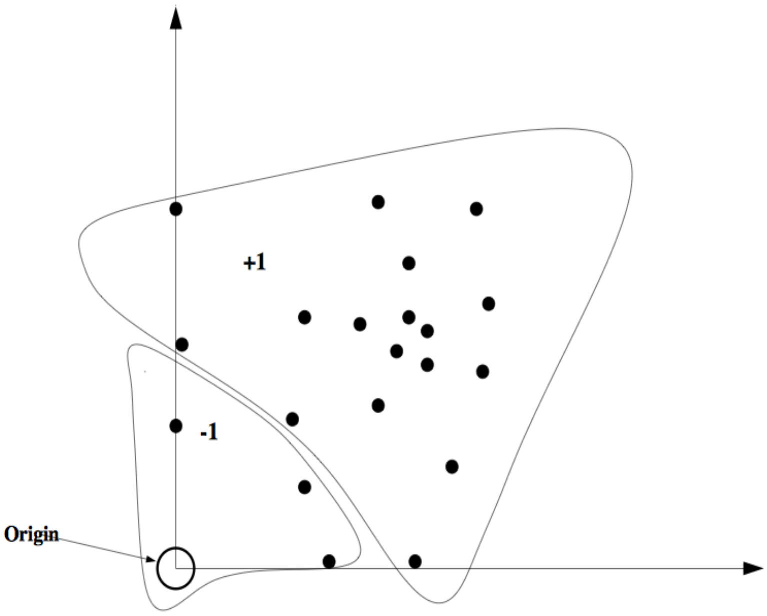
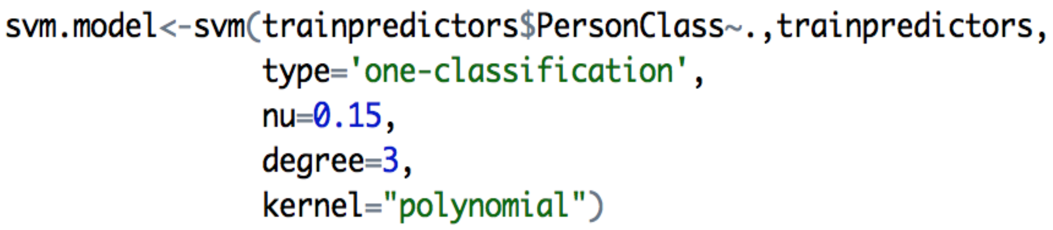


FIGURE 4.7: SVM: One Class

4.5.2.3 OC-SVM: The implementation: LibSVM in R

LibSVM is a popular open source machine learning library which implements the algorithm for kernelled support vector machines, supporting classification and regression.

In R, the package e1071 makes use of the LibSVM library to provide the *svm*() function which is used to train a support vector machine. Following is the R script I used with the *svm*() function and its parameters I needed, which I will explain in detail.



```
svm.model<-svm(trainpredictors$PersonClass~.,trainpredictors,
               type='one-classification',
               nu=0.15,
               degree=3,
               kernel="polynomial")
```

Figure 4.8: SVM: Implementation in R

The parameters which I am using represent the following:

1. trainpredictors$PersonClass : these are the labels for every row/component - in this case the label is always the same, as I am having only one reference class.

2. trainpredictors : these are all the rows/components of the matrix with the features which have to be learned by the SVM so that a suited model is created.

3. type='one-classification' : *svm*() can be used for different types of classification or regression or for novelty detection, which is what I am using by specifying the 'one-classification' type.

4. nu=0.15 : this is a mandatory parameter used for the type of classification I chose, which is controlling the tradeoff between a wide margin and the classifier error. "nu" is a number between 0 and 1, generally from .1 to .8, which controls the ratio of support vectors to data points. When "nu" is .1, the margin is small, the number of support vectors will be a small percentage of the number of data points. When "nu" is .8, the margin is very large and most of the points will fall in the margin. I chose a small number for the "nu" parameter, because the data I am classifying shows very small differences from a person to another, and that is why I don't want to have a large margin for the SV. This also means that it can have an impact on the training error rate, which will be a bit higher, but in this case, it will increase the chance that the testing error rate will be smaller.

5. kernel="polynomial" : after more calculations and analysing the types of kernels I could use, I chose the polynomial kernel as the best fit for the current problem.
6. degree=3: this represents the degree of the polynomial function used to define the kernel.

After the SVM model is built, I use the predict() function to see the prediction values for the training and for the testing data. The predict function needs the created model as a parameter and then the data to be classified.

4.5.2.4 Results and Interpretation

The training prediction results recorded with data from different data sets were between 80% to 84%. The testing prediction results have to be as small as possible if the testing data is from another person, and as close to a higher point if the testing data is from the same person. In the first testing case: when the testing data is form a different person than the person from the training data, the results varied between 40% to 60%. In the second case, when the testing data was from the was from the same person as the training data, the result also varied from 70% to 85%.

These results seem to be approximately good, but I would not say that they can be completely reliable, because there is a small difference between the prediction results which come when testing with the same or with a different person. As my aim of this work was to classify people using a small amount of testing data, it could happen, that the above implemented approach would sometimes fail.

My interpretation on why the OC-SVM does not get better results is that, although it creates the hypersphere only around the training data of a person, it is still complicated to decide if some new data which is actually coming from another person, is or is not from the same person, as it is not taking into consideration the patterns which can appear and which could show that although some data points are very close, they are not from the same person.

In the end, the result of the classifier can be considered good, but some further research is needed. Or maybe some other type of classifier, like the nearest neighbour for one-class classification would be a better fit. This is a topic for the future research which can be done in a next phase of this prototype project.

4.5.3 Approach 2: Algorithm Implementation using the Markov Model Principle

For the second approach, I chose not to use a ”classic” modified classifier. Instead I chose to use my own original mathematical approach, implementing an algorithm based on the Markov Model Principle.

4.5.3.1 The Markov Model: Generalities

The easiest way to treat sequential data would be simply to ignore the sequential aspects and treat the observations as independent and identically distributed (i.i.d.), corresponding to a simple graph without any links. Such an approach, however, would fail to exploit the sequential patterns in the data, such as correlations between observations that are close in the sequence. Suppose, for instance, that we observe a binary variable denoting whether on a particular day it rained or not. Given a time series of recent observations of this variable, we wish to predict whether it will rain on the next day. If we treat the data as i.i.d., then the only information we can glean from the data is the relative frequency of rainy days. However, we know in practice that the weather often exhibits trends that may last for several days. Observing whether or not it rains today is therefore of significant help in predicting if it will rain tomorrow. [12] This means that such an effect can be expressed in a probabilistic model. Without loss of generality, the product rule can be used to express the joint distribution for a sequence of observations:

$$p(x_1, ..., x_n) = \prod_{i=1}^{n} p(x_n|x_1, ..., x_{n-1}) \tag{4.12}$$

A markov chain is a process that undergoes transitions from one state to another on a state space. Markov chains have many applications as statistical models of real world processes: speech recognition, information sciences, Internet applications, economics and finance, genetics, games, mathematical biology etc.

This is how a slightly complex, but actually very simple Markov Chain System looks like:

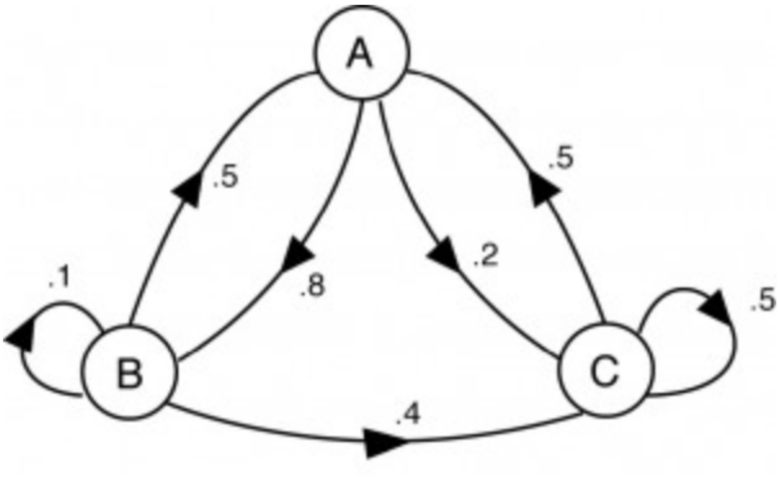


FIGURE 4.9: Markov Model: Markov Chain Example

The 3-state Markov System pictured in the figure above has multiple paths between each state. For example, the formula for calculating the probability of transitioning into state looks like this:

$$p(C) = Pr(C|A)Pr(A) + Pr(C|B)Pr(B) + Pr(C|C)Pr(C)$$

From the probabilities of transitioning from a state to another, we build the transition matrix A_{jk}, and then we can define the conditional distribution with the following formula:

$$p(z_n|z_{n-1}, A) = \prod_{k=1}^{K} \prod_{j=1} KA_{jk}^{z_{n-1,j}, z_{nk}} \text{ , where K = dimension of the variables} \tag{4.13}$$

4.5.3.2 The implementation: Algorithm in Java

Before I start explaining the algorithm I implemented, I will show the pseudocode for this algorithm and then the explanations referring to it. The pseudocode is presented in a simplified manner.

Algorithm pseudocode

```
$trainingAttributes \leftarrow extractAttributes(trainingData)$
$trainingProbabilities \leftarrow extractProbabilities(trainingAttributes)$
$testingAttributes \leftarrow extractAttributes(testingData)$
$testingProbabilities \leftarrow extractProbabilities(testingAttributes)$
for each testingProbability do
    $probabilitiesProportion \leftarrow calculateProbabilitiesProportion($
    $testProbability, trainProbability)$
    $differenceBetweenProbabilities \leftarrow calculateDifferenceBetweenProbabilities($
    $testProbability, trainProbability, probabilitiesProportion)$
    $similarityPercent \leftarrow calculateSimilarityPercent(getPercentages($
    $testProbability, trainProbability))$
    $inclusionPercent \leftarrow calculateInclusionPercent($
    $differenceBetweenProbabilities, similarityPercent)$
end for
$adjustedInclusionPercentList \leftarrow adjustInclusionPercent(inclusionPercentList)$
$percentResult \leftarrow calculatePercent(adjustedInclusionPercentList)$
```

The parameters: trainingData and testingData in the algorithm above represent the attributes for the features extracted and selected in the previous steps from the heart rate data sets.

I implemented the *extractAttributes*() function to be the basis for the implementation of the Markov model in this algorithm. So what it does is: for every data in the training/testing data set, along with the current component, it saves the next state, having the mean heart rate as the reference feature, but also saving all the other features which are needed in the further steps of the algorithm.

In the *extractProbabilities*() function, I calculate the probability with which for every mean heart rate of a component, a next value appears, by looking at the reference feature: the mean heart rate. So this means that after calling this function, I will have a Markov Chain System like the one in the figure 4.9, but much more complex, because the states in which a heart rate can go from one value to another can be in a wider range.

Then, I make a loop which goes through all the existing states in the Markov Chain System. For every testing component, first I calculate the proportion of the probabilities according to its corresponding training component. This means that for each training and testing component I am looking at how the probabilities for the next states are distributed.

In the next step of the loop, I calculate the difference between every testing probability state and its corresponding training probability state. This means that for every current testing and training state I am calculating the difference between every pair of probability states.

Next, I calculate the similarity percent. This value is determined by taking into consideration all the percentages of the corresponding pair testing and training values. Also to achieve a better result for this similarity value, now is the moment when I am also taking into consideration, for the pairs of training and testing probabilities, the values for all the features they have, and the differences between every pair consisting of a testing and a training feature.

The last step in the loop decides the inclusion percent. This means that according to the difference between the probabilities calculated before and similarity percent just determined, an inclusion percent which can be in the $[-100, 100]$ range is also determined and saved in a list where all the inclusion percentages are saved for each testing component.

After the looping through the components ends, the inclusion percentages are adjusted according to some specific parameters and thresholds.

Then, the last step is to find out the actual result I am looking for: the percentage which shows how much the testing data is like the training data, so, if the testing data is from the same person from which the training data is, or not. This value is calculated by determining the mean value of the adjusted inclusion percentages.

4.5.3.3 Results and Interpretation

The algorithm receives an input data set, called the training data set, from which it learns the Markov Chain System with all the features and percentages. After it has finished this learning phase, the algorithm can be tested with any data set from the same person or not from the same person.
When I test the algorithm with a data set with components which belong to the same person for which the algorithm was trained, the resulting percentages testing with different data sets are between 80% to even 100% recognition rate.
In the other case, when I am testing with a data set containing entries from another person, the expected result should be as small as possible, and it is. The result varies from 0% to 10%. So this means that we can be sure that is is another person.

These results are very good. The difference between the result for the recognition of the same person and the result for the recognition of a different person is high enough so that we can trust the result to be reliable.

The algorithm I implemented works very good with this kind of data because it learns the features of the heart rate signal and the probabilities of the states in which the heart rate can be found along with its specific features. This means that although the data from two different people seems to be very similar, like the charts in figure 4.3 show us, if the important patterns are learned, these can make the difference between those similar charts.

4.5.4 Comparing the results

The table below gives a better overview of the results received using the two approaches.

Person Data Set Implementation	Same	Other
OC-SVM	70%-85%	40%-60%
Alg. using Markov Chains	80%-100%	0%-10%

Chapter 5

Conclusion

Now that I have presented the concept and the architecture of this system along with its implementation and the technologies used, the advantages presented in the introduction are obvious.

The main focus of this work was to present a prototype of a system which can help the modern societies in a lot of ways.

On the one hand, the main part of the project tackles the subject about how the system can be implemented so that it is suited for the insurance companies, as well as for the insured people, or simply for some people which want to be up to date with their heath development and their health habits.

On the other hand, the other subject of a high importance is the analytics subject which is approached in this work. The analytics problem this project deals with, classifying people by their heart rate, is trying to stop, or at least minimise the impact of fraud attempts when it comes to implementing the system in the context of an insurance company.

5.1 Insurance and Insured

The insurance company has a new way of ”being closer” to the insured users they have. Also the insurance company can use this system to create better personalised care plans for every user, because everyone has his/her specific needs and wishes. With this new

system, the insurance company can also implement a bonus program described in more detail in the introduction of this work. The bonus program can also be personalised according to the doctor's suggestions and orders for every patient.
The bonus program could be, of course, defrauded if the insured person wants to do that and sends inaccurate or false data through the system. This is a problem which is solved using analytics.

5.2 Research

Analytics takes care of the arising fraud problem in this system in the context of the insurance company.

Regarding the approach to classify a person by his heart rate data, the literature does not mention exactly such a way of person classification.
This is an original way of solving this classification problem, which produces very good results.

5.3 Future outlook

As this project tackles a lot of important problems, it has a high extension potential as well.
In the future, an extension domain represents the use cases for every person which wants to have or achieve a more healthy lifestyle.
Another further development area, is regarding the insurance company. An insurance company can come with new ideas for new features to implement. For example, when other types of data is retrieved from the user, health care plans can be further improved.
Research is of course one of the main areas in which this project can be further taken for achieving new interesting results.

5.3.1 Insurance and Insured

As I said also in the implementation chapter, the GUI I implemented is just a small prototype. Any type of GUI, also a mobile one can be added, because the REST web service allows this. The features for the end user can be extended. For example, if a user wants to set specific goals for a period of time, or if his doctor prescribed him a specific routine he/she can follow the updates in the app so that he/she can better and easier achieve these goals.

If the user agrees, the insurance can also use this information for improving the services they provide for their insured people. This is a high help also for the insurance company, because if a person is healthy, it is much easier to maintain this status, and prevent possible problems, than to treat an ill person.

5.3.2 Research

The data from a persons vitals and activity bring a huge amount of information about someone.
Anomalies and health problems can be detected at an earlier stage. An earlier detection of a health related problem can significantly help its treatment process.
Also, prevention is another subject which can be tackled in this area.

Each year, millions of people die preventable deaths. A 2004 study showed that about half of all deaths in the United States in 2000 were due to preventable behaviours and exposures. Leading causes included cardiovascular disease, chronic respiratory disease, unintentional injuries, diabetes, and certain infectious diseases. This same study estimates that 400,000 people die each year in the United States due to poor diet and a sedentary lifestyle. According to estimates made by the World Health Organisation (WHO), about 55 million people died worldwide in 2011, two thirds of this group from non-communicable diseases, including cancer, diabetes, and chronic cardiovascular and lung diseases. This is an increase from the year 2000, during which 60% of deaths were attributed to these diseases. Preventive healthcare is especially important given the worldwide rise in prevalence of chronic diseases and deaths from these diseases. [21]
This shows that the prevention stage needs a lot of research and a lot of help, a part which can be also provided by further developing this area of the project.

Bibliography

[1] *Bluetooth, Wikipedia*

[2] *Bluetooth Low Energy, Wikipedia*

[3] *Bluetooth Smart Technology: Powering the Internet of Things, Bluetooth*

[4] *Health Kit, Apple Developer*

[5] Alexa Huth, James Cebula, *The Basics of Cloud Computing*

[6] J. Chris Anderson, Jan Lehnardt, Noah Slater, *CouchDB: The Definitive Guide*

[7] *Analytics, Wikipedia*

[8] *Heart Rate, Wikipedia*

[9] *Heart Rate Variability, Aha Journals*

[10] *Standard Deviation and Variance, Math is Fun*

[11] *Standard Density Distribution, Federation of Cardiology*

[12] Christopher M. Bishop, *Pattern Recognition and Machine Learning*

[13] Bernhard Schölkopf, Robert Williamson, Alex Smola, John Shawe-Taylor, John Platt, *Support Vector Method for Novelty Detection*

[14] E. Hon, S. Lee, *Electronic evaluation of the fetal heart rate. VIII. Patterns preceding fetal death, further observations*

[15] Task Force of the European Society of Cardiology and the North American Society of Pacing and Electrophysiology, *Heart rate variability: standards of measurement, physiological interpretation and clinical use*

[16] G. Casolo, E. Balli, T. Taddei, J. Amuhasi and C. Gori, *Decreased spontaneous heart rate variability in congestive heart failure*

[17] R. A. Thuraisingham, *reprocessing RR interval time series for heart rate variability analysis and estimates of standard deviation of RR intervals.*

[18] A. E. Aubert, D. Ramaekers, F. Beckers, R. Breem, C. Denef, and F. Van deWerf *The analysis of heart rate variability in unrestrained rats. Validation ofmethod and results*

[19] *Cubic Spline Interploation, Utah - Physics*

[20] *Heart Rate Variability, Wikipedia*

[21] *Preventive Healthcare, Wikipedia*