FACULTEIT ECONOMIE EN BEDRIJFSKUNDE

# CLASSIFYING RECORDED HEART SOUNDS

## A DATA MINING CASE STUDY.

Aantal woorden: < 00.000 >

Jasper Bracke
Stamnummer : 000131075894

Promotor: Prof. Dr. Els Clarysse

Masterproef voorgedragen tot het bekomen van de graad van:

Master in de handelswetenschappen: management en informatica

Academiejaar: 2018-2019

UNIVERSITEIT GENT

**VERTROUWELIJKHEIDSCLAUSULE/ CONFIDENTIALITY AGREEMENT**

**PERMISSION**

Ondergetekende verklaart dat de inhoud van deze masterproef mag geraadpleegd en/of gereproduceerd worden, mits bronvermelding.
I declare that the content of this Master's Dissertation may be consulted and/or reproduced, provided that the source is referenced.

Naam student/name student :……

Handtekening/signature

NEDERLANDSTALIGE SAMENVATTING

Deze thesis is een casestudie met als onderwerp classificatie van opgenomen hartgeluiden, deze kunnen normaal of abnormaal zijn.

In de introductie worden drie actueel problemen aangereikt waar machine learning classificatie een oplossingen kan bieden. Deze zijn: de training van medici, screening van hart en vaatziekten en het monitoren van hartslagen bij foetussen. Vervolgens wordt ook de gebruikte hartslag databases al kort vermeldt, deze komt voort uit de PhysioNet Challenge. Het doel van deze uitdaging is het creëer van een robuust classificatiemodel.

De literatuurstudie begin met een kleine samenvatting welke literatuur inzake classificatie aan bod zal komen. Een eerste deel van de literatuur gaat over het hart, de fases, geluiden en hun frequenties. Vervolgens worden de methodes om geluidssignalen te verwerken besproken. Het deel segmentatie behandelt 4 verschillende technieken om hartopnames op te delen in de fases van een hartcyclus. Dit deel wordt opgevolgd door de effectieve classificatiemethodes en hun performantie op de PhysioNet database. Hier worden ook optimalisatie algoritmes en een methode om de scheve verdeling van abnormale tegenover normale hartslagen in de database recht te trekken (oversampling). De literatuurstudie wordt afgesloten met software en bibliotheken die gebruikt worden in het datamining proces.

in het gedeelte case study. Wordt met behulp van de literatuur een onderzoeksvraag geponeerd Deze luidt als volgt: Leidt het combineren van oversampling met spectrale kenmerken anders dan lineaire predictieve coderingscoëficienten tot een beter classificatiemodel? De relevantie van deze onderzoeksvraag wordt ook kort besproken.

De methodologie beschrijft het crisp-dm raamwerk waarmee dit dataminingprobleem werd aangepakt. 1. Business understanding omvat het doel, de problemen en slaagcriteria voor het te creëren model. 2. Data understanding onderzoekt de database, en meer specifiek de lengte van de opnames en het aantal hartslagen in de database. 3. Data preparation beschrijft het opschonen van de data, de verkregen kenmerken en de finaal gekozen kenmerken. Er werden 3 groepen van kenmerken gecreëerd. Een groep met als basis het Springer segmentatie algoritme. Een tweede groep van algemene spectrale eigenschappen voor elke opname. En finaal ook een clusteringmethode die gebruikt word elke bestand te beschrijven als een combinatie van verschillende hartslagtypes. 4. Modelling behandeld de verkenning van modellen, hun ontwerp en creatie. De getrainde modellen zijn rotation forest, J48 en SMO. Deze werden getraind op verschillende “oversampled” datasets. En vervolgens ook geoptimaliseerd door de combinatie met andere algoritmes. 5. Evaluation, deze sectie bevat de evaluatiemethode van de uitdaging zoals gegeven door PhysioNet 6. Deployment presenteert de performantie van de getrainde modellen op een aparte validatie dataset.

De resultaten koppelen terug naar de gecreëerde modellen, oversampling en optimalisatie van de modellen. Deze komen ook terug in de discussie, daar wordt worden ook de spectrale features. Dit besproken. De modellen worden vergeleken met eerdere oplossingen van het probleem uit de literatuur. Hier komen ook de limitatie van het onderzoek aan bod. De conclusie beantwoord de onderzoeksvraag. Het combineren van oversampling met spectrale kenmerken anders dan lineaire predictieve coderingscoëficienten leidt tot een beter classificatie model. Het ontworpen model “Smote 100 AdaBoost J48” presteert beter dan alle voorgangers uit de literatuur. De relevantie hiervan en de aanbeveling voor toekomstig onderzoek worden ook nog uitgediept.

# PREFACE

The choice for this dissertation originated in my passion for data science. I want to thank my promotor: Prof. Dr. Els Clarysse for lighting this spark during the business intelligence lectures. I'm grateful to have been able to work on a thesis that requires a more hands on approach. Although it has to be said that this thesis was a real challenge. I might have chosen a topic that was a bit too far out of the scope of the It-management master, and my own comfort zone. The signal processing proved to be more difficult than expected. Nevertheless I achieved promising results, in the end it was a great learning experience.

"If you *torture* the *data long enough*, it will *confess*." — Ronald H. Coase

# ABSTRACT

This data mining case study covers the “Classification of Normal/Abnormal Heart Sound Recordings: the PhysioNet/Computing in Cardiology Challenge 2016”. From the literature study the research question *“Does combining oversampling with spectral features other than Linear Predictive Coding coefficients lead to a better classifier?” arose. The Crisp-DM framework was used to tackle this classification problem. The main proportion of this dissertation was extraction good features out of the audio files. In the end three feature groups where extracted. The first one being the Springer segmentation algorithm features provide by the challenge. The second group consists out of spectral features like MFCCs calculated on each recording. The third classifying feature was achieved by splitting the sound files into beats using an onset method. These beats where clustered to 128 beat types. That way the recordings could be defined as the combination different beat types. This features proved to be the best classifier. Three base algorithms where chosen (Random Forest, J48 and SMO) and combined with different boosting and bagging algorithm. These where trained on SMOTE adjusted datasets at different oversampling rates. After deploying these models on the Validation set provided by the challenge. One model was picked based on its accuracy and balanced sensitivity and specificity. The model the C4.5 algorithm (J48 in Weka), combined with AdaBoost and trained on a dataset oversampled at a SMOTE value of 100 percent.*

*Keywords:* PhysioNet, Cardiology, PCG, Classification, MFCC, Onset, SMOTE, AdaBoost, J48

# TABLE OF CONTENTS

## Contents

PREFACE ..... I
ABSTRACT ..... II
TABLE OF CONTENTS ..... III
LIST OF ABBREVIATIONS ..... VI
LIST OF FIGURES ..... VII
LIST OF TABLES ..... VIII
1 Introduction ..... 1
2 Literature review ..... 2
2.1 About heart sound classification ..... 2
2.2 The Anatomy of the heart ..... 2
2.3 The Cardiac Cycle ..... 4
2.3.1 First heart sound (S1) ..... 4
2.3.2 Systolic phase ..... 4
2.3.3 Second heart sound (S2) ..... 5
2.3.4 Diastolic phase ..... 5
2.4 Abnormal heart sounds ..... 6
2.4.1 Aortic stenosis ..... 6
2.4.2 Aortic regurgitation ..... 6
2.4.3 Pulmonary regurgitation ..... 6
2.4.4 Pulmonary stenosis ..... 6
2.4.5 Atrial septal defect ..... 6
2.4.6 Ventricular septal defect ..... 6
2.4.7 Mitral regurgitation ..... 6
2.4.8 Tricuspid Regurgitation ..... 7
2.4.9 Mitral Stenosis ..... 7
2.4.10 Tricuspid stenosis ..... 7
2.4.11 Mitral valve prolapse ..... 7
2.4.12 Patent ductus arteriosus ..... 7
2.4.13 Flow murmur ..... 7
2.5 Heart sound Frequencies ..... 8
2.6 Signal processing ..... 9
2.6.1 Nyquist frequency ..... 9
2.6.2 Butterworth band pass ..... 9

2.6.3 Energy and RMSE ..... 9
2.6.4 Zero Crossing Rate ..... 9
2.6.5 Fast Fourier Transform (FFT) ..... 9
2.6.6 Short-time Fourier Transform (STFT) ..... 10
2.6.7 Constant-Q Transform ..... 10
2.6.8 Chroma ..... 10
2.6.9 Spectral centroid ..... 10
2.6.10 Onset Detection ..... 10
2.6.11 Mel-Frequency Cepstral Coefficients ..... 10
2.6.12 Linear Prediction Coding Coefficients ..... 10
2.7 Heart sound segmentation ..... 11
2.7.1 Envelope-based methods ..... 11
2.7.2 Feature-based methods ..... 12
2.7.3 Machine-learning methods ..... 12
2.7.4 Hidden Markov model methods ..... 12
2.8 Heart sound classification ..... 14
2.8.1 Neural networks (NN) ..... 14
2.8.2 Support vector machines (SVM) ..... 14
2.8.3 K-Nearest neighbours (k-NN) ..... 14
2.8.4 Classification trees ..... 14
2.8.5 Boosting algorithms ..... 14
2.9 Software and libraries ..... 16
2.9.1 Jupyter notebook ..... 16
2.9.2 Librosa ..... 16
2.9.3 Scpypi ..... 16
2.9.4 Numpy ..... 16
2.9.5 Matplotlib ..... 16
2.9.6 Pandas ..... 16
2.9.7 Scikit-learn ..... 16
2.9.8 Matlab ..... 16
2.9.9 Weka ..... 16
3 Case Study ..... 17
3.1 Research question ..... 17
3.2 Significance ..... 17
3.3 Methodology ..... 18
3.3.1 Crisp DM ..... 18

3.4 Business understanding ..... 20
3.5 Data understanding ..... 21
3.5.1 The dataset ..... 21
3.5.2 Sound duration ..... 21
3.5.3 Number of beats ..... 22
3.6 Data Preparation ..... 23
3.6.1 Data cleaning ..... 23
3.6.2 Feature extraction. ..... 24
3.6.3 Feature selection ..... 27
3.7 Data modelling ..... 28
3.7.1 Exploration ..... 28
3.7.2 Development ..... 28
3.7.3 Built models ..... 30
3.8 Evaluation ..... 31
3.9 Deployment ..... 32
4 Results ..... 33
4.1 Base algorithms ..... 33
4.2 Oversampling ..... 34
4.3 Boosting ..... 34
5 Discussion ..... 35
5.1 Findings ..... 35
5.1.1 Spectral features ..... 35
5.1.2 Base algorithms ..... 35
5.1.3 Oversampling ..... 35
5.1.4 Boosting ..... 35
5.2 Comparison to other models ..... 36
5.3 Limitations ..... 36
6 Conclusion ..... 37
REFERENCES ..... 38

# LIST OF ABBREVIATIONS

| | |
|---|---|
| AHA | American Heart Association |
| AV | Atrioventricular valves |
| CRISP- DM | Cross-industry standard process for data mining |
| CVD | Cardiovascular disease |
| ECG | Echocardiogram |
| FFT | Fast Fourier Transform |
| Hz | Hertz |
| k-NN | k-nearest neighbours |
| LPCC | Linear Prediction Coding Coefficients |
| MFCC | Mel frequency cepstral coefficient |
| NN | Neural network |
| PCG | Phonocardiogram |
| RF | Random Forest |
| RMSE | Root Mean Squared of the Energy |
| S1 | First heart sound |
| S2 | Second heart sound |
| SEE | Shannon energy envelope |
| Se | Sensitivity |
| SMOTE | Synthetic Minority Over-sampling Technique |
| Sp | Specificity |
| STFT | Short-time Fourier Transform |
| SVM | Support Vector Machine |

# LIST OF FIGURES

FIGURE 1 THE CIRCULATORY SYSTEM (OPENSTAX, 2019) ..... 3
FIGURE 2 THE CARDIOVASCULAR CYCLE AS PCG AND ECG (SPRINGER, TARASSENKO, & CLIFFORD, 2016) ..... 4
FIGURE 3 CARDIOVASCULAR CYCLE (OPENSTAX, 2019) ..... 4
FIGURE 4 BAND PASS FILTER (ANALOGICTIPS, 2017) ..... 9
FIGURE 5 HILBERT ENVELOPE (MATHWORKS, 2019A) ..... 11
FIGURE 6 RMS ENVELOPE (MATHWORKS, 2019C) ..... 12
FIGURE 7 CRISP DM ..... 18
FIGURE 8, FILE DURATION IN SECONDS ..... 22
FIGURE 9, A0001.WAV SPLIT INTO BEATS ..... 22
FIGURE 10 FEATURE EXTRACTION PROCESS ..... 24
FIGURE 11 BOXPLOT OF ABSOLUTE DIFFERENCE BETWEEN SE AND SP ..... 34

# LIST OF TABLES

TABLE 1 FRECUENCY RANGE OF HEART RATE SOUNDS (THIYAGARAJA ET AL., 2018) ..... 8
TABLE 2, TRANING SET CLASSIFICATION ..... 21
TABLE 3, NUMBER OF BEATS ..... 22
TABLE 4 GAIN RATIO OF FEATURE SETS VS. F-SCORE ..... 27
TABLE 5 LOWEST RANKED FEATURES ON NORMAL VS. SMOTE 301 FEATURE SET ..... 27
TABLE 6 SMOTE VALUES WITH J48 ..... 28
TABLE 7 SMOTE VALUES WITH RF ..... 29
TABLE 8 BUILT MODELS ..... 30
TABLE 9 PHYSIONET SCORE REFERENCE LABELS (PHYSIONET, 2016) ..... 31
TABLE 10 DEPLOYED MODELS ..... 32
TABLE 11 TOP 15 MODELS ..... 33
TABLE 13 TOP EIGHT ENTRIES PHYSIONET CHALLENGE (G. D. CLIFFORD ET AL., 2016) ..... 36

# 1 Introduction

*"Claims that the stethoscope is dead are entirely false. In fact, with its new digital capabilities, the stethoscope is healthier than ever"* (Fuster, 2016).

MD, PhD Fuster is the editor-in-chief in chief of the *Journal of the American College of Cardiology*, which rakend third in the category cardiac and cardiovascular systems in 2017. MD, PhD Fuster pleads for the utilisation of the stethoscope in combination with new technologies. The Stethoscope is a tool for acoustic auscultation. A recording of the auscultation of the circulatory system is called a phonocardiogram (PCG). Unfortunately medical professionals seem to lose the skill to auscultate the heart correctly (Vukanovic-Criley et al., 2010). This is a major issue. Cardiovascular diseases (CVDs) are the main cause of death in the world, resulting in more than 17 million fatalities in 2016 alone (WHO, 2019a). According to the World Health Organisation this number will rise to 22.2 million by 2030 (WHO, 2019b).This is a precarious situation. They propose a global strategy to overcome this lethal trend. An important element of their strategy is the cost effectiveness of heart screenings. In the Western countries Electrocardiograms (ECG) are a common and effective method to evaluate heart sounds, they are based on an electrical signal rather than the acoustical signal of the PCG. In developing countries this technique is not as widespread. The development of a stethoscope extension for smartphones is a cost effective solution to this problem (Bhaskar, 2012; Thiyagaraja et al., 2018). The use of PCG's is also in further development to detect foetal heart rates since it is a non-invasive monitoring method (Cesarelli, Ruffo, Romano, & Bifulco, 2012). All these issues can be addressed by machine learning algorithms. These can be used to train medical professionals (Legget et al., 2018), aid in the screening of CVDs (Gari D. Clifford et al., 2017) and are already implemented in the detection of foetal heart rates. To encourage the development of machine learning models on PCG's PhysioNet released the "open access database for the evaluation of heart sound algorithms "in 2016 (Liu et al., 2016). This database is the first large open access heart sound recordings collection. The goal of this dissertation is to train a machine learning model on this dataset capable of distinguishing abnormal from normal heart sounds.

# 2 Literature review

## 2.1 About heart sound classification

In data science there are two categories of techniques: supervised and unsupervised learning. In supervised learning the goal is to get an output that is be labelled. E.g. in this dissertation, the input information for each heart sound recording is used in a function that returns a label that is ether normal or abnormal. This is in essence classification, the recording belongs to the class “normal” or “abnormal”. Unsupervised learning on the other hand is used to group or associate multiple inputs based on their similarities. The output is not a label or a value, it is a collection of multiple inputs that are similar. In order to build a good classification model it is important to understand what the input is and how it corresponds to the output. Hence the first step of this literature review focusses on the cardiac cycles , it’s phases and sounds. This section is follow by signal processing techniques that can be used to turn the audio signals in useful information. This knowledge can be leveraged in the next step, the segmentation. Extracting the phases out of the signal is a strenuous task. Four types of segmentation are discussed in this section. Once the phases are determined features can be created, also by using signal processing. Based on these features the actual classification can be done. The available literature on the different classification models and their performance is discussed extensively under point 2.8. The final part of this literature review contains an overview of software and libraries used in data science.

Before kicking of the actual literature review it is important to discuss the significance and impact of the open access database for the evaluation of heart sound algorithms. Before the release of this database there was already a lot of literature concerning segmentation and classification of heart sounds. Nevertheless is was hard to understand the value of new contributions to this field. Often the techniques where tested on small or private or databases. Most of these databases consisted solely out of healthy patients. The open database is a bliss, it is both large with 4,430 recordings and contains normal as well as abnormal heart sound recordings.

## 2.2 The Anatomy of the heart

Anatomy and Psychology (OpenStax, 2019) is a textbook used to teach majors at Rice University. The book is frequently updated and freely available online as a courtesy by the Bill and Melinda Gates Foundation. Unit 4: Fluids and transport introduces us to the anatomy of the heart. The human heart consists of four chambers and two linked circulation circuits.

The chambers are the left atrium, left ventricle, right atrium and right ventricle. The atriums receive blood, contract and push blood to the ventricle. The ventricle provides blood to the lungs and the rest of the body.

The two circuits are the pulmonary and the systemic circuits. The pulmonary circuit regulates the transport to and from the lungs. There the blood gets oxygen and the carbon dioxide is extracted to be exhaled. The oxygenated blood is then transported by the systemic circuit. Almost all body tissues get this oxygen rich blood. The body tissues (partially) deplete the oxygen and produce carbon dioxide as a waste product. The deoxygenated blood with carbon dioxide is then transported back to the heart by the systemic circuit. Finally the heart sends it back to the pulmonary system.

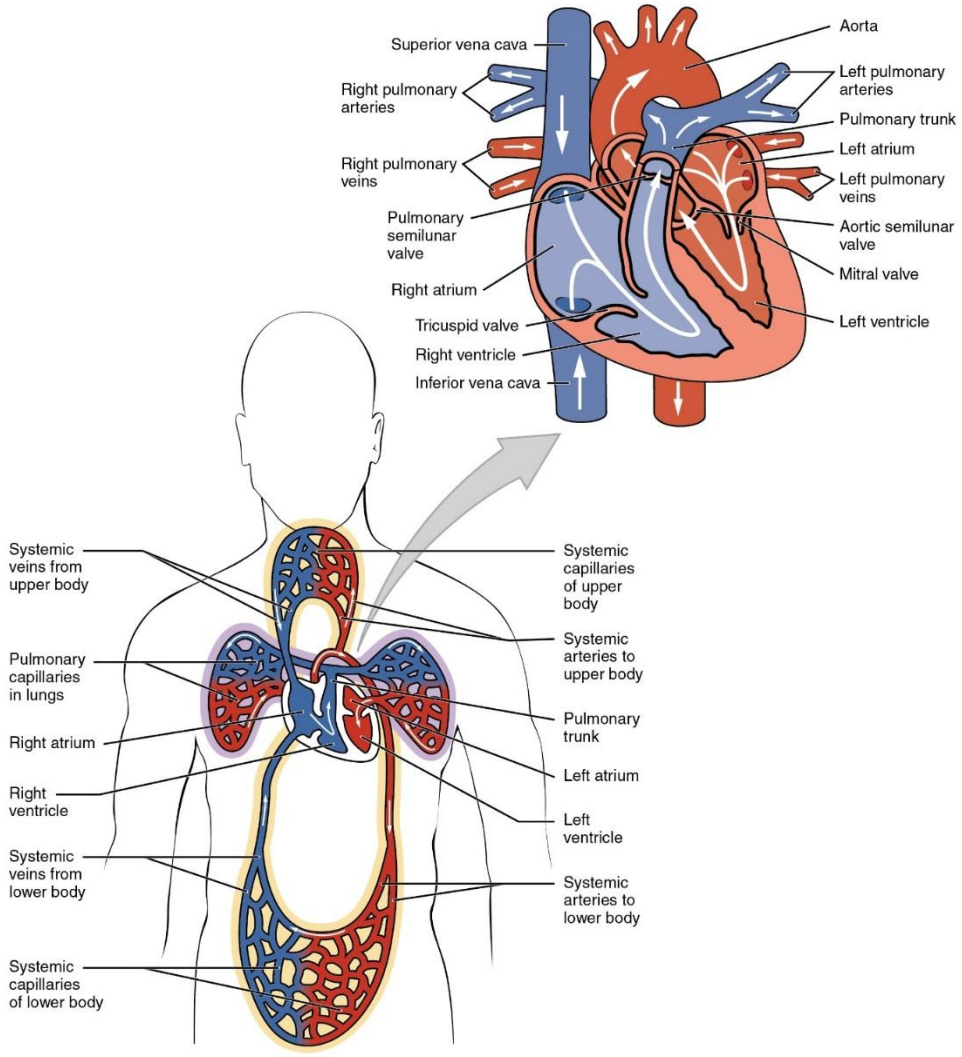


*Figure 1 the circulatory system (OpenStax, 2019)*

## 2.3 The Cardiac Cycle

Chapter 19, “The Cardiovascular System: The Heart” (OpenStax, 2019) describes the cardiac cycle. A normal cycle consists out of two audible beats and is subdivided in the following four phases: S1, Systole, S2 and Diastole.

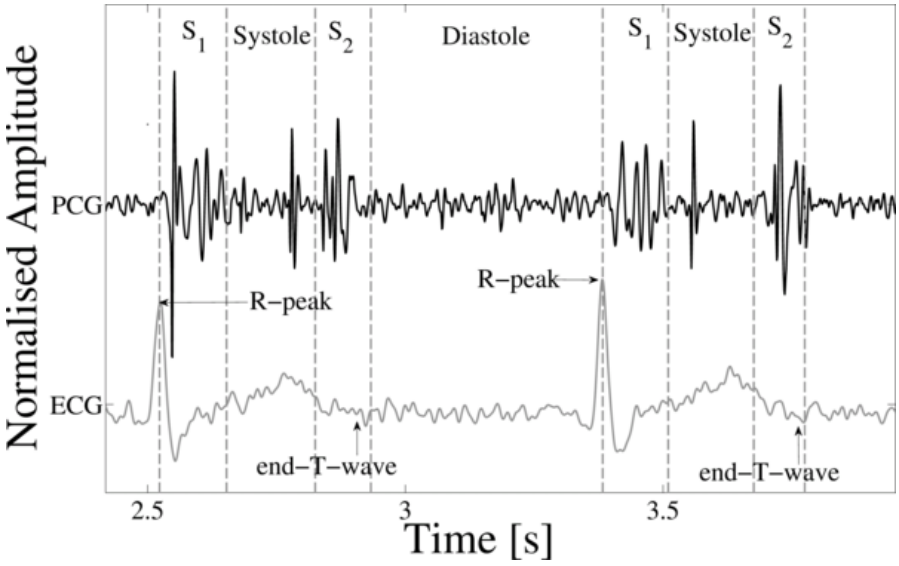


*Figure 2 the cardiovascular cycle as PCG and ECG (Springer, Tarassenko, & Clifford, 2016)*

### 2.3.1 First heart sound (S1)

The first beat of a normal cardiac cycle consists is referred to as S1. It is often called the “Lub” sound. The sound is created by the closing of the atrioventricular (AV) valves.

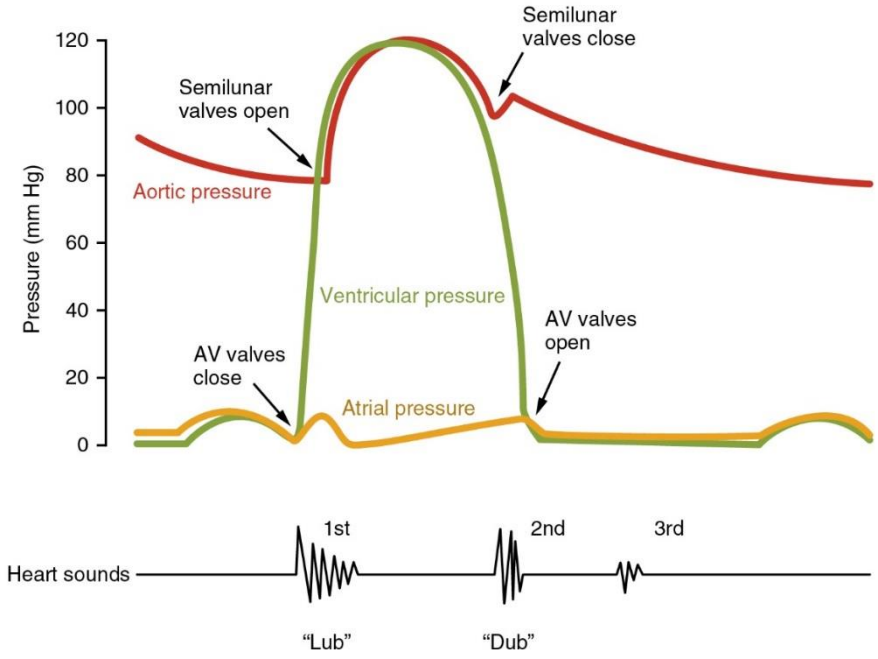


*Figure 3 cardiovascular cycle (OpenStax, 2019)*

### 2.3.2 Systolic phase

The phase after the first beat is called systole. During this phase the AV stays closed and the Semilunar valves open. During this phase the heart muscle contracts and the blood flows out.

### 2.3.3 Second heart sound (S2)

After the systolic phase the seconds heart beats occurs (S2), this is the “dub” sound. The S2 sound is created by the closing of the Semilunar valve and is followed by the diastolic phase.

### 2.3.4 Diastolic phase

During the Diastolic phase the Semilunar valves is closed and the aortic valve opens. The heart relaxes and blood flow in. After the diastolic phase one cycle is finished.

## 2.4 Abnormal heart sounds

The American Heart Association is a non-profit organisation, founded in 1924. They fund academic research and provide education and care to battle cardiovascular disease. Every month they publish "The Journal of the American Heart Association" which is open access and peer reviewed. In 2017 it ranked 36th out of the 128 journals in the Cardiac and cardiovascular systems category with a journal impact factor of 4.450. On their website heart.org they provide clear and comprehensive definitions of different heart problems.

### 2.4.1 Aortic stenosis

*"Aortic stenosis is one of the most common and most serious valve disease problems. Aortic stenosis is a narrowing of the aortic valve opening. Aortic stenosis restricts the blood flow from the left ventricle to the aorta and may also affect the pressure in the left atrium (AHA, 2019e)."*

### 2.4.2 Aortic regurgitation

*"Aortic regurgitation is leakage of the aortic valve each time the left ventricle relaxes. A leaking (or regurgitant) aortic valve allows blood to flow in two directions. Oxygen-rich blood either flows out through the aorta to the body — as it should — but some flows backwards from the aorta into the left ventricle when the ventricle relaxes (AHA, 2019d)."*

### 2.4.3 Pulmonary regurgitation

*"Pulmonary regurgitation (PR, also called pulmonic regurgitation) is a leaky pulmonary valve. This valve helps control the flow of blood passing from the heart to the lungs. A leaky pulmonary valve allows blood to flow back into the heart chamber before it gets to the lungs for oxygen (AHA, 2019i)."*

### 2.4.4 Pulmonary stenosis

*"Pulmonary stenosis is a condition caused by a narrowing of the pulmonary valve opening. Pulmonary stenosis restricts blood flow from the lower right chamber (called the ventricle) to the pulmonary arteries, which delivers blood to the lungs. It is most commonly the result of a congenital heart defect. However, rarely PS can develop as a result of infections like rheumatic fever or carcinoid syndrome (AHA, 2019j)."*

### 2.4.5 Atrial septal defect

*"A "hole" in the wall that separates the top two chambers of the heart. This defect allows oxygen-rich blood to leak into the oxygen-poor blood chambers in the heart. ASD is a defect in the septum between the heart's two upper chambers (atria). The septum is a wall that separates the heart's left and right sides (AHA, 2019a)."*

### 2.4.6 Ventricular septal defect

*"VSD is a hole in the wall separating the two lower chambers of the heart. In normal development, the wall between the chambers closes before the foetus is born, so that by birth, oxygen-rich blood is kept from mixing with the oxygen-poor blood. When the hole does not close, it may cause higher pressure in the heart or reduced oxygen to the body (AHA, 2019m)."*

### 2.4.7 Mitral regurgitation

*"Mitral regurgitation is leakage of blood backward through the mitral valve each time the left ventricle contracts. Watch an animation of mitral valve regurgitation. A leaking mitral valve allows blood to flow in two directions during the contraction. Some blood flows from the ventricle through the aortic valve — as it should — and some blood flows back into the atrium (AHA, 2019g)."*

### 2.4.8 Tricuspid Regurgitation

*"Tricuspid regurgitation is leakage of blood backwards through the tricuspid valve each time the right ventricle contracts As the right ventricle contracts to pump blood forward to the lungs, some blood leaks backward into the right atrium, increasing the volume of blood in the atrium. As a result, the right atrium can enlarge, which can change the pressure in the nearby chambers and blood vessels (AHA, 2019k)."*

### 2.4.9 Mitral Stenosis

*"Mitral stenosis is a narrowing of the mitral valve opening. Mitral stenosis restricts blood flow from the left atrium to the left ventricle (AHA, 2019h)."*

### 2.4.10 Tricuspid stenosis

*"Tricuspid stenosis is a narrowing of the tricuspid valve opening. Tricuspid stenosis restricts blood flow between the upper and lower part of the right side of the heart, or from the right atrium to the right ventricle (AHA, 2019l)."*

### 2.4.11 Mitral valve prolapse

*"Mitral valve prolapse is a condition in which the two valve flaps of the mitral valve do not close smoothly or evenly, but instead bulge (prolapse) upward into the left atrium. Mitral valve prolapse is also known as click-murmur syndrome, Barlow's syndrome or floppy valve syndrome (AHA, 2019f)."*

### 2.4.12 Patent ductus arteriosus

*"An unclosed hole in the aorta. Before a baby is born, the foetus's blood does not need to go to the lungs to get oxygenated. The ductus arteriosus is a hole that allows the blood to skip the circulation to the lungs. However, when the baby is born, the blood must receive oxygen in the lungs and this hole is supposed to close. If the ductus arteriosus is still open (or patent) the blood may skip this necessary step of circulation. The open hole is called the patent ductus arteriosus (AHA, 2019c)."*

### 2.4.13 Flow murmur

*"Not every murmur is associated with valve disease. Murmurs can also be caused by conditions that may temporarily increase blood flow such as pregnancy, fever, hyperthyroidism, anaemia and rapid growth spurts in children (AHA, 2019b)."*

## 2.5 Heart sound Frequencies

In “A novel heart-mobile interface for detection and classification of heart sound” (Thiyagaraja et al., 2018) the frequency range for the normal and abnormal heart sounds (Ut supra) was calculated. This was achieved by using Fast Fourier Transform and Short Time Fourier Transform (Ut infra). The lowest minimum Frequency is 45 Hz for mitral stenosis, aortic stenosis has the highest maximum frequency, 450 Hz. Other researchers keep the maximum frequency on 400 Hz, (Kudriavtsev, Polyshchuk, & Roy, 2007) and (Schmidt, Holst-Hansen, Graff, Toft, & Struijk, 2010).

| Cardiac Signal | Minimum Frequency (Hz) | Maximum Frequency (Hz) |
|---|---|---|
| First heart sound | 100 | 200 |
| Second heart sound | 50 | 250 |
| Aortic regurgitation | 60 | 380 |
| Pulmonary regurgitation | 90 | 150 |
| Aortic stenosis | 100 | 450 |
| Pulmonary stenosis | 150 | 400 |
| Atrial septal defect | 60 | 200 |
| Ventricular septal defect | 50 | 180 |
| Mitral regurgitation | 60 | 400 |
| Tricuspid Regurgitation | 90 | 400 |
| Mitral Stenosis | 45 | 90 |
| Tricuspid stenosis | 90 | 400 |
| Mitral valve prolapse | 45 | 90 |
| Patent ductus arteriosus | 90 | 140 |
| Flow murmur | 85 | 300 |

*Table 1 Frecuency range of heart rate sounds (Thiyagaraja et al., 2018)*

## 2.6 Signal processing

Signal processing analyses, synthesis and modifies signals. For this datamining case study the signal is audio. Resampling and filtering a signal results in faster computation times but at the cost of loss of data. It is important to understand some the signal processing techniques and concepts, in order to ensure that no value data is lost.

### 2.6.1 Nyquist frequency

When sampling a signal at a certain sampling rate the data in between two sample points is lost. So choosing the right sampling rate is a crucial task. To reliably reconstruct a signal the sampling rate has to larger than twice the highest frequency in the signal (Stiltz, 1961). This frequency is called the folding frequency of a sampling system or the Nyquist frequency. The book "*Digital signal processing using MATLAB for students and researchers"* by leis provides - aside of an excelent explanation –a more practical example of the nyquiqt Theorem. *"The fundamental theorem of sampling states that we must sample at least twice as fast as the highest frequency component that we want to maintain after sampling. In practical terms, the sampling rate is chosen to be higher than this two - times rule above would suggest. For example, CD audio systems are designed to handle frequencies up to approximately 20 kHz. The sample rate is 44.1 kHz — a little more than double the highest expected frequency" (Leis, 2011).*

### 2.6.2 Butterworth band pass

The Butterworth band pass (Butterworth, 1930) filters out the frequencies that lie beyond a defined passband. In the case of heart sound filtering it can be used to filter out noise or such as sound s of the respiratory system or background noises.

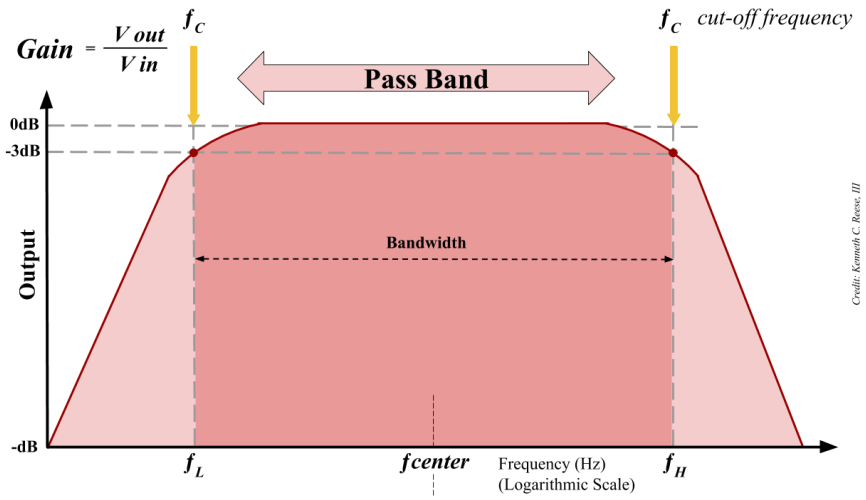


*Figure 4 Band pass filter (AnalogICTips, 2017)*

### 2.6.3 Energy and RMSE

The energy or total magnitude of a signal measures how loud an audio signal is. RMSE is the Root Mean Squared of the Energy.

### 2.6.4 Zero Crossing Rate

The zero crossing rate indicates the number of times that a signal crosses over the x-axis.

### 2.6.5 Fast Fourier Transform (FFT)

The Fast Fourier Transform transforms time-domain signal into the frequency domain (Heideman, Johnson, & Burrus, 1984).

### 2.6.6 Short-time Fourier Transform (STFT)

Short-time Fourier Transform is calculated by taking the Fourier transform of a number of successive frames (Sejdić, Djurović, & Jiang, 2009).

### 2.6.7 Constant-Q Transform

The Constant-Q (Brown, 1991) is a transformation like FFT but with a logarithmic frequency scale.

### 2.6.8 Chroma

Chroma (Shepard, 1964) calculates a 12 element-vector for each pitch from C-major to B.

### 2.6.9 Spectral centroid

The spectral centroid (Grey & Gordon, 1978) is a weighted method to calculate the central frequency of a signal.

#### *2.6.9.1 Centroid Bandwidth*

The bandwidth is the difference between the lowest and highest frequency in the spectrum.

#### *2.6.9.2 Centroid Spectrum*

Spectral contrast (Jiang et al., 2002) considers the spectral peak, the spectral valley, and their difference in each frequency sub band.

#### *2.6.9.3 Centroid Roll-off*

Spectral roll-off is the frequency below which a specified percentage of the total spectral energy, e.g. 85%, lies (Eerola, Ferrer, & Alluri, 2012).

### 2.6.10 Onset Detection

An onset marks the position at which the beginning of the transient part of a sound, or the earliest moment at which a transient can be reliably detected (Bello et al., 2005).

### 2.6.11 Mel-Frequency Cepstral Coefficients

The Mel frequency cepstral coefficients (MFCCs) describes the shape of a spectral envelope using the Mel scale. The Mel scale represents pitch in a logarithmic manner (Umesh, Cohen, Nelson, & Ieee, 1999). They are often used in natural language processing.

### 2.6.12 Linear Prediction Coding Coefficients

Linear Prediction Coding Coefficients (LPCCs) are features from the speech recognition domain, their linearity makes them effective even after resampling data (Deng & O'Shaughnessy, 2018).

## 2.7 Heart sound segmentation

"An open access database for the evaluation of heart sound algorithms" (Liu et al., 2016) reviews four methods to segment heart sound recordings. The authors also add a table with the performance of these methods performed in different research papers. Their accuracy can be used as an indication for performance. Yet it is equally important to notice that the majority of these papers use a different heart sound database.

### 2.7.1 Envelope-based methods

The envelope of a signal is a smooth curve outlining its extremes (Johnson, Sethares, & Klein, 2011).

#### *2.7.1.1 Shannon energy*

*"Shannon energy envelope (SEE), is the average spectrum of energy and is better able to detect peaks"* (Beyramienanlou & Lotfivand, 2017). This technique is mainly used on ECG-signals because the S1-phase starts with a R-peak, as can be seen in figure 2.

#### *2.7.1.2 Hilbert envelope*

*"The Hilbert transform can be considered to be a filter which simply shifts phases of all frequency components of its input by -π/2 radians"* (Feldman, 1997). The Hilbert envelope is calculated on the real part of the signal, its instantaneous frequency is derived from the imaginary part (Liu et al., 2016).

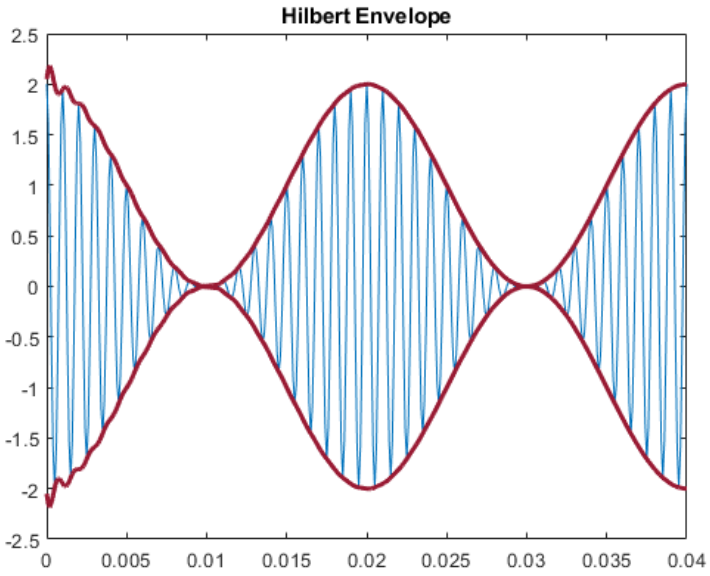


*Figure 5 Hilbert envelope (MathWorks, 2019a)*

#### *2.7.1.3 Squared envelope*

The squared envelope calculates the upper and lower root-mean-square envelopes of a signal. This method is used in "A robust heart sound segmentation algorithm for commonly occurring heart valve diseases." (Ari, Kumar, & Saha, 2008).

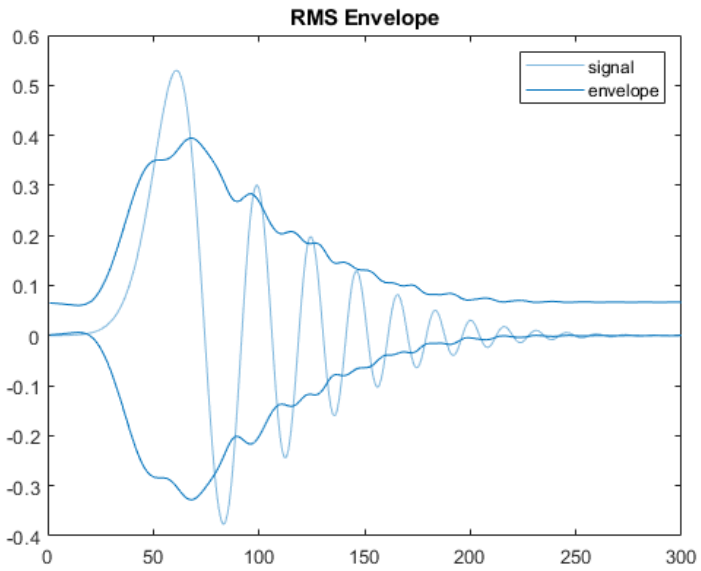


*Figure 6 RMS envelope (MathWorks, 2019c)*

### 2.7.2 Feature-based methods

There are two notable Feature based methods to segment heart sounds. The first one utilises frequency and amplitude (Naseri, Homaeinezhad, & Pourkhajeh, 2013). The latter is an instantaneous phase boundary determination feature after applying a Shannon energy envelope (Varghees & Ramachandran, 2014).

### 2.7.3 Machine-learning methods

"Detection of the first heart sound using a time-delay neural network"(Oskiper & Watrous, 2002) uses a neural network to segment PCG signals. The network achieves a 96.2% accuracy on data from 30 healthy persons. Neural networks will be further discussed under section 2.8. K-means clustering is also often used in machine learning. This algorithm uses the distance between points to measure their similarity. (Chen, Kuan, Celi, & Clifford, 2010) and (Gupta, Palaniappan, Swaminathan, & Krishnan, 2007) use this method in combination with other machine learning algorithms. More about clustering can also be read under section 2.8.

### 2.7.4 Hidden Markov model methods

A hidden Markov model is an adaption of the Markov model that determines probabilities when the system state is only partially observable. "Hidden" refers to that unobservable state. Gamero and Waltrous where the first to use this model for heart sound segmentation (Gamero & Watrous, 2003). In "Detection and identification of heart sounds using homomorphic envelogram and self-organizing probabilistic model" (Gill, Gavrieli, & Intrator, 2005) timing durations are added to the Hidden Markov model. Schmidt et al. improve this method by using a hidden semi-Markov model to estimate the duration of the heart phases within the hidden Markov model (Schmidt et al., 2010). The Markov model is also combined with a Viterbi algorithm. This maximises the likelihood of occurrence of a sequence of observed states (Jurafsky & Martin, 2014). Springer et al. adapts the work of Schmidt et al. and adds logistic regression to the hidden semi Markov model. The logistic regression copes with the problem that real world recording of heart sounds contain noise and background sounds. Springer also modifies the Viterbi algorithm. On a dataset of 112 patients his algorithm scores a stunning F1-score of 95.63%.

$$F1 = 2.\left(\frac{Precision \cdot Recall}{Precision + Recall}\right)$$

This outperforms the previous highest scoring model by more than 10%. *"Therefore, this method is regarded as the state-of-the-art method in heart sound segmentation studies"(Liu et al., 2016).*

## 2.8 Heart sound classification

In this section frequently used machine learning algorithms for classifying heart sounds as normal or abnormal are touched upon. Their performance on the "PhysioNet/Computing in Cardiology Challenge 2016" dataset is also compared.

### 2.8.1 Neural networks (NN)

The highest accuracy on the "PhysioNet/Computing in Cardiology Challenge 2016" is achieved by using a Convolution Neural Network (Potes, Parvaneh, Rahman, & Conroy, 2016). Potes et al. scores an accuracy of 86,02%. This method also utilizes a boosting algorithm called Adaboost (Ut infra). Another convolutional neural network by Rubin et al. scores 83,99% (Rubin et al., 2016). A 'simple' neural network has 3 types of layers and input layer, a set of layers that contains neurons called the hidden layer and the final layer, the output layer. Basic neural networks are feed forward, so information goes from layer to layer in the order as specified above. The neurons in the hidden layer get a weight assigned to indicate their importance in classifying the end result. In convolutional neural network multiple copies of the weighted neurons in the are stored.

Kay et al. uses a Regularized Neural Network (Kay & Agarwal, 2016) and scores 85,20%. Regularized neural networks are networks that try to reduce overfitting by correcting neurons that are assigned a disproportionate weight.

### 2.8.2 Support vector machines (SVM)

The support vector machine of Zabihi et al. scores 84,59% (Zabihi, Rad, Kiranyaz, Gabbouj, & Katsaggelos, 2016). A support vector machine separates the classes in classification by using a hyperplane. A hyperplane has one dimension less than its surrounding space. So in a n-dimensional space the hyperplane has a dimension of n-1. The support vectors are the points that are closest to the hyperplane. The best hyperplane to divide the classes is the one that has the largest margin relative to the support vectors

### 2.8.3 K-Nearest neighbours (k-NN)

Boboli uses a K-nearest neighbours algorithm in combination with MFCC's (Bobillo, 2016; Homsi et al., 2016). This method achieves a score of 84,54%. k-NN is a clustering algorithm, it uses a distance function to determine which are the k (number) points closest to a starting point. The MFCC's are a well proven method in speech recognition systems (Zheng, Zhang, & Song, 2001).

### 2.8.4 Classification trees

The random forest model of Homsi et al. results in an accuracy of 84.48% (Homsi et al., 2016). A boosting algorithm called LogitBoost (Ut infra) is also used. Random forest (RF) creates multiple decision trees per instance, the output of each tree will be a certain class. A voting calculation determines which of these classes the instance most likely belongs to. Another tree algorithm is the C4.5 decision tree, in Weka a freeware datamining tool it is implemented as J48. It is described to be *"a landmark decision tree program that is probably the machine learning workhorse most widely used in practice to date"* (Ian H. Witten, 2019).

### 2.8.5 Boosting algorithms

AdaBoost is an algorithm that is used in combination with another Machine learning algorithm (Kégl, 2013). It focusses on the weak classifiers, those with a low correlation or low predictive power. The weak classifiers are combined and used to assign a weight to the final output. This can result in a boost for instances that were misclassified.

LogitBoost (Friedman, Hastie, & Tibshirani, 2000) is in essence the Adaboost algorithm regulated by a cost function based on logic regression.

Ibarra-Hernández et al, introduce the SMOTE method to the PhysioNet database (Ibarra-Hernández, Bertin, Alonso-Arévalo, & Guillén-Ramírez, 2018). This algorithm uses oversampling to cope with imbalanced classes. Only 20% of the PhysioNet database are abnormal heart sound recordings vs. 80% normal. They also use Linear Prediction Coding coeficients.

## 2.9 Software and libraries

### 2.9.1 Jupyter notebook

Jupyter is a browser based tool in which documents called “notebooks” can be creates . In these notebooks cells can be added. Cells can contain Python code, text or visualisations. It is an excellent tool to share code, ideas or to keep a track of your work (jupyter.org, 2019)

### 2.9.2 Librosa

Librosa is a Python package for audio and music analysis (McFee et al., 2019).

### 2.9.3 Scpypi

The SciPy library contains functions for numerical integration, interpolation, optimization, linear algebra and statistics in Python (SciPy.org, 2019).

### 2.9.4 Numpy

Numpy allows to do array based scientific computing in Python (NumPy, 2019).

### 2.9.5 Matplotlib

Matplotlib is a visualisation library for Python (Matplotlib, 2019).

### 2.9.6 Pandas

Pandas is an open source library providing ideal for manipulating data (Project, 2019).

### 2.9.7 Scikit-learn

Scikit-learn is the Machine Learning library in Python if you don’t want to reinvent the wheel, this is the way to go (scikit-learn, 2019).

### 2.9.8 Matlab

Matlab is software with its own matrix based programming language, perfect to transform and analyse large datasets (MathWorks, 2019b).

### 2.9.9 Weka

Weka is open source machine learning software (Witten, Frank, Hall, & Pal, 2016).

# 3 Case Study

## 3.1 Research question

Based on the insights gained from the literature review the following question is posed:

*"Does combining oversampling with spectral features other than Linear Predictive Coding coefficients lead to a better classifier?"*

## 3.2 Significance

The literature review reveals that the highest achieved accuracy of classification on the PhysioNet dataset is 0.86.This not high enough to be deployed in a medical context. As a comparison a Belgian smartphone app for Heart Rhythm Monitoring that recently got approved by the American Food and Drug Administration scored an accuracy of 0.96 in 2016 (Mortelmans, 2016). Ibarra Hernández et al succesfuly implement an oversampeling method, they do use spectral features (LPCC) in there model, but not the same as for example Homsi et al. This begs the question if the combination of SMOTE with other spectral features like MFCCs will lead to a better model.

## 3.3 Methodology

### 3.3.1 Crisp DM

The Cross Industry Standard Process for Data Mining is a framework designed to tackle datamining problems. It was conceptualised by leading data mining users and suppliers and partly sponsored by the European Commission (Wirth & Hipp, 2000). Solving these problems is a creative process, it consists out of six steps which have been modelled in a cyclical way. Most problems require more than one cycle to be solved. Often within one cycle new insights are discovered, that's why some steps loop back to previous steps.

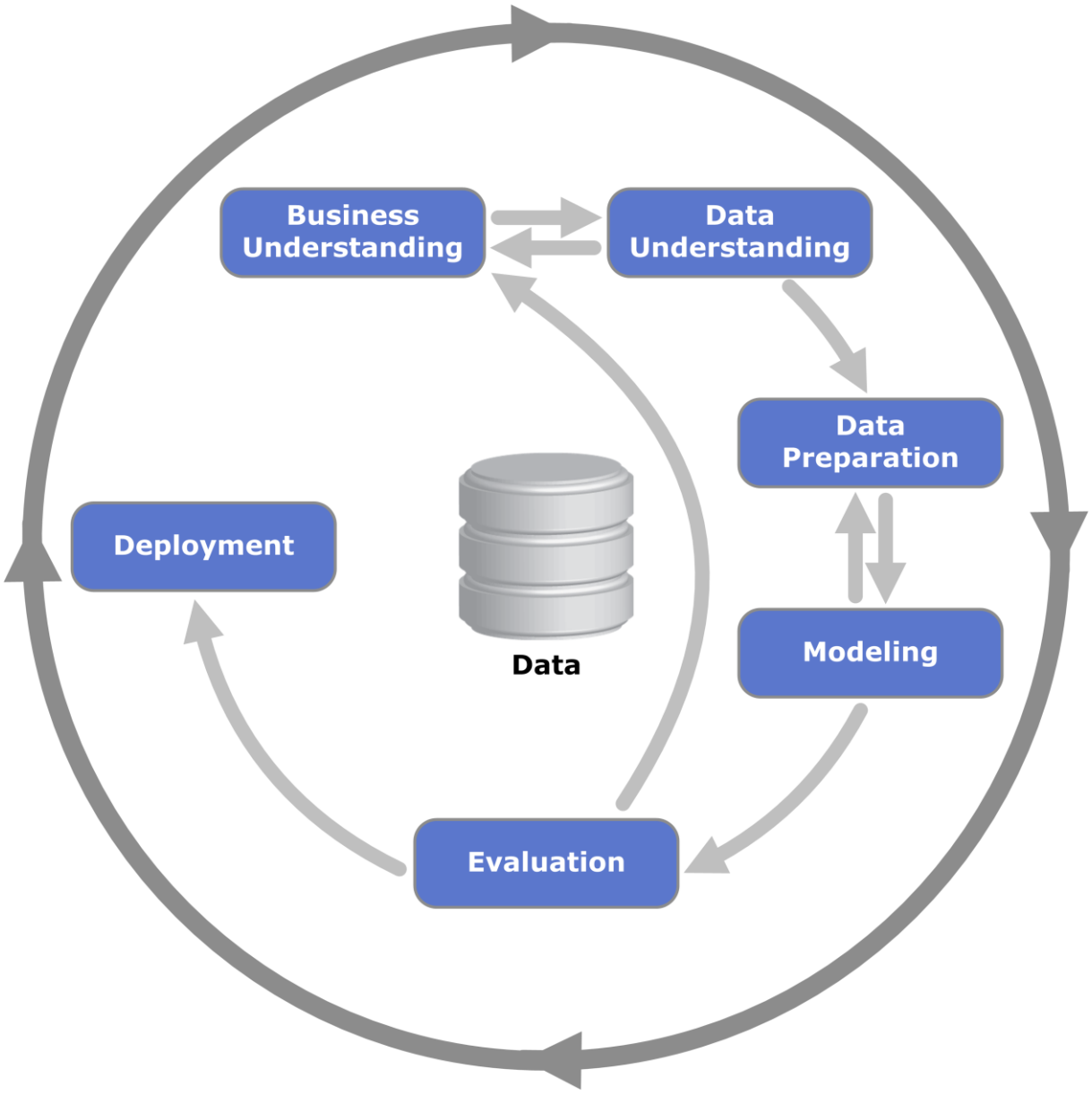


*Figure 7 Crisp DM*

#### 3.3.1.1 Business understanding

Business understanding comprises the goal, problems and criteria of success of this case study.

#### 3.3.1.2 Data understanding

In data understanding the dataset is explored in general. Also the sound duration and number of beats are elaborated upon.

#### 3.3.1.3 Data preparation

Data preparation describes the process of cleaning the data. Followed by extracting three groups of features, Springer, Spectral and beat types. Finally the features to train the model are chosen.

#### 3.3.1.4 Modelling

Modelling handles the exploration, development and building of classification methods.

#### 3.3.1.5 Evaluation

This section covers the evaluation mechanism of the PhysioNet Challenge.

#### 3.3.1.6 Deployment

Under deployment the performance of the deployed models is presented.

## 3.4 Business understanding

The goal of this case study is to classify the recorded heart sounds in the “PhysioNet/Computing in Cardiology Challenge 2016” dataset. There are two classification classes, normal and abnormal. The PhysioNet Challenge provides the Springer segmentation algorithm and a baseline model. Both of them are written in MATLAB. The baseline model is a simple voting algorithm applied to features created by the Springer method. This model can be used as a benchmark, its accuracy (MAcc) is 0.7057, the sensitivity (Se) is 0.6545 and the specificity (Sp) 0.7569 (with MAcc = (Se+Sp)/2).

There are 4 known problems that will needed to be faced in this challenge. The first being the class balance. Cardiovascular diseases are responsible for the major cause of death in the world. Nevertheless there are still more people with a healthy heart. So the recordings will not likely be balanced between the two classes. This problem might be addressed by the use of class balancers or oversampling. The next problem are the recordings in itself, they can be distorted by background noises or even the respiratory system of the patient. These problems can be tackled by frequency filtering or techniques from the natural language processing field like MFCCs. The third problem is the balance of the sensitivity and the specificity. Not only the accuracy is important in this clinical context. Guarding these values is also crucial when deciding on the best model. The final problem is one that every datamining project has to face, the risk of overfitting. In every step of this process 10 fold cross validation will be used in combination with a holdout dataset to evaluate the risk of overfitting.

The main self-imposed criteria of success for this challenge is building a robust classification model that can compete with the top 8 of know entries ( these have a MAcc ranging from 0.8399 to 0.8602). Robust meaning not prone to overfitting and a sensitivity and specificity that are balanced.

## 3.5 Data understanding

### 3.5.1 The dataset

The “PhysioNet/Computing in Cardiology Challenge 2016” dataset contains 3240 record heart sounds. Both healthy and pathological patients are represented in the dataset. The recordings originate from different sources and where taken in a clinical or non-clinical setting (e.g. home visits) (G. D. Clifford et al., 2016). “*The heart sound recordings were collected from different locations on the body. The typical four locations are aortic area, pulmonic area, tricuspid area and mitral area, but could be one of nine different locations* “ (Physionet, 2016).

The dataset consists out of six folders containing .wav files. Each folder also contains a “.hae” file for each of the recordings, and one “REFERENCE.cvs” file. The reference file has two columns: filename and classification. In the classification column a negative one value signifies a normal heart sound, a value of one is used for the abnormal ones. In total the dataset is split into 79,48 percent normal and 20,52 abnormal.

There is also a validation set which contains 300 files that are randomly chosen out of the training set.

| Folder | training-a | training-b | training-c | training-d | training-e | training-f | **Total** |
|---|---|---|---|---|---|---|---|
| Number of files | 409 | 409 | 31 | 55 | 2141 | 114 | **3240** |
| Abnormal | 292 | 104 | 24 | 28 | 183 | 34 | **665** |
| Normal | 117 | 368 | 7 | 27 | 1958 | 80 | **2575** |

*Table 2, Traning set classification*

### 3.5.2 Sound duration

The duration of the recordings varies between a minimum duration of 5 and a maximum duration of 122 seconds. The median duration is 20.82 seconds, the mean length is 22. 46 with a standard deviation of 12.38 seconds. The histogram show that the distribution is skewed to the left. Notable frequencies within the histogram are the frequencies with a duration between 7.25 and 9.19 seconds (546 recordings), and the range between 34.48 and 36.42 seconds (298 recordings).

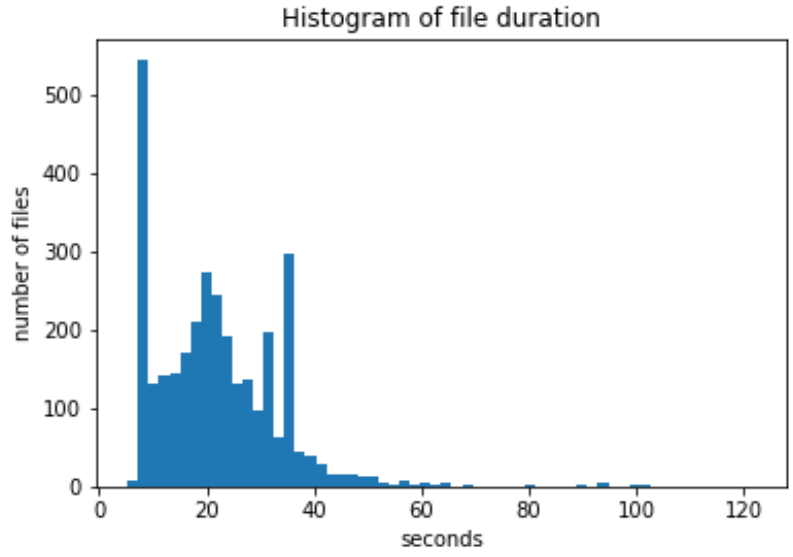

*Figure 8, File duration in seconds*

### 3.5.3 Number of beats

Because of the great variety in the length of the sound files, the total number of beats and their position was also extracted out of the recordings. The part before the first and after the last beat of each file was removed. This was done in order to ensure no irrelevant audio frames or half beats where used in further modelling. The total dataset contains 82444 beats.

| Folder | training-a | training-b | training-c | training-d | training-e | training-f | **Total** |
|---|---|---|---|---|---|---|---|
| Number of beats | 14652 | 3662 | 1874 | 874 | 56854 | 4528 | **82444** |

*Table 3, Number of beats*

The position of each beat was calculated after applying a Butterworth band pass filter. This filter reduces the noise in the high and low frequencies. The signal was filtered with a fourth order Butterworth band pass filter with cut-off frequencies at 25 Hz and 400 Hz (Schmidt, (Schmidt et al., 2010).

Using the Librosa library (McFee et al., 2019) an onset envelope was calculated on the percussive part of each beat. Onset detection is the process of finding the starting points of all musically relevant events in an audio performance (Sebastian & Widmer, 2013). The onset detection has the best results on percussive sounds(Downie, 2012). The location of the onset frames was used to track the beginning of each beat. The beat track function is configured for a variable tempo, since the tempo of a heartbeat can be irregular. The mean beat and standard deviation where chosen to be 65.9 and 9.7 (Moser et al., 1994), these parameters are the starting point to estimate the tempo and location of the heart beats.

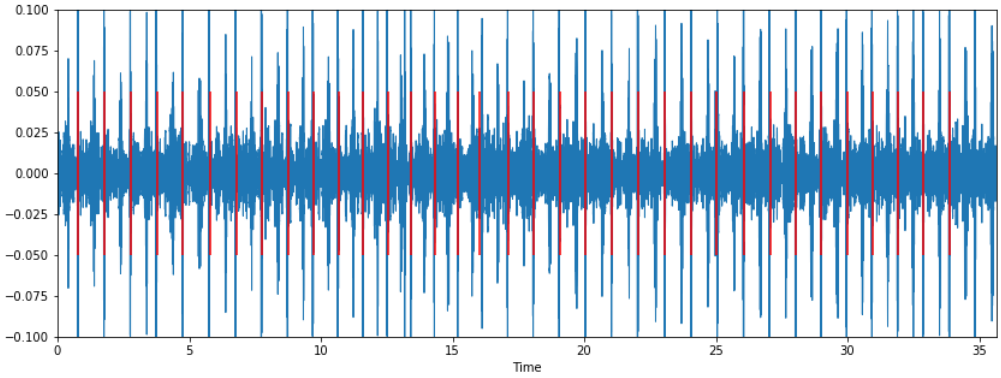


*Figure 9, a0001.wav split into beats*

## 3.6 Data Preparation

### 3.6.1 Data cleaning

The audio files where resampled to 2024 Hz, this was achieved by using the Librosa library (McFee et al., 2019) in Python. Springer et al., ref resamples to 1 kHz (Springer et al., 2016).

2024 Hz was choses with the Nyquist theorem in mind. The next step was the prepossessing method of Schmidt et al. (Schmidt et al., 2010). A fourth order Butterworth band pass filter was applied with 25 and 400 Hz as cut-off frequencies. The frequencies above can be distorted by the sounds of human respiration. The implementation of Butterworth in Python was duplicated from Stack overflow (Weckesser, 2019). Schmidt et al. also defines the method to remove the spikes.

> *"* *(1) The recording is divided into 500 ms windows.*
>
> *(2) The maximum absolute amplitude (MAA) in each window is found.*
>
> *(3) If at least one MAA exceeds three times the median value of the MAA's, the following steps were carried out. If not continue to point 4.*
>
> *(a) The window with the highest MAA was chosen.*
>
> *(b) In the chosen window, the location of the MAA point was identified as the top of the noise spike.*
>
> *(c) The beginning of the noise spike was defined as the last zero-crossing point before the MAA point.*
>
> *(d) The end of the spike was defined as the first zero-crossing point after the maximum point.*
>
> *(e) The defined noise spike was replaced by zeroes.*
>
> *(f) Resume at step 2.*
>
> *(4) Procedure completed."* (Schmidt et al., 2010).

This algorithm was programmed In Python and applied to all the files, with the only difference that windows of 512ms where chosen. This is the standard hop length of the window function in Librosa. The zero crossing function was also available in Librosa and consequently applied . Numpy provided the methods for the array operations like replacing the spike with zero's.

### 3.6.2 Feature extraction.

The features extracted from the audio files can be divided into three categories. Firstly there are the Springer Features, the second created feature group contains the spectral features and finally there are the beat type features.

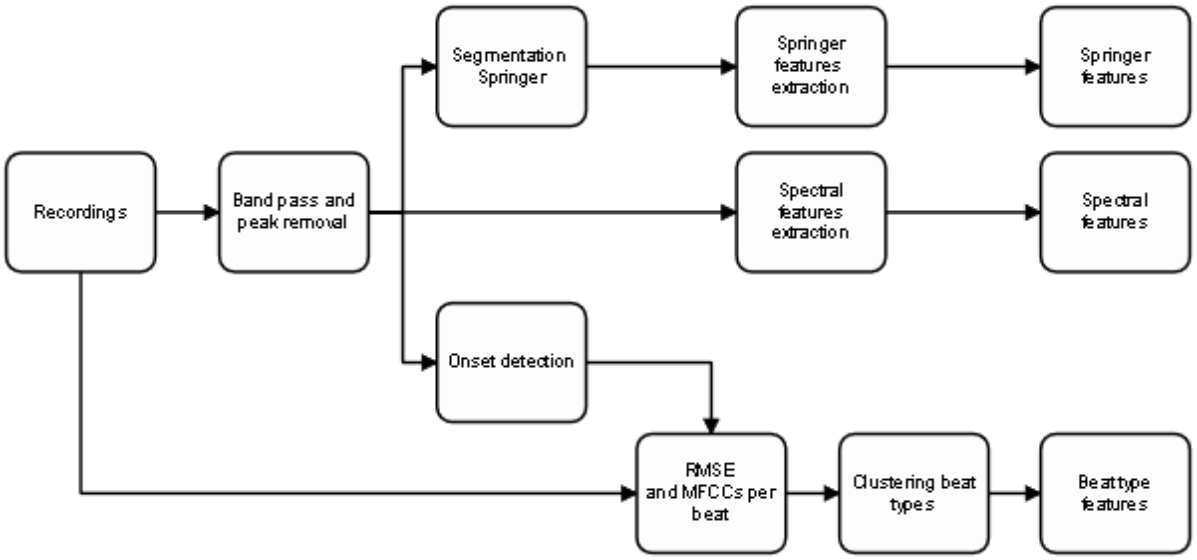


*Figure 10 Feature extraction process*

#### *3.6.2.1 Springer features*

The PhysioNet challenge issues the Springer segmentation algorithm (Springer et al., 2016) to get the challenge started. The code is written in MATLAB and contains the Springer algorithm. A way to read and write audio files and apply the codes still needed to be added. Using the algorithm the following 20 features were created. They represent information based on the position of S1, systole, S2 and diastole.

"1. ***m_RR:*** *mean value of RR intervals*

2. ***sd_RR:*** *standard deviation (SD) of RR intervals*

3. ***m_IntS1:*** *mean value of S1 intervals*

4. ***sd_IntS1:*** *SD of S1 intervals*

5. ***m_IntS2:*** *mean value of S2 intervals*

6. ***sd_IntS2:*** *SD of S2 intervals*

7. ***m_IntSys:*** *mean of systolic intervals*

8. ***sd_IntSys:*** *SD of systolic intervals*

9. ***m_IntDia:*** *mean of diastolic intervals*

10. ***sd_IntDia:*** *SD of diastolic intervals*

11. ***m_Ratio_SysRR:*** *mean of the ratio of systolic interval to RR of each heart beat*

12. ***sd_Ratio_SysRR:*** *SD of the ratio of systolic interval to RR of each heart beat*

13. ***m_Ratio_DiaRR:*** *mean of ratio of diastolic interval to RR of each heart beat*

14. ***sd_Ratio_DiaRR:*** *SD of ratio of diastolic interval to RR of each heart beat*

15. ***m_Ratio_SysDia:*** *mean of the ratio of systolic to diastolic interval of each heart beat*

16. ***sd_Ratio_SysDia:*** *SD of the ratio of systolic to diastolic interval of each heart beat*

17. ***m_Amp_SysS1:*** *mean of the ratio of the mean absolute amplitude during systole to that during the S1 period in each heart beat*

18. ***sd_Amp_SysS1:*** *SD of the ratio of the mean absolute amplitude during systole to that during the S1 period in each heart beat*

19. ***m_Amp_DiaS2:*** *mean of the ratio of the mean absolute amplitude during diastole to that during the S2 period in each heart beat*

20. ***sd_Amp_DiaS2:*** *SD of the ratio of the mean absolute amplitude during diastole to that during the S2 period in each heart beat"* (Liu et al., 2016).

#### *3.6.2.2 Spectral features*

In Python the 30 features were extracted for each file. The code to calculate these spectral features was available in the Librosa package. Most of the features returned a vector or matrix, as they are mostly used to plot spectral signals on a graph. For the MFccs the mean for each of the 13 vectors was used, this number is based on Bobilo's tensor approach to the classification problem (Bobillo, 2016). The 12 Chroma's (C, C♯, D, D♯, E ,F, F♯, G, G♯, A, A♯, B) are represented by their standard deviation, as are the other spectral features.

21-33. **m_MFccs:** mean of thirteen Mel-frequency cepstral coefficients.

33-45 **sd_chroma:** SD of twelve Chroma features.

46. **tempo:** the estimated tempo.

47. **sd_spectral_centroid:** SD of spectral centroid for each frame in the signal

48. **sd_zero_crossings**: SD of number of times the signal crosses the horizontal axis

49. **sd_spec_bw:** SD of spectral bandwidth

50. **sd_flatness:** SD of spectral flatness

51. **sd_rolloff:** SD of spectral roll-off

#### *3.6.2.3 Beat type features*

The onset detection method (as described in the data understanding) was used to extract the location of each beat in every recording. The "PhysioNet/Computing in Cardiology Challenge 2016" (G. D. Clifford et al., 2016) reports that there are 84425 beats in the database (after hand correcting the springer segmentation result). The hand corrections where done by experts in the field. The onset detection method extracts 82443 beats from the database this is a difference of 0.0235 . The onset method is less accurate, but results in much lower computation times.

For each of the beats the mean RMS spectral energy and 13 mean MFccs where calculated, this was done on the original unfiltered data. The rationale being that the fact that a beat is distorted also is valuable information to build a model. Nevertheless they were also computed on the filtered data.

Afterwards the beats where clustered into beat types. As clustering is an unsupervised method, it is unknown what the different beat types stand for. The beats where clustered with k-nearest

neighbours into 8, 16,32 and 64, 128, 256, 512 and 1052 clusters, both for the filtered and unfiltered data. The presence of each type of beat in each recording was calculated for these 16 models. This was done using a PivotTable in Excel where the rows represent the recording and the columns the clusters. The value in each cluster-column is the percentage of each beat type in accordance to the rows (recordings) total. Based on the gain ratio in Weka the unfiltered 128-clusters proved to be the best attributes .

52 – 179. **beat_typeX:** percentage of presence for the one hundred twenty-eight different beat types.

### 3.6.3 Feature selection

In order to select features the gain ratio was calculated in Weka. With the gain ratio as threshold different low ranked attributes where removed from the feature set. The effects of the removal of parameters can be observed in table 4. The F-scores of three algorithms went down slightly as more attributes where removed. Given the unbalanced nature of the database and the idea to use AdaBoost in the data modelling stage, it was decided to select all the created features. When comparing the 20 lowest ranked attributes before and after applying oversampling it becomes evident that saving the zero ranked attributes can result in gains. For both datasets the lowest ranked attributes are cluster and Chroma. This does not imply that the beat type clusters are bad features. The top 20 ranked features are all clusters, with an average merit ranging from 0.317 to 0.108. the order of the rankings is different after applying oversampling. In general it can bet concluded that some have a higher clusters and m_MFccs have a higher gain ratio than the springer features.

| Gain Ratio | attributes | F-score Random forest | F-score Logistic | F-score SMO |
|---|---|---|---|---|
| >0.05 | 53 | 0.903 | 0.881 | 0.869 |
| >0.04 | 81 | 0.908 | 0.886 | 0.876 |
| >0.03 | 97 | 0.899 | 0.896 | 0.886 |
| >0.01 | 126 | 0.902 | 0.895 | 0.888 |
| >0 | 131 | 0.906 | 0.896 | 0.889 |
| >=0 | 179 | 0.907 | 0.898 | 0.903 |

*Table 4 Gain ratio of feature sets vs. F-score*

| Normal dataset | | | Oversampled at SMOTE 301 | | |
|---|---|---|---|---|---|
| average merit | average rank | attribute | average merit | average rank | attribute |
| 0 0 | 155.3 +- 0.9 | 157 cluster91 | 0.015 +- 0.001 | 159.4 +- 2.24 | 40 chroma_7 |
| 0 0 | 158.6 +- 4.94 | 107 cluster98 | 0.015 +- 0.002 | 160 +- 1.67 | 35 chroma_2 |
| 0 0 | 161.1 +- 1.04 | 66 cluster15 | 0.01 +- 0.03 | 161.8 +-35.27 | 57 cluster6 |
| 0 0 | 161.9 +- 1.37 | 61 cluster11 | 0.012 +- 0.001 | 161.9 +- 1.81 | 45 chroma_12 |
| 0 0 | 163.1 +- 8.81 | 106 cluster95 | 0.011 +- 0.001 | 163.7 +- 1.9 | 36 chroma_3 |
| 0 0 | 163.6 +- 0.66 | 59 cluster8 | 0.011 +- 0.017 | 164.2 +-16.19 | 59 cluster8 |
| 0.002 +- 0.007 | 163.9 +-18.66 | 102 cluster75 | 0.01 +- 0.001 | 164.2 +- 1.89 | 38 chroma_5 |
| 0 0 | 164.8 +- 0.4 | 58 cluster7 | 0.009 +- 0.005 | 164.5 +- 5.08 | 88 cluster71 |
| 0 0 | 165.6 +- 3.77 | 68 cluster29 | 0.009 +- 0.006 | 164.9 +- 4.7 | 37 chroma_4 |
| 0 0 | 165.9 +- 0.54 | 57 cluster6 | 0.008 +- 0.001 | 165.8 +- 1.47 | 41 chroma_8 |
| 0 0 | 166.9 +- 0.54 | 54 cluster3 | 0.007 +- 0.022 | 167.7 +-26.34 | 77 cluster32 |
| 0 0 | 167.2 +- 2.86 | 73 cluster31 | 0+- 0 | 168.7 +- 1.9 | 116 cluster77 |
| 0 0 | 167.6 +- 3.75 | 75 cluster21 | 0+- 0 | 168.8 +- 1.25 | 148 cluster126 |
| 0 0 | 172.9 +- 1.37 | 99 cluster37 | 0+- 0 | 168.9 +- 1.37 | 162 cluster107 |
| 0 0 | 173 +- 2.9 | 77 cluster32 | 0+- 0 | 169.9 +- 3.33 | 106 cluster95 |
| 0 0 | 174.2 +- 0.98 | 98 cluster114 | 0.003 +- 0.009 | 170.9 +- 9.42 | 99 cluster37 |
| 0 0 | 174.6 +- 1.85 | 88 cluster71 | 0.003 +- 0.009 | 173.3 +- 7.32 | 42 chroma_9 |
| 0 0 | 176.5 +- 2.29 | 78 cluster22 | 0+- 0 | 174 +- 1.79 | 85 cluster103 |
| 0 0 | 177.3 +- 0.46 | 85 cluster103 | 0+- 0 | 175.2 +- 2.18 | 61 cluster11 |
| 0 0 | 179 +- 0 | 90 cluster33 | 0+- 0 | 179 +- 0 | 90 cluster33 |

*Table 5 Lowest ranked features on normal vs. Smote 301 feature set*

## 3.7 Data modelling

### 3.7.1 Exploration

The data modelling was kicked-off by using the classification learner in MATLAB. There 71 different classification models where trained, using a 10 fold cross validation technique to avoid overfitting. From their results it was evident that different trees, SVM and an ensemble of k-NN's where the algorithms that could be further developed. The next step was double checking the modelled algorithms against a holdout dataset of 300 instances. This showed that the SVM and k-NN's where still prone to overfitting. The k-NN's where emitted from the data modelling process while the SVM's where kept in (Ut infra).

### 3.7.2 Development

#### *3.7.2.1 Base algorithms*

In Weka three algorithms where picked to be further developed. The first one is the random forest decision tree (RF), this choice was inspired by the work of Ibarra Hernández et al. (Ibarra-Hernández et al., 2018). These authors show that the random forest algorithm performs best on the PhysioNet database when using the SMOTE oversampling technique. The second algorithm is J48, also a tree model. J48 was often used to make decisions during this case study and proved to be a reliable classifier. Finally SMO - a support vector machine – was added, although it seems that it resulted in overfitting. The reason being that Ibarra Hernández et al. show that rotation forrest trees improve and SMV the get worse when using oversampling (Ibarra-Hernández et al., 2018). It has to be noted that their findings were deducted from tests on the same dataset, but the features used in their report and this case study obviously differ. So by adding both SM0 and RF, it can be checked whether their theory still holds up for other features.

#### *3.7.2.2 Oversampling*

The effect of the SMOTE oversampling technique was tested for J48 and RF and SMO. Different SMOTE values where applied to the dataset with nearest neighbours set on five. The SMOTE value, represents the percentage of oversampling. A zero value corresponds to the original dataset, where the abnormal instances still make up 20 percent of the database. At an oversampling percentage of 301 the dataset is oversampled to an equal or fifty-fifty division between normal and abnormal. For both the algorithms the F-score for 10-fold cross validation and MAcc for the holdout dataset are higher when oversampling is used. No conclusion other than the variation of results can be drawn from the amount of SMOTE applied for the J48 tree. For random forest is seems that more SMOTE leads to higher accuracies, although the relation between sensitivity and specificity fluctuates.

| J48 | | | | | | |
|---|---|---|---|---|---|---|
| | | 10-fold CV | | Holdout | | |
| SMOTE | abnormal instances | F-score | Acc | Se | Sp | MAcc |
| 301.00 | 0.50 | 0.92 | 0.84 | 0.89 | 0.77 | 0.83 |
| 250.00 | 0.47 | 0.90 | 0.83 | 0.83 | 0.82 | 0.83 |
| 200.00 | 0.43 | 0.89 | 0.84 | 0.87 | 0.80 | 0.84 |
| 150.00 | 0.38 | 0.89 | 0.81 | 0.80 | 0.82 | 0.81 |
| 100.00 | 0.33 | 0.89 | 0.83 | 0.86 | 0.79 | 0.83 |
| 0.00 | 0.20 | 0.89 | 0.81 | 0.78 | 0.84 | 0.81 |

*Table 6 SMOTE values with J48*

| RF | | | | | | |
|---|---|---|---|---|---|---|
| | | 10-fold CV | | Holdout | | |
| SMOTE | abnormal instances | F-score | Acc | Se | Sp | MAcc |
| 301.00 | 0.50 | 0.96 | 0.85 | 0.89 | 0.81 | 0.85 |
| 250.00 | 0.47 | 0.97 | 0.85 | 0.89 | 0.80 | 0.85 |
| 200.00 | 0.43 | 0.95 | 0.86 | 0.86 | 0.84 | 0.85 |
| 150.00 | 0.38 | 0.95 | 0.84 | 0.81 | 0.86 | 0.84 |
| 100.00 | 0.33 | 0.93 | 0.84 | 0.81 | 0.86 | 0.84 |
| 0.00 | 0.20 | 0.91 | 0.82 | 0.72 | 0.93 | 0.82 |

*Table 7 SMOTE values with RF*

#### *3.7.2.3 Boosting*

In the top 5 of classification models on the PhysioNet database, two entries use a boosting algorithm(G. D. Clifford et al., 2016). Potes et al., who claims the top spot uses AdaBoost, Homsi et al. ranked fifth uses LogitBoost combined with random forest. The Rotation Forest algorithm proved to be a good ensemble method when tested on different types of datasets in 2013. Both with and without Logitboost it improved the accuracy on three different heart sound datasets (Kotsiantis, 2013). As a result all of the techniques above where taken into consideration when developing the final model. The Rotation Forest model was added to Weka via the package manager. In table 8 a combination the results of different models is displayed. These were all calculated using 10-fold cross validation. Unfortunately all the combinations of Rotation Forest and LogitBoost caused problems and where skipped during the modelling process. Consequently they are not in the table, the build of Rotation Forest and AdaBoost with J48 also failed multiple times on the 150 SMOTE database.

### 3.7.3 Built models

| Oversampling | | RF | | | J48 | | | SMO | |
|---|---|---|---|---|---|---|---|---|---|
| Performance boosting | Precision | Recall | F-Score | Precision | Recall | F-Score | Precision | Recall | F-Score |
| 0 SMOTE | | | | | | | | | |
| - | 0.911 | 0.913 | 0.907 | 0.886 | 886 | 0.886 | 0.903 | 0.907 | 0.903 |
| AdaBoost | 0.905 | 0.908 | 0.903 | 0.906 | 0.909 | 0.907 | 0.895 | 0.898 | 0.896 |
| LogitBoost | 0.917 | 0.919 | 0.916 | - | - | - | - | - | - |
| Rotation Forest | 0.91 | 0.912 | 0.907 | 0.915 | 0.918 | 0.915 | - | - | - |
| Rotation Forest + Adaboost | 0.909 | 0.911 | 0.906 | 0.926 | 0.928 | 0.926 | - | - | - |
| Bagging | 0.903 | 0.906 | 0.9 | 0.909 | 0.912 | 0.91 | - | - | - |
| Bagging + Logitboost | 0.909 | 0.911 | 0.907 | - | - | - | - | - | - |
| Bagging + Adaboost | 0.916 | 0.918 | 0.915 | 0.944 | 0.944 | 0.944 | - | - | - |
| 100 SMOTE | | | | | | | | | |
| - | 0.934 | 0.935 | 0.934 | 0.889 | 0.889 | 0.889 | 0.905 | 0.905 | 0.905 |
| AdaBoost | 0.938 | 0.938 | 0.938 | 0.929 | 0.93 | 0.93 | 0.895 | 0.896 | 0.896 |
| LogitBoost | 0.948 | 0.948 | 0.948 | - | - | - | - | - | - |
| Rotation Forest | 0.939 | 0.939 | 0.939 | 0.94 | 0.94 | 0.94 | - | - | - |
| Rotation Forest + Adaboost | 0.937 | 0.938 | 0.937 | 0.949 | 0.949 | 0.949 | - | - | - |
| Bagging | 0.932 | 0.932 | 0.932 | 0.924 | 0.923 | 0.923 | - | - | - |
| Bagging + Logitboost | 0.94 | 0.94 | 0.94 | - | - | - | - | - | - |
| Bagging + Adaboost | 0.932 | 0.932 | 0.932 | 0.944 | 0.944 | 0.944 | - | - | - |
| 150 SMOTE | | | | | | | | | |
| - | 0.945 | 0.945 | 0.945 | 0.891 | 0.891 | 0.891 | 0.906 | 0.906 | 0.906 |
| AdaBoost | - | - | - | 0.935 | 0.935 | 0.935 | 0.901 | 0.902 | 0.901 |
| LogitBoost | 0.952 | 0.952 | 0.952 | - | - | - | - | - | - |
| Rotation Forest | 0.946 | 0.946 | 0.946 | 0.942 | 0.941 | 0.941 | - | - | - |
| Rotation Forest + Adaboost | 0.945 | 0.944 | 0.945 | [1] | | | - | - | - |
| Bagging | 0.94 | 0.94 | 0.94 | 0.917 | 0.916 | 0.917 | - | - | - |
| Bagging + Logitboost | | | | - | - | - | - | - | - |
| Bagging + Adaboost | 0.938 | 0.938 | 0.938 | 0.949 | 0.949 | 0.949 | - | - | - |
| 301 SMOTE | | | | | | | | | |
| - | 0.959 | 0.958 | 0.958 | 0.915 | 0.915 | 0.915 | 0.91 | 0.909 | 0.908 |
| AdaBoost | 0.958 | 0.957 | 0.957 | 0.949 | 0.949 | 0.949 | 0.901 | 0.9 | 0.9 |
| LogitBoost | 0.966 | 0.965 | 0.965 | - | - | - | - | - | - |
| Rotation Forest | 0.958 | 0.957 | 0.957 | 0.952 | 0.952 | 0.952 | - | - | - |
| Rotation Forest + Adaboost | 0.959 | 0.958 | 0.958 | 0.964 | 0.964 | 0.964 | - | - | - |
| Bagging | 0.955 | 0.954 | 0.954 | 0.936 | 0.935 | 0.935 | - | - | - |
| Bagging + Logitboost | 0.959 | 0.958 | 0.958 | - | - | - | - | - | - |
| Bagging + Adaboost | 0.956 | 0.955 | 0.955 | 0.963 | 0.962 | 0.962 | - | - | - |

*Table 8 Built models*

[1] The build of this model failed.

## 3.8 Evaluation

The PhysioNet Challenge provides an evaluation method for classification on their dataset. The score is calculated by taking the average of the combined sensitivity specificity. The challenge allows to label noisy recordings as unsure, to cope with distorted recordings. If a recording is labeled with a zero value it is considered uncertain. These are handled in a different manner when calculating the final score. Whether or not this label is used was a free choice. In this model it was chosen not to add these.

| Reference label | Normal (-1) | Uncertain (0) | Abnormal (1) |
|---|---|---|---|
| Normal, clean | $Nn_1$ | $Nq_1$ | $Na_1$ |
| Normal, noisy | $Nn_2$ | $Nq_2$ | $Na_2$ |
| Abnormal, clean | $An_1$ | $Aq_1$ | $Aa_1$ |
| Abnormal, noisy | $An_2$ | $Aq_2$ | $Aa_2$ |

*Table 9 PhysioNet score reference labels (Physionet, 2016)*

Weights defined for clean and noisy records:

$$wa_1 = \frac{\text{clean abnormal records}}{\text{total abnormal records}} \qquad na_1 = \frac{\text{clean normal records}}{\text{total normal records}}$$

$$wa_2 = \frac{\text{noisy abnormal records}}{\text{total abnormal records}} \qquad na_2 = \frac{\text{noisy normal records}}{\text{total normal records}}$$

Weighted modified Sensitivity and specificity:

$$Se = wa_1 \frac{Aa_1}{Aa_1 + Aq_1 + An_1} + wa_2 \frac{Aa_2 + Aq_2}{Aa_2 + Aq_2 + An_2}$$

$$Sp = wn_1 \frac{Nn_1}{An_1 + Nq_1 + Nn_1} + wn_2 \frac{Nn_2 + Nq_2}{Na_2 + Nq_2 + Nn_2}$$

Since there are no instances classified a noisy in this model the formula becomes:

$$Se = \frac{Aa_1}{Aa_1 + An_1}$$

$$Sp = \frac{Nn_1}{An_1 + Nn_1}$$

The challenges overall score is defined as MAcc, the average of the combined Sensitivity and Specificity :

$$\text{MAcc} = \frac{\text{Se} + \text{Sp}}{2}$$

## 3.9 Deployment

Based on the F-score a top forty of created models were deployed on the 300 instance validation set. Such a large number of models was deployed for the simple reason that almost all the created models scored an F-measure above 90 percent. The training F-score of the forty models ranges from 0.9150 to 0.9650. None of the support vector machines made it to this stage.

| Model | Se | Sp | MAcc | Rank based on training F-score |
|---|---|---|---|---|
| Smote 100 AdaBoost J48 | 0.8609 | 0.8776 | 0.8692 | 34 |
| Smote 150 Rotation Forest RF | 0.8609 | 0.8707 | 0.8658 | 17 |
| LogitBoost RF | 0.8013 | 0.9252 | 0.8632 | 37 |
| Smote 100 Rotation Forest + Adaboost J48 | 0.8344 | 0.8912 | 0.8628 | 13 |
| Smote 301 AdaBoost J48 | 0.8742 | 0.8503 | 0.8623 | 14 |
| Smote 150 Bagging + Adaboost J48 | 0.8278 | 0.8912 | 0.8595 | 15 |
| Smote 150 AdaBoost RF | 0.8278 | 0.8912 | 0.8595 | 27 |
| Smote 150 LogitBoost RF | 0.8543 | 0.8639 | 0.8591 | 11 |
| Rotation Forest + AdaBoost RF | 0.8543 | 0.8639 | 0.8591 | 18 |
| Smote 301 Rotation Forest RF | 0.8940 | 0.8231 | 0.8586 | 4 |
| Rotation Forest J48 | 0.8079 | 0.9048 | 0.8564 | 38 |
| Smote 100 LogitBoost RF | 0.8278 | 0.8844 | 0.8561 | 16 |
| Smote 100 Rotation Forest + Adaboost J48 | 0.8013 | 0.9048 | 0.8530 | 36 |
| Smote 100 Bagging + Adaboost J48 | 0.8278 | 0.8776 | 0.8527 | 20 |
| Smote 150 Bagging + Adaboost J48 | 0.8278 | 0.8776 | 0.8527 | 21 |
| Smote 100 Rotation Forest RF | 0.8278 | 0.8776 | 0.8527 | 26 |
| Smote 100 Bagging RF | 0.8344 | 0.8707 | 0.8526 | 23 |
| Smote 150 AdaBoost J48 | 0.8411 | 0.8639 | 0.8525 | 30 |
| Smote 301 LogitBoost RF | 0.8609 | 0.8435 | 0.8522 | 1 |
| Smote 150 Bagging + Logitboost RF | 0.8344 | 0.8639 | 0.8492 | 24 |
| Smote 100 AdaBoost J48 | 0.8609 | 0.8367 | 0.8488 | 35 |
| Smote 301 Rotation Forest + Adaboost RF | 0.8808 | 0.8163 | 0.8486 | 5 |
| Smote 301 - RF | 0.8874 | 0.8095 | 0.8485 | 6 |
| Bagging + AdaBoost RF | 0.8079 | 0.8844 | 0.8462 | 32 |
| Smote 301 Rotation Forest + Adaboost RF | 0.8212 | 0.8707 | 0.8460 | 29 |
| Smote 100 Bagging RF | 0.8079 | 0.8776 | 0.8427 | 33 |
| Smote 301 Bagging + Adaboost J48 | 0.8278 | 0.8571 | 0.8425 | 3 |
| Smote 100 Rotation Forest J48 | 0.8278 | 0.8571 | 0.8425 | 25 |
| Smote 301 Bagging + Adaboost RF | 0.8344 | 0.8503 | 0.8424 | 28 |
| Smote 301 Rotation Forest + Adaboost J48 | 0.8344 | 0.8503 | 0.8424 | 2 |
| AdaBoost RF | 0.8543 | 0.8299 | 0.8421 | 8 |
| Smote 100 Rotation Forest J48 | 0.8609 | 0.8163 | 0.8386 | 22 |
| Smote 301 Bagging + Logitboost RF | 0.8742 | 0.8027 | 0.8384 | 7 |
| Smote 301 Bagging RF | 0.8742 | 0.8027 | 0.8384 | 10 |
| Bagging + AdaBoost RF | 0.7616 | 0.9116 | 0.8366 | 39 |
| Smote 150 - RF | 0.8079 | 0.8639 | 0.8359 | 19 |
| Smote 100 - RF | 0.8079 | 0.8639 | 0.8359 | 31 |
| Smote 301 Rotation Forest J48 | 0.8543 | 0.8095 | 0.8319 | 12 |
| Smote 301 Bagging + Adaboost RF | 0.8742 | 0.7891 | 0.8316 | 9 |
| Smote 301 - J48 | 0.8940 | 0.7687 | 0.8314 | 40 |

*Table 10 Deployed models*

# 4 Results

This section elaborates on the top 15 deployed algorithms, this is the maximum number of models that was allowed as a challenge entry.

| Model | Se | Sp | MAcc | Rank based on training F-score |
|---|---|---|---|---|
| Smote 100 AdaBoost J48 | 0.8609 | 0.8776 | 0.8692 | 34 |
| Smote 150 Rotation Forest RF | 0.8609 | 0.8707 | 0.8658 | 17 |
| LogitBoost RF | 0.8013 | 0.9252 | 0.8632 | 37 |
| Smote 100 Rotation Forest + Adaboost J48 | 0.8344 | 0.8912 | 0.8628 | 13 |
| Smote 301 AdaBoost J48 | 0.8742 | 0.8503 | 0.8623 | 14 |
| Smote 150 Bagging + Adaboost J48 | 0.8278 | 0.8912 | 0.8595 | 15 |
| Smote 150 AdaBoost RF | 0.8278 | 0.8912 | 0.8595 | 27 |
| Smote 150 LogitBoost RF | 0.8543 | 0.8639 | 0.8591 | 11 |
| Rotation Forest + AdaBoost RF | 0.8543 | 0.8639 | 0.8591 | 18 |
| Smote 301 Rotation Forest RF | 0.8940 | 0.8231 | 0.8586 | 4 |
| Rotation Forest J48 | 0.8079 | 0.9048 | 0.8564 | 38 |
| Smote 100 LogitBoost RF | 0.8278 | 0.8844 | 0.8561 | 16 |
| Smote 100 Rotation Forest + Adaboost J48 | 0.8013 | 0.9048 | 0.8530 | 36 |
| Smote 100 Bagging + Adaboost J48 | 0.8278 | 0.8776 | 0.8527 | 20 |
| Smote 150 Bagging + Adaboost J48 | 0.8278 | 0.8776 | 0.8527 | 21 |

*Table 11 Top 15 models*

## 4.1 Base algorithms

As mentioned before none of the support vector machines made it to the deployment stage. So they are also absent in the top 15. Although it has to be pointed out that only 8 of them were created. The 15 J48 and RF algorithms score a MAcc in the range of 0.8527 to 0.8692. There seems to be no correlation between the F-score in training and the final MAcc- score The top five is made up out of algorithms scoring a MAcc above 0.86. Three of them are J48 based, two RF. Originally there were 24 J48-based models created versus 32 RF.

## 4.2 Oversampling

Five out of 15 models use SMOTE 100, SMOTE 150 makes up another quintet . SMOTE 301 is with 2 appearances less present in the top 15. The same goes for the non-SMOTE models, only two of them made it to the top. One of them does takes the third spot. This leads us to another conclusion. The top 15 can be plotted as a boxplot based on absolute difference between Sensitivity and Specificity. This results in a Q1 of 0.0167, Q2 of 0.0566 and a Q3 of 0.0709. Three values lie above Q3, two of them belong to the two models that don't use SMOTE. The model which is ranked third has a specify that is 0.1239 higher than is sensitivity (0.8013).

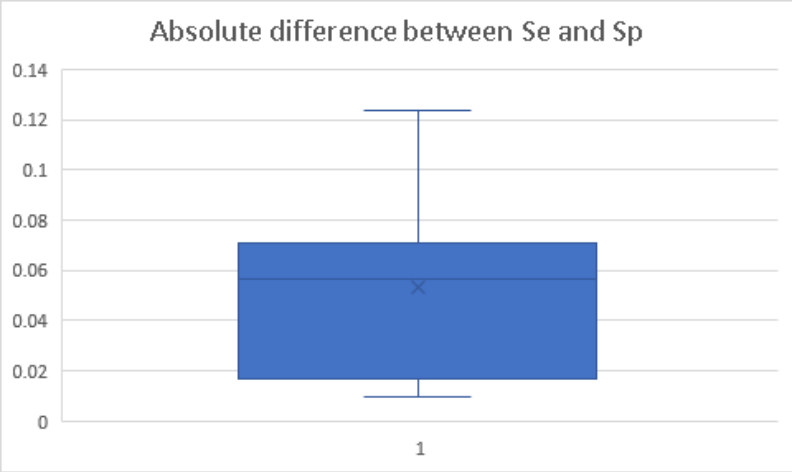


*Figure 11 Boxplot of absolute difference between Se and Sp*

## 4.3 Boosting

All of the top tier decision trees use some kind of additional algorithm. Of the 8 J48- based models 7 are used in combination with AdaBoost. The Rotation Forest J48 model is the only one that does not use AdaBoost or SMOTE. It scores a bit (0.0027) less than its AdaBoost counterpart (0.8591). Of the 7 RF models only two are combined with Adaboost. Six ensembles, made by the Rotation Forest algorithm are also good performers. The bagging algorithms only got into the top 15 when they use a combination of Adaboost and J48.

# 5 Discussion

## 5.1 Findings

### 5.1.1 Spectral features

The MFCCs already proved their worth in the feature selection phase, no other spectral feature added as much gain to the model. So the first part of the posed research question was already answered in an early iteration of the Crisp-DM process. This also explains why they are used in both the spectral features and the beat types cluster. The beat type clusters where mainly created to serve as a vehicle for improving the classifying merit of the MFCCs.

### 5.1.2 Base algorithms

The absence of in SVM does not confirm whether or not they result in over fitted models. They just lack the performance to successfully classify heart sound recordings with the given feature set. Decision trees J48 and random forest are the right classification methods. Even the lowest ranked methods have proven their worth. There was no distinguishable relation training F-score in training and the final MAcc- score. This might indicate that there is overfitting, or that the validation set differs much from the training set. The fact that J48 is omnipresent in the top results, despite the lower numbers of trained models can indicate two things. The first being that it is the better more consistent algorithm in this case study. The latter that random forest, in itself an kind of bagging algorithm does not handle the combination with algorithms other than LogitBoost as well.

### 5.1.3 Oversampling

It is certain that the introduction of SMOTE overfitting contributes a lot to this classification problem. Not only does it lead to higher accuracies. It also results into a smaller difference between sensitivity and specificity. The models that were not trained on a SMOTE adjusted dataset had a specificity that was much larger than the sensitivity. This is bad for this medical classification problem, since the detection of abnormal heart rates is more important. Of course a sensitivity that outweighs the specificity is also wrong. Diagnosing a healthy patient with a disease is bad, but is can be considered the lesser evil. The balance between sensitivity and specificity is a vital part of the built model .Due to the lack of performance of the SVM's the effect of SMOTE is unknown. While training, the F-score went up slightly with a higher SMOTE. This contradicts the results if Ibarra-Hernández et al (Ibarra-Hernández et al., 2018) where SMOTE has a larger negative impact on SVMs. But since the 6 SMV's where not deployed it is hard to draw conclusions.

### 5.1.4 Boosting

The deployed models show that boosting performs better than random forest. Random Forest is better than bagging which in itself is more effective than a single tree. It is not clear where rotating forest fit in this order. There is a disadvantage to rotating forests that can't be derived from this case study. They harvest a lot of time and computing power. This is logical because in some cases an ensemble was created for an algorithm that boost another algorithm that votes about multiple instances of a single tree.

## 5.2 Comparison to other models

| Rank | Entrant | Se | Sp | MAcc | Method |
|---|---|---|---|---|---|
| 1 | Potes et al. | 0.9424 | 0.7781 | 0.8602 | AdaBoost & CNN |
| 2 | Zabihi et al. | 0.8691 | 0.849 | 0.859 | Ensemble of SVMs |
| 3 | Kay & Agarwal | 0.8743 | 0.8297 | 0.852 | Regularized Neural Network |
| 4 | Bobillo | 0.8639 | 0.8269 | 0.8454 | MFCCs, Wavelets, Tensors & KNN |
| 5 | Homsi et al. | 0.8848 | 0.8048 | 0.8448 | Random Forest + LogitBoost |
| 6 | Maknickas | 0.8063 | 0.8766 | 0.8415 | no publication |
| 7 | Plesinger et al. | 0.7696 | 0.9125 | 0.8411 | Probability-distribution based |
| 8 | Rubin et al. | 0.7278 | 0.9521 | 0.8399 | Convolutional NN with MFCs |

*Table 12 Top eight entries PhysioNet Challenge (G. D. Clifford et al., 2016)*

The created models can be compared to the results listed in the challenge's proceedings. These are the scores of the top eight out of the 348 models that were submitted by 48 participants. The best model in this case study (Smote 100 AdaBoost J48) scores a MAcc of 0.8692 this is higher than the top model by Potes et al. The model of Potes et al. combines AdaBoost and CNN and achieves a MAcc of 0. 8602. The lowest ranking model in the top 15 of deployed models (Smote 150 Bagging + Adaboost J48) still scores better than the third best model (by Kay & Agarwal). In this case study a deployed model uses LogitBoost RF this was the same approach as Homsi et al. who scores 0.8448. The deployed model scores better (0.8632), this might indicate that the features in this case study are better. The model of Homsi et al. does seem to encounter the same problem concerning the higher sensitivity, as was discussed in the sections above. It has to be mentioned that these authors had to commit a "blind" entry to the challenge. Although there was already a lot of literature available on the matter, the heart sound dataset itself was brand new at the moment their models were conceived. So they couldn't benefit from the findings and literature that resulted from this challenge. Kudos to the pioneers the results in this case study also reflect their rigorous efforts .

## 5.3 Limitations

The segmentation of the beats for the beat type clustering was done using an onset method. This choice was made to leverage faster computation times. Although the number of beats extracted from the dataset was in the same magnitude as the springer algorithm, the real performance of this method is unknown. The method has not yet been verified to acquire the right segments of a heartbeat. The exact accuracy was in this case not the main goal of this method. Since clustering is an unsupervised method and the primary goal was to segment the audio files into comparable parts. It must be acknowledged that the springer method is the right way to extract phases from heart beat recordings

While building the model the combination of rotating forest for a "logitboosted" random forest failed due to the lack of computing power. This is a pity since the rotating forest for a normal random forest algorithm has proven to be the second best algorithm. The third place was achieved by the combination of LogitBoost and random forest. It makes one wonder whether the combination of place two and three would result into a superior model.

An endeavour was undertaken to improve the best ranking models. Starting at the tree level different parameters like confidence level where adjusted in order to optimize the algorithm during training. These efforts did not yet yield better results, the limitation in this affair: a lack of time. The main power of this case study lies in the feature extraction part. Good features are the best gateway to a model that performs well. Hence most of the available time was invested in this step.

# 6 Conclusion

The model proposed in this case study can be used to carry out the classification of heart sound recordings into normal and abnormal. The answer to the research question: *"Does combining oversampling with spectral features other than Linear Predictive Coding coefficients lead to a better classifier?"* is affirmative. The main spectral feature improving this model are the Mel Frequency Coefficients, the positive effect of oversampling is also confirmed. However although the proposed model's accuracy outperforms previous attempts to solve this problem, the accuracy is not high enough to be used successfully in a professional medical context. It can be concluded that a J48 decision tree performs best in combination with AdaBoost after oversampling the database with a SMOTE rate of 150. The SMOTE oversampling improves the gain ratio of the features with less classifying abilities. Adaboost succeeds in levering the power of combining these attributes The use of SMOTE results in a higher accuracy and also significantly lowers the difference between the sensitivity and specificity. From here on forward several proposals for future research can be done. The first one being implementing the clustering of beat types with the Springer segmentation algorithm. This implementation needs to be tested against the onset detection method for creating clustered beats. A second proposal is testing different algorithms. The combination of rotation forest and LogitBoost on random forest that failed to build can be tested. Another model that might achieve better results is a convolutional neural network, like the one used by Potes et al. The last proposal is the further optimisation of the parameters of the current model.

# REFERENCES


AHA, A. H. A. (2019a). Atrial Septal Defect (ASD). Retrieved on the 20th of april 2019 from https://www.heart.org/en/health-topics/congenital-heart-defects/about-congenital-heart-defects/atrial-septal-defect-asd

AHA, A. H. A. (2019b). Heart Murmurs and Valve Disease. Retrieved on the 20th of april 2019 from https://www.heart.org/en/health-topics/heart-valve-problems-and-disease/heart-valve-problems-and-causes/heart-murmurs-and-valve-disease

AHA, A. H. A. (2019c). Patent Ductus Arteriosus (PDA). Retrieved on the 20th of april 2019 from https://www.heart.org/en/health-topics/congenital-heart-defects/about-congenital-heart-defects/patent-ductus-arteriosus-pda

AHA, A. H. A. (2019d). Problem: Aortic Valve Regurgitation. Retrieved on the 20th of april 2019 from https://www.heart.org/en/health-topics/heart-valve-problems-and-disease/heart-valve-problems-and-causes/problem-aortic-valve-regurgitation

AHA, A. H. A. (2019e). Problem: Aortic Valve Stenosis. Retrieved on the 20th of april 2019 from https://www.heart.org/en/health-topics/heart-valve-problems-and-disease/heart-valve-problems-and-causes/problem-aortic-valve-stenosis

AHA, A. H. A. (2019f). Problem: Mitral Valve Prolapse. Retrieved on the 20th of april 2019 from https://www.heart.org/en/health-topics/heart-valve-problems-and-disease/heart-valve-problems-and-causes/problem-mitral-valve-prolapse

AHA, A. H. A. (2019g). Problem: Mitral Valve Regurgitation. Retrieved on the 20th of april 2019 from https://www.heart.org/en/health-topics/heart-valve-problems-and-disease/heart-valve-problems-and-causes/problem-mitral-valve-regurgitation

AHA, A. H. A. (2019h). Problem: Mitral Valve Stenosis. Retrieved on the 20th of april 2019 from https://www.heart.org/en/health-topics/heart-valve-problems-and-disease/heart-valve-problems-and-causes/problem-mitral-valve-stenosis

AHA, A. H. A. (2019i). Problem: Pulmonary Valve Regurgitation. Retrieved on the 20th of april 2019 from https://www.heart.org/en/health-topics/heart-valve-problems-and-disease/heart-valve-problems-and-causes/problem-pulmonary-valve-regurgitation

AHA, A. H. A. (2019j). Problem: Pulmonary Valve Stenosis. Retrieved on the 20th of april 2019 from https://www.heart.org/en/health-topics/heart-valve-problems-and-disease/heart-valve-problems-and-causes/problem-pulmonary-valve-stenosis

AHA, A. H. A. (2019k). Problem: Tricuspid Valve Regurgitation. Retrieved on the 20th of april 2019 from https://www.heart.org/en/health-topics/heart-valve-problems-and-disease/heart-valve-problems-and-causes/problem-tricuspid-valve-regurgitation

AHA, A. H. A. (2019l). Problem: Tricuspid Valve Stenosis. Retrieved on the 20th of april 2019 from https://www.heart.org/en/health-topics/heart-valve-problems-and-disease/heart-valve-problems-and-causes/problem-tricuspid-valve-stenosis

AHA, A. H. A. (2019m). Ventricular Septal Defect (VSD). Retrieved on the 20th of april 2019 from https://www.heart.org/en/health-topics/congenital-heart-defects/about-congenital-heart-defects/ventricular-septal-defect-vsd

AnalogICTips. (2017). Basics of bandpass filters. In: AnalogICTips.

Ari, S., Kumar, P., & Saha, G. (2008). A robust heart sound segmentation algorithm for commonly occurring heart valve diseases. *Journal of Medical Engineering & Technology, 32*(6), 456-465.

Bello, J. P., Daudet, L., Abdallah, S., Duxbury, C., Davies, M., & Sandler, M. B. (2005). A tutorial on onset detection in music signals. *Ieee Transactions on Speech and Audio Processing, 13*(5), 1035-1047.

Beyramienanlou, H., & Lotfivand, N. (2017). Shannon's Energy Based Algorithm in ECG Signal Processing. *Computational and Mathematical Methods in Medicine*.

Bhaskar, A. (2012). A simple electronic stethoscope for recording and playback of heart sounds. *Advances in physiology education, 36*(4), 360-362.

Bobillo, I. J. D. (2016, 11-14 Sept. 2016). *A tensor approach to heart sound classification.* Paper presented at the 2016 Computing in Cardiology Conference (CinC).
Brown, J. C. (1991). CALCULATION OF A CONSTANT-Q SPECTRAL TRANSFORM. *Journal of the Acoustical Society of America, 89*(1), 425-434.
Butterworth, S. (1930). On the Theory of Filter Amplifiers. In (Vol. 7, pp. 536–541): Wireless Engineer.
Cesarelli, M., Ruffo, M., Romano, M., & Bifulco, P. (2012). Simulation of foetal phonocardiographic recordings for testing of FHR extraction algorithms. *Computer Methods and Programs in Biomedicine, 107*(3), 513-523
Chen, T., Kuan, K., Celi, L. A. G., & Clifford, G. D. (2010). Intelligent Heartsound Diagnostics on a Cellphone Using a Hands-Free Kit. *AAAI Spring Symp. on Articial Intelligence for Development (Stanford University)*, 26–31.
Clifford, G. D., Liu, C., Moody, B., Springer, D., Silva, I., Li, Q., & Mark, R. G. (2017). Classification of normal/abnormal heart sound recordings: The PhysioNet/Computing in Cardiology Challenge
Clifford, G. D., Liu, C. Y., Moody, B., Springer, D., Silva, I., Li, Q., & Mark, R. G. (2016). *Classification of Normal/Abnormal Heart Sound Recordings: the PhysioNet/Computing in Cardiology Challenge 2016.*
Deng, L., & O'Shaughnessy, D. (2018). *Speech processing: a dynamic and optimization-oriented approach*: CRC Press.
Downie, J. S. (2012). 13th International Society for Music Information Retrieval Conference Retrieved on the 5th of may 2019 from https://www.music-ir.org/mirex/wiki/2012:MIREX_Home
Eerola, T., Ferrer, R., & Alluri, V. (2012). TIMBRE AND AFFECT DIMENSIONS: EVIDENCE FROM AFFECT AND SIMILARITY RATINGS AND ACOUSTIC CORRELATES OF ISOLATED INSTRUMENT SOUNDS. *Music Perception, 30*(1), 49-70.
Feldman, M. (1997). Non-linear free vibration identification via the Hilbert transform. *Journal of Sound and Vibration, 208*(3), 475-489.
Friedman, J., Hastie, T., & Tibshirani, R. (2000). Additive logistic regression: a statistical view of boosting (with discussion and a rejoinder by the authors). *The annals of statistics, 28*(2), 337-407.
Fuster, V. (2016). The Stethoscope's Prognosis Very Much Alive and Very Necessary. *Journal of the American College of Cardiology, 67*(9), 1118-1119.
Gamero, L. G., & Watrous, R. (2003, 17-21 Sept. 2003). *Detection of the first and second heart sound using probabilistic models.* Paper presented at the Proceedings of the 25th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (IEEE Cat. No.03CH37439).
Gill, D., Gavrieli, N., & Intrator, N. (2005, 25-28 Sept. 2005). *Detection and identification of heart sounds using homomorphic envelogram and self-organizing probabilistic model.* Paper presented at the Computers in Cardiology, 2005.
Grey, J. M., & Gordon, J. W. (1978). PERCEPTUAL EFFECTS OF SPECTRAL MODIFICATIONS ON MUSICAL TIMBRES. *Journal of the Acoustical Society of America, 63*(5), 1493-1500. Retrieved from <Go to ISI>://WOS:A1978FB40800030. doi:10.1121/1.381843
Gupta, C. N., Palaniappan, R., Swaminathan, S., & Krishnan, S. M. (2007). Neural network classification of homomorphic segmented heart sounds. *Applied Soft Computing, 7*(1), 286-297.
Heideman, M., Johnson, D., & Burrus, C. (1984). Gauss and the history of the fast fourier transform. *IEEE ASSP Magazine, 1*(4), 14-21.
Homsi, M. N., Medina, N., Hernandez, M., Quintero, N., Perpiñan, G., Quintana, A., & Warrick, P. (2016, 11-14 Sept. 2016). *Automatic heart sound recording classification using a nested set of ensemble algorithms.* Paper presented at the 2016 Computing in Cardiology Conference (CinC).
HuFF, D. A. R. R. E. L. L., & How, T. O. (1954). How to Lie with Statistics. W W. *W. Norton & Co., Inc., New York*.

Ian H. Witten, E. F., Mark A. Hal. (2019). Data Mining: Practical Machine Learning Tools and Techniques. Retrieved on the 5th of may 2019 from https://www.cs.waikato.ac.nz/~ml/weka/book.html

Ibarra-Hernández, R. F., Bertin, N., Alonso-Arévalo, M. A., & Guillén-Ramírez, H. A. (2018). *A benchmark of heart sound classification systems based on sparse decompositions* (Vol. 10975).

Jiang, D. N., Lu, L., Zhang, H. J., Tao, J. H., Cai, L. H., & Ieee. (2002). Music type classification by spectral contrast feature. *Ieee International Conference on Multimedia and Expo, Vol I and Ii, Proceedings*, 113-116.

Johnson, C. R., Sethares, W. A., & Klein, A. G. (2011). *Software Receiver Design: Build your Own Digital Communication System in Five Easy Steps*: Cambridge University Press.

jupyter.org. (2019). The Jupyter Notebook 6.0.0.dev documentation. Retrieved on the 5th of may 2019 from https://jupyter-notebook.readthedocs.io/en/latest/notebook.html

Jurafsky, D., & Martin, J. H. (2014). *Speech and language processing* (Vol. 3): Pearson London.

Kay, E., & Agarwal, A. (2016, 11-14 Sept. 2016). *DropConnected neural network trained with diverse features for classifying heart sounds.* Paper presented at the 2016 Computing in Cardiology Conference (CinC).

Kégl, B. (2013). The return of AdaBoost. MH: multi-class Hamming trees. *arXiv preprint*

Kotsiantis, S. (2013). Rotation Forest with Logitboost. *International Journal of Innovative Computing, Information and Control (IJICIC), 9*(3), 1087-1094.

Kudriavtsev, V., Polyshchuk, V., & Roy, D. L. (2007). Heart energy signature spectrogram for cardiovascular diagnosis. *Biomedical Engineering Online, 6*, 22. Retrieved from <Go to ISI>://WOS:000247454700001. doi:10.1186/1475-925x-6-16

Legget, M. E., Toh, M., Meintjes, A., Fitzsimons, S., Gamble, G., & Doughty, R. N. (2018). Digital devices for teaching cardiac auscultation-a randomized pilot study. *Medical education online, 23*(1), 1524688.

Leis, J. (2011). *Digital signal processing using MATLAB for students and researchers*: Wiley Online Library.

Liu, C., Springer, D., Li, Q., Moody, B., Abad Juan, R., J Chorro, F., . . . D Clifford, G. (2016). *An open access database for the evaluation of heart sound algorithms* (Vol. 37).

MathWorks. (2019a). Envelope Extraction - MATLAB & Simulink - MathWorks Benelux. Retrieved on the 5th of may 2019 from https://nl.mathworks.com/help/signal/ug/envelope-extraction-using-the-analytic-signal.html

MathWorks. (2019b). MATLAB - MathWorks. Retrieved on the 5th of may 2019 from https://nl.mathworks.com/products/matlab.html

MathWorks. (2019c). Signal envelope - MATLAB envelope - MathWorks Benelux. Retrieved on the 5th of may 2019 from https://nl.mathworks.com/help/signal/ref/envelope.html

Matplotlib. (2019). Matplotlib 3.1.0 documentation. Retrieved from https://matplotlib.org/

McFee, B., Raffel, C. A., Liang, D., Ellis, D. P. W., McVicar, M., Battenberg, E., & Nieto, O. (2019). librosa: Audio and Music Signal Analysis in Python.

Mortelmans, C. (2016). Validation of a new smartphone application ("FibriCheck") for the diagnosis of atrial fibrillation in primary care. In: Leuven.

Moser, M., Lehofer, M., Sedminek, A., Lux, M., Zapotoczky, H. G., Kenner, T., & Noordergraaf, A. (1994). HEART-RATE-VARIABILITY AS A PROGNOSTIC TOOL IN CARDIOLOGY - A CONTRIBUTION TO THE PROBLEM FROM A THEORETICAL POINT-OF-VIEW. *Circulation, 90*(2), 1078-1082.

Naseri, H., Homaeinezhad, M. R., & Pourkhajeh, H. (2013). Noise/spike detection in phonocardiogram signal as a cyclic random process with non-stationary period interval. *Computers in Biology and Medicine, 43*(9), 1205-1213.

NumPy. (2019). NumPy v1.16 Manual. Retrieved on the 5th of may 2019 from https://docs.scipy.org/doc/numpy/user/whatisnumpy.html

OpenStax. (2019). Anatomy and Physiology - OpenStax CNX. Retrieved from https://cnx.org/contents/FPtK1zmh@15.2:IsP5aaud/19-3-Cardiac-Cycle.
Oskiper, T., & Watrous, R. (2002). Detection of the first heart sound using a time-delay neural network. *Computers in Cardiology 2002, Vol 29, 29*, 537-540
Physionet. (2016). Classification of Normal/Abnormal Heart Sound Recordings: the PhysioNet/Computing in Cardiology Challenge 2016. Retrieved from https://physionet.org/challenge/2016/
Potes, C., Parvaneh, S., Rahman, A., & Conroy, B. (2016, 11-14 Sept. 2016). *Ensemble of feature-based and deep learning-based classifiers for detection of abnormal heart sounds.* Paper presented at the 2016 Computing in Cardiology Conference (CinC).
Project, T. p. (2019). pandas: Python Data Analysis Library. Retrieved on the 15th of march 2019from https://pandas.pydata.org/about.html
Rubin, J., Abreu, R., Ganguli, A., Nelaturi, S., Matei, I., & Sricharan, K. (2016, 11-14 Sept. 2016). *Classifying heart sound recordings using deep convolutional neural networks and mel-frequency cepstral coefficients.* Paper presented at the 2016 Computing in Cardiology Conference (CinC).
Schmidt, S. E., Holst-Hansen, C., Graff, C., Toft, E., & Struijk, J. J. (2010). Segmentation of heart sound recordings by a duration-dependent hidden Markov model. *Physiological Measurement, 31*(4), 513-529.
scikit-learn. (2019). scikit-learn 0.21.1 documentation. Retrieved on the 15th of march 2019 from https://scikit-learn.org/stable/index.html#
SciPy.org. (2019). SciPy. Retrieved on the 15th of march 2019 from https://scipy.org/scipylib/index.html
Sebastian, B., & Widmer, G. (2013). *Maximum Filter Vibrato Suppression for Onset Detection.* Paper presented at the Proceedings of the 16th International Conference on Digital Audio Effects (DAFx-13), Maynooth, Ireland.
Sejdić, E., Djurović, I., & Jiang, J. (2009). Time–frequency feature representation using energy concentration: An overview of recent advances. *Digital Signal Processing, 19*(1), 153-183. Shepard, R. N. (1964). CIRCULARITY IN JUDGMENTS OF RELATIVE PITCH. *Journal of the Acoustical Society of America, 36*(12), 2346
Springer, D. B., Tarassenko, L., & Clifford, G. D. (2016). Logistic Regression-HSMM-Based Heart Sound Segmentation. *IEEE Transactions on Biomedical Engineering, 63*(4), 822-832.
Stiltz, H. L. (1961). *Aerospace Telemetry*: Prentice-Hall.
Thiyagaraja, S. R., Dantu, R., Shrestha, P. L., Chitnis, A., Thompson, M. A., Anumandla, P. T., . . . Dantu, S. (2018). A novel heart-mobile interface for detection and classification of heart sounds. *Biomedical Signal Processing and Control, 45*, 313-324.
). *Fitting the Mel scale.* Paper presented at the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP 99), Phoenix, Az.
Varghees, V. N., & Ramachandran, K. I. (2014). A novel heart sound activity detection framework for automated heart sound analysis. *Biomedical Signal Processing and Control, 13*, 174-188.
Vukanovic-Criley, J. M., Hovanesyan, A., Criley, S. R., Ryan, T. J., Plotnick, G., Mankowitz, K., . . . Criley, J. M. (2010). Confidential Testing of Cardiac Examination Competency in Cardiology and Noncardiology Faculty and Trainees: A Multicenter Study. *Clinical Cardiology, 33*(12), 738-745.
Weckesser, W. (2019). How to implement band-pass Butterworth filter with Scipy.signal.butter. Retrieved on the 12th of march 2019 from https://stackoverflow.com/questions/12093594/how-to-implement-band-pass-butterworth-filter-with-scipy-signal-butter
WHO. (2019a). Global Atlas on cardiovascular disease prevention and contro. *WHO*. Retrieved on the 12th of march 2019 from https://www.who.int/cardiovascular_diseases/hearts/en/.

WHO. (2019b). Hearts: technical package for cardiovascular disease management in primary health care. *WHO*. Retrieved on the 12th of march 2019 from https://www.who.int/cardiovascular_diseases/hearts/en/.

Wirth, R., & Hipp, J. (2000). *CRISP-DM: Towards a standard process model for data mining.* Paper presented at the Proceedings of the 4th international conference on the practical applications of knowledge discovery and data mining.

Witten, I. H., Frank, E., Hall, M. A., & Pal, C. J. (2016). *Data Mining: Practical machine learning tools and techniques*: Morgan Kaufmann.

Zabihi, M., Rad, A. B., Kiranyaz, S., Gabbouj, M., & Katsaggelos, A. K. (2016). *Heart sound anomaly and quality detection using ensemble of neural networks without segmentation.* Paper presented at the 2016 Computing in Cardiology Conference (CinC).

Zheng, F., Zhang, G., & Song, Z. (2001). Comparison of different implementations of MFCC. *Journal of Computer Science and Technology*, 16(6), 582-589.