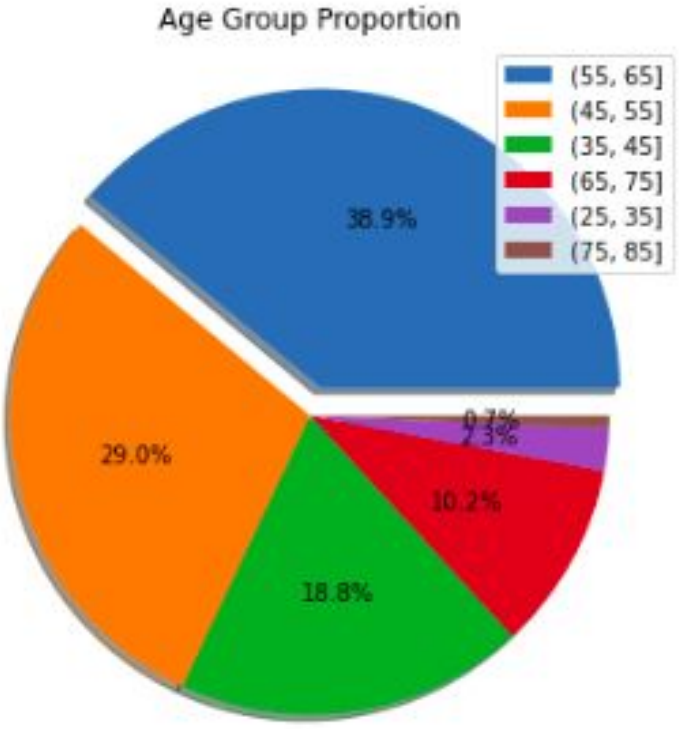
Heart Disease

1. The main goal of this project was to investigate what are the signs of heart disease and what can cause it. Main questions were: “Does age have an influence on heart problems?”, “Is there a dependence between age and heart disease?”, “Can a chest pain type be the sign for diagnosis of heart problem?” and so on. To answer these questions I used python programming language, hypothesis testing and statistical analysis.
2. Analysis:
   2.1. “Does age have an influence on Heart Disease?
      2.1.1. Methods:

Chi square testing, pie plot of age groups

      2.1.2. Analysis:



As we can notice, people between 55 - 65 age visit doctors with heart problems more often than other age groups(38.9% of all dataset). Second place is for the 45-55 age group, it is smaller then first place but

is also very big(29% of all dataset). Third place is the 35-45 age group(18.8% of all dataset).

Let’s do a hypothesis testing to check, does actually all age groups have the same problems with heat.

The null hypothesis is: “All age groups does not have significant differences in heart disease“

Alternative hypothesis: “All age groups have significant difference in heart problem”

After applying chi square testing we can reject null hypothesis and accept alternative hypothesis that age groups have significant difference in heart problem

2.1.3. Conclusions:

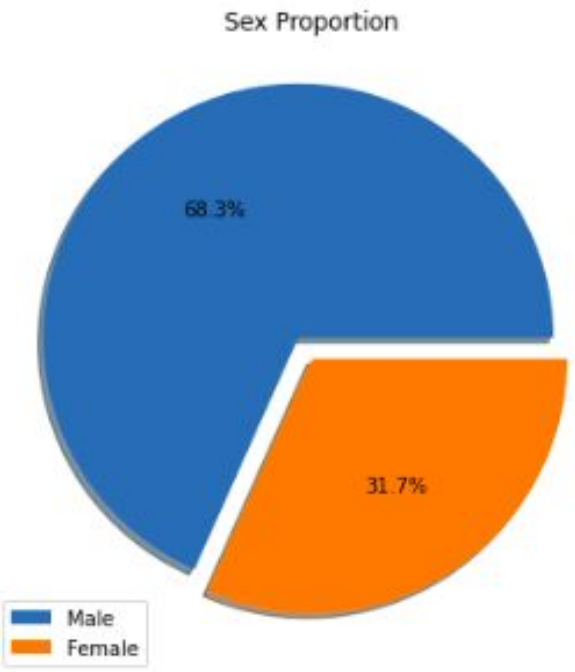
Age may have an influence on heart disease, people between 55-65 age complain about heart the most.

2.2. “Does gender have an influence on heart problems?”

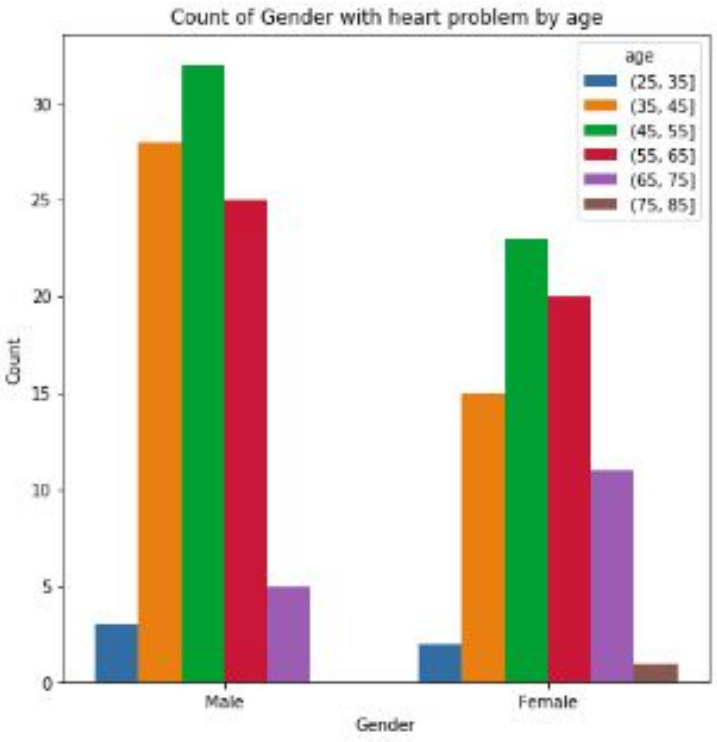
2.2.1. Methods:

Chi square test, count plot with age group partitioning

2.2.2. Analysis:



Here we can observe that male tend to complain about their hearts more than females. Let's investigate the count plot of gender and their age group in those who actually had confirmed heart disease.



In both genders almost the same age group suffer from heart problems equally. From the graph we also can admit that male have

more confirmed heart disease than females. Let's carry out another chi square test to understand is the difference in confirmed heart disease for genders significant.

The null hypothesis is: "Both genders does not have significant differences in heart disease"

Alternative hypothesis: "Both genders have significant difference in heart problem"

After applying chi square testing we can reject null hypothesis and accept alternative hypothesis that both genders have significant difference in heart problem

2.2.3. Conclusions:

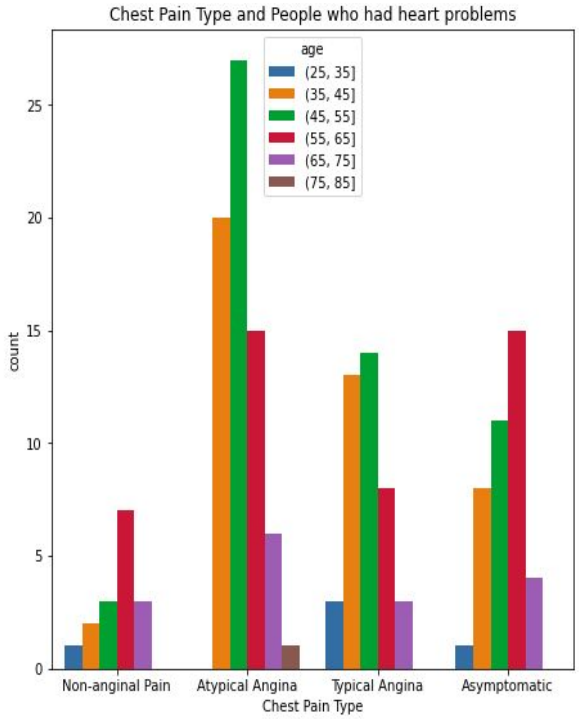
Male complain about their heart more often than females. Both genders tend to be diagnosed with heart disease at the age between 45-55 the most.

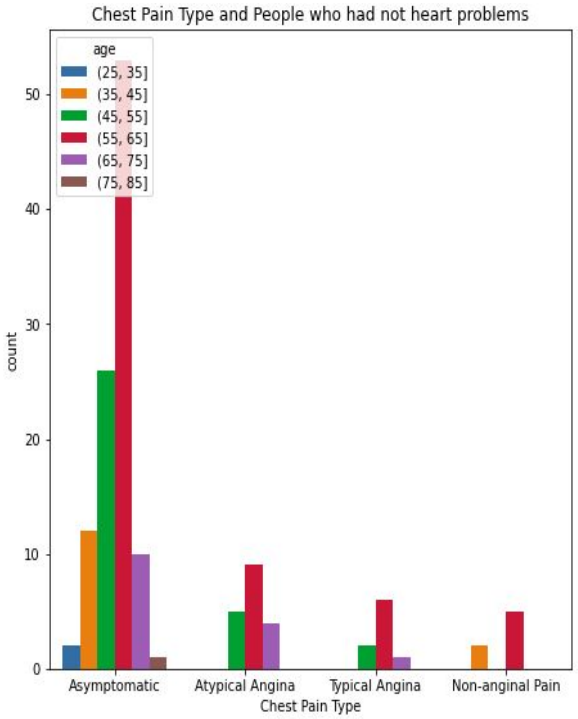
2.3. "Can chest pain type be the sign of heart disease?"

2.3.1. Methods:

Comparing two count plots of people who went to the doctor and did not have heart problems and people who were diagnosed with heart problems with age partitioning.

2.3.2. Analysis:





From this chart we can see two categories of people who had diagnosed heart problems and people who had not that. This chart shows that asymptomatic chest pain type can not be the sign of heart disease, because a lot of people complained of heart having asymptomatic pain type and this diagnosis was not approved. But what can be the sign? I think, we can admit that atypical anginal and typical anginal pain can tell us more about heart, because these are the dominant pain types among people who had heart problem, while asymptomatic chest pain type can also be the sign of heart problem, but it is probably can be the sign of different illness and can not say exactly does the person have problems with heart

2.3.3. Conclusions:

Asymptomatic pain type does not guarantee a heart disease, while if you have atypical or typical anginal pain you probably have a problem with heart.

2.4. “What heart rate can tell us about the health of the heart?”

2.4.1. Methods:

T-test to compare significant difference in mean of heart rate, distribution plot, which shows us frequency of heart rate in different categories.

2.4.2. Analysis:

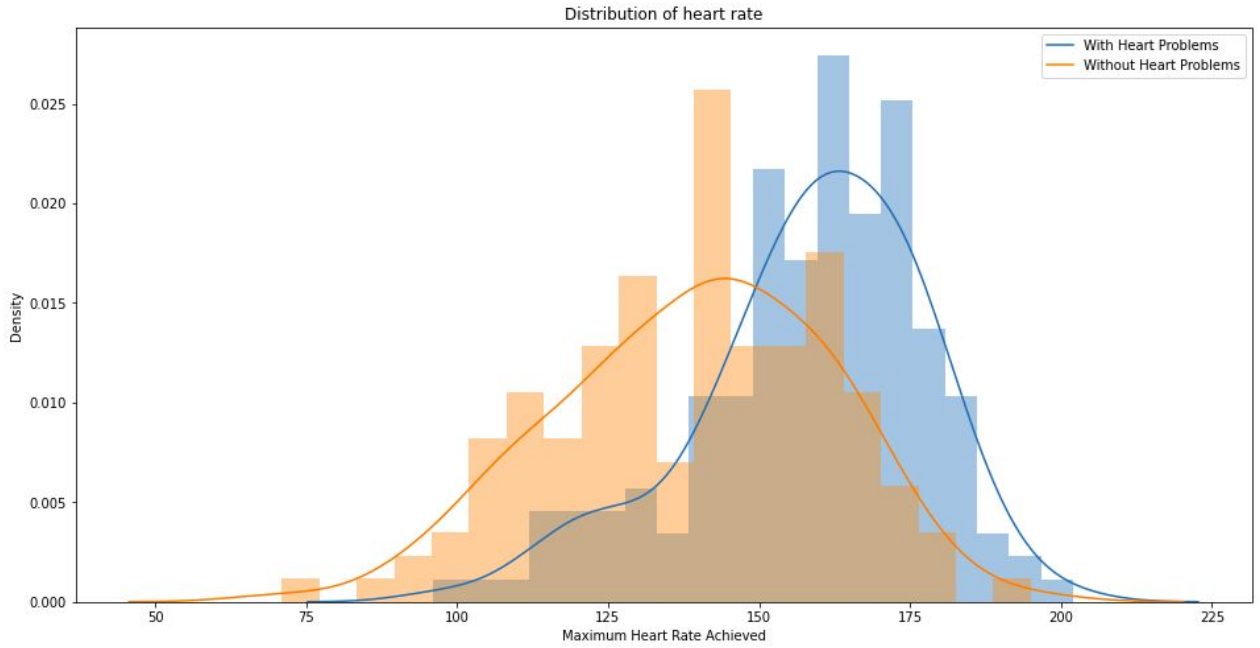
First of all we need to answer the simple question: “Is the heart rate different in people who were diagnosed with heart disease and people who don’t?”. In order to answer I would like to apply t-test.

Mean value of heart rate in people who had disease is 139, and 158 in people who were healthy and t-test showed that it is a significant difference. Let’s explore more! We can also compare it among age groups.

Age group	Mean heart rate in people who had heart problem	Mean heart rate in people who had not heart problem
26-35	184	143
36-45	168	146
46-55	158	138
56-65	152	139
66-75	141	129

We may notice that heart rate also depends on age, but the main thing we must accept is that the mean heart rate in people with heart problems is greater in all ages. It says that heart rate can be a good indicator of upcoming or already existing heart disease in a person.

Let’s compare a distribution plot which shows us the frequency and most common values of heart rate in a people.



As we can see the most common heart rate in people who had not heart problem but complained about heart is around 130 and 160 and heart rate between 155 and 180 is in people who had heart problems.

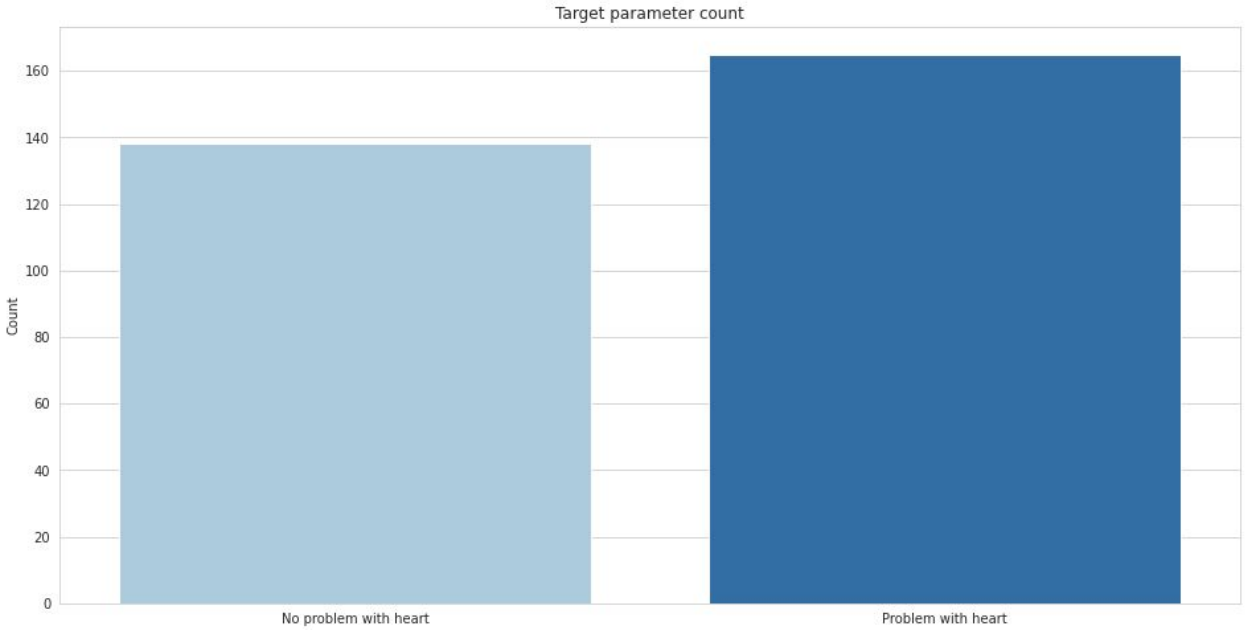
2.4.3. Conclusions:

Heart rate can tell us a lot about heart, people in all age group, who were diagnosed with heart disease, had higher heart rate than people who don't have those.

3. Neural Network predicting heart disease

3.1. "Are our target values equal?"

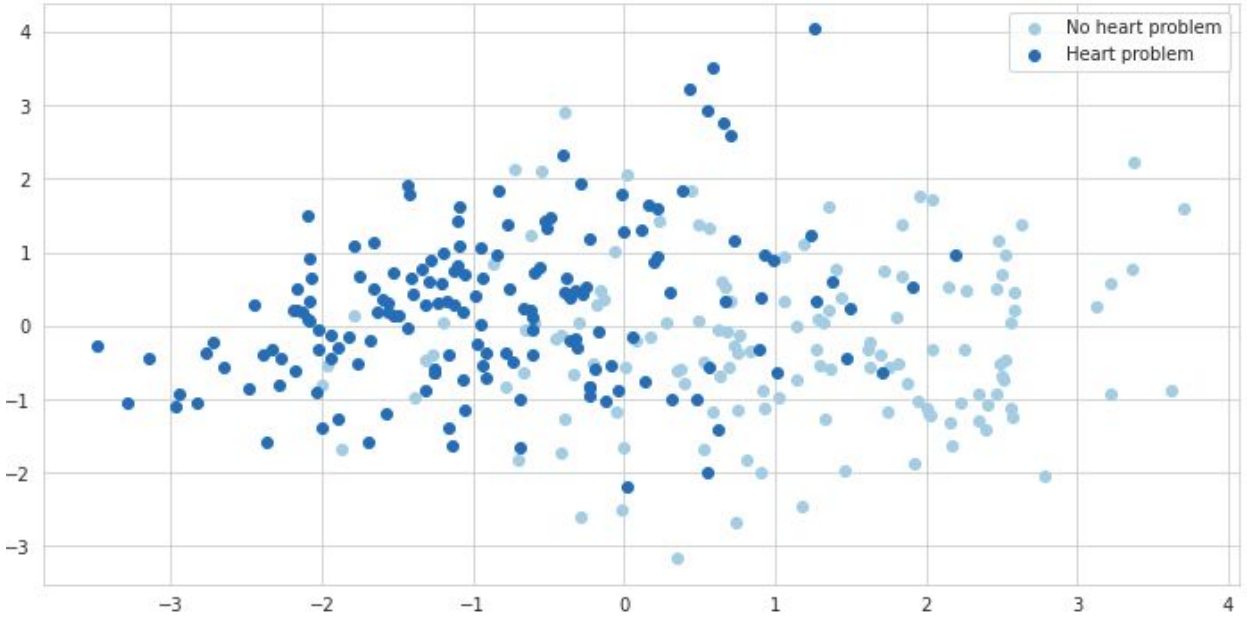
First of all before predicting we should investigate what is the difference in count between people who have heart problems and who don't. Let's show this on our graph!



It seems that they are almost equal, the number of people who have heart disease is 162 and the number of people who don't have problems is 142.

3.2. Plot the dataset on graph, in order to investigate the difference between those two groups

In order to this we will apply Principal Component Analysis(PCA) in order to reduce the amount of parameters and using this we can easily plot the data

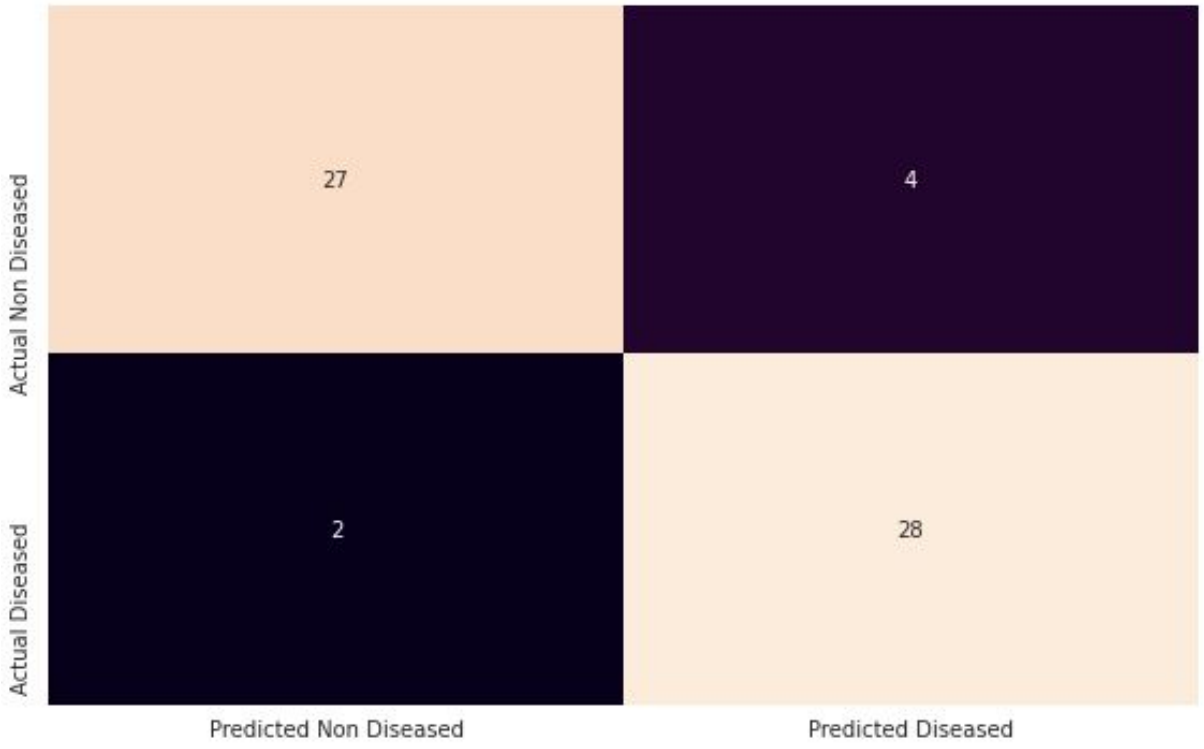


As we can observe in this plot, both these categories have a lot in common, but they are quite different from each other. It will be quite hard to predict heart disease, because of the amount of values in these groups that are quite similar to each other but let's try!

3.3. Prediction

After preprocessing the data we will apply Neural Network on this data in order to create a model for prediction of heart disease. In order to create the model we need to choose the right parameters to achieve the best accuracy of our prediction.

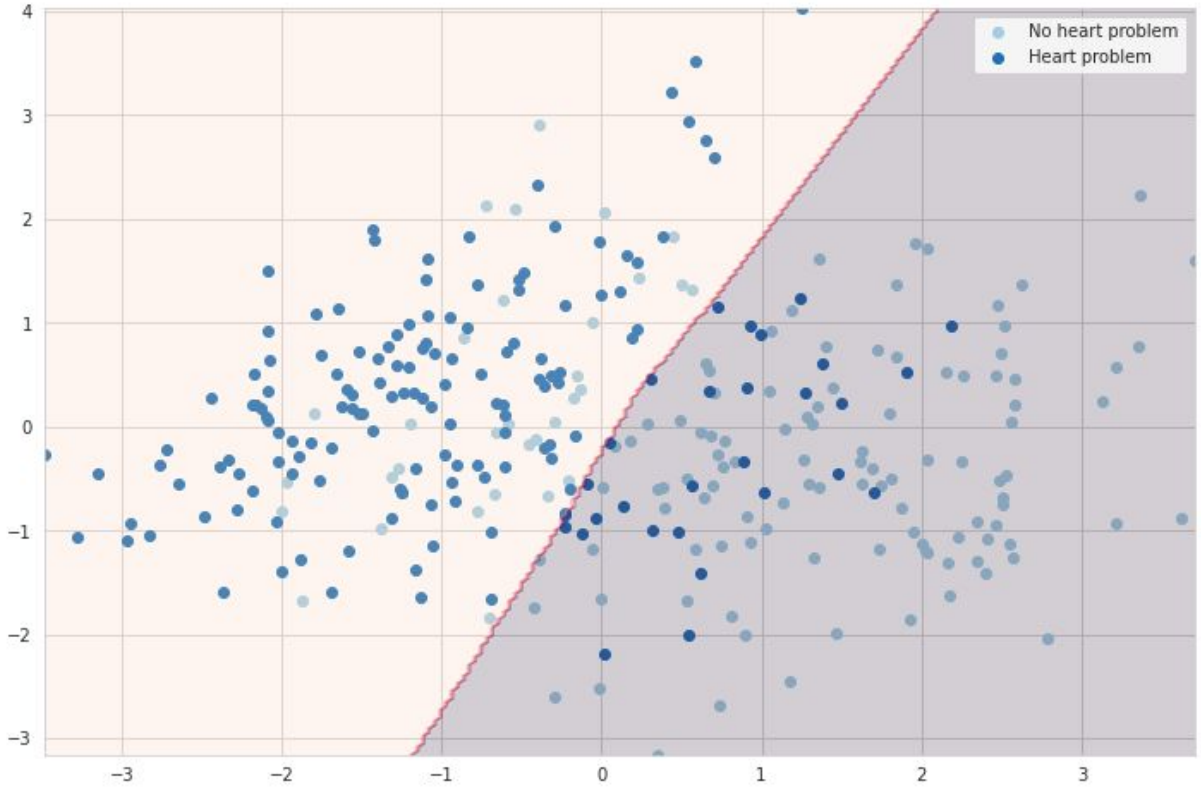
Using the GridSearch algorithm I found that a Neural network with 128 neurons and learning rate of 0.001 showed the best accuracy. The scores of this neural network: on the train set it showed 88% of accuracy, while on test set it showed 90% of accuracy. The score of sensitivity(recall score) of our model is 88% and specificity(precision) of our model is 93% and F1 score is 90%. It seems to be a good prediction model, let's explore where it made a mistake on our testing.



As we can notice it made a mistake 6 times and 2 times of it our model predicted an absence of heart problem while the person had a

problem with heart and 4 times it predicted disease while the person had not any problems with heart.

Bonus! We can also plot a decision boundary in order to check how our model separate two categories of people who have heart problems and who don't



4. Conclusion:

In this article we explore people with heart disease, their chest pain type, heart rate, age and sex in order to get to know who tends to have heart disease and what are the signs of diseases. Besides this we tried to create a neural network in order to predict heart disease or its absence.