|  |  |  |
| --- | --- | --- |
| metin, grafik, grafik tasarım, poster içeren bir resim  Yapay zeka tarafından oluşturulan içerik yanlış olabilir. | **T.C.**  **MANISA CELAL BAYAR UNIVERSITY**  **FACULTY OF ENGINEERING**  **DEPARTMENT OF COMPUTER ENGINEERING** | metin, grafik, grafik tasarım, poster içeren bir resim  Yapay zeka tarafından oluşturulan içerik yanlış olabilir. |

**Python Project**

**PREPARED BY**

**Mehmet Toygun Tutuk 230316003**

**Ragıp Egehan Işık 220316058**

**Zeynep Çağıl 210316020**

**İsmail Esat Uluöz 220316041**

**İbrahim Umut Sezginalp 220316087**

Contents

[HeartAttackRisk2.0 2](#_Toc201611724)

[project.py 2](#_Toc201611725)

[gui.py 3](#_Toc201611726)

[Heart Attack Data Set 3](#_Toc201611727)

[Prediction Result: 4](#_Toc201611728)

# HeartAttackRisk2.0

## project.py

**Purpose:**

The aim is to develop a machine learning model that can predict the risk of heart attack using medical data such as age, cholesterol, and blood pressure of individuals. After data cleaning, visualization, and preprocessing steps, Logistic Regression, SVM, and LightGBM algorithms were applied; the most successful model was selected and recorded. This model was integrated with the user interface to create a practical health prediction system.

Dataset: Heart Attack Data Set.csv

We started the project by examining the dataset; we checked the missing data and cleaned the extreme values ​​with the IQR method. We scaled the numerical data with StandardScaler and made the categorical data suitable for the model with the one-hot encoding method. We used visualization methods such as correlation map and pairplot to understand the relationships between the variables. Then, we separated the data into training and testing and ran tests with Logistic Regression, SVM and LightGBM models. We selected the SVM model that gave the best results and optimized the model parameters using GridSearchCV. Finally, we saved the most successful model and scaler in .pkl format and made it ready to be used in the user interface application.

Hyperparameter optimization was applied to the SVC (Support Vector Classifier) ​​model to provide better results. Initially, the model was trained and tested with default settings; however, it was thought that the accuracy rate obtained in this state could be improved. Therefore, different parameter combinations were tried using the GridSearchCV method and the settings that gave the best results were determined. Thus, the success of the model in heart attack risk predictions was increased, and more accurate and reliable results were obtained. This step increased the overall accuracy of the model, contributing to its becoming more reliable, especially in critical predictions in the field of health.

## gui.py

**Purpose:**

It is a Streamlit-based web interface designed to present the trained machine learning model to the user and to predict heart attack risk.

We developed a user interface that can predict heart attack risk using a machine learning model that we previously trained and recorded. We created the interface with the Streamlit library and designed it to receive numerical values ​​such as age, blood pressure, cholesterol, as well as categorical information such as gender, chest pain type, and ECG results from the user. The entered data is first converted to a suitable format and scaled, then the model makes predictions. According to the result, the user is given feedback as "High risk" or "Low risk" and a pie chart visualization is provided to support this. Thus, the user can evaluate their own heart attack risk through an easy-to-use interface.

## Heart Attack Data Set

The dataset we used includes various medical information related to heart attack risk. In addition to numerical variables such as age, gender, cholesterol, blood pressure, heart rate, the dataset also includes categorical features such as chest pain type, ECG result, and exercise-induced angina. All of these variables provide information that will help determine whether individuals have heart disease.

When we examined the data, we saw that there was no missing data and that there were different numbers of unique values ​​in the columns. Thanks to correlation analyses, we noticed the relationship of variables such as maximum heart rate (thalach) and ST depression (oldpeak) with the target class. Apart from this, there were some outliers in the data, and since we thought that these values ​​could negatively affect the accuracy of the model, we applied the IQR method to clean it.

This dataset was the main source for training and testing the models throughout the project. It also allowed us to practice understanding, visualizing, and preparing the data correctly before modeling. As a result, this dataset was quite suitable and instructive for a machine learning project that could predict the risk of heart attack.

## Prediction Result:

metin, ekran görüntüsü, yazılım, multimedya yazılımı içeren bir resim

Yapay zeka tarafından oluşturulmuş içerik yanlış olabilir.

It shows the prediction made according to the patient information entered into the model. In the visual, when the data of a 45-year-old female patient was analyzed, the model determined that the person's risk of having a heart attack was high and gave the warning "High Risk. Please see a doctor." In addition, this situation was graphically visualized with a pie chart as 70% "At Risk" and 30% "Safe". This result emphasizes that the model can effectively predict risk based on personal health data and that early precautions should be taken.

metin, ekran görüntüsü, yazılım, multimedya yazılımı içeren bir resim

Yapay zeka tarafından oluşturulmuş içerik yanlış olabilir.