Heart Failure Prediction

COURSE PROJECT REPORT

18CSE398J -Machine Learning - Core Concepts with Applications

(2018 Regulation)

III Year/ VI Semester

Academic Year: 2022 -2023 (EVEN)

By

Hardik Khandelwal - RA2011003010227

Samaresh Samanta - RA2011003010247

Aryaman Sharma - RA2011003010667

Under the guidance of

Dr. V. Vijayalakshmi

Professor

Department of Data Science and Business Systems

DEPARTMENT OF DATA SCIENCE AND BUSINESS SYSTEMS

FACULTY OF ENGINEERING AND TECHNOLOGY

SRM INSTITUTE OF SCIENCE AND TECHNOLOGY

Kattankulathur, Kancheepuram

MAY 2023

TABLE OF CONTENT

S. NO.	TITLE	PAGE NO.
1	Abstract	3
2	Introduction	4
3	Dataset	5
4	Methods	6
5	Experiments and Results	7
6	Conclusions and Future work	8
7	References	9

Abstract

Cardiovascular diseases (CVDs) are a leading cause of death worldwide, with heart failure being a common event caused by CVDs. Early detection and management are crucial for people with CVDs or who are at high cardiovascular risk. This heart failure prediction project aims to develop a machine learning model to predict mortality by heart failure using a dataset containing 12 features. The dataset was created by combining five heart datasets, making it the largest heart disease dataset available for research purposes.

Several machine learning models were trained and tested using different techniques, and the best performing model was selected based on various evaluation metrics. The results indicate that the developed model can accurately predict mortality by heart failure, which can aid in early detection and management of CVDs.

Introduction

Cardiovascular diseases (CVDs) are the number 1 cause of death globally, taking an estimated 17.9 million lives each year, which accounts for 31% of all deaths worldwide. Four out of 5 CVD deaths are due to heart attacks and strokes, and one-third of these deaths occur prematurely in people under 70 years of age. Heart failure is a common event caused by CVDs and this dataset contains 11 features that can be used to predict a possible heart disease.

People with cardiovascular disease or who are at high cardiovascular risk (due to the presence of one or more risk factors such as hypertension, diabetes, hyperlipidaemia or already established disease) need early detection and management wherein a machine learning model can be of great help. Machine learning has shown great potential in predicting heart failure mortality, which can aid in early detection and management of CVDs. In this heart failure prediction project, we aim to develop a machine learning model to predict mortality by heart failure using the dataset given

Dataset

The dataset used in this heart failure prediction project contains 11 features, including :-

1. Age: age of the patient [years]
2. Sex: sex of the patient [M: Male, F: Female]
3. ChestPainType: chest pain type [TA: Typical Angina, ATA: Atypical Angina, NAP: Non-Anginal Pain, ASY: Asymptomatic]
4. RestingBP: resting blood pressure [mm Hg]
5. Cholesterol: serum cholesterol [mm/dl]
6. FastingBS: fasting blood sugar [1: if FastingBS > 120 mg/dl, 0: otherwise]
7. RestingECG: resting electrocardiogram results [Normal: Normal, ST: having ST-T wave abnormality (T wave inversions and/or ST elevation or depression of > 0.05 mV), LVH: showing probable or definite left ventricular hypertrophy by Estes' criteria]
8. MaxHR: maximum heart rate achieved [Numeric value between 60 and 202]
9. ExerciseAngina: exercise-induced angina [Y: Yes, N: No]
10. Oldpeak: oldpeak = ST [Numeric value measured in depression]
11. ST_Slope: the slope of the peak exercise ST segment [Up: upsloping, Flat: flat, Down: downsloping]
12. **HeartDisease: output class [1: heart disease, 0: Normal]**

Link: https://www.kaggle.com/datasets/fedesoriano/heart-failure-prediction

Methods

The dataset was analyzed and Exploratory Data Analysis was done using python libraries such as numpy, pandas, matplotlib and seaborn.

Before applying machine learning models to the dataset EDA was performed to increase the accuracy of the models. These were:-

- Checking and removing null data points
- Label Encoding for few columns in the dataset
- Distribution plot using seaborn and matplotlib
- Finding Correlations between features
- Removing outliers from dataset

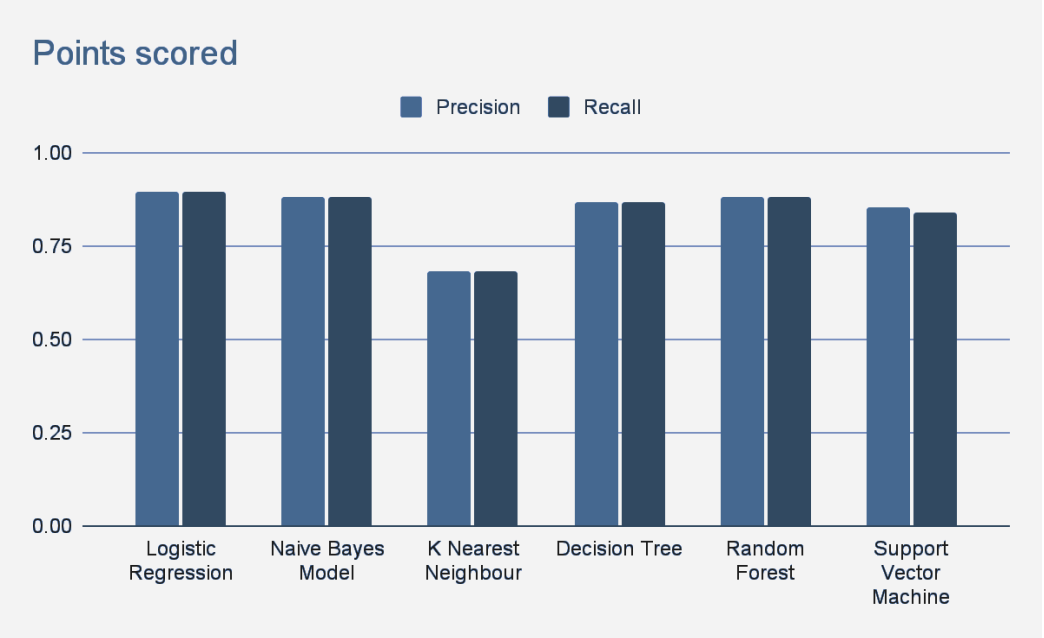
After the EDA was done , the dataset was divided train data and test data in a ratio of 80:20 and after that several machine learning models were trained and tested using different techniques, including Logistic Regression, K-Nearest Neighbors, Decision Tree, Random Forest, and Support Vector Machine. The models were evaluated using various evaluation metrics, including Accuracy, Precision, Recall, F1-score, and Specificity.

Experiments and Results

The experiments were conducted using a 80:20 train-test split, with the best performing model selected based on various evaluation metrics.

The table below represents the accuracy, precision and some other data's achieved by different models:

Model	Precision	Recall	F1-Score	Specificity	Accuracy
Logistic Regression	**0.89603**	**0.8958**	**0.8927**	**0.8958**	**89.58%**
Naive Bayes Model	0.88308	0.8819	0.8821	0.8923	88.19%
K Nearest Neighbour	**0.68329**	**0.6805**	**0.6807**	**0.6578**	**68.06%**
Decision Tree	0.86934	0.8680	0.8676	0.9078	86.81%
Random Forest	0.88192	0.8819	0.8818	0.8947	88.19%
Support Vector Machine	0.85194	0.8407	0.8412	0.7951	84.03%



Conclusions and Future work

Heart attack prediction is an important area of research that has the potential to improve patient outcomes by identifying high-risk individuals and providing early interventions. Based on the work done on this project, the following conclusions and future work can be suggested:

Conclusions:

- The accuracy of the heart attack prediction model can be improved by including more features or by applying more advanced machine learning algorithms.
- The model's performance can be further evaluated using larger datasets, including more diverse patient populations, and with the help of clinical experts.
- The model can be integrated into clinical decision-making systems to assist healthcare providers in identifying patients at high risk of heart attack and making personalized treatment plans.

Future work:

- Explore the use of other machine learning algorithms such as deep learning, reinforcement learning, or hybrid models to improve the accuracy of the heart attack prediction model.
- Incorporate additional features, such as patient lifestyle and behavioral factors, to provide more comprehensive risk assessments.
- Create a user-friendly application or web-based interface that healthcare providers can easily use to access the model's predictions and integrate them into patient care plans.

References

- https://www.w3schools.com/python/python_ml_logistic_regression.asp
- https://www.geeksforgeeks.org/ml-naive-bayes-scratch-implementation-using-python/
- https://www.geeksforgeeks.org/k-nearest-neighbor-algorithm-in-python
- https://www.w3schools.com/python/python_ml_decision_tree.asp
- https://www.geeksforgeeks.org/random-forest-regression-in-python/
- https://www.geeksforgeeks.org/classifying-data-using-support-vector-machinessvms-in-python/

Github Links:-

- **https://github.com/hardik-kh/ML-Sem-Project-Heart-attack-prediction-**
- **https://github.com/nectro/HeartDeseasePrediction**
- **https://github.com/Aryaman9/heart-attack-prediction**