

Capstone Project Submission

Instructions:

- i) Please fill in all the required information.
- ii) Avoid grammatical errors.

Team Member's Name, Email and Contribution:

Individual

NAVED MANSURI - navedmansuri165@gmail.com

Please paste the GitHub Repo link.

GitHub Link: - <https://github.com/NavedMansuri/Cardiovascular-Risk-Prediction>

Please write a short summary of your Capstone project and its components. Describe the problem statement, your approaches and your conclusions. (200-400 words)

Since a few years ago, the prevalence of cardiovascular disease has been rising quickly around the world. Numerous studies have been carried out in an effort to identify the most important risk factors for heart disease and to precisely estimate the total risk. Heart disease is also referred to as a silent killer since it causes a person to pass away without any evident signs. In high-risk individuals, an early diagnosis of heart disease is crucial for helping them decide whether to modify their lifestyle, which lowers consequences. By evaluating patient data that is used to categorise whether or not they have heart disease using machine-learning algorithms, this study seeks to forecast the likelihood of developing heart disease in the future. We must evaluate several classification methods to prediction of cardio CHD.

we trained 5 Machine Learning models, and hyperparameter adjustment was utilised models to increase model performance. The training dataset was oversampled using SMOTE to reduce bias on one outcome, missing values were handled, feature engineering, and feature selection were performed. Cardiovascular heart disease affects a similar number of smokers and non-smokers. Age, total cholesterol, systolic blood and diastolic blood pressure, BMI, heart rate, and glucose are the main factors in determining a person's 10-year chance of having cardiovascular heart disease.

We chose the oversampling technique because the data provided to us had fewer records. But since there will be a lot of unbalanced and large amounts of health data, we can try to work on cost-sensitive learning, which, rather than changing the data records, only gives more weight to the minority and focuses on the individuals at high risk for heart disease.

The K Nearest Neighbour is proved to be best algorithms can be used for the risk prediction of Cardiovascular heart disease.