Capstone Project Submission

Team Member's Name, Email and Contribution:

Contributor's Role:

1. **Parth Sharma**
   Email: parthsharma01197@gmail.com.
   Contribution:
   - Understanding the Coronary Heart Disease Prediction Dataset.
   - Data Wrangling on the Coronary Heart Disease Prediction Dataset.
   - Discovering the underlying patterns.
   - Data Visualization on the Coronary Heart Disease Prediction Dataset.
   - Drawing conclusions from the visualizations.
   - Feature Engineering and Feature Transformation on the data set.
   - Testing different ML classification models on the data set.
   - Checking for the model giving the confirming results.
   - Conclusion and Model Suggestion.

Please paste the GitHub Repo link.

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

The load of cardiovascular disease is rapidly increasing all over the world from the past few years. Many researches have been conducted in attempt to pinpoint the most influential factors of heart disease as well as accurately predict the overall risk. Heart Disease is even highlighted as a silent killer which leads to the death of the person without obvious symptoms. The early diagnosis of heart disease plays a vital role in making decisions on lifestyle changes in high-risk patients and in turn reduces the complications.

This project aims to predict the chance of future Heart Disease by analyzing data of patients which classifies whether they have heart disease or not using machine-learning algorithm.
We need to test different classification algorithms and suggest the one that could predict the risk of Coronary Heart Disease the best.

The document discusses a comparative approach to the classification of coronary heart disease datasets using machine learning (ML) algorithms.
Starting with basic understanding of the data we then move on to clean our data from different anomalies like missing values, outliers, duplicates. Successfully treating the missing values using iterative imputation next comes the outliers that are eliminated using the capping. Then comes the encoding of categorical variables .Selecting the best parameters in order to assess the impact on two distinct classes on the performance metrics one of the challenging step as it requires some expert advice but I have relied on the Chi-score. The dataset was then subjected to SMOTE to handle unbalanced classes.
Then come the stage where I passed my transformed data into various classification models and evaluate them of the basis of different classification metrics.

The results show that K-Nearest Neighbours produced the highest performance score on the original dataset compared to the other algorithms employed. In conclusion, this study suggests that K-Nearest Neighbours on a well-processed and standardized dataset can predict coronary heart disease with greater robustness than the other algorithms. The metrics the KNN provides us are :

- Accuracy-0.883
- Precision-0.847,
- Recall-0.933.

Concluding our study we can say that, Heart diseases are a silent killer, The early prognosis of heart disease can aid in making decisions on lifestyle changes in high-risk patients and in turn reduce the complications.

Hence, we can say that Machine Learning can help save lives of many and help them to switch over to a healthy lifestyle to chop off any health related issue.