**Capstone Project Submission**

**Instructions:**

i) Please fill in all the required information.

ii) Avoid grammatical errors.

|  |
| --- |
| **Team Member’s Name, Email and Contribution:** |
| Name: Mohammed Javeed  Email: [mohammedjaveed407@gmail.com](mailto:mohammedjaveed407@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.** |
| GitHub Link: <https://github.com/Javee01/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)**  **In this project, we aim to predict the probability of future Heart Disease through the analysis of data from patients which classifies whether or not they have heart disease.**  **To determine which classification algorithm could best predict Coronary Heart Disease risk, we need to test different algorithms.**  **Problem Statement:**  **There has been an increase in cardiovascular disease around the world in recent years. There has been a great deal of research done to identify the most influential causes of heart disease as well as accurately predict the overall risk of developing the disease. Heart disease is even referred to as a silent killer because symptoms tend not to be visible before death. Identifying heart disease at an early stage provides an opportunity to make lifestyle changes and thus reduce complications in high-risk patients.**  **A machine learning algorithm is used to analyze data of patients and classify whether or not they have heart disease in order to predict the chances of future Heart Disease.The best classification algorithm for Coronary Heart Disease predictability requires testing multiple classification algorithms.**  **Approach:**  **Based on machine learning (ML) algorithms, the document presents a comparative approach to the classification of coronary heart disease datasets. Once we understand the data, we move on to clean it from different anomalies, such as missing values, outliers, and duplicates. Iterative imputation is used to treat missing values, followed by a capping process to eliminate outliers. In this step, categorical variables are encoded. One of the challenging steps was selecting the best parameters to assess their effect on the performance metrics. However, I used the Chi-score in this step. SMOTE was then used to handle unbalanced classes in the dataset. Once my transform data were passed to various classification models, I evaluated them based on various classification metrics.**  **Conclusions:**  **Comparing K-Nearest Neighbours and the other algorithms employed, K-Nearest Neighbours produced the highest performance score.**  **According to this study, K-Nearest Neighbors can be more accurately predicted coronary heart disease when based on a well-processed and standardized dataset.**  **KNN provides us with the following metrics:**  **Accuracy-0.89**  **Precision-0.89**  **Recall-0.87**  **As a result, machine learning can help people adopt healthier lifestyles** |