Name: **Kanaishk Garg**

ID: rishi.gupta90@gmail.com

Title: **Predicting Heart Disease Risk using Machine Learning: A Comparative Analysis of Logistic Regression and Random Forest Models**

Heart disease is a leading cause of death globally, and it is essential to identify early signs of the disease to prevent adverse outcomes. This project aims to build a model to predict the risk of heart disease using various features such as age, sex, chest pain type, maximum heart rate achieved, exercise-induced angina, and several other factors.

The dataset used in this project is the Cleveland Heart Disease dataset, which contains 303 samples with 14 features. The project's first step was to perform exploratory data analysis (EDA) to understand the distribution and correlation between features. The EDA revealed some interesting insights, such as the relationship between age and heart disease, the correlation between chest pain type and heart disease, and the correlation between exercise-induced angina and heart disease.

After completing the EDA, we built two baseline models using logistic regression and random forest classification algorithms. We split the dataset into training and testing sets, and the baseline models achieved a maximum accuracy of 85.000% and 88.333%, respectively.

We then performed feature selection using logistic regression to identify the most significant features in predicting the risk of heart disease. The logistic regression model identified nine significant features, including age, sex, chest pain type, maximum heart rate achieved, exercise-induced angina, and several other factors. We used these features to train logistic regression and random forest models, and they achieved a maximum accuracy of 88.333% and 85.000%, respectively.

Finally, we evaluated the models' performance using confusion matrices and accuracy scores. We also explored ways to improve the model's accuracy by tuning hyperparameters and trying out different algorithms.

Overall, this project demonstrates the importance of data analysis and feature selection in building predictive models. The models built in this project could be used to identify individuals at risk of heart disease and provide early intervention to prevent adverse outcomes.