**Capstone Project Submission**

**Instructions:**

i) Please fill in all the required information.

ii) Avoid grammatical errors.

|  |
| --- |
| **Team Member’s Name, Email and Contribution:** |
| 1. Pankaj Kumar Yadav ( [pankajkumaryadav1500420@gmail.com](mailto:pankajkumaryadav1500420@gmail.com) )  * Visualizing Data * Eda * Splitting * SUPPORT VECTOR MACHINE * XG Boost Classification  1. Harshada Gore ( [harshada1442@gmail.com](mailto:harshada1442@gmail.com) )    * Data cleaning    * Eda    * Logistic Regression    * Data Analyses 2. Prince Kumar Jha ([pjha3913@gmail.com](mailto:pjha3913@gmail.com) )    * Data exploration    * Feature engineering    * Data Balancing    * Random Forest 3. Satyajit Mohanty ([msatyajit143@gmail.com](mailto:msatyajit143@gmail.com) )    * Feature engineering    * Data Analysis    * Collab formation    * Model score    * Classification |
| **Please paste the GitHub Repo link.**  <https://github.com/princekjha/Cardiovascular-Risk-Prediction> |
| **Google Drive Link**  <https://drive.google.com/drive/folders/1xnipiKR0dG9V80TcfyCCDnFtceDlbUSS?lfhs=2> |

|  |
| --- |
| **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 dataset is from an ongoing cardiovascular study on residents of the town of Framingham, Massachusetts. The classification goal is to predict whether the patient has a 10-year risk of future coronary heart disease (CHD). The dataset provides the patients’ information. It includes over 4,000 records and 15 attributes..   * **Sex**: male or female * **Age**: Age of the patient * **Is\_smoking**: patient is a current smoker * **Cigs Per Day**: the number of cigarettes that the person smoked on average in one day. * **BP Meds**: blood pressure medication * **Prevalent Stroke**: previously had a stroke * **Prevalent Hyp**: patient was hypertensive * **Diabetes**: the patient had diabetes * **Tot Chol:** total cholesterol level * **Sys BP**: systolic blood pressure * **Día BP**: diastolic blood pressure * **BMI**: Body Mass Index * **Heart Rate**: heart rate * **Glucose**: glucose level * **TenYearCHD**: 10-year risk of coronary heart disease   The objective of this project is to predict the overall risk of heart disease using Classification regression:   1. Loading the data into data frame 2. Cleaning the data 3. Extracting statistics from the dataset 4. Exploratory analysis and visualizations 5. Feature Engineering 6. Train Test Split 7. Logistic Regression 8. Random Forrest 9. XGBoost 10. Support Vector Machine   **Conclusion**  That's how we have accomplished our team work in CARDIOVASCULAR RISK PREDICTION Project. Throughout the project we learn many new things right from taking problem statement to understand the technical side of a product to analysis. We deal with imbalanced data for train and test at the time of data modeling. |