**General Healthcare Prediction**

**COURSE PROJECT REPORT**

**18CSE398J -Machine Learning - Core Concepts with Applications**

**(2018 Regulation)**

**III Year/ VI Semester**

**Academic Year: 2022 -2023 (EVEN)**

By

**Uday Goel - RA2011003010544**

**Chandrashekhar - RA2011003010591**

**Nishant Jain - RA2011003010610**

Under the guidance of

**Dr. V. Vijayalakshmi**

**Assistant Professor**

**Department of Data Science and Business Systems**



**DEPARTMENT OF DATA SCIENCE AND BUSINESS SYSTEMS**

**FACULTY OF ENGINEERING AND TECHNOLOGY**

**SRM INSTITUTE OF SCIENCE AND TECHNOLOGY**

**Kattankulathur, Kancheepuram**

**MAY 2023**

General Healthcare Prediction

LIST OF STUDENTS (NAME, MAIL ID), V. Vijayalakshmi, vijayalv3@srmist.edu.in

1. **Abstract**
2. **Introduction**
3. **Dataset**
4. **Methods**
5. **Experiments and results**
6. **Conclusions and future work**
7. **References (min 20)**

**Github Link of the project work:**

**Uday Goel:** <https://github.com/Deadalus1901/Healthcare-Prediction>

**Chandrashekhar S:** <https://github.com/chan279>

**Nishant Jain:** <https://github.com/chan279>

**ABSTRACT**

The objective of this machine learning project is to predict general healthcare outcomes based on patient information such as demographics, medical history, and lifestyle factors. The model is trained on a large dataset of electronic health records to identify patterns and correlations between patient characteristics and health outcomes. The project aims to provide a valuable tool for healthcare professionals to make more informed decisions and provide personalized care to patients. The model's accuracy will be evaluated using various metrics such as precision, recall, and F1-score. The algorithms we are using are Logistic regression, Random forest, KNN, SVM, Decision Trees, Naive Bayes and Gradient Boosting. The results of this project have the potential to contribute to the development of better healthcare strategies and improve patient outcomes.

**INTRODUCTION**

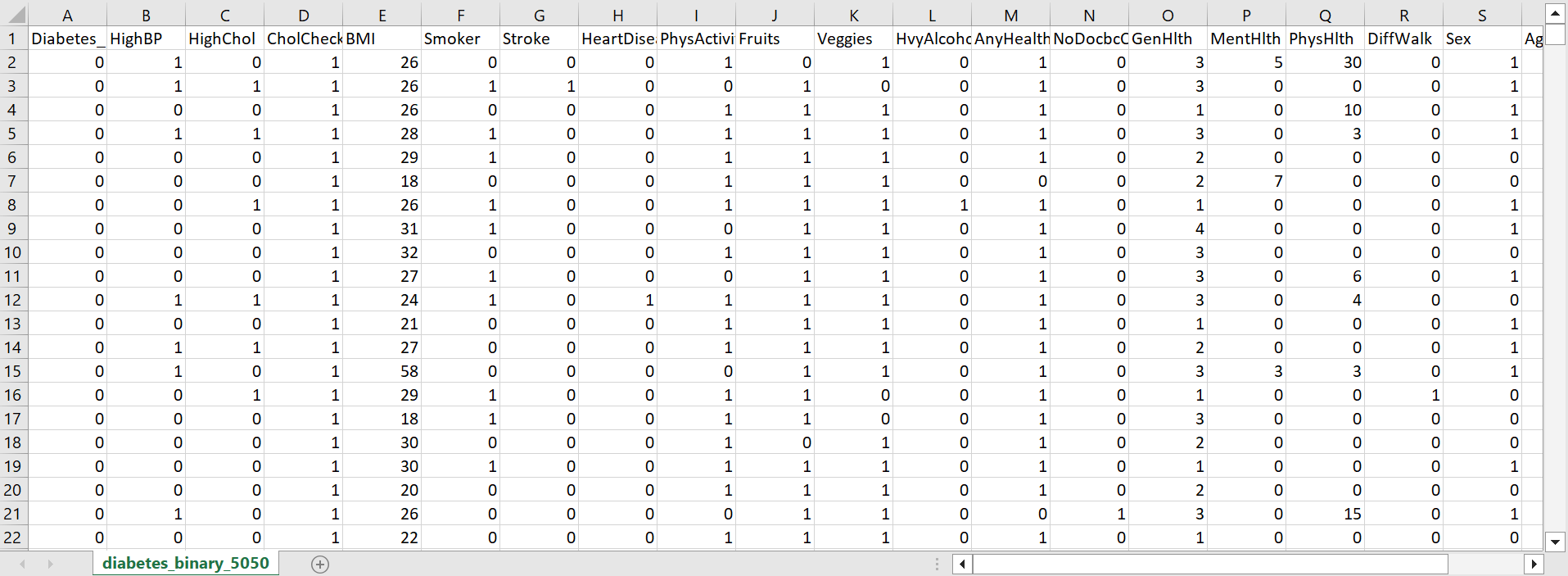
The healthcare industry is one of the most important and rapidly evolving sectors in the world. The increasing demand for healthcare services, along with rising costs, presents significant challenges for healthcare providers, policymakers, and patients. To address these challenges, there is a growing interest in leveraging the power of machine learning to improve healthcare outcomes. Machine learning algorithms can be trained to identify patterns in large datasets, enabling healthcare professionals to predict patient outcomes and provide personalized care.

This project focuses on developing a machine learning model that can predict general healthcare outcomes based on patient information such as demographics, medical history, and lifestyle factors. The model will be trained on a large dataset of electronic health records, which contains information from tens of thousands of patients. By analyzing this data, the model can identify patterns and correlations between patient characteristics and health outcomes. For example, the model could identify patients who are at high risk of developing a particular condition, allowing healthcare professionals to intervene early and prevent the onset of the disease.

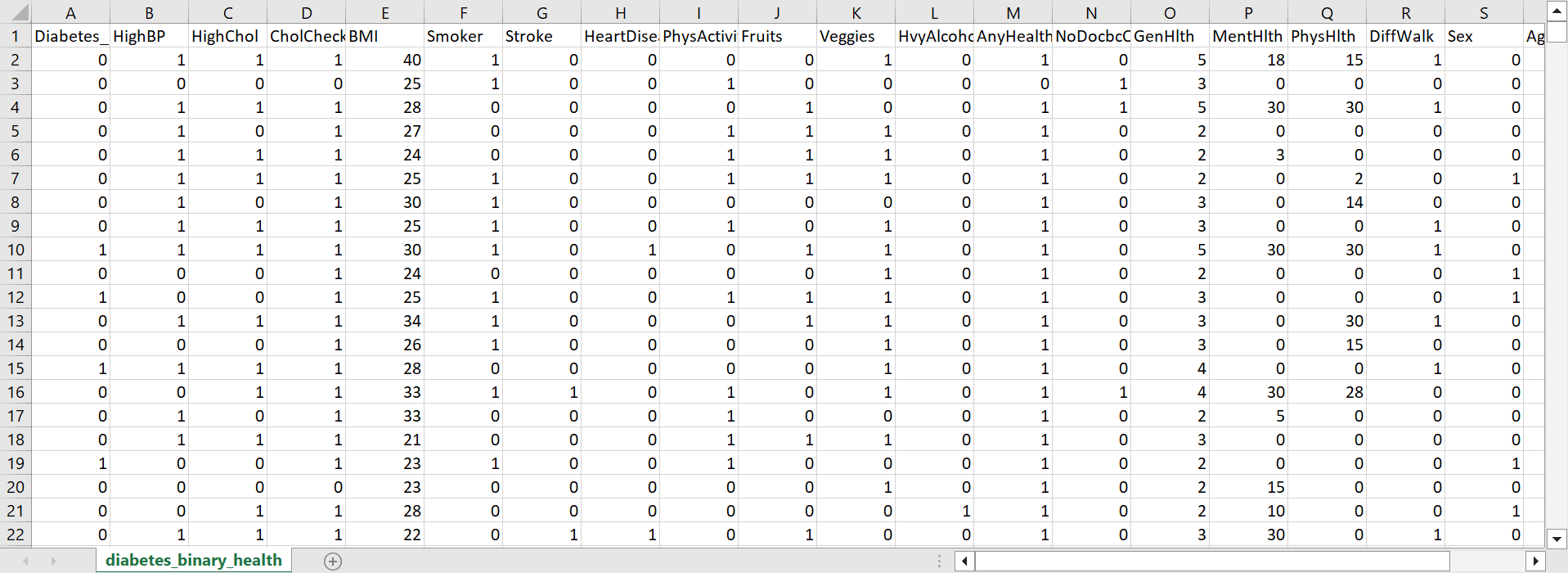
The accuracy of the machine learning model will be evaluated using various metrics such as precision, recall, and F1-score. These metrics will help to ensure that the model is reliable and can be trusted by healthcare professionals. The ultimate goal of this project is to provide healthcare professionals with a powerful tool to make more informed decisions and provide personalized care to patients. By leveraging the power of machine learning, this project has the potential to revolutionize healthcare by improving outcomes and reducing costs.

**DATASET**

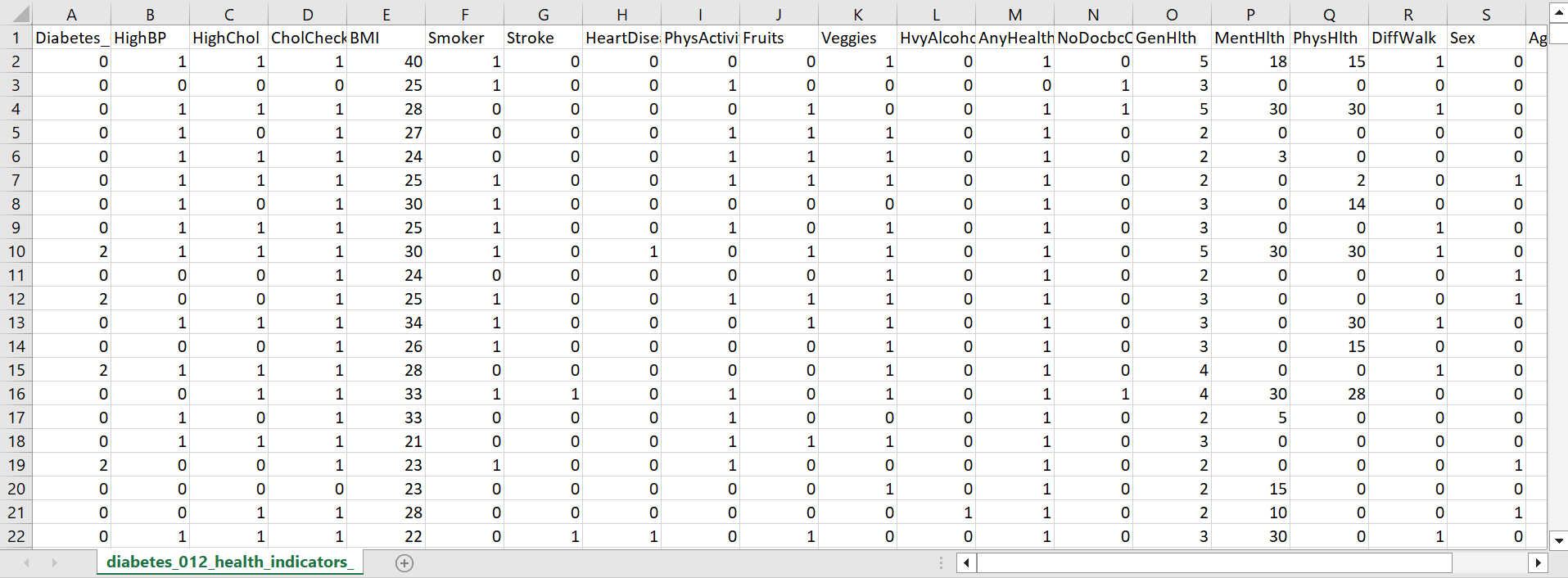
[Diabetes\_binary\_1](file:///C:\Users\NANDNI%20JAIN\OneDrive\Desktop\diabetes_binary_5050.csv)



[Diabetes\_binary\_2](file:///C:\Users\NANDNI%20JAIN\OneDrive\Desktop\diabetes_binary_health.csv)



[Diabetes\_bianry\_3](file:///C:\Users\NANDNI%20JAIN\OneDrive\Desktop\diabetes_012_health_indicators_BRFSS2015.csv)



**METHODS**

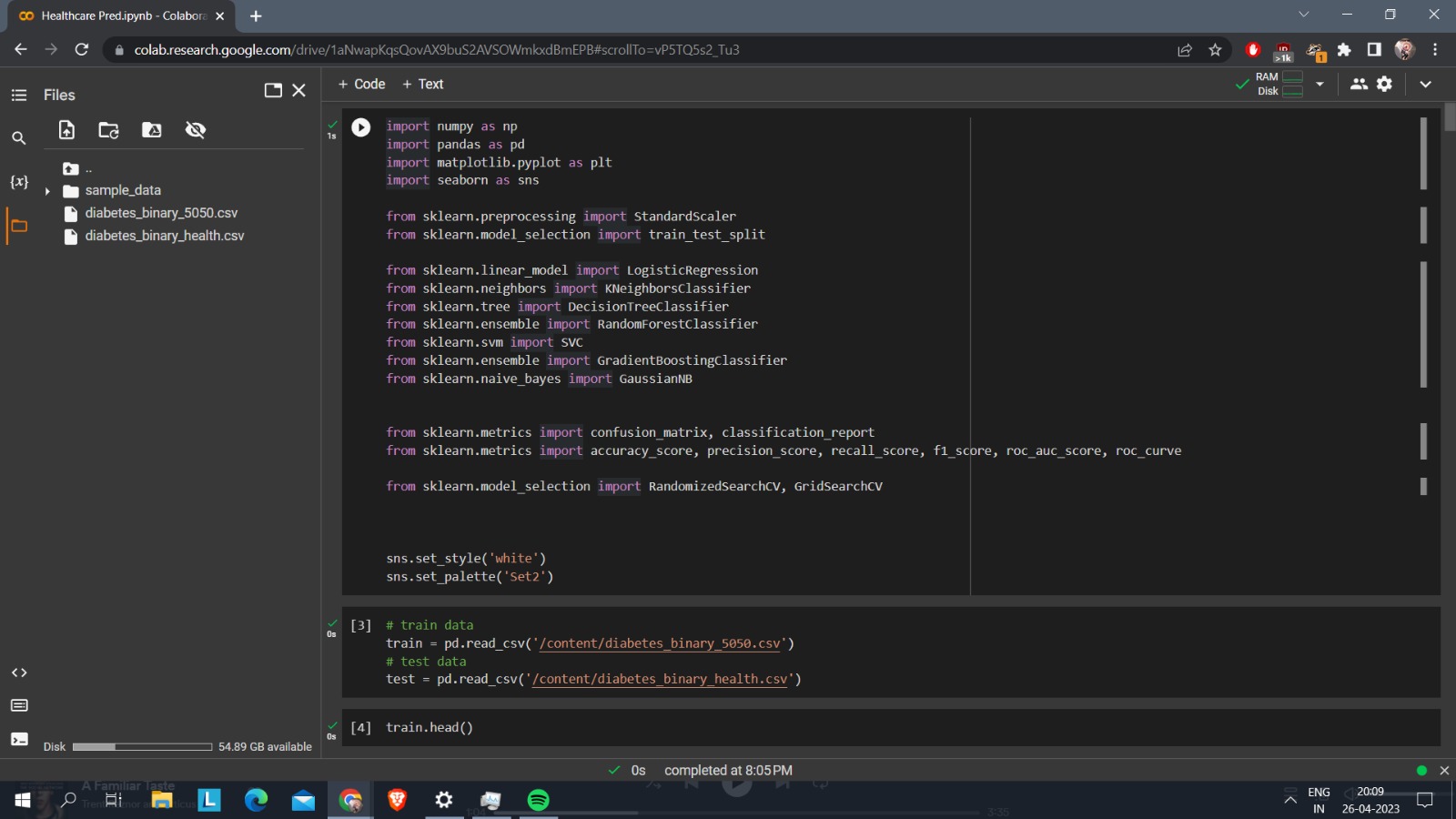
We have used the following algorithms to train our model :

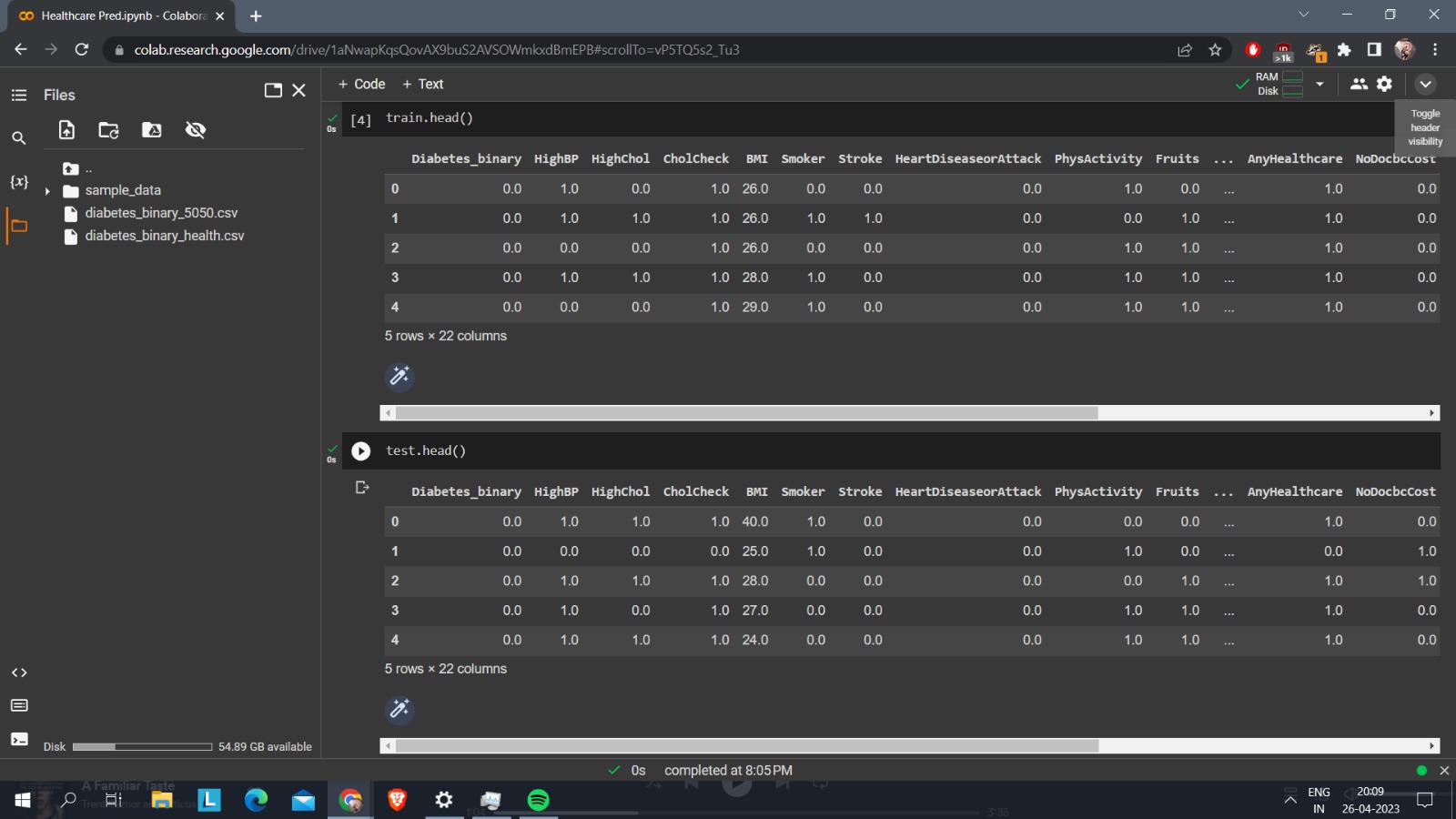
1. Logistic Regression: It is a statistical model used to predict a binary outcome (yes/no, 0/1, true/false, etc.) based on one or more input variables. It calculates the probability of an event occurring and classifies the observation into one of two classes based on a threshold value.
2. Decision Tree: It is a tree-like model where each internal node represents a feature or attribute, each branch represents a decision or rule, and each leaf node represents a class label. It is used for both classification and regression problems.
3. Random Forest: It is an ensemble learning method that creates a large number of decision trees, each using a random subset of the features and observations. The final prediction is made by taking the majority vote of all the trees.
4. KNN (K-Nearest Neighbors): It is a lazy learning algorithm that uses a distance metric to find the K nearest neighbors to a given data point and then predicts the class based on the majority class of the neighbors.
5. SVM (Support Vector Machine): It is a binary classification algorithm that finds the best hyperplane to separate the two classes in the feature space. It is often used in high-dimensional spaces and with non-linearly separable data.
6. Naive Bayes: It is a probabilistic algorithm that makes predictions based on Bayes' theorem. It assumes that the presence or absence of a particular feature is independent of the presence or absence of any other feature.
7. Gradient Boosting: It is a boosting ensemble method that combines weak learners (usually decision trees) in a stepwise manner to improve the overall prediction accuracy. It builds the model by iteratively adding new weak learners that minimize the error of the previous learners.

**EXPERIMENTS AND RESULTS**

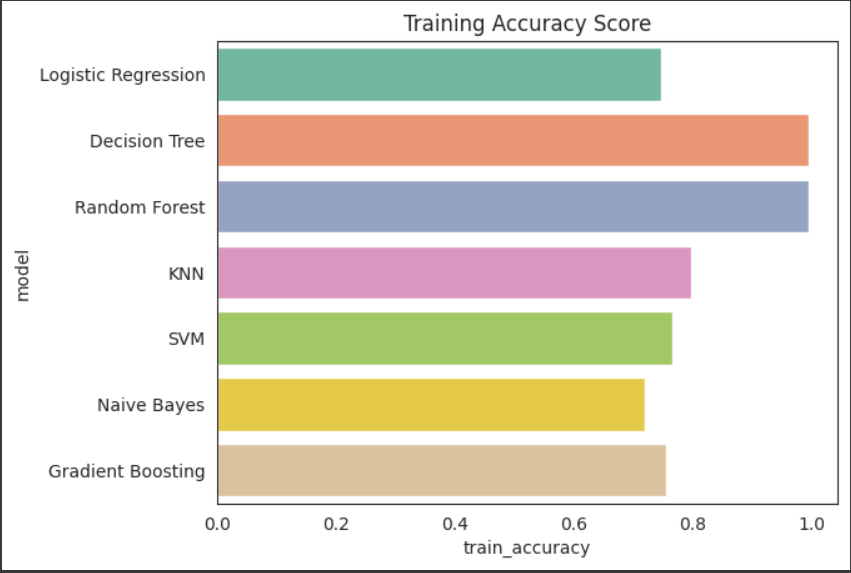
The following codes ran for our experiment are in the link inserted below

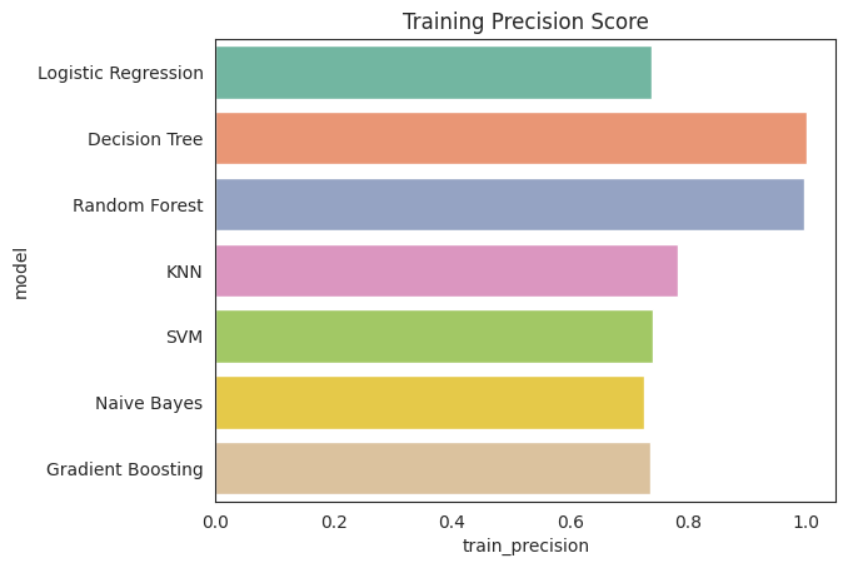
[Healthcare Prediction](file:///C:\Users\NANDNI%20JAIN\OneDrive\Desktop\Healthcare%20Prediction.pdf)

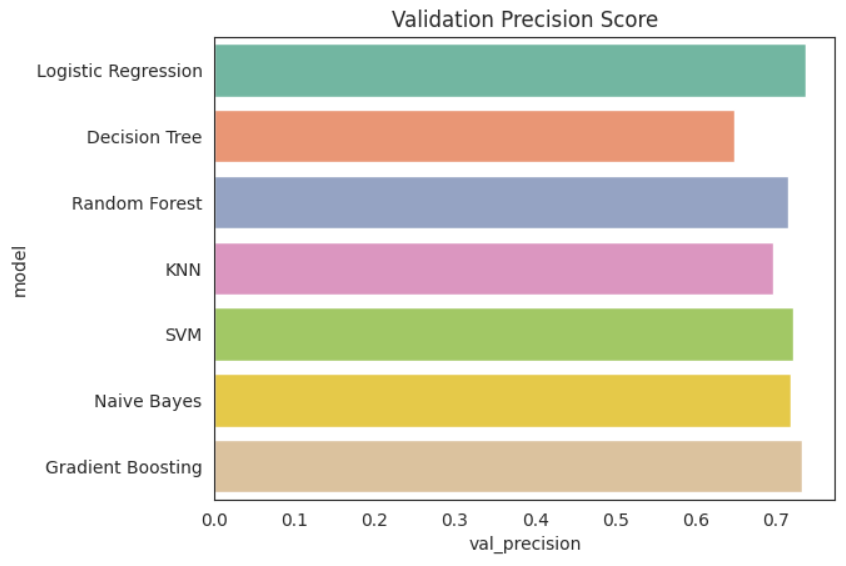
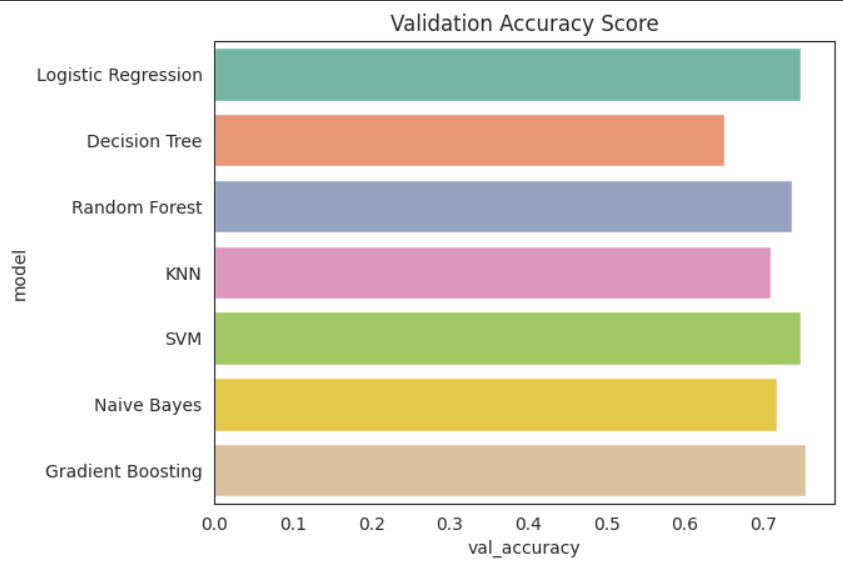


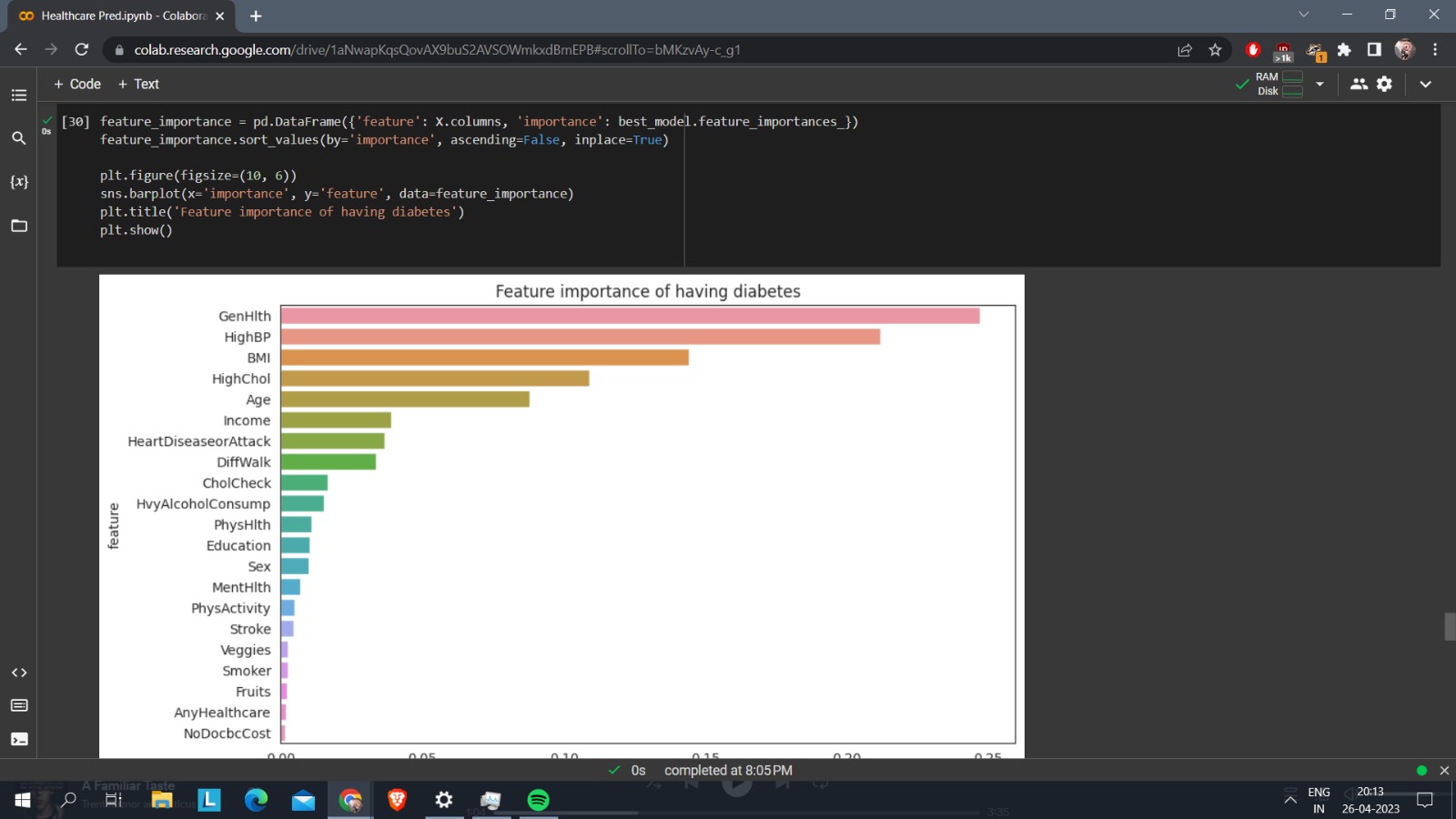


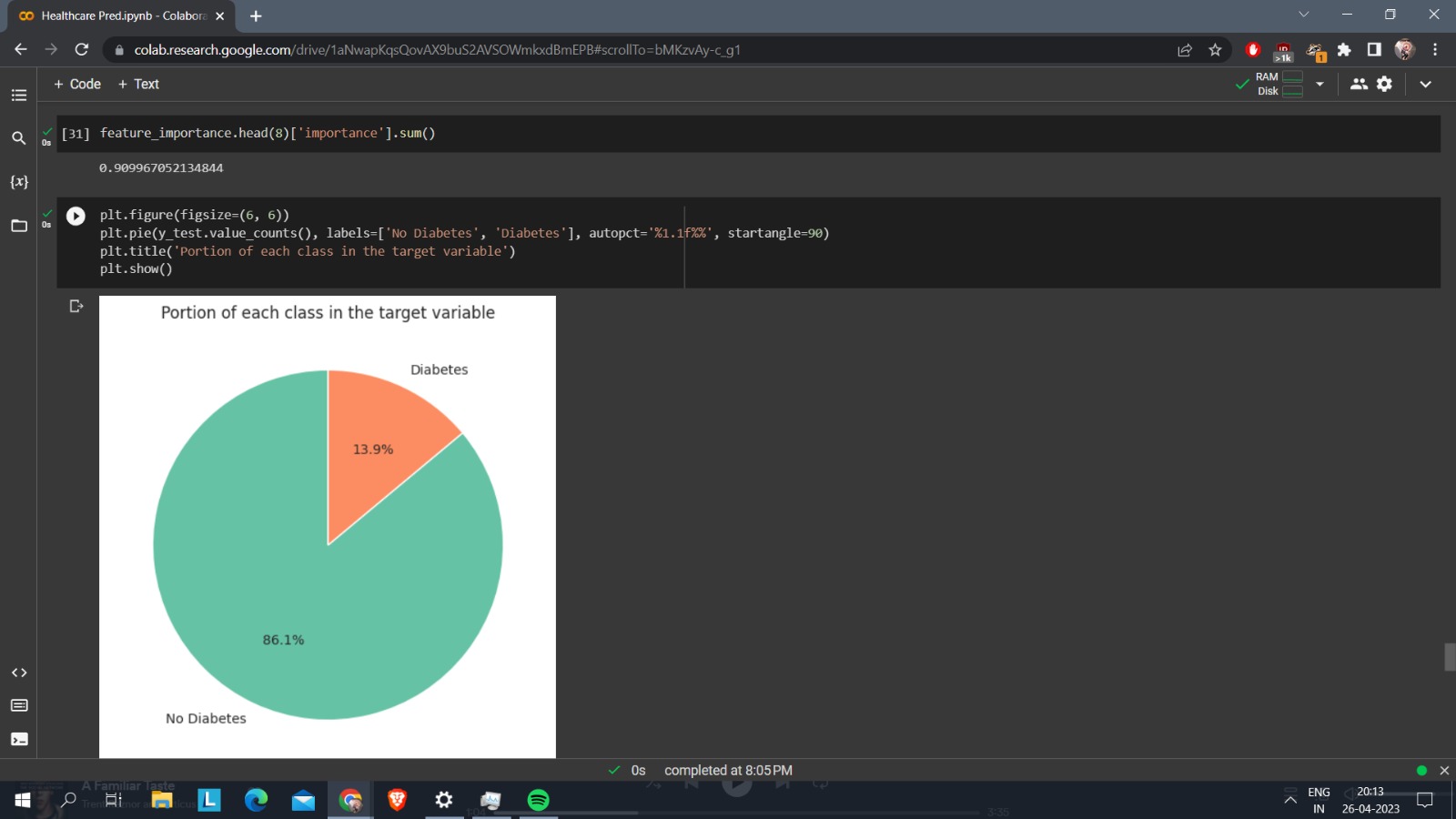
**Outcomes**

****

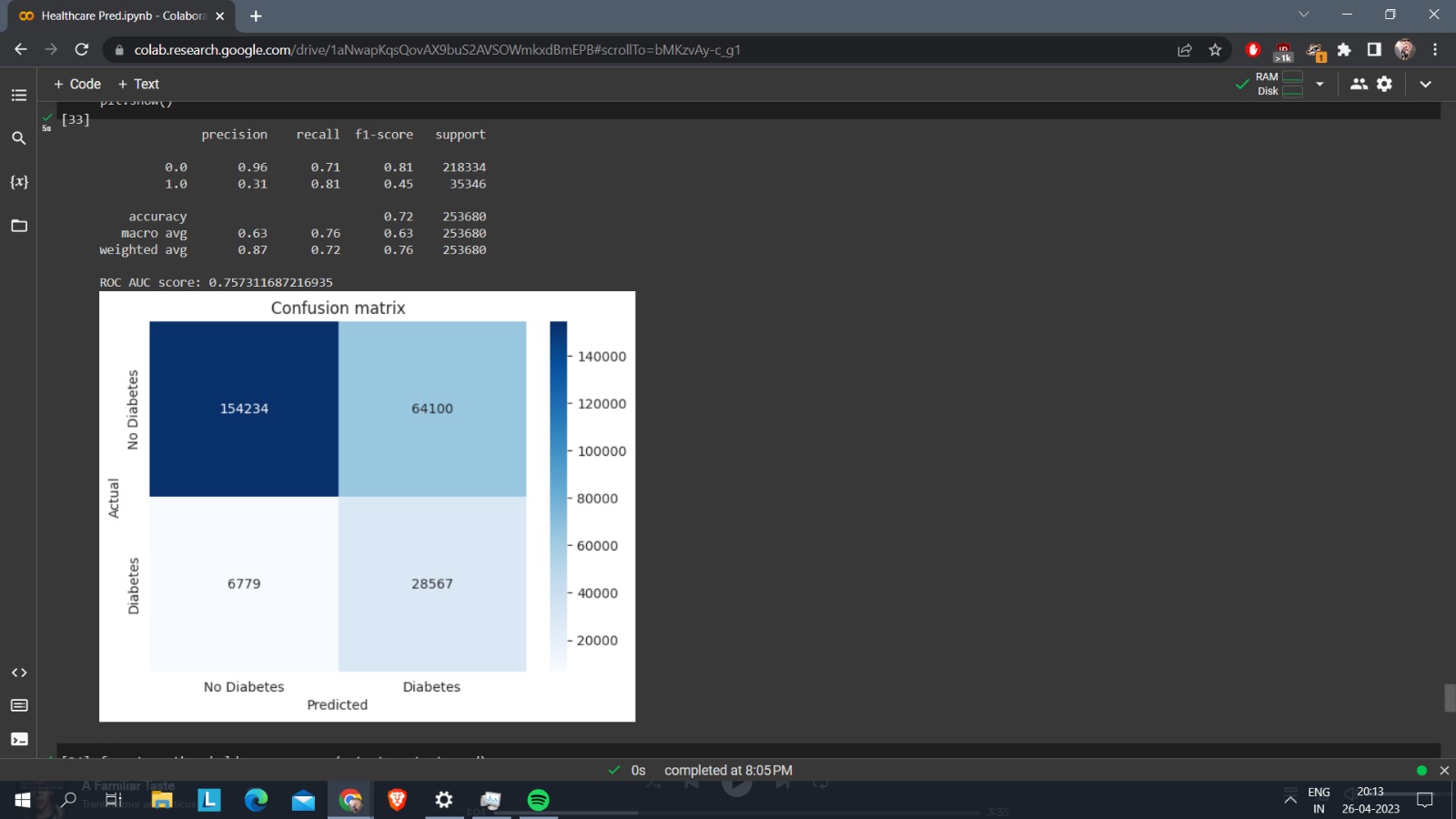
****

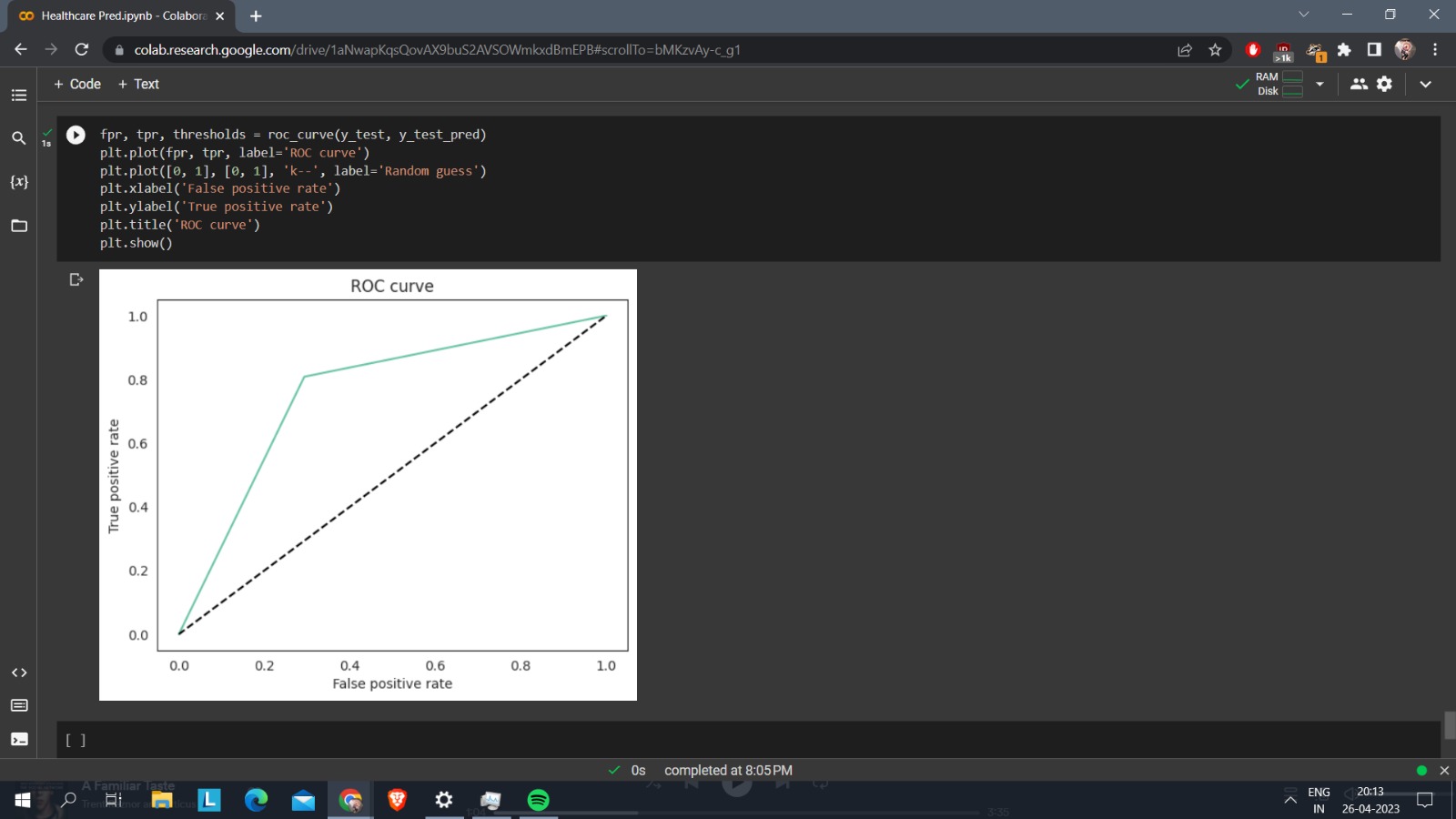
****





****





**CONCLUSION AND FUTURE WORK**

In conclusion, this machine learning project focused on developing a model to predict general healthcare outcomes based on patient information. By training the model on a large dataset of electronic health records, the project aimed to identify patterns and correlations between patient characteristics and health outcomes. The accuracy of the model was evaluated using various metrics, which helped to ensure its reliability and validity.

The results of this project have the potential to revolutionize healthcare by providing healthcare professionals with a powerful tool to make more informed decisions and provide personalized care to patients. By predicting patient outcomes, healthcare professionals can intervene early and prevent the onset of diseases, resulting in improved outcomes and reduced costs. The project also highlights the potential of machine learning in the healthcare industry and the need to continue investing in this field.

However, the success of this project also depends on the quality and availability of healthcare data. The development of robust data infrastructure and standardization of data collection can enable the healthcare industry to fully leverage the power of machine learning to improve patient outcomes.

The ultimate goal of this project is to provide healthcare professionals with a powerful tool to make more informed decisions and provide personalized care to patients. By leveraging the power of machine learning, this project has the potential to revolutionize healthcare by improving outcomes and reducing costs.

Overall, this project provides valuable insights into the potential of machine learning in healthcare and serves as a stepping stone for future research and development in this field.