|  |  |
| --- | --- |
|  | PREDICTION OF HEART ATTACK |
|  |  |
|  | TEAM – B12  03-02-2024 |





|  |  |
| --- | --- |
| **S.NO** | **CHAPTER NAME** |
| 1. | Problem Statement |
| 2. | Abstract |
| 3. | Introduction |
| 4. | Dataset |
| 5. | Preprocessing of Data |
| 6. | Predictive modelling approach |
| 7. | Results and Findings |
| 8. | Conclusion |

**TABLE OF CONTENTS**

**PROBLEM STATEMENT**

Predicting the likelihood of Heart Attack in individuals based on various demographic, lifestyle and health-related factors. Early detection of individuals at risk of a heart attack is crucial for timely intervention and prevention. Despite advances in medical research and technology, existing predictive models often lack the necessary precision and comprehensiveness to reliably identify individuals susceptible to a heart attack.

The objective of this model is to develop a machine learning model in order to tackle the predictive model that considers the multifactorial nature of heart attack.

**ABSTRACT**

This problem statement highlights the complexities hindering the development of accurate predictive models for early detection of heart attacks.

The proposed model uses a diverse set of features, including age, gender, family history and alcohol consumption, to train a predictive algorithm. Data will be collected from a variety of sources, generating data set by using python library (Pandas). The model will be designed to handle numerical data, by using ranges, ensuring its applicability to a wide array of input features.

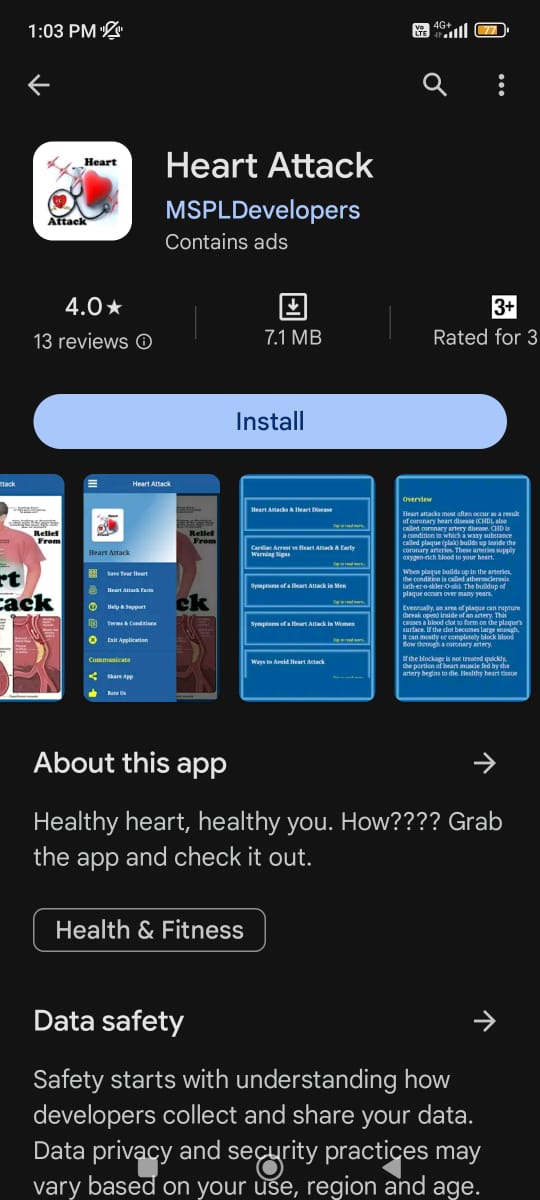
Several machine learning algorithms, such as linear regression, decision trees, and, will be employed and fine-tuned to optimize predictive accuracy. The model's performance will be repeatedly evaluated through cross-validation and validation on an independent dataset to ensure generalizability.

This research seeks to contribute to the ongoing conversation about using technology to enhance transparency, accessibility, and fairness in the prediction of heart attack.

**INTRODUCTION**

The way that predicting heart attack are shaped by technology is fundamentally changing the healthcare industry. Here comes machine learning (ML), a powerful instrument that has the potential to completely transforming the daily life features into the decision-making of prediction of heart attack. ML models can forecast the possibility of prediction of heart attacks by evaluating past data and taking into various variables includes gender, age, smoking, family history,

To overcome these challenges, a multidisciplinary approach is proposed, involving collaboration between healthcare professionals, data scientists, and technology experts. The development of a predictive model must consider a wide array of risk factors, leverage emerging technologies like machine learning, and adapt to dynamic changes in individual risk profiles over time.

**SOURCE :**

* Heart Attack App
* New Atlas.com
* Apple Watches ( smart watches )

**FEATURE SELECTION AND ENGINEERING**

For the purpose of predicting healthcare costs, feature selection and engineering are essential. To improve the models' capacity for prediction, they entail the discovery, transformation, and production of pertinent features. Here is further information on feature engineering and selection for the prediction of heart attack:

**Identification of Relevant Features:** To capture the elements that help in a prediction of heart attack, the appropriate set of features must be identified.

Some elements to think about could be: Data about the users age, gender, smoking, drugs, and stress.

Feature engineering is the process of changing existing features or developing new ones to more accurately depict the underlying patterns in the data. The following are some methods for feature engineering:

**Binning:** By establishing bins or ranges, continuous data, such as height or weight, can be transformed into categorical variables.

**Interaction Terms:** By mixing two or more predictors, interaction characteristics are produced, for example the interplay between state and country is done under one column known as zip code.

**FEATURE ENGINEERING**

**STEP 1-> FEATURE SELECTION :**

* Age
* Gender
* Height
* Weight
* BP
* Diabetes
* Blood Thickness
* Family History
* Calories Burn
* Working hours6
* Sleeping hours
* Type of work
* Smoking
* Alcohol Consumption
* Eating Planning
* Voices
* Stress
* Exercise
* Chest Pain
* Heart rate
* Drugs

**STEP 2-> FEATURE TRANSFORMATION:**

* Age
* Gender
* BMI
* BP
* Family history
* Blood thickness
* Stress
* Smoking
* Alcohol consumption
* Heart rate

**STEP 3-> FEATURE CREATION:**

* No feature creation

**STEP 4-> FEATURE ENCODING:**

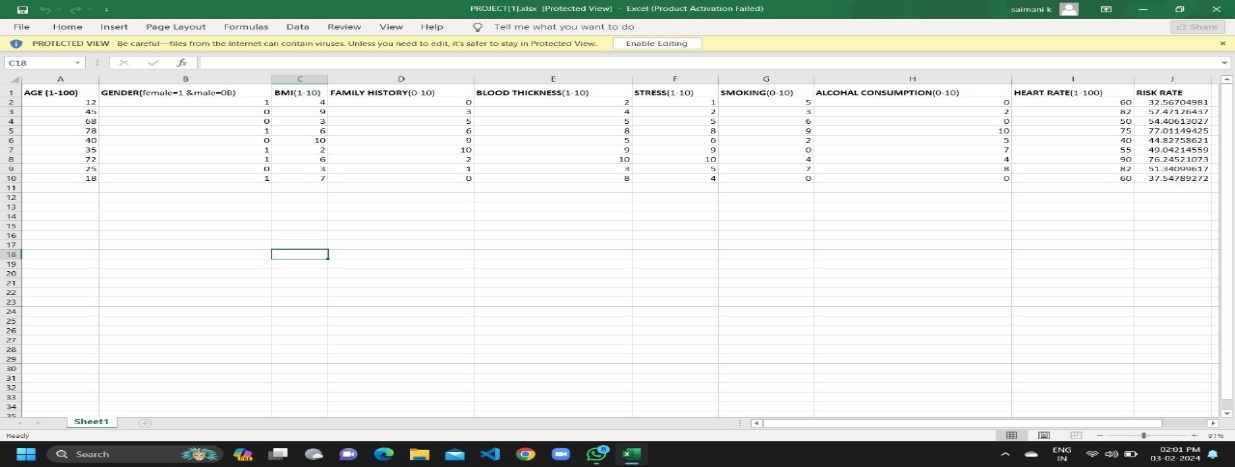
* No feature encoding

**STEP 5-> FEATURE EXTRACTION:**

* Age
* Gender
* BMI
* Family History
* Blood Thickness
* Stress
* Smoking
* Alcohol consumption
* Heart rate

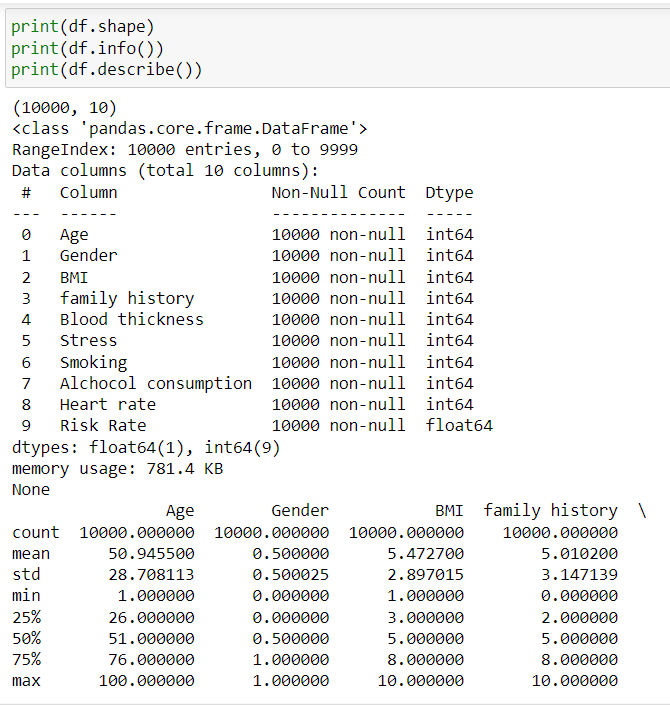
**DATASET**

A dataset is an ordered set of data that is usually arranged tabularly, with rows denoting specific observations or occurrences and columns denoting variables or characteristics related to those reports. We obtained the authentic sample data for this model from official sources and the necessary information by using search engines like Google. Next, a dataset has been constructed. The sample csv file is shown below:



**DATA DISCOVERY:** It explains the basic properties of data, like Rows x Columns =1001 x 9

**DATA EXPLORATION:**



**DATA CLEANSING:** No data cleansing

**DATA STRUCTURING:** No data structuring

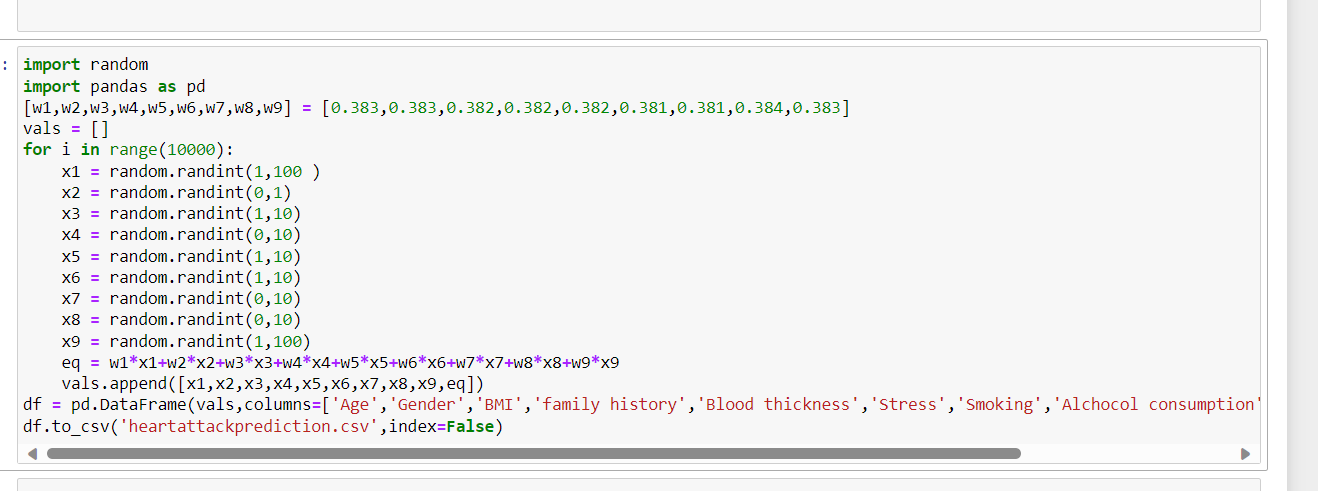
**DATA TRANSFORMATION:** No data transformation

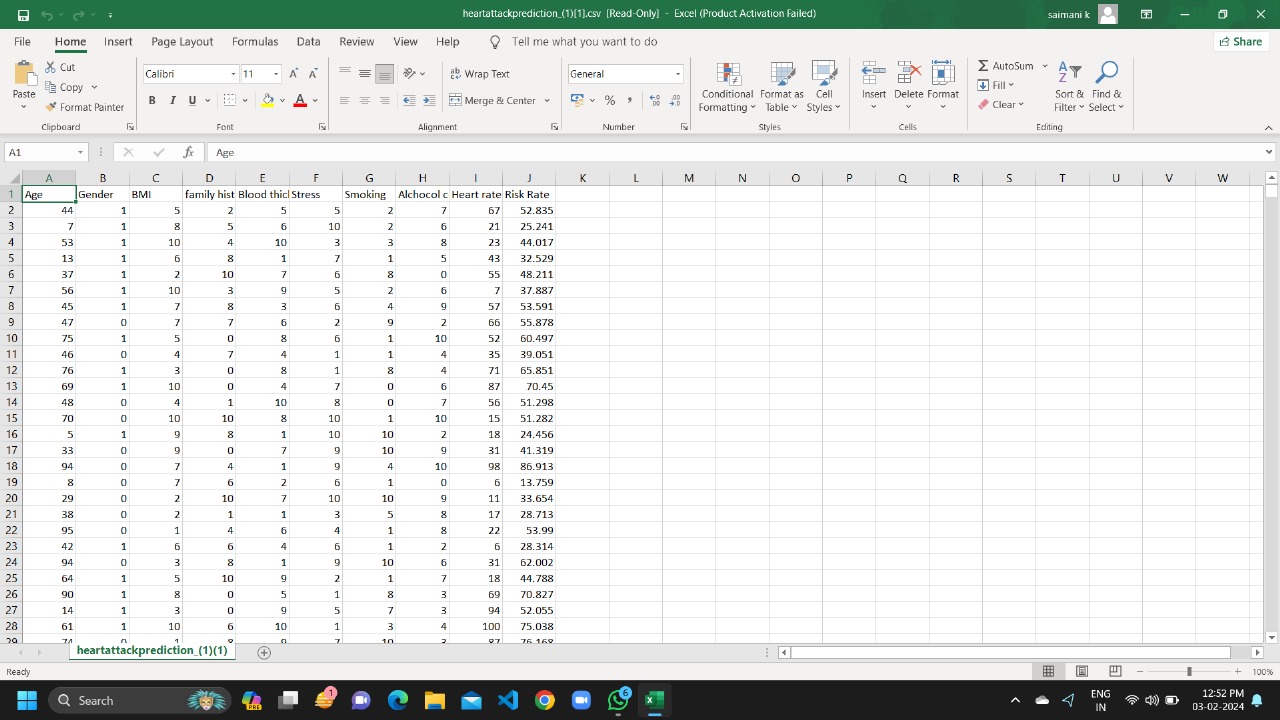
**Preprocessing of data**

According to the table the dataset consists of 9 variables. Each of these features help us recommend the prediction of heart attack. To effectively apply the data to the ML algorithms, it is examined and updated in this stage.

In order to represent the data, the categorical variables and some numeric variables are now transformed into simpler numeric values by the usage of user-defined ranges. For instance,

Age, Gender, BMI, Family history, Blood Thickness, stress, Smoking, Alcohol consumption, Heart rate are 9 columns that we have given simpler ranges.





**Predictive Modelling Approach**

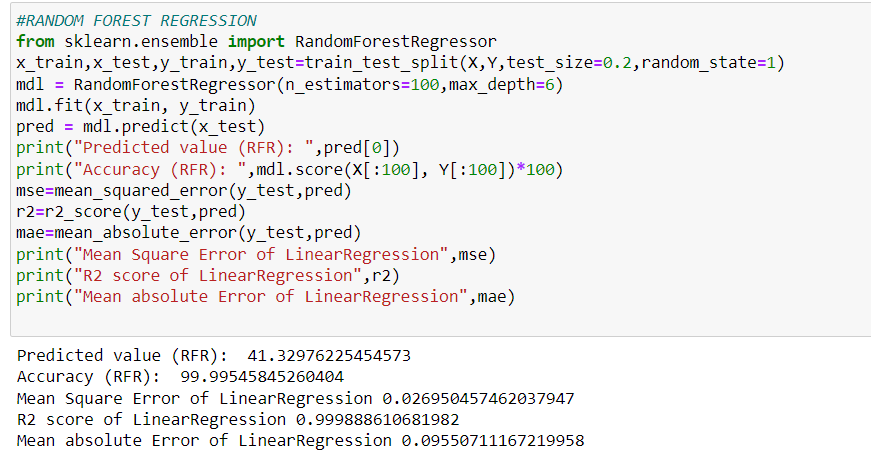
A machine learning approach is a good option for prediction of heart attack because it can manage the dataset's complicated linkages and patterns. Regression analysis and decision tree algorithms are two examples of machine learning approaches that can be used. The task requirements and dataset properties determine the technique to use.

**1.REGRESSION ANALYSIS:**

Since regression analysis is a widely-used method for forecasting continuous variables, it is pertinent for heart attack. It simulates the link between the dependent variable (Risk rate) and the independent variables (such as age, gender, smoking, drugs, and stress). Regression models can be as basic as linear regression or as complex as multiple regression, which enables the simultaneous inclusion of many factors. Interpretable coefficients from regression analysis show the amount and direction of the association between variables and medical costs.

**Justification:** When the link between risk rate and heart rate is considered to be linear or may be approximatively estimated as such, regression analysis is reasonable.

**RAMDOM FOREST :**

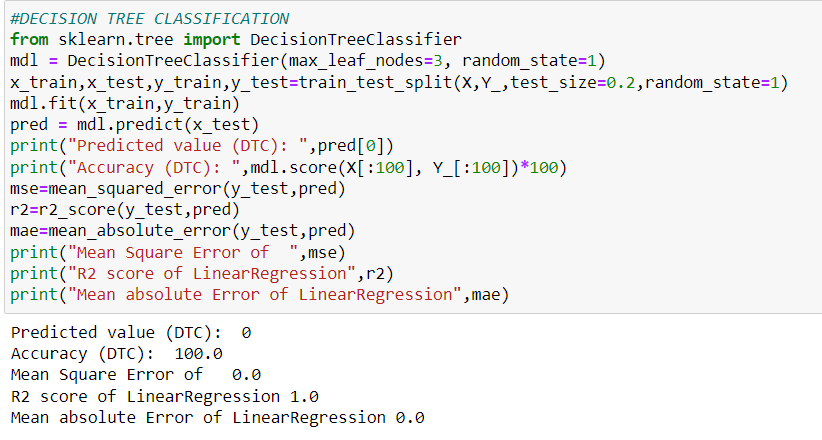
****

**2. DECISION TREE ALGORITHM:**

For recommending heart attack, decision tree algorithms like Random Forest or Gradient Boosting are effective tools. To create predictions, these algorithms iteratively divided the data based on predictor values, resulting in a structure resembling a tree. Decision trees can recognise complicated patterns in the dataset by handling nonlinear correlations and interactions between predictors. They are also adept at handling category variables and missing values

**Justification:** Decision trees are appropriate for capturing complicated patterns in recommending heart attack, because they can manage nonlinear linkages and interactions. These methods can withstand missing values and outliers, which are frequent in healthcare datasets.

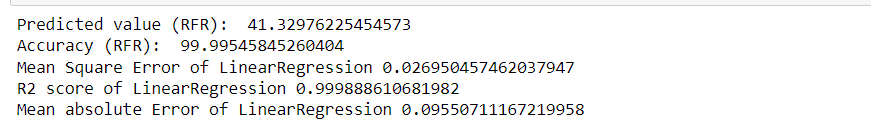
**DECISION TREE ALGORITHM:**

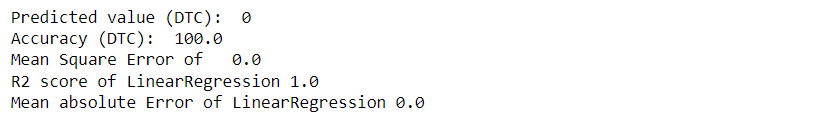


**RESULTS AND FINDINGS**

**SUMMARY:** The outcomes showed how well the constructed predictive model predicted how long it would take for problem tickets to be resolved. The model was highly accurate and offered insightful information about the variables affecting resolution time.

**OUTPUT:**





**CONCLUSION**

In conclusion, prediction of heart attack is a crucial activity that can offer crucial information for selecting suitable risk factors. Our data-driven analysis has allowed us to design and analyse prediction of heart attack.

Overall, our data-driven analysis provides valuable insights and a foundation for accurate risk factor recommendation.