

**CREDIT SCORE PREDICTION**

**Group Members:**

Jainam Shah (SUID: 785948053)

Anuj Purandare (SUID: 801847949)  
 Janvi Saddi (SUID: 440902034)

**Introduction:**

The financial sector has been expanding rapidly with the use of new technologies and innovations. Banks and financial institutions have collected a large amount of data over the years, including bank details and credit-related information. An individual's credit score plays a significant role in assessing their financial stability and creditworthiness. However, manually evaluating and segregating customers into different credit score brackets can be time-consuming and prone to errors. To address this issue, a global finance company plans to leverage data science to build an intelligent system that can automatically segregate customers into credit score brackets, streamlining the credit assessment process and reducing manual efforts.

The main goal of data scientists is to create a robust and precise credit score classification model that can efficiently segregate customers into credit score brackets based on their credit-related information. This model will take into account various factors such as credit history, outstanding debts, repayment behavior, and other relevant data points. By utilizing data science, the global finance company aims to provide a powerful tool that can significantly enhance its credit assessment process and reduce manual efforts.

**Problem Context:**

Credit scoring prediction is based on the business operations of banks and financial institutions. Traditionally, assessing an individual's creditworthiness has been done manually, which can be a time-consuming and error-prone process. However, with the increasing availability of data, there is a need for a more efficient way to evaluate a customer's creditworthiness. This is where credit scoring models come in.

Credit scoring models use data points such as credit history, outstanding debts, repayment behavior, and other relevant information to predict an individual's creditworthiness. By accurately predicting an individual's creditworthiness, these models can help streamline the credit assessment process and reduce manual efforts for financial institutions. This is important because the faster and more accurate the credit assessment process is, the better decisions can be made regarding lending money.

Furthermore, credit scoring models can help financial institutions make more informed lending decisions. By leveraging the power of data science, businesses can gain insights into patterns and trends that may not be visible with traditional assessment methods. This allows them to make lending decisions that are more accurate and less risky, ultimately leading to better business outcomes.

**Problem Statement:**

The traditional process of manually assessing an individual's creditworthiness can be time-consuming and error-prone. With the increasing amounts of data available, banks and financial institutions need a more efficient way to evaluate a customer's creditworthiness. To address this issue, we aim to develop a credit scoring model that can accurately predict an individual's creditworthiness based on their credit-related information. This model will take into account various factors such as credit history, outstanding debts, repayment behavior, and other relevant data points. The objective of this project is to create a robust and accurate credit score classification model that can efficiently segregate customers into credit score brackets, providing banks and lenders with a powerful tool to enhance their credit assessment process and reduce manual efforts.

**Objective/Outcome:**

The main objective of this project is to build an intelligent system that can segregate customers into different credit score brackets based on their credit-related information. The outcome of this project will be a robust and accurate credit score classification model that can efficiently and automatically classify customers into credit score brackets.

The model will be designed to take into account various factors such as credit history, outstanding debts, repayment behavior, and other relevant data points to generate a credit score for each customer. The benefits of this project include minimizing the errors associated with manual classification, and streamlining the credit assessment process.

Overall, the outcome of this project will be a powerful tool that can significantly enhance the credit assessment process.

**Research scope:**

The literature review and preliminary investigation are crucial components of any research project, and they provide a foundation for the development of a credit score classification model. The following areas will be covered in the literature review and preliminary investigation of this project:

Credit Scoring Models: A review of different credit scoring models that have been developed and used in the industry. This review will cover traditional scoring models, such as FICO and VantageScore, as well as newer models that leverage machine learning algorithms.

Feature Engineering Techniques: A review of different feature engineering techniques that can be used to extract and engineer relevant features from the available credit-related data. This review will cover techniques such as principal component analysis, feature scaling, and feature selection.

Machine Learning Algorithms: A review of different machine learning algorithms that can be used to develop a credit score classification model. This review will cover algorithms such as logistic regression, decision trees, random forests.

Ethics: A review of ethical considerations related to the development and deployment of a credit score classification model. This review will cover topics such as bias in machine learning algorithms, fairness, transparency, and accountability.

Industry Trends: A review of current industry trends related to credit assessment and scoring. This review will cover topics such as the use of alternative data sources, the impact of artificial intelligence and machine learning, and emerging technologies in the credit scoring domain.

Overall, the literature review and preliminary investigation will help identify best practices and challenges related to credit score classification and inform the development of an accurate and efficient credit score classification model for our global finance company.

**Solution Scope:**

The solution scope of this project includes the development of a credit score classification model that can accurately classify customers into different credit score brackets based on their credit-related information. The following components will be included in the project:

Data Exploration: Exploring the available credit-related data and identifying relevant data points that can be used to develop a credit score classification model.

Feature Engineering: Developing and engineering relevant features that can be used to train the credit score classification model.

Model Selection: Evaluating different machine learning algorithms and selecting the best algorithm that can accurately classify customers into credit score brackets.

Model Training and Validation: Training and validating the selected machine learning algorithm using relevant credit-related data to develop an accurate credit score classification model.

Model Interpretation: Interpreting the developed credit score classification model to understand the key factors that influence a customer's credit score and creditworthiness.

Deployment and Integration: Deploying the developed credit score classification model into the existing credit assessment process of our global finance company and integrating it with other relevant systems.

Performance Evaluation: Evaluating the performance of the developed credit score classification model by comparing it with existing manual credit assessment methods and analyzing the efficiency and accuracy of the automated credit assessment process.

The following components will be excluded from the project:

Data Collection: The project will assume that the credit-related data has already been collected and will not include any data collection activities.

Data Cleaning: The project will assume that the credit-related data has been cleaned and will not include any data cleaning activities.

User Interface Design: The project will not include the design and development of a user interface for the credit score classification model.

Legal and Regulatory Compliance: The project will assume that all legal and regulatory compliance requirements have been met and will not include any compliance activities.

Hardware and Software Infrastructure: The project will assume that the necessary hardware and software infrastructure is already in place and will not include any infrastructure setup activities.

By excluding these components from the project, we can focus our efforts on developing a robust and accurate credit score classification model that can be seamlessly integrated into the existing credit assessment process of our global finance company.

**Project Methodology:**

Identify a dataset that includes current and past employee records.

Clean the dataset, manage missing data and determine new features whenever needed.

Remove outliers and Null values from Dataset.

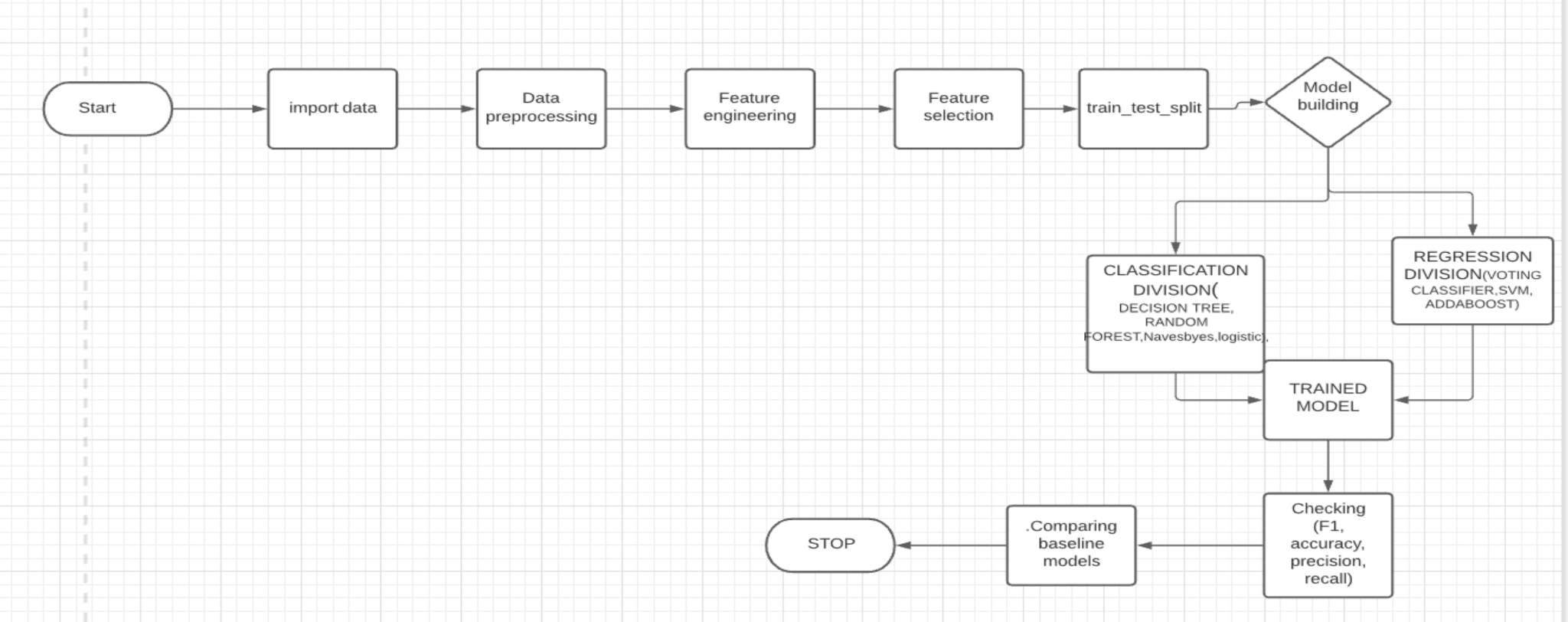
Select features among the individuals data that are appropriate for the attrition prediction.

Use the feature selection method, then select the most convenient features to assess an individual's creditworthiness.

Try applying classification, regression algorithms and store the precision, recall, Accuracy and F-measure results.

Compare the results of baseline models and find out the best model to predict employee attrition by comparing each other with their precision, recall, accuracy and f-measure results.

Apply hyperparameter tuning on the selected model to increase accuracy and precision even more.



**Resources Required:**

Software Requirements:

Python 3

Hardware Requirements:

1. Os : windows 7 and above recommended: Windows 10/11

2. Cpu: 1.60 Ghz processor and above

3. Disk Storage: 8gb of free disk space

4. Architecture: 64-bit

For Execution:

Jupyter Notebook/Google Colab

Development Time:

1-3 months

**References:**