IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.301
Volume 3, Issue 16, May 2023

A Research Paper on Credit Card Approval, Using Machine Learning models.

Sawan Nihatkar[1], Aditya Shahari[2], Tushar Patil[3]Atharva Kulkarni[4], Ms. Suvarna Potdukhe[5]
BE Student, Department of Information Technology[1,2,3,4]
Professor, Dept. of IT Engineering [5]
RMD Sinhgad School of Engineering, Pune, India

Abstract*:* This study focuses on the use of machine learning (ML) approaches to forecast consumer eligibility for a credit card in order to prevent any future credit risk that might influence the bank's financial stability and credit performance. A credit card is a credit facility provided to a consumer by banks and financial organisations all over the world. Banks and financial institutions face credit risk as a result of the credit facility. Repayments are rarely guaranteed, and the loan frequently becomes a non-performing credit facility (NPL). To reduce credit risk, banks analyse applicants' creditworthiness and eligibility before issuing a credit facility. Traditional credit scoring techniques are used to make the judgement, and credit worthiness is not always correct.Using predictive models, this initiative intends to assist banks and financial institutions in identifying and interacting with creditworthy consumers. To create models, we employed Stochastic Gradient Descent (SGD) and Logistic Regression. SGD Classifier, Logistic Regression, SVC, Decision Tree Classifier, Random Forest Classifier, Gaussian NB, K Neighbours Classifier, Gradient Boosting Classifier, Linear Discriminant Analysis, Bagging Classifier, MLP Classifier, Ada Boost Classifier, Extra trees Classifier. Both linear SVM and nonlinear SVM models were utilised to determine the optimal SVM approach. Statistical approaches are used for feature selection in filter-based feature selection methods. For training and test data, model accuracy was assessed using Mean Absolute Error, Confusion Matrix, and Area Under Curve (AUC). We tested three classifiers and discovered that Nonlinear SVM outperformed the others.

Keywords: SGD Classifier, Logistic Regression, SVC, Decision Tree Classifier, Random Forest Classifier, Gaussian NB, K Neighbours Classifier, Gradient Boosting Classifier, Linear Discriminant Analysis, Bagging Classifier, MLP Classifier, Ada Boost Classifier, Extra trees Classifier

I. INTRODUCTION

In different ways, commercial banks contribute to economic growth. The interest charged on loan is one of the most important revenue streams for any banking or financial organisation. Banks must bear the greatest credit risk in all of their lending. Banks provide a variety of loan solutions to their consumers. Credit cards, on the other hand, are one of the most important lending instruments that any bank could possibly have. Almost all financial institutions across the world are experiencing difficult times and credit risk when it comes to providing loans to their end consumers. Repayments are rarely guaranteed, and the loan frequently becomes a non-performing credit facility (NPL). Due to the credit risk aspect included in the credit card, banks and financial institutions are severely analysing eligibility for a credit facility before giving facility to the consumer. This procedure comprises verification, validation, and approval and may result in a delay in issuing a facility, which is detrimental to both the applicant and the bank. Credit officers decide whether borrowers can meet the conditions for a facility, and their judgements and projections are never correct. Credit scoring is a classic way of determining a customer's or entity's trustworthiness when applying for a bank credit facility.

When issuing a credit card to a client in the past, banks had to rely on the applicant's background and history to determine the applicant's creditworthiness. The procedure includes scrutinising application data using reference papers, which was not always precise, and both consumers and the bank had difficulty approving the credit card. However, with the digital transition, there has been an increase in Artificial Intelligence and Machine 2 Learning Technology

ISSN
2581-9429
IJARSCT

during the last two decades. As a result, machine learning approaches are being utilised to analyse credit risk and automate credit scoring by properly predicting consumer eligibility using customer demographic data and past transactional data.

1.1 Background:

The credit card industry plays a crucial role in the financial sector, providing individuals with convenient access to credit. However, the process of approving credit card applications involves assessing the creditworthiness and risk associated with each applicant. Traditional approaches to credit card approval often rely on manual evaluation, which can be time-consuming and subjective.With the advancements in machine learning and data analytics, there has been a growing interest in developing predictive models for credit card approval. These models aim to automate the decision-making process, enhance efficiency, and improve accuracy in assessing creditworthiness.The primary objective of this research is to develop a machine learning-based model for credit card approval prediction. By analysing historical data and relevant features of applicants, the model aims to identify patterns and relationships that can effectively predict the likelihood of credit card approval.Machine learning algorithms have the capability to analyse large volumes of data, identify complex patterns, and make accurate predictions. By leveraging these algorithms, credit card issuers can streamline their approval process, minimize the risk of default, and make more informed decisions.This study utilizes a comprehensive dataset containing various applicant attributes such as income, employment status, credit history, and demographic information. The dataset is collected from a reliable source and represents a diverse range of applicants.The research involves implementing and evaluating several machines learning algorithms, including logistic regression, support vector machines, decision trees, random forests, and neural networks. These algorithms will be trained on the dataset to learn patterns and build predictive models.The performance of the developed models will be assessed using appropriate evaluation metrics such as accuracy, precision, recall, and F1-score. This will provide insights into the effectiveness and reliability of the models in predicting credit card approvals.The successful development of a reliable credit card approval prediction model can have significant implications for both credit card issuers and applicants. It can lead to more efficient and fair credit evaluation processes, faster turnaround times for applicants, reduced risks for lenders, and improved customer satisfaction.It is important to acknowledge the limitations and challenges associated with credit card approval prediction using machine learning. These may include data quality issues, class imbalance, feature selection, and interpretability of the models. Addressing these challenges will be crucial to ensure the accuracy and reliability of the developed models.

II. MOTIVATION

The motivation behind using machine learning for credit card approval prediction is to enhance the efficiency, accuracy, and automation of the decision-making process. Traditional credit card approval methods often rely on manual assessment and subjective judgment, which can be time-consuming, prone to human errors, and limited in their ability to analyse complex patterns and large amounts of data.By leveraging machine learning algorithms, we can create predictive models that analyse historical applicant data and relevant features to identify patterns and make data-driven decisions on credit card approvals. This approach offers several benefits:

1. Improved Efficiency: Machine learning models can process large volumes of data quickly and efficiently. By automating the credit card approval process, it reduces the time and effort required for manual review, resulting in faster turnaround times for applicants.
2. Enhanced Accuracy: Machine learning algorithms can learn from historical data patterns and identify subtle correlations and indicators of creditworthiness. This enables more accurate predictions and reduces the risk of human biases and errors in the decision-making process.
3. Scalability: Machine learning models can scale to handle large datasets and adapt to changing trends and patterns in credit card applications. As the volume of applications increases, the models can handle the workload without compromising accuracy or efficiency.

The ML models used in this credit card approval prediction project include:

- SGD Classifier: A linear classifier trained using stochastic gradient descent.
- Logistic Regression*:* A linear model that predicts the probability of a binary outcome using logistic function.

- *SVC*: Support Vector Machines, a classifier that finds an optimal hyperplane to separate data points into different classes.
- Decision Tree Classifier: A classifier that uses a decision tree to make predictions based on features.
- Random Forest Classifier: An ensemble classifier that combines multiple decision trees for improved accuracy.
- Gaussian NB: A classifier based on the Gaussian distribution, assuming independence between features.
- K Neighbours Classifier: A classifier that makes predictions based on the k nearest neighbors in the feature space.
- Gradient Boosting Classifier*:* An ensemble classifier that combines multiple weak classifiers to create a strong classifier.
- Linear Discriminant Analysis*:* A classifier that finds linear combinations of features that maximize class separability.
- Bagging Classifier: An ensemble classifier that combines multiple classifiers trained on different subsets of the data.
- MLP Classifier: Multi-layer Perceptron, a classifier that uses artificial neural networks for prediction.
- Ada Boost Classifier: An ensemble classifier that combines multiple weak classifiers to create a strong classifier using adaptive boosting.
- Extra Trees Classifier: An ensemble classifier that combines multiple decision trees with random feature selection.

By utilizing these machine learning models, we aim to build robust and accurate credit card approval prediction systems that can provide valuable insights to financial institutions, reduce the risk of defaults, and ensure fair and efficient credit card approval processes.

III. GOALS AND OBJECTIVES

The goals and objectives of credit card approval prediction using machine learning are centred around improving the efficiency, accuracy, and fairness of the credit card approval process. Here is a detailed explanation of the goals and objectives, one of the key objectives is to streamline and expedite the credit card approval process. Traditional methods often involve manual review and assessment, which can be time-consuming and lead to delays in processing applications. By implementing machine learning models, the goal is to automate the decision-making process and reduce the time and effort required for manual reviews. This would result in faster turnaround times for applicants and improve overall operational efficiency. Another objective is to enhance the accuracy of credit card approval decisions. Machine learning algorithms can analyse large volumes of historical applicant data and identify patterns, correlations, and indicators of creditworthiness. By leveraging these patterns, the models aim to make more accurate predictions and reduce the risk of human biases and errors. The goal is to improve the precision and reliability of credit card approval decisions, ensuring that deserving applicants are approved while minimizing the risk of defaults. Fairness is a crucial objective in credit card approval prediction. Machine learning models can help mitigate potential biases and ensure a fair evaluation of credit card applications. By using a diverse range of features and training the models on representative datasets, the goal is to eliminate or reduce discriminatory practices that may be present in manual decision-making processes. The objective is to create a system that treats all applicants fairly and evaluates their creditworthiness based on objective criteria, regardless of personal attributes such as gender, race, or ethnicity. Machine learning models can identify relevant risk factors that contribute to credit card defaults. By analysing historical data, the models can uncover patterns and correlations that indicate potential risks. The goal is to provide insights to financial institutions about the specific factors that increase the likelihood of defaults, such as high debt-to-income ratios or past payment delinquencies. This information can assist in making informed decisions, setting appropriate credit limits, and managing credit card portfolios more effectively.

Another objective is to optimize resource allocation within financial institutions. By accurately predicting credit card approvals, the models can help allocate resources efficiently by focusing on promising applicants and minimizing efforts on applications that are likely to be declined. This can lead to cost savings and better utilization of human

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.301
Volume 3, Issue 16, May 2023

resources, allowing the institution to allocate their resources more effectively and serve customers in a timely manner.In summary, the goals and objectives of credit card approval prediction using machine learning revolve around improving efficiency, enhancing accuracy, ensuring fairness, identifying risk factors, and optimizing resource allocation. By leveraging machine learning algorithms, financial institutions can make data-driven and objective credit card approval decisions, ultimately benefiting both the institution and the applicants.

II. LITERATURE SURVEY

While there are a few systems and models currently in place that use Support Vector Machine and Artificial neural network ANN models for prediction of approval. They lack the desired accuracy and hence there are a few drawbacks that we need to assess.

Limitations in the current model are: -

Deepak Ishwar Gouda [1] The distribution of loans is practically every bank's fundamental operation. The majority of the bank's assets are directly derived from the profit gained on loans provided by the bank. The primary goal of banking environment is to put their funds in safe hands wherever they are. Today, many banks/financial companies approve loans after a lengthy process of verification and validation, but there is no guarantee that the chosen applicant is the most deserving of all applicants. We can forecast if a given applicant is safe or not using this approach, and the entire feature validation process is automated using machine learning techniques. Loan Prediction is extremely beneficial to both bank employees and applicants. The goal of this Paper is to give a quick, uncomplicated approach to choose qualified applicants. It may give the bank with specific benefits. The Loan Prediction System can automatically determine the weight of each feature involved in loan processing, and the same characteristics are processed with respect to their associated weight on new test data. A time restriction might be imposed for the applicant to determine whether or not his or her loan can be approved. The Loan Prediction System allows you to skip to a single application and check it on a priority basis. SVM, on the other hand, employs a statistical learning model for prediction categorization. To assess the suggested technique, a dataset from the UCI repository with 21 characteristics was used. Experiments revealed that, rather than independent classifier performances (NB and SVM), the integration of NB and SVM resulted in an effective loan prediction classification. During the data analysis, the following major factors were concentrated: annual income versus loan purpose, customer trust, loan tenure versus delinquent months, loan tenure versus credit category, loan tenure versus number of years in current job, and chances for loan repayment versus house ownership. Finally, the current work resulted in inferring the constraints on the customer who is applying for the loan, followed by a prediction regarding the repayment. Furthermore, the results revealed that customers preferred to take out short-term loans over long-term loans.

Ms. Kathe Rutika Pramod [2] Small loans are an important part of our daily lives because they allow aspiring entrepreneurs to get started on ideas that could grow into businesses; they allow curious students to afford higher education that would otherwise be unavailable without a stable income; and, most importantly, they allow ordinary people who have no friends or relatives for support to obtain short-term financial assistance and get back on their feet to fight for the American Dream. However, as with any loan, there is the possibility of default. Default is a financial phrase that describes the failure to satisfy a loan's legal obligation - paying back the principal and interest. It's a widespread issue in the financial industry and one of the key dangers of lending. Overall, default is a fact of life, and most financial institutions have a well-established practice for mitigating its impact and absorbing the loss. But what if, instead of a single bank, the loan is made up of cash contributed by several investors? Lending Club is one of several peer-to-peer lending companies that have contributed to this unusual predicament. In simple terms, a peer-to-peer lending organization operates as a middleman between borrowers and investors. The firm develops a platform on which borrowers may generate modest unsecured personal loans, and investors can search for these loans and choose which ones to invest in. There is little question that Lending Club already has a process in place for approving loans submitted on their website. This study will investigate the process and outcome of developing a new machine learning model that can forecast loan default; however, the model will focus on minimising the total loss in investment of bad loans in order to reduce the burden placed on individual investors. In addition, the article will investigate privacy-preserving mechanisms for sensitive information obtained from the borrower's credit report. The ultimate purpose is to analyse a

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.301
Volume 3, Issue 16, May 2023

simpler version of RAPPOR (Randomised Aggregately Privacy Preserving Ordinal Response) to see if data hashed by this technique can still be used to predict loan default as indicated.

E. Chandra Blessie [3] Almost every bank's business strategy revolves around finance raising and lending for real estate, consumer, mortgage, and commercial loans. The primary source of credit risk is lending money to unsuitable clients. The majority of the bank's assets are obtained directly from the profits made on its loans. However, banking companies face a dual challenge in distinguishing possible deliberate defaulters from applicants, as well as the biased nature of a few bank employees who have been at the behest of developers of defaulting companies for many years. The basic purpose of the banking community is to invest their cash securely. Many NBFCs and banks now approve loans after thorough verification and authentication. NB and Support Vector Machines (SVM) approaches were used to develop a loan prediction model. Nave Bayes is an independent speculation technique that incorporates probability theory in data categorization. SVM, on the other hand, employs a statistical learning model for prediction categorization. To assess the suggested technique, a dataset from the UCI repository with 21 characteristics was used. Experiments revealed that, rather than separate classifier performances (NB and SVM), the integration of NB and SVM resulted in an efficient categorization of loan forecasts. The loan sanctioning prediction process is based on the NB approach combined with the K-Nearest Neighbour (KNN) and binning algorithms. Income, age, occupation, current loan with duration, amount, and approval status were the seven criteria examined. (1) Pre-processing (managing missing values with KNN and data refining using binning technique), (2) Classification using NB approach, and (3) Frequent dataset updates result in adequate improvement in the loan prediction process. Experimentation led to the conclusion that integrating KNN and binning algorithms with NB resulted in enhanced loan sanctioning process prediction. Using the R programme, I proposed a risk analysis approach for authorizing a loan for the clients. Data selection, pre-processing, feature extraction and selection, creating the model, prediction, and assessment are among the modules. The dataset for this method's assessment was obtained from the UCI repository. The pre-processing operation contains the following sub-processes to fine-tune the prediction accuracy: identification, ranking, and removal of outliers, elimination of imputation, and balancing of dataset via proportionate bifurcation related testing and training process. Additionally, the feature selection process improves prediction accuracy. When tested, the DT model had a prediction accuracy of 94.3%.

Yu Li [4] The scoring method based on logistic regression algorithm has the following advantages: (1) simple algorithm and mature technology; (2) robust estimation of probability under given data conditions; (3) strong explanatory power of variables and models; and (4) easy detection and deployment of models. But, at the same time, the traditional credit scoring system has some flaws: (1) The model has a restricted number of variables and may be exploited by fraudsters. (2) the logistic regression algorithm must fulfil certain assumptions, but the actual business may not meet the related assumptions; (3) the logistic regression method's differentiation capacity is difficult to enhance. Other supervised learning methods in machine learning, such as neural networks, nearest neighbor methods, and support vector machines, have been rapidly developed over the last 40 years, in addition to the logistic regression algorithm [1]. Logical regression dates back to the 1950s. Three traditional decision tree implementations. In this article, the observation period is defined as the most recent credit report received within two months of the application date. The performance term is 15 months from the date of application. There are two types of response variables: (1) poor samples are 90 days or more late within 15 months; (2) excellent samples are no more than 89 days late within 15 months. To assure the predictive modelling impact and the speed of algorithm operation, stratified sampling was used to provide model samples, which included 45,000 excellent samples, 5,000 poor samples, and a 9:1 good sample to bad sample ratio. The credit risk prediction using machine learning is investigated in this study. When the model discrimination, model interpretability, and model stability of the logistic regression model and the XGBoost model are compared, it is clear that the XGBoost model has significantly higher model discrimination and model stability than the logistic regression model, which can effectively improve the identification ability of personal fast credit risk.

Archana Gahlaut [5] Credit allows an individual or corporation to "buy ahead of ability" or "desire to pay." Banks make loans to people based on their requirements in the agricultural, industrial, and commercial sectors. Furthermore, when people use their brilliant minds and entrepreneur skills in the presence of credit, it results in overall economic growth, thereby strengthening the country's economy. With today's developing country growth, it might be risky for banks to extend credit to all of its clients without knowing whether or not they will be able to repay it on time. The

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.301
Volume 3, Issue 16, May 2023

same might be a huge problem for all banks nowadays, since the overall loss will be enormous if clients do not return credit loans with interest on time, resulting in bankruptcy. We offered various data mining models to lessen the probability of failure. These models will aid in determining a customer's ability to repay credit loans on time by using credit scoring and categorizing them as a 'Good credit'- customer has a good score and no faulty or defaulter past credit records- or a 'Bad credit'- customer has a bad score and may have faulty past records. Banks will be able to provide decent credit, which will eventually result in a profit in their annual income. Many studies in the banking and insurance analytics sectors have discussed related issues within the data mining framework. For example, Jin et al. used a data-driven approach and a data mining approach to predict loan risk and compared data mining models: decision trees, support vector machines, and neural networks, using a 10-fold cross-validation approach and a high average percent hit ratio to demonstrate the better prediction. To assess the quality, a cumulative lift curve analysis is performed. [1] The Support Vector Machine fared the best.

ZHANG Lei-lei [6] Credit scoring has piqued the curiosity of many researchers in the literature. The credit rating manager frequently analyses the consumer's credit based on intuition. The manager, on the other hand, can precisely analyse the applicant's credit score with the help of the credit categorization model. Support Vector Machine (SVM) classification is a recent research topic that effectively addresses classification issues in a variety of disciplines. In order to create a model with more explanatory power, this article applies support vector machines (SVM) to the problem. For the Australian and German credit datasets from UCI, we used backpropagation neural network (BNN) as a benchmark and obtained prediction accuracy of around 80% for both BNN and SVM methods. Credit scoring is a way of estimating the risk of credit applications. In general, it combines statistical methodologies and historical data to provide a score that financial organizations may use to assess the risk of credit applications.[1] Credit scoring models have been widely employed for credit admittance evaluation due to the fast expansion of the credit business. Several quantitative methods for credit admission decision making have been developed over the last two decades. Credit scoring algorithms are designed to categories applicants as either approved or refused based on their factors such as age, income, and marital status. [2-4] lending officers are faced with the challenge of growing lending volume without increasing their risk of default. SVMs evolved from statistical learning theory, with the goal of addressing only the problem of interest without having to tackle a more complex problem as an intermediary step.

SVMs are founded on the notion of structural risk minimization, which is closely connected to regularization theory. [15-17]

This concept includes capacity management to prevent over-fitting and so provides a partial solution to the bias-variance trade-off quandary. The mathematical programming approaches and kernel functions are two critical components in the implementation of SVM. Rather than addressing a non-convex, unconstrained optimization problem, the parameters are determined by solving a quadratic programming problem with linear equality and inequality constraints. Because of the versatility of kernel functions, the SVM may explore a wide range of hypothesis spaces.

Mohammad Ahmad Sheikh [7] Banks have numerous goods to sell in our banking system, but the major source of income for each bank is its credit line. As a result, they may profit from the interest on the loans they credit. A bank's profit or loss is heavily influenced by loans, namely whether clients repay the loan or fail on it. The bank can lower its Nonperforming Assets by forecasting loan defaulters. As a result, research into this phenomenon is critical. Previous research in this age has revealed that there are several techniques for studying the subject of loan default control. Logistic Regression models have been run, and various performance indicators have been obtained. The models are compared using performance metrics such as sensitivity and specificity. The final findings demonstrated that the model produces diverse outcomes. Model is marginally better because it includes variables (personal attributes of a customer such as age, purpose, credit history, credit amount, credit duration, and so on) other than checking account information (which shows a customer's wealth) that should be taken into account to correctly calculate the probability of loan default. As a result, using a logistic regression technique, the ideal clients to target for loan giving may be simply identified by analysing their chance of loan default. The suggested model predicts whether a bank will give a loan to a customer. Classification is the goal for constructing the model, hence Logistic Regression with a sigmoid function is employed to create it. Pre-processing is the most time-consuming component of the model, followed by Exploratory Data Analysis, Feature Engineering, and Model Selection. Feeding the model with two distinct datasets, followed by the model. Logistic regression is a statistical machine learning technique/algorithm used to categorize data by analysing

DOI: 10.48175/IJARSCT-10986

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.301
Volume 3, Issue 16, May 2023

outcome variables on extreme ends and attempting to draw a logarithmic line that differs between them. Logistic Regression may be used to make predictions in this manner.

Md. Golam Kibria [8] The growing number of credit card defaulters has compelled businesses to exercise caution before approving credit applications. Credit card firms typically use their discretion to choose whether or not to offer a credit card to a consumer who meets specific requirements. Some machine learning algorithms were also used to help make the decision. The primary goal of this study is to develop a deep learning model based on UCI (University of California, Irvine) data sets that can aid in credit card acceptance decisions. Second, the created model's performance is compared to that of two more standard machine learning algorithms: logistic regression (LR) and support vector machine (SVM). Our findings reveal that our deep learning model performs somewhat better overall. The emergence of the internet has resulted in a considerable increase in credit card usage. It is becoming one of the most popular payment options. Credit card fraud is expanding at an alarming rate as the global economy grows [1]. It is also clear that the number of credit card defaulters has grown dramatically. As a result, credit card companies are becoming more cautious when giving credit cards to clients. Furthermore, the downturn of financial institutions in the United States and Europe during the subprime mortgage crisis in the United States and the European sovereign crisis in Europe has raised concerns about proper risk management [2]. As a result, scholars and practitioners have paid close attention to these difficulties. To handle credit card-related challenges, a variety of statistical and machine learning approaches have been developed (see [1]-[7]). It has been discovered that machine learning approaches outperform other traditional statistical techniques when it comes to credit rating [8–11]. Deep learning, in particular, is a popular and accurate classification approach that outperforms other machine learning models (for example, logistic regression (LR), linear discriminant analysis (LDA), multiple discriminant analysis (MDA), k-nearest neighbour (k-NN), decision trees, and so on). [12]. Specifically, the DL achieves the highest F1-measure score of .886, indicating the overall performance of the model, based on the F1-measure. The F1-measure value for SVM is.863, whereas it is.861 for LR. These two methods gave nearly identical F1-measure results. The SVM beat the other two algorithms in terms of false positive rate, with 12.80% for SVM, 16.10% for LR, and 16% for DL. Based on all accuracy indicators except FP in Table V, we can infer that the deep learning model outperforms the other two models.

Ambika and Santosh Biradar [9] People desire to apply for loans through the Internet as data volumes increase due to banking sector digitalization. Artificial intelligence (AI), as a common approach to information exploration, is gaining popularity. Individuals from diverse industries are using AI calculations to solve problems based on their sector knowledge. Banks are experiencing substantial difficulties in loan approval. Every day, there are several apps that bank staff must maintain, and the likelihood of errors is great. Most banks benefit from loans, but selecting eligible consumers from a large number of applications is hazardous. A single error might result in a significant loss for a bank. Loans have simplified our lives by giving us with financial leverage that goes beyond our wages. Loans, whether Credit Card, Home Loan, Personal Loan, or Auto Loan, are credit granted to us by lenders based on certain critical characteristics. Obtaining a loan in India, on the other hand, might be a time-consuming procedure for the uninitiated, but not for those with a decent credit score. Banks use your CIBIL Score and Report to analyse your credit history and credit eligibility whenever you ask for a loan. The higher your score, the more likely your loan application will be accepted. This application is operational and meets all Banker criteria. This component is easily pluggable into a variety of different systems. It functions properly, meets all banking criteria, and may be linked to a variety of different systems. There were several computer faults, content problems, and weight correction in computerized forecast systems. In the near future, banking software might be more dependable, accurate, and dynamic, and it could be integrated with an automated processing unit. There have been several incidents of computer glitches, content mistakes, and most importantly, the weight of features has been addressed in automated prediction system to make it more secure, dependable, and dynamic weight adjustment. The system is trained using old training datasets, however future software may be designed such that new testing data can be included in training data after a certain period of time. Machine learning aids in understanding the elements that have the greatest influence on certain results. Other methods, like as neural networks and discriminate analysis, can be employed alone or in combination to improve prediction reliability and accuracy.

IJARSCT

ISSN (Online) 2581-9429

International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)

International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal

Impact Factor: 7.301

Volume 3, Issue 16, May 2023

Aboobyda Jafar Hamid [10] Customer segmentation and profitability, high risk loan applicants, predicting payment default, marketing, credit analysis, ranking investments, fraudulent transactions, optimizing stock portfolios, cash management and forecasting operations, most profitable Credit Card Customers, and Cross Selling are just a few examples of how data mining can be used in the financial sector. When seeking to borrow money, there are many different sorts of loans to choose, and it's critical to understand your alternatives. Loan categorization is the process of evaluating loan collections and grouping or grading loans based on perceived risk and other loan attributes. Loans come in a variety of forms, including: Open-ended loans are loans that can be extended indefinitely. The most well-known kinds of open-ended loans are credit cards and lines of credit. You have a credit limit that you may use to purchase items with one of these two sorts of loans. Your available credit will drop whenever you may purchase automatically. As you spend, your cash on hand grows, allowing you to use the credit more and more. Closed-ended loans cannot be borrowed again once they have been repaid. When you make payments on closed-ended loans, the loan balance decreases. However, you do not have any existing credit to use on closed-ended loans. Secured loans are those that are backed by an asset. In the event of loan default, the lender may seize the asset and utilize it to satisfy the amount. Secured loan well-being rates may be lower than unsecured loan well-being rates. Before you may get a secured loan, the asset may need to be examined. Unsecured loans may be more difficult to get and have higher concern rates. Unsecured loans rely only on your credit history and income to fulfil the lending conditions. If you fail to repay an unsecured loan, the lender must exhaust collection options such as debt collectors and file a claim to recover the payment. Prediction and description are the two most significant aims for data mining. Prediction entails utilizing certain variables in a data set to forecast the unknown values of other variables. Description focuses on identifying patterns in data that can be comprehended by humans. Data mining is the process of identifying hidden patterns in massive amounts of data in order to make write judgements. The generated information must be novel, not obvious, relevant, and applicable in the field where it was obtained. It is also the technique of obtaining usable information from unstructured data. Data Mining is a fascinating and important field of study that aims to extract information from massive amounts of amassed data sets. Data mining is growing more popular in the banking industry since there is a need for effective analytical methodologies for finding unknown and important information in bank data. Skills and expertise are crucial requirements for completing Data Mining tasks since the success or failure of Data Mining is heavily dependent on the person leading the process owing to the lack of a standard framework.

III. METHODOLY

This software predicts whether or not a credit card applicant will be accepted. Each time there is a hard inquiry, your credit score suffers. This programme predicts your chances of approval without hurting your credit score. Applicants who wish to check out if they will be approved for a credit card without impacting their credit score can use this app. Gradient boosting was the final model employed for this assignment. Recall was utilised as a metric.

Why use recall as a metric: Because the goal of this challenge is to reduce the risk of a loan default, the metrics to employ are determined by the present economic situation:

People feel prosperous and employed during a bull market (when the economy is rising). Money is typically inexpensive, and the danger of default is low due to economic stability and low unemployment. Because the financial institution can handle the risk of default, it is not too cautious when it comes to credit. The financial institution can manage a few problematic customers as long as the majority of credit card holders are excellent customers (those who pay back their credit on time and in full). In this instance.

People lose employment and money in the stock market and other investment venues during a bear market (when the economy is declining). Many people have difficulty meeting their financial commitments. As a result, financial institutions tend to be more conservative when it comes to granting credit or loans. The financial organisation cannot afford to extend credit to many consumers who will be unable to repay it. The financial institution would prefer to have fewer good clients, even if it means turning down some good ones. In this instance, high accuracy (specificity) is preferred All paragraphs must be indented.

It should be noted that there is always a trade-off between precision and memory. Choosing the appropriate measurements is dependent on the task at hand. According to the findings of this experiment, the three most predictive factors in predicting whether an application would be authorised for a credit card are income, family member number,

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.301
Volume 3, Issue 16, May 2023

and employment length. Other factors such as age and employment status are also useful. The kind of residence and automobile ownership are the least relevant characteristics.

When reviewing the application profile, the idea is to pay more attention to the most predictive aspects and less attention to the least predictive features.

The methodology for credit card approval prediction using the mentioned machine learning models typically involves the following steps:

1. Data Pre-processing:
    - Data Cleaning: Remove any duplicate entries, handle missing values, and correct inconsistent or erroneous data.
    - Feature Selection: Identify relevant features that can contribute to the prediction task and remove irrelevant or redundant ones.
    - Feature Encoding: Convert categorical features into numerical representations using techniques such as one-hot encoding or label encoding.
    - Data Split: Divide the dataset into training and testing sets to evaluate the performance of the models.

2. Model Training:
    - Initialize each machine learning model with appropriate hyperparameters.
    - Fit the models on the training data, allowing them to learn patterns and relationships between the input features and the target variable (credit card approval).

3. Model Evaluation:
    - Assess the performance of each model using evaluation metrics such as accuracy, precision, recall, and F1 score. These metrics help determine the effectiveness of the models in predicting credit card approval.
    - Compare the performance of different models to identify the most suitable one.

4. Model Selection and Fine-tuning:
    - Select the model that performs best based on the evaluation metrics.
    - Fine-tune the selected model by adjusting hyperparameters using techniques such as cross-validation or grid search to optimize its performance.

5. Model Deployment:
    - Deploy the selected and fine-tuned model to make predictions on new, unseen data.
    - Continuously monitor the model's performance and retrain/update it periodically to maintain its accuracy.

It is important to note that the specific implementation details and techniques may vary depending on the specific requirements of the credit card approval prediction task and the characteristics of the dataset. Additionally, feature engineering, model interpretation, and model explain ability techniques can be employed to gain insights into the factors influencing credit card approval decisions.

There are six phases involved in the credit card approval prediction project, following the Data Science Process Alliance framework. Each phase plays a crucial role in ensuring the success of the project. Let's take a closer look at each phase:

- Business Understanding: In this phase, the project team gains a thorough understanding of the project objectives and requirements. They develop a detailed plan that outlines the focus and scope of the project. This phase sets the foundation for the entire project and ensures alignment with the business goals.
- Data Understanding: During this phase, the team focuses on identifying, collecting, and analysing the data. They explore the data format and fields, and use visualization techniques to identify relationships and patterns. Additionally, they assess the quality of the data, determining if it is clean or contains inconsistencies that need to be addressed.

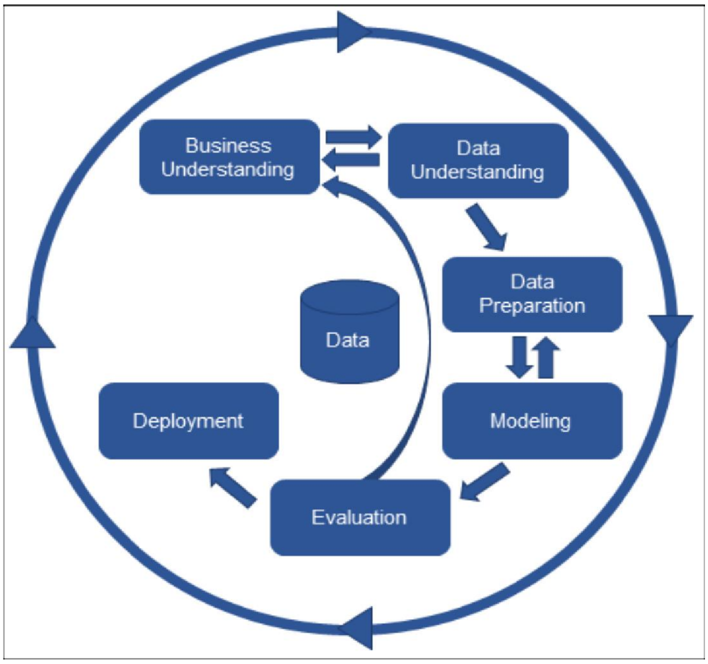

Fig 1. Application Model

- Data Preparation: Referred to as "data munging" or "data wrangling," this phase involves preparing the data for further analysis. The team selects the relevant data, cleans it by removing errors or duplicates, constructs new variables if needed, integrates data from different sources, and formats it in a way that is suitable for modelling.
- Modelling: In this phase, the team determines the appropriate algorithms to use for the credit card approval prediction task. They design the tests, build the models using the selected algorithms, and assess their performance. Model selection and assessment are critical activities in this phase to ensure that the models are capable of meeting the business requirements.
- Evaluation: The evaluation phase focuses on identifying the model that best fits the business requirements. The team evaluates the results generated by the models, reviews the entire modelling process, and determines the next steps. This phase helps in making informed decisions about whether to proceed with model deployment or iterate further to improve the models.
- Deployment: The deployment phase emphasizes making the model outputs and results accessible for stakeholders. It involves creating a deployment plan, monitoring and maintaining the deployed models, producing a final report summarizing the project findings, and conducting a thorough project review. This phase ensures that the project outcomes are effectively implemented and aligned with the business objectives.

By following these six phases, the credit card approval prediction project can progress systematically and increase the chances of achieving accurate and reliable predictions.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.301
Volume 3, Issue 16, May 2023

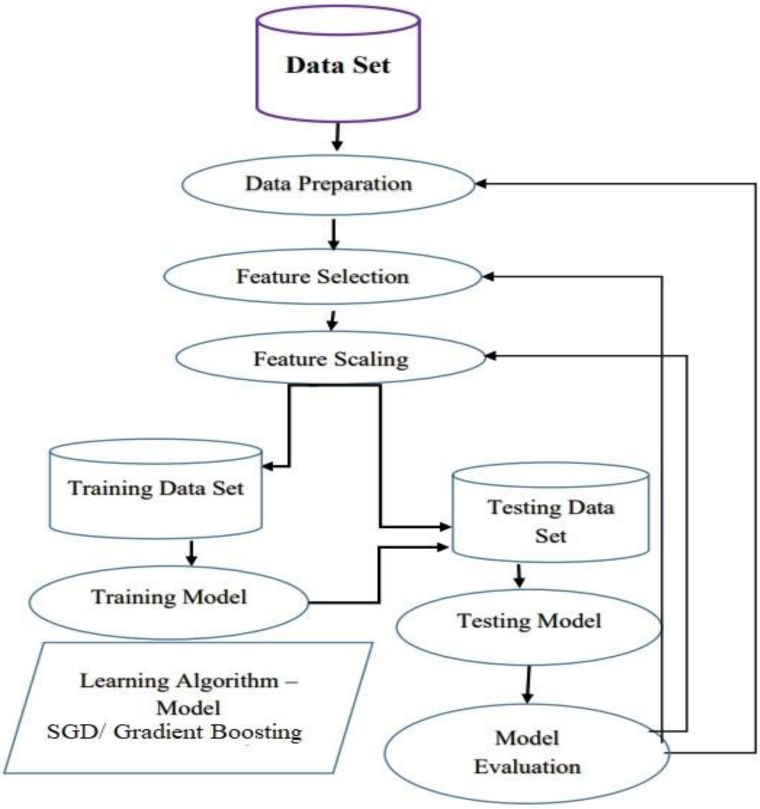

Fig 2. Modelling Work Flow

[A] SGD Classifier

SGDClassifier, which stands for Stochastic Gradient Descent Classifier, is a machine learning algorithm commonly used for classification tasks. It is a variation of the traditional gradient descent algorithm that can efficiently handle large datasets and high-dimensional feature spaces.The SGDClassifier works by iteratively updating the model parameters to minimize a specified loss function. It performs the updates in a stochastic (random) manner, using a single training example at a time rather than the entire dataset. This stochastic approach makes SGDClassifier particularly suitable for large-scale datasets, as it allows for faster training times and lower memory requirements compared to batch gradient descent algorithms.One of the key advantages of SGDClassifier is its flexibility in handling various types of loss functions. It can be used with different loss functions depending on the classification problem at hand, such as logistic loss for binary classification or hinge loss for multi-class classification. This adaptability allows SGDClassifier to be applied to a wide range of classification tasks.SGDClassifier also supports regularization techniques such as L1 and L2 regularization, which help prevent overfitting by adding a penalty term to the loss function. Regularization helps to control the complexity of the model and improve its generalization ability on unseen data.In terms of implementation, SGDClassifier is available in popular machine learning libraries such as scikit-learn in Python. It provides a flexible and efficient framework for training and deploying classification models.Overall, SGDClassifier is a versatile algorithm for classification tasks, offering advantages such as scalability, flexibility in loss functions, and regularization capabilities. It is commonly used in various domains where large-scale classification problems need to be addressed.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.301
Volume 3, Issue 16, May 2023

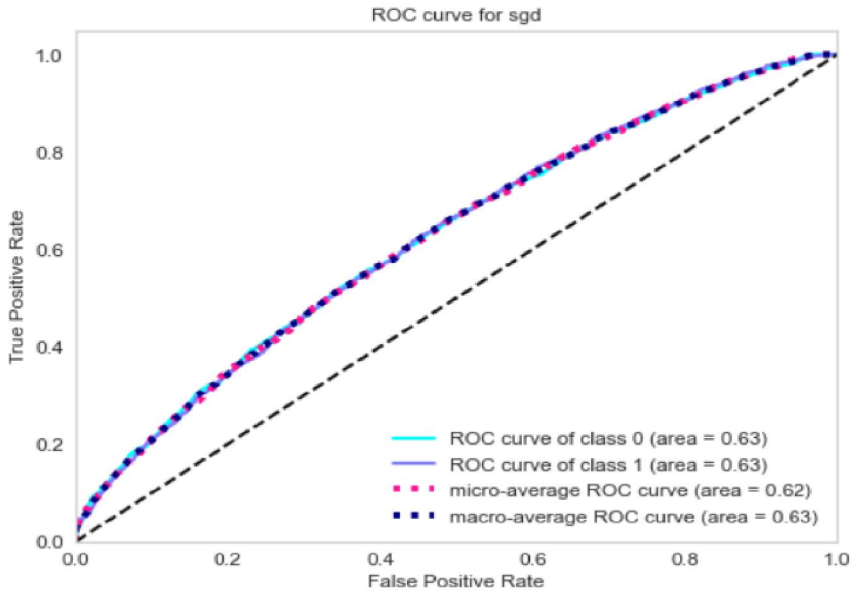

Fig 3. ROC Curve for SGD

[B] Logistic Regression

Logistic regression is a statistical model used for binary classification tasks, where the goal is to predict the probability of an event occurring. It is a popular and widely used algorithm in machine learning and has applications in various domains such as healthcare, finance, and marketing.In logistic regression, the dependent variable or the target variable is binary, meaning it can take only two possible values (e.g., 0 or 1, true or false). The independent variables, also known as features or predictors, can be either continuous or categorical. The objective of logistic regression is to find the best-fitting relationship between the independent variables and the probability of the event occurring.The logistic regression model uses the logistic function, also called the sigmoid function, to map the linear combination of the independent variables to a value between 0 and 1. The logistic function ensures that the predicted probability falls within this range, which is essential for binary classification. The model's parameters are estimated using maximum likelihood estimation, which involves finding the values that maximize the likelihood of observing the actual outcomes given the predictors.Interpreting the coefficients in logistic regression is different from linear regression. The coefficients represent the change in the log-odds of the event for a one-unit change in the corresponding independent variable, holding all other variables constant. The odds ratio can be derived from the coefficients, indicating how the odds of the event change with respect to the independent variables.To train a logistic regression model, the data is typically split into a training set and a testing/validation set. The model is trained on the training set, adjusting the parameters to minimize the difference between the predicted probabilities and the actual outcomes. The performance of the model is then evaluated on the testing/validation set, using metrics such as accuracy, precision, recall, and F1 score.Logistic regression has several advantages. It is a simple and interpretable model that provides insights into the relationship between the independent variables and the probability of the event. It can handle both categorical and continuous predictors and can be extended to handle multiclass classification problems. Logistic regression is also computationally efficient and can handle large datasets.However, logistic regression has some limitations. It assumes a linear relationship between the independent variables and the log-odds of the event. Non-linear relationships may require additional feature engineering or the use of more complex models. Logistic regression is also sensitive to outliers and multicollinearity among the independent variables, which can impact the model's performance.In conclusion, logistic regression is a widely used algorithm for binary classification tasks. It provides a probabilistic framework for estimating the likelihood of an event occurring based on the independent variables. By understanding the principles and assumptions of logistic regression, practitioners can effectively apply it to solve binary classification problems and gain insights from the model's coefficients and predictions.

IJARSCT

ISSN (Online) 2581-9429

International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)

International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal

Impact Factor: 7.301

Volume 3, Issue 16, May 2023

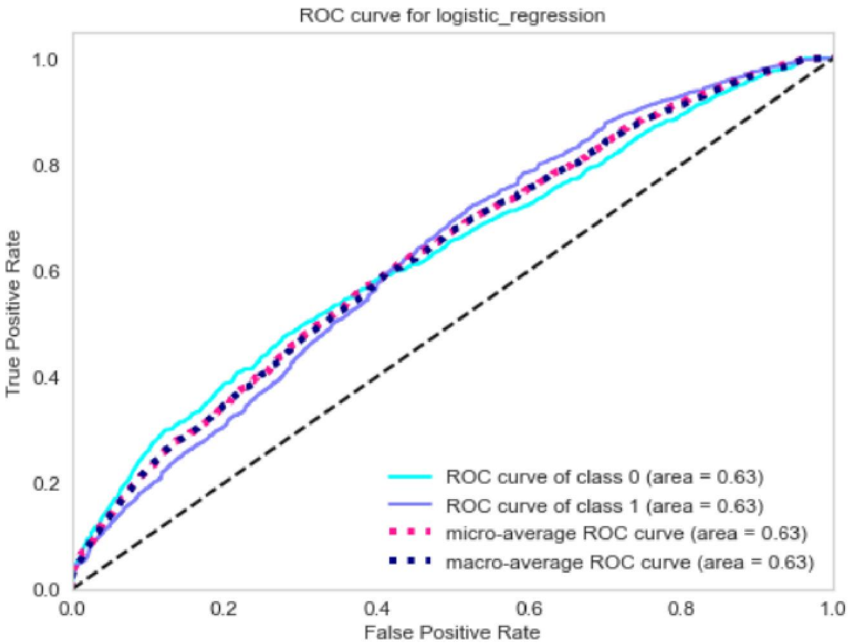

Fig 3. ROC For Logistic Regression

[C] Decision Tree Classifier

A decision tree is a versatile and intuitive machine learning algorithm used for both classification and regression tasks. It is a non-parametric model that learns simple decision rules from the data and represents them in the form of a tree-like structure. Each internal node of the tree represents a decision based on a specific feature, and each leaf node represents a class or a predicted value.The decision tree algorithm works by recursively partitioning the data based on the values of the features. At each step, it selects the feature that best splits the data and creates a decision node. The splitting criterion is often based on metrics like Gini impurity or information gain, which measure the homogeneity of the target variable within the resulting subsets. The process continues until a stopping criterion is met, such as reaching a maximum depth or a minimum number of samples in each leaf.

One of the main advantages of decision trees is their interpretability. The resulting tree structure can be easily visualized and understood, making it useful for explaining the decision-making process. Decision trees also handle both categorical and numerical features, as well as missing values, without requiring extensive data pre-processing. They are robust to outliers and can capture non-linear relationships between features and the target variable.However, decision trees are prone to overfitting, especially when the tree becomes too complex and captures noise or irrelevant patterns in the data. To address this, techniques like pruning and setting constraints on the tree's size can be applied. Additionally, decision trees may struggle with capturing complex relationships that require interactions between multiple features. To mitigate these limitations, ensemble methods like random forests and gradient boosting are often used, which combine multiple decision trees to make more accurate predictions. These methods aggregate the predictions of individual trees to achieve better generalization and reduce overfitting.In summary, decision trees are powerful and interpretable models that can handle both classification and regression tasks. They provide a clear and visual representation of the decision-making process, making them useful for understanding and explaining the underlying patterns in the data. While they have some limitations, such as overfitting, these can be addressed through techniques like pruning and ensemble methods. Decision trees are widely used in various domains due to their simplicity, flexibility, and ability to handle diverse types of data.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.301
Volume 3, Issue 16, May 2023

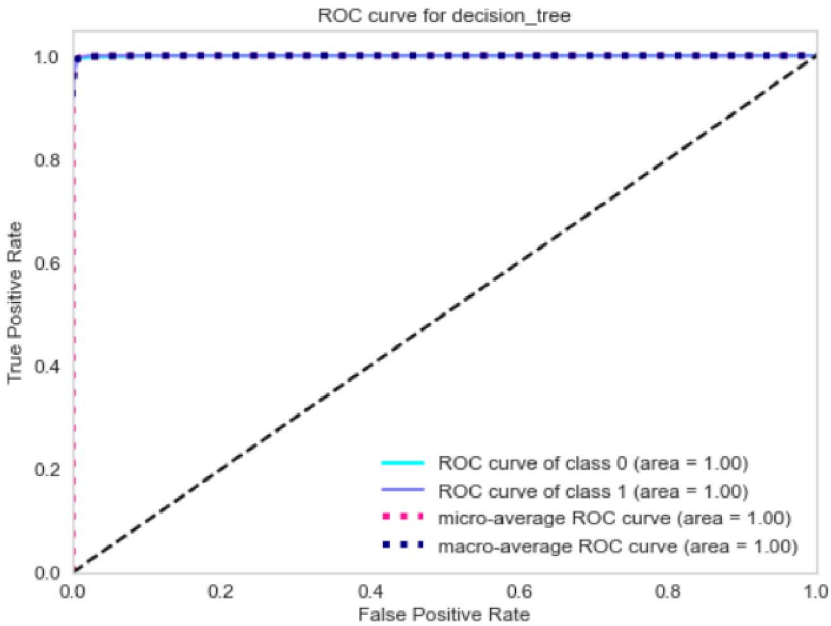

Fig 4. ROC For Decision Tree

[D] Feature Importance

In the context of credit card approval prediction using machine learning models, feature importance refers to the assessment of the relevance or contribution of each input feature in determining the prediction outcome. It helps in understanding which features have the most significant impact on the model's decision-making process.Feature importance is derived from the ML model's internal mechanisms, such as decision trees, random forests, or gradient boosting algorithms. These models assign importance scores to features based on various criteria, such as how frequently a feature is used for splitting nodes in decision trees or how much a feature reduces the impurity or error in the model's predictions.By analysing feature importance, we gain insights into the factors that influence credit card approval decisions. It helps us identify the key drivers or indicators of creditworthiness and risk. Features with higher importance scores indicate a stronger influence on the model's predictions, while features with lower scores have relatively less impact.Understanding feature importance can offer several benefits. It helps in feature selection, where we can focus on the most important features and discard less relevant ones. This can lead to simpler and more interpretable models, reducing the risk of overfitting and improving model performance. Feature importance also aids in identifying potential biases or data quality issues if certain features have unexpectedly high or low importance scores.Moreover, feature importance assists in explaining and justifying the model's predictions to stakeholders. It provides a transparent way of highlighting which factors contribute to a credit card application being approved or rejected. This information can be valuable for financial institutions, regulators, or customers seeking transparency and fairness in the decision-making process.It's important to note that the calculation of feature importance may vary depending on the specific ML model used. Techniques like permutation importance, Gini importance, or mean decrease impurity are commonly employed to quantify feature importance. Additionally, feature importance should be interpreted in the context of the particular dataset and problem domain, as the relevance of features can vary in different scenarios.In summary, feature importance in credit card approval prediction models helps identify the key factors influencing creditworthiness decisions. It aids in feature selection, model interpretation, and transparency, allowing for better-informed decision-making and understanding of the credit evaluation process.

DOI: 10.48175/IJARSCT-10986

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.301
Volume 3, Issue 16, May 2023

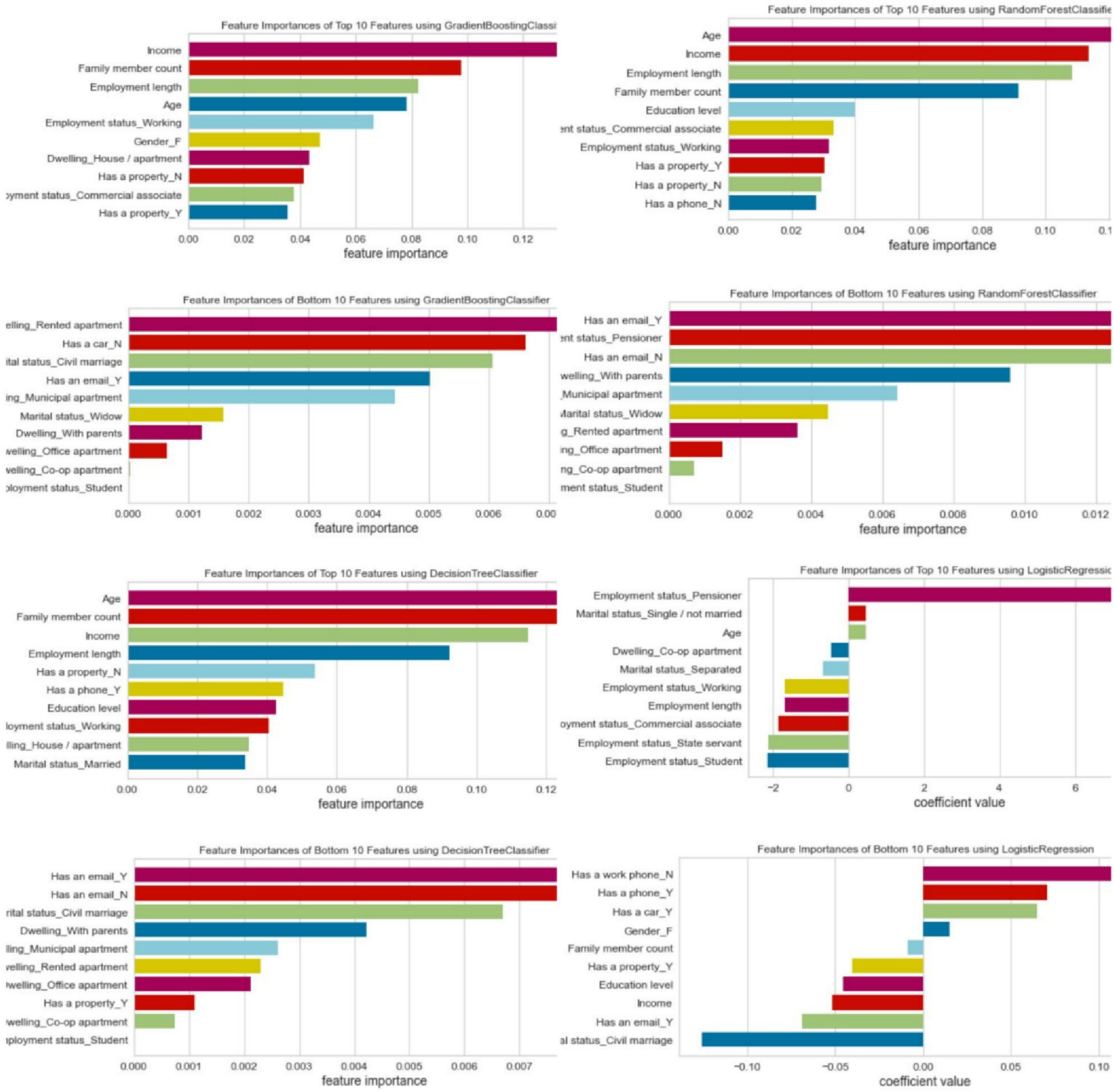

Fig 5. Feature Extraction

[A] Gradient Boosting Algorithm

Gradient Boosting is a machine learning algorithm that works by combining multiple weak learners to create a strong learner. The algorithm iteratively trains decision trees on the residuals (errors) of the previous tree, gradually improving the model's accuracy.

In Gradient Boosting, each decision tree focuses on the residuals of the previous tree, rather than the raw data. This allows the model to learn from its mistakes and adjust its predictions accordingly. The algorithm also uses a technique called boosting, which assigns more weight to misclassified data points in order to correct for errors.

The main advantage of Gradient Boosting is its ability to handle complex, non-linear relationships between variables. It is particularly effective for problems where the input data is high-dimensional and there are many interacting variables.

IJARSCT

ISSN (Online) 2581-9429

International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)

International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal

Impact Factor: 7.301

Volume 3, Issue 16, May 2023

Gradient Boosting is a powerful machine learning algorithm that can be used to improve the accuracy of credit card approval models. It works by iteratively training decision trees on the residuals of the previous tree, gradually improving the model's accuracy.

In the context of credit card approval, Gradient Boosting can be used to predict the probability of default for each applicant. This is done by training the model on historical data that includes information about each applicant's credit history, income, employment status, and other relevant factors. Once the model has been trained, it can be used to make predictions about the creditworthiness of new applicants. The output of the model is a probability score, indicating the likelihood that an applicant will default on their credit card payments. Based on this score, the lender can make an informed decision about whether to approve or deny the applicant's request for credit.

By using Gradient Boosting to improve the accuracy of credit card approval models, lenders can make more informed decisions about who to approve for credit. This can help to reduce the risk of default and improve the overall profitability of the lending business.

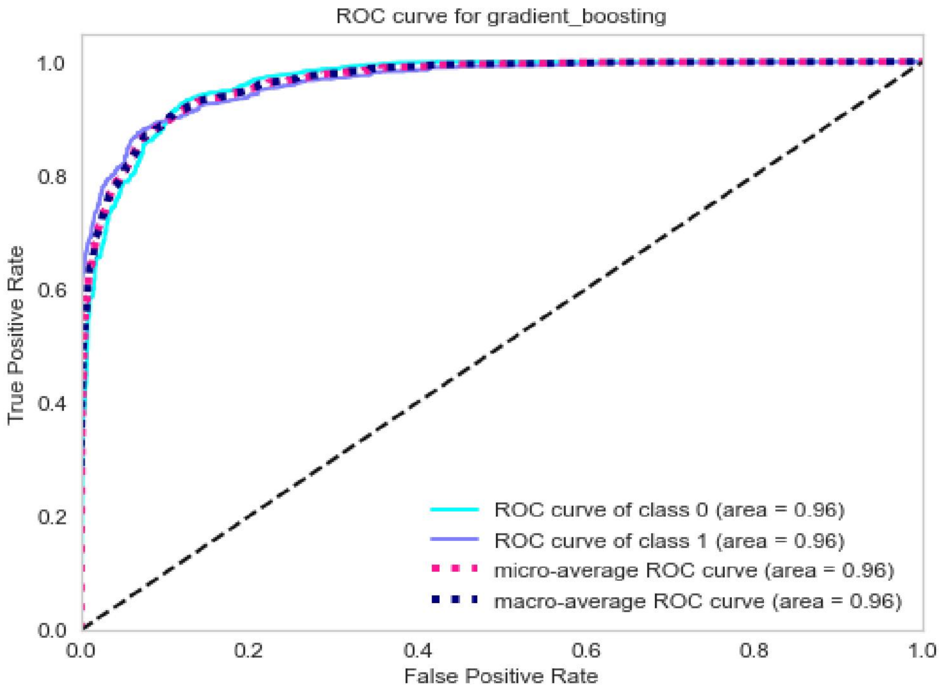

Fig. 2 ROC Curve for Gradient Boosting

[B] Online User Interface

The online UI for credit card approval prediction is designed to provide a user-friendly interface for users to input their personal and financial information in order to predict whether their credit card application will be approved or not. The UI is designed with simplicity and ease of use in mind, with clear labels and instructions to guide users through the process.

The UI begins with a welcome screen, inviting the user to enter their personal information, including their name, age, and contact details. The user is then prompted to provide details about their occupation, education, income, family size, and marital status. The UI also includes fields for employment status and work experience, which are important factors in determining creditworthiness.

The next section of the UI focuses on the user's financial information, including their monthly income, monthly expenses, and any outstanding debt or loans. The UI also includes fields for the user's credit score and credit history, which are key indicators of creditworthiness and play a major role in determining whether a credit card application will be approved.

The final section of the UI asks the user to provide their phone number and email address, which will be used to contact them with the results of their credit card application. The UI also includes a disclaimer and privacy policy to

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.301
Volume 3, Issue 16, May 2023

ensure that users are aware of how their personal information will be used. Once the user has provided all of the necessary information, the UI uses a machine learning algorithm to predict the likelihood that their credit card application will be approved. The algorithm takes into account a variety of factors, including the user's personal and financial information, credit score, and credit history, to generate an accurate prediction.

The results of the credit card approval prediction are displayed on the UI, along with an explanation of how the prediction was generated. If the prediction is positive, the user is given information on how to apply for a credit card and what steps they can take to improve their credit score. If the prediction is negative, the user is given advice on how to improve their creditworthiness and increase their chances of getting approved in the future.

Overall, the online UI for credit card approval prediction is designed to provide a simple and intuitive way for users to determine their creditworthiness and increase their chances of getting approved for a credit card. By taking into account a wide range of personal and financial factors, the UI provides a comprehensive and accurate prediction that can help users make informed decisions about their credit card applications.

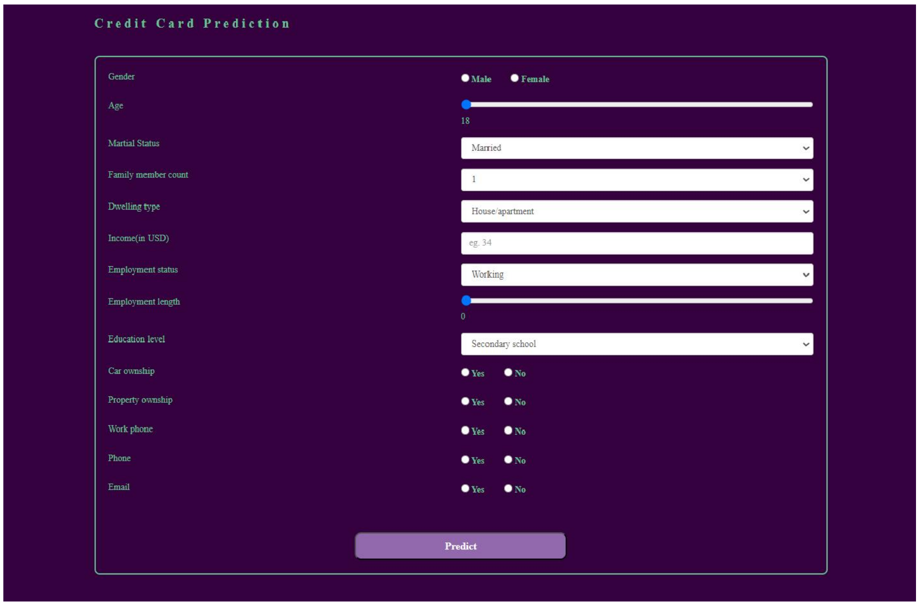

Fig 7. User Interface Website

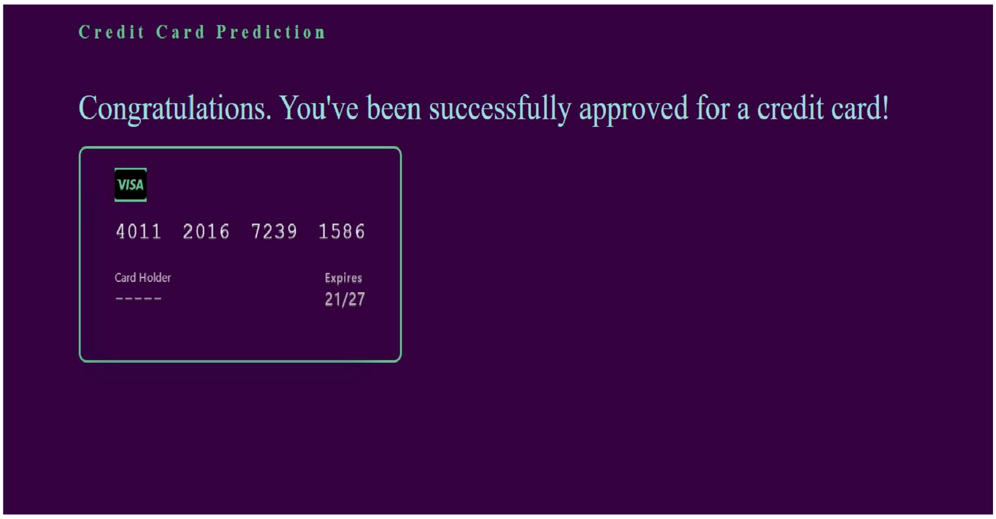

Fig 8. Credit Card Approval Message

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.301
Volume 3, Issue 16, May 2023

IV. CONCLUSION

We obtained a publicly available dataset and performed exploratory data analysis to gain insights into the data. We carried out essential data preparation steps such as data preprocessing, feature selection, and feature scaling to ensure accurate results. The dataset was divided into training and test sets to validate the model's accuracy. We implemented three predictive models:SGDClassifier, LogisticRegression, SVC, DecisionTreeClassifier, RandomForestClassifier, Gaussian NB, KNeighboursClassifier, GradientBoostingClassifier, LinearDiscriminantAnalysis, BaggingClassifier, MLPClassifier, AdaBoostClassifier, Extra treesClassifierwe used various metrics such as accuracy, precision, recall, and area under the curve (AUC).

In the linear SVM model, we achieved an accuracy of 0.71, precision of 0.83, recall of 0.88, and AUC of 0.89. However, in the nonlinear SVM model, we observed improved performance with an accuracy of 0.88, precision of 0.88, recall of 0.90, and AUC of 0.89. The nonlinear SVM model outperformed both the ANN and linear SVM models in terms of accuracy, precision, and recall. A recall rate of 0.90 indicates that the model correctly predicts positive classes 90% of the time.

In the Gradient Boosting, the accuracy was 0.75, with precision of 0.85 and recall of 0.90 Based on our evaluation, we found that the Gradient Boosting model performed the best among the three classifiers. Its high accuracy level suggests its potential applicability in real-world scenarios. It is important to note that the dataset used contained demographic data specific to the context of a particular region. Therefore, considering the local context can be beneficial when applying this model.

REFERENCES

[1]. Agarwal, A., Rana, A., Gupta, K., Verma, N. (2020). A Comparative Study and enhancement of classification techniques using Principal Component Analysis for credit card dataset. In 2020 International Conference on Intelligent Engineering and Management (ICIEM) [Conference paper]. Presented at the 2020 International Conference on Intelligent Engineering and Management (ICIEM), IEEE, London, United Kingdom, pp. 443–448.
[Link](https://doi.org/10.1109/ICIEM48762.2020.9160230)

[2]. Antonakis, A.C., Sfakianakis, M.E. (2009). Assessing naïve Bayes as a method for screening credit applicants. Journal of Applied Statistics, 36, 537–545.
[Link](https://doi.org/10.1080/02664760802554263)

[3]. Banasik, J., Crook, J., Thomas, L. (1999). Not if but when will borrowers default. Journal of the Operational Research Society, 50, 6.

[4]. Bhatore, S., Mohan, L., Reddy, Y.R. (2020). Machine learning techniques for credit risk evaluation: a systematic literature review. Journal of Banking and Financial TechnologyL, 4, 111–138.
[Link](https://doi.org/10.1007/s42786-020-00020-3)

[5]. Blagus, R., Lusa, L. (2013). SMOTE for high-dimensional class-imbalanced data. BMC Bioinformatics, 14, 106.

[6]. [Link](https://doi.org/10.1186/1471-2105-14-106)

[7]. Chornous, G., Nikolskyi, I. (2018). Business-Oriented Feature Selection for Hybrid Classification Model of Credit Scoring. In 2018 IEEE Second International Conference on Data Stream Mining & Processing (DSMP) [Conference paper]. Presented at the 2018 IEEE Second International Conference on Data Stream Mining & Processing (DSMP), IEEE, Lviv, pp. 397–401.
[Link](https://doi.org/10.1109/DSMP.2018.8478534)

[8]. Data Science Process Alliance. (n.d.). What is CRISP DM? [Website]. URL: https://www.datascience-pm.com/crisp-dm-2/ (accessed 11.20.20).

[9]. Elreedy, D., Atiya, A.F. (2019). A Comprehensive Analysis of Synthetic Minority Oversampling Technique (SMOTE) for handling class imbalance. Information Sciences, 505, 32–64.

[10]. Link](https://doi.org/10.1016/j.ins.2019.07.070)

[11]. Fernando, J. (2021). Delinquent Account Credit Card [Website]. Loan Basis. URL: https://www.investopedia.com/terms/d/delinquent-account-credit-card.asp (accessed 4.28.21).

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.301
Volume 3, Issue 16, May 2023

[12]. Galarnyk, M. (2018). Understanding Boxplots [Website]. URL: https://towardsdatascience.com/understanding-boxplots-5e2df7bcbd51 (accessed 1.20.21).

[13]. Hussein, A.S., Li, T., Yohannese, C.W., Bashir, K. (2019). A-SMOTE: A New Preprocessing Approach for Highly Imbalanced Datasets by Improving SMOTE. International Journal of Computational Intelligence Systems, 12(2), 11.

[14]. Jafar Hamid, A., Ahmed, T.M. (2016). Developing Prediction Model of Loan Risk in Banks Using Data Mining. MLAIJ, 3, 1–9.
[Link](https://doi.org/10.5121/mlaij.2016.3101)

[15]. Karthiban, R., Ambika, M., Kannammal, K.E. (2019). A Review on Machine Learning Classification Technique for Bank Loan Approval. In 2019 International Conference on Computer Communication and Informatics (ICCCI) [Conference paper]. Presented at the 2019 International Conference on Computer Communication and Informatics (ICCCI), IEEE, Coimbatore, Tamil Nadu, India, pp. 1–6.
[Link](https://doi.org/10.1109/ICCCI.2019.8822014)

[16]. Kumar, B. (2021). 10 Techniques to deal with Imbalanced Classes in Machine Learning [Website]. URL: https://www.analyticsvidhya.com/blog/2020/07/10-techniques-to-deal-with-class-imbalance-in-machine-learning/

[17]. Kumar Gupta, D., Goyal, S. (2018). Credit Risk Prediction Using Artificial Neural Network Algorithm. IJMECS, 10, 9–16.
[Link](https://doi.org/10.5815/ijmecs.2018.05.02)

[18]. Lee, T., Chen, I. (2005). A two-stage hybrid credit scoring model using artificial neural networks and multivariate adaptive regression splines. Expert Systems with Applications, 28, 743–752.
[Link](https://doi.org/10.1016/j.eswa.2004.12.031)

[19]. Leo, M., Sharma, S., Maddulety, K. (2019). Machine Learning in Banking Risk Management: A Literature Review. Risks, 7, 29.
[Link](https://doi.org/10.3390/risks7010029)

[20]. Madyatmadja, E.D., Aryuni, M. (2005). Comparative Study of Data Mining Model for Credit Card Application Scoring in Bank. Journal of Theoretical and Applied Information Technology, 59, 6.

[21]. Marqués, A.I., García, V., Sánchez, J.S. (2012). Two-level classifier ensembles for credit risk assessment. Expert Systems with Applications, 39, 10916–10922.
[Link](https://doi.org/10.1016/j.eswa.2012.03.033)

[22]. Munkhdalai, L., Munkhdalai, T., Namsrai, O.-E., Lee, J., Ryu, K. (2019). An Empirical Comparison of Machine-Learning Methods on Bank Client Credit Assessments. Sustainability, 11, 699.
[Link](https://doi.org/10.3390/su11030699)

[23]. Nellur, S. (2020). The Neural Network at its Simplest [Website]. URL: https://medium.com/analytics-vidhya/neural-networks-in-nutshell-7d1cc3ae6443 (accessed 2.25.21).

[24]. Oreski, S., Oreski, D., Oreski, G. (2012). Hybrid system with genetic algorithm and artificial neural networks and its application to retail credit risk assessment. Expert Systems with Applications, 39, 12605–12617. [Link](https://doi.org/10.1016/j.eswa.2012.05.023)

[25]. Payments and Settlements Department. (2020). Payments Bulletin - Third Quarter (Quarterly Report) [Report]. Payment Bulletin. Central Bank of Sri Lanka