**A PROJECT REPORT**

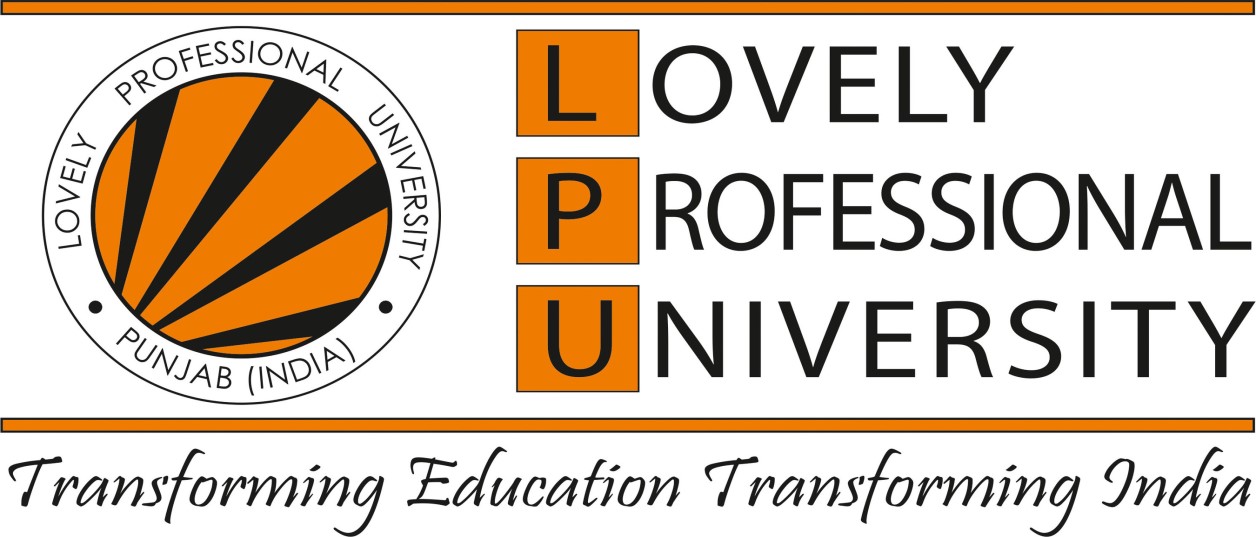
**OF THE COURSE**

**Machine Learning [INT 214]**

**ON THE TOPIC OF**

**CREDIT CARD**

**APPROVAL PREDICTION**



**PROGRAMME NAME:** *Bachelor of Technology (Computer Science and Engineering)*

*Under the guidance of* ***Dr. Dhanpratap Singh***

**SUBMITTED BY**

|  |  |
| --- | --- |
| *HARI PRASAD BODDEPALLI* | *12112271* |
| *AYUSH GUPTA* | *12105138* |
| *ABHINANDHAN* | *12108947* |

# DECLERATION

I hereby declare that the project work entitled “Credit card Approval Prediction” is an authentic record of my own work carried out as requirements of Project for the award of B. Tech degree in Computer Science and Engineering from Lovely Professional University, Phagwara, under the guidance of Dr Dhanpratap Singh, during August to December 2023. All the information furnished in this project report is based on my own intensive work and is genuine

2th November 2023

HARI PRASAD BODDEPALLI 12112271

AYUSH GUPTA 12105138

ABHINANDHAN 12108947

# CERTIFICATE

This is to certify that the declaration statement made by this student is correct to the best of my knowledge and belief. He has completed this Project under my guidance and Supervision. The present work is the result of his original investigation, effort and study. No part of the work has ever been submitted for any other degree at any University. The Project is fit for the submission and partial fulfilment of the conditions for the award of B. Tech degree in Computer Science and Engineering from Lovely Professional University, Phagwara.

Dr Dhanpratap Singh

School of Computer Science and Engineering,

Lovely Professional University,

Phagwara, Punjab

Date: 2th November, 2023

# TABLE OF CONTENTS

1. Introduction
2. What is Credit Card?
3. Credit Card Scoring and Decisioning Models
4. Effects of ML credit card customer experience
5. Comparison of different supervised machine learning classifiers to predict credit card approvals
6. Predicting credit card approval of customers through customer profiling using machine learning
7. Advantages of a credit card
8. Disadvantages of a credit card
9. Challenges of credit card approval using machine learning
10. Model generation
11. Conclusion and future scope
12. References

# INTRODUCTION

Credit approval, such as for credit cards, is crucial to the modern economy. In today's interconnected global economy, credit approval, such as for credit cards, plays a pivotal role. Even in developing nations like India, the accessibility of credit cards has become a reality. However, for financial institutions, the challenge lies in accurately assessing whether a consumer represents an acceptable credit risk for granting credit. This challenge is particularly pronounced in emerging economies, where conventional rules and models from more developed nations may not be directly applicable.

In this report, we present our investigation into the prediction of credit card approval for applicants using various machine learning algorithms. Our research journey begins with data preprocessing and comprehensive exploratory data analysis(EDA) to gain deeper insights into the critical factors influencing the model's training. Subsequently, we apply ten distinct machine learning algorithms to the preprocessed data, aiming to identify the model that strikes the optimal balance between accuracy and precision-recall trade-off.

The report is structured as follows: In Section II, we share the insights derived from our literature review, shedding light on the existing knowledge in this domain. In Section III, we provide a detailed exposition of our entire system, covering data preprocessing, model selection, and evaluation. We also present and analyze the results obtained, comparing them from different perspectives. Finally, in the concluding section, we summarize our findings and observations.

# ****WHAT IS A CREDIT CARD?****



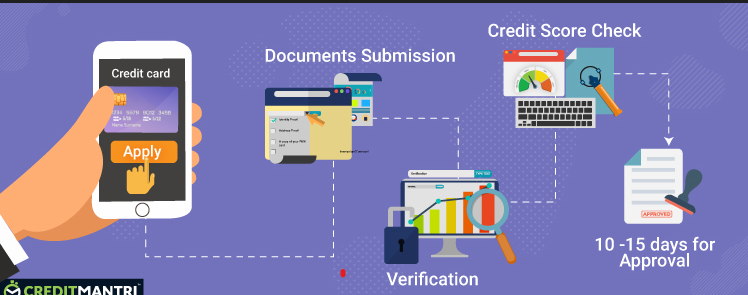
Credit cards are offered by banks and NBFCs (Non-Banking Financial Company) to carry out seamless, cashless transactions online and offline. The credit card lender sets a pre-decided limit on your credit card based on your income, credit score and other factors. You can use the [credit card](https://navi.com/blog/what-is-credit-card/) for various transactions up to the maximum credit limit. The credit card bill is regenerated at a regular interval. If you pay the bill within the due date, interest will not be imposed. However, after this period, the bank will charge a certain interest rate.

# CREDIT SCORING AND DECISIONING MODELS

### **Credit Scoring**

Credit scoring is a statistical analysis performed by financial lenders to determine the creditworthiness of an individual or a business. This creditworthiness is influenced by several key factors:

1. **Credit History:** The primary source of information is usually the applicant’s credit history, which is a detailed record of all past borrowing and repayment, including late payments and defaults. Credit bureaus compile this data and generate credit reports.
2. **Personal Information:** Personal data, such as income level, employment status, and length of credit history, are vital pieces of information. This information is often self-reported by the applicants on their credit application.
3. **Other Data Sources:** In recent years, alternative data has become increasingly important. This could include utility bills, rent payments, bank account information, and even social media activity. All these sources provide a more holistic view of the applicant’s financial behavior.



Based on these data points, models generate a credit score. This credit score is then used to decide whether to extend credit to a borrower and at what interest rate. Credit scores typically range from 300 to 850, with higher scores indicating less risk to the lender.

### **Credit Decisioning**

Credit decisioning, on the other hand, is the process by which financial institutions make decisions on credit applications. It involves evaluating the borrower’s creditworthiness and the likely profitability of the loan to the lender. This margin may differ based on the firm’s risk appetite and lending strategy.

For example, a lender specializing in subprime loans might accept higher default probabilities than a traditional bank. Additionally, firms must comply with federal and state financial regulations.

In another example, lenders cannot discriminate based on age, race, religion, nationality, or marital status, and such factors must be handled appropriately by the model. Ultimately, the decisioning process may result in acceptance, rejection, or a request for additional information.

# EFFECTS OF ML CREDIT CARD CUSTOMER EXPERIENCE

Machine learning and artificial intelligence are leading a new digitisation renaissance across multiple global industries. In the financial landscape, machine learning and AI are not new technologies. Banks, financial institutions, and more have relied on machine learning algorithms for critical processes. But now more and more banks are moving towards utilising the emerging prowess of newer machine learning models towards the credit card industry.

ML is perfect for the credit card industry as the technology can help companies make sense of humongous sets of data and give insights about every single customer within the ecosystem. From bolstering fraud detection to fine-tuning personalized financial trajectories, this cutting-edge technology is orchestrating a paradigm shift in financial services.



Machine learning operates as a type of artificial intelligence, enabling computers to learn like humans do, by building on past experiences. It's like teaching a computer to recognize patterns in information without much human help. This technology delves into data, finding patterns on its own, and doesn't need constant human guidance.

Any task that follows certain patterns or rules from data can be automated using machine learning. This means things like handling customer calls, managing accounts, or even going through resumes can be done by machines. Machine learning systems tackle large amounts of data, finding important patterns within it.

# COMPARISON OF DIFFERENT SUPERVISED MACHINE LEARNING CLASSIFIERS TO PREDICT CREDIT CARD APPROVALS :-

This study contrasts various supervised machine learning models to forecast the likelihood that a credit card request will be accepted based on various criteria like Precision, Recall, Time, Accuracy, and F1 Score. The aim here was to identify the best classifier for automatically predicting credit card approval based on the characteristics of credit card applications. The analysis also demonstrates that every classifier performs better in one or more metrics. To improve the performance of each model, the method used hyperparameter optimization based on GridSearchCV to optimise certain parameters. The UCI Machine Learning Repository dataset used in this work was unbalanced and hence F1 score was relied on up to test the models. The classifiers used in this study are Logistic Regression, Random Forest, Decision Tree, XGBoost, Gradient Boost, Support Vector Machine (SVM), and Sequential Neural Network. Finally, based on F1 Score and AUC value, the research finds that Random Forest classifier is the best model for predicting Credit Card approvals with a F1 score of 86%. Although the research tries multiple machine learning models to test the dataset, there was no attempt made to balance out the data for better results.

# PREDICTING CREDIT CARD APPROVAL OF CUSTOMERS THROUGH CUSTOMER PROFILING USING MACHINE LEARNING

This study focuses on forecasting credit card approval for users using a limited number of algorithms. The data was taken from bank customers in 2 ways, primary data and secondary data and then combined into one. These customer datasets were fully gathered, evaluated, and trained. These trained datasets helped in predicting whether credit card applications from customers will be approved. Since only a small number of variables were employed to determine the final decision, the training and testing accuracy of both decision tree and k nearest neighbour algorithms were roughly 99.7% and 99.6%, respectively. The training and testing accuracies of the decision and knn algorithms would alter in real time as more datasets are trained and tested and as the variables for the final choice are raised. This study however falls short in testing various other classification algorithms that could show better results in the future when more variables are taken in consideration.

# **ADVANTAGES OF A CREDIT CARD**

A credit card has several benefits with thw help of ml as well . They are as follows:

#### **1. EMI Option :**

Credit cards are excellent for purchasing goods and services with a low monthly EMI. This alleviates the burden of having to pay the money in one lump sum. Furthermore, EMI payments via credit cards may be more convenient than obtaining a personal loan.

#### **2. Exciting Offers and Cashbacks :**

Most banks offer credit cards with a variety of offers, cashbacks, and rewards. These offers and rewards are available whenever you make an online or offline purchase. You can also get cheaper air tickets, train tickets, hotel reservations, grocery shopping, and so on.

#### **3. Protection of Purchase :**

Credit cards provide extra security in the form of insurance for card purchases that are lost, damaged, or stolen. If you want to file a claim, you can use the credit card statement to back it up.

#### **4. Improving Credit Score :**

Credit cards allow you to build up a credit line. This is critical because it allows banks to view an active credit history based on your card repayments and card usage. Banks and financial institutions frequently use credit card usage to assess a potential loan applicant’s creditworthiness, making your credit card important for future loan or rental applications.

**5. Enhanced Accuracy:**

Machine learning algorithms have the ability to analyze vast amounts of data and identify patterns that may not be apparent to human analysts. By incorporating various data points, including credit history, income, employment, and spending patterns, machine learning models can make more accurate predictions regarding an individual's creditworthiness. This leads to better-informed credit card approval decisions, reducing the risk of defaults and improving overall portfolio performance.

**6. Faster Processing:**

Traditional credit card approval processes can be time-consuming, involving manual reviews, paperwork, and extensive documentation. Machine learning streamlines this process by automating many of the tasks. By leveraging algorithms and predictive models, financial institutions can expedite credit card approvals, providing customers with faster access to credit facilities.

**7.Personalized Offerings:**

 Machine learning enables lenders to personalize credit card offerings based on individual profiles and preferences. By analyzing customer data and behavior, machine learning algorithms can identify specific needs, spending patterns, and risk profiles. This allows lenders to tailor credit card features, such as interest rates, credit limits, rewards programs, and promotional offers, to match the unique requirements of each customer.

**8.Risk Mitigation:**

The use of machine learning algorithms in credit card approval helps mitigate risks associated with lending. By accurately assessing creditworthiness and identifying high-risk applicants, financial institutions can make informed decisions on interest rates, credit limits, and terms of repayment. This not only protects lenders from potential losses but also ensures responsible lending practices and safeguards the financial well-being of customers.

# ****DISADVANTAGES OF A CREDIT CARD :-****

Here are a few disadvantages of using a Credit Card:

**1. Hidden Costs :-**

Credit cards may appear easy and straightforward initially, but they comprise numerous hidden costs that can increase the expense amount by a high margin. These extra charges can come in the form of late payment costs, renewal fees, processing fees etc. Nevertheless, if you miss any payment, it can cause a penalty and diminish your credit history.

**2. Restricted Drawings :-**

Credit cards, unlike debit cards, do not offer as many benefits when it comes to cash withdrawals. This is due to the fact that some credit cards charge an additional fee in addition to an annual interest rate of approximately 40%.

# CHALLENGES OF CREDIT CARD APPROVAL USING MACHINE LEARNING

**1. Data Privacy and Security:** The use of machine learning in credit card approval requires access to vast amounts of sensitive customer data. It is crucial for financial institutions to implement robust data privacy and security measures to protect this information from unauthorized access or misuse. Strict compliance with data protection regulations and encryption techniques is essential to ensure the confidentiality and integrity of customer data.

**2. Model Interpretability and Transparency:** Machine learning algorithms can be complex, making it challenging to interpret and explain the decisions they make. This lack of interpretability can pose challenges in terms of regulatory compliance and consumer trust. Efforts must be made to develop transparent models that provide clear explanations for credit card approval decisions, ensuring fairness and accountability.

**3. Bias and Fairness:** Machine learning algorithms are susceptible to bias as they learn from historical data that may contain inherent biases. This can lead to discriminatory practices in credit card approval, impacting certain demographic groups unfairly. It is important to continuously monitor and evaluate machine learning models to ensure fairness and mitigate any bias that may arise.

# MODEL GENERATION

## Step 1: Loading datasets

The code begins by importing essential libraries for data manipulation, visualization, and machine learning. These libraries include Pandas (pd), NumPy (np), Seaborn (sns), Matplotlib (plt), and specific modules from scikit-learn. The primary goal is to load and process two datasets: "application\_record.csv" and "credit\_record.csv."

## Step 2: Merging datasets

In this step, the two datasets are combined into a single DataFrame (data) using the Pandas merge function. The merging is based on the 'ID' column, which is assumed to be a common identifier in both datasets. This merging allows us to have a comprehensive dataset containing both application details and credit records.

## Step 3: Exploring the data

1. \*Basic Statistics\*: The code uses the data.describe() function to compute and display basic statistics for the dataset. This includes statistics such as mean, standard deviation, minimum, maximum, and quartiles for numeric columns. It gives an overview of the data's central tendencies.
2. \*First Few Rows\*: The data.head() function is used to display the first few rows of the merged dataset. This is useful for understanding the structure and contents of the data.

## Step 4: Data Cleaning

Data cleaning is a crucial step to ensure the quality and integrity of the dataset:

* \*Missing Values Check\*: The code uses data.isnull().sum() to count and display missing values in each column. This is important for identifying columns with missing data.
* \*Removing Rows with Missing Target Variable\*: Rows with missing values in the 'STATUS' column are removed using data.dropna(subset=['STATUS']). This is done to ensure that only records with credit status information are retained for the modeling process.

Step 5: Defining features

A set of features that will be used as input for the machine learning model is defined. These features represent various attributes and characteristics related to the applicants, such as gender, income, education, and more. These features are stored in the 'features' list.

## Step 6: Selecting features and target variable

In this step, the code selects the independent variables (features) and the dependent variable (target) to prepare the data for modeling:

- X contains the selected features from the dataset.

- y contains the target variable, which is 'STATUS,' representing the credit status.

Step 7: Converting categorical variables to numerical using one-hot encoding

Many machine learning algorithms require numerical input, so categorical variables need to be converted into a numerical format. One-hot encoding is used for this purpose. It creates binary columns for each category within the categorical variables, indicating the presence or absence of each category.

Step 8: Data Visualization

Data visualization is an essential part of data exploration. The code plots a histogram of income distribution using Seaborn and Matplotlib. This visualization provides insights into the distribution of income levels in the dataset.

## Step 9: Splitting the data into training and testing sets

To assess the model's performance, the dataset is divided into training and testing subsets. The train\_test\_split function from scikit-learn is used for this purpose. It allocates 80% of the data for training (X\_train and y\_train) and 20% for testing (X\_test and y\_test).

Step 10: Initializing and training the Random Forest Classifier

A Random Forest Classifier is initialized with 100 decision trees. It is a machine learning model used for classification tasks. The model is then trained on the training data (X\_train and y\_train) to learn patterns and relationships within the data.

## Step 11: Making predictions on the test set

The trained Random Forest Classifier is used to make predictions on the test set (X\_test). The predicted values are stored in the 'y\_pred' variable.

Step 12: Evaluating the model using classification report

To assess the model's performance, a classification report is generated using the classification\_report function from scikit-learn. This report includes various metrics such as precision, recall, F1-score, and support for each class, allowing a detailed assessment of the model's ability to classify different credit statuses.

Step 13: Displaying the classification report

Finally, the classification report is displayed, providing a comprehensive summary of the model's performance. It includes accuracy and other classification metrics for different credit statuses, helping to evaluate the model's effectiveness in predicting credit outcomes.

This code represents a complete machine learning workflow, from data preprocessing to model evaluation, in the context of predicting credit statuses based on applicant and credit record data.

# FUTURE SCOPE

To further improve our system, we can use deep learning models as it can increase our accuracy. Neural networks can be used as it can discover hidden patterns and correlations in raw data, cluster and classify it, and continuously learn and improve over time. In the future, this credit card approval system will be able to be optimized and implemented in an artificial intelligence environment. By displaying the prediction result on a web or desktop application, the system can also be automated. Thus, this work has a good future scope and can be enhanced by adding other various feature for better predictions.

# CONCLUSION

In this paper, we have mentioned various machine learning methods to predict whether a credit card will be approved for an individual or not. Several parameters were taken into consideration as these parameters make the model more effective and help institutions make better decisions to avoid fraud and losses. We applied a lot of data pre-processing techniques as good amount of data pre-processing contributes effectively to developing better performance of traditional machine learning models. During Exploratory Data Analysis, we plotted a lot of graphs and charts to study the dataset deeply so that we can get a better understanding of the dataset. This was done so that we can decide which models to apply which can perform well on this dataset and can correctly predict whether to approve a credit card or not. This prediction system can be helpful to various banks as it makes their task easier and increases efficiency as compared to the manual system which is currently used by many banks and this system is cost effective.

# REFERENCES

* [***www.google.com***](http://www.google.com)
* [***www.wikipedia.com***](http://www.wikipedia.com)
* [***www.ijraset.com***](http://www.ijraset.com)
* [***https://github.com/AyushGupta16/Credit-Card-Approval-Prediction.git***](https://github.com/AyushGupta16/Credit-Card-Approval-Prediction.git)