Loan Prediction Analysis using Machine Learning Techniques

DAFFODIL INTERNATIONAL UNIVERSITY

DHAKA, BANGLADESH

DECEMBER 2023

LOAN PREDICTION ANALYSIS USING MACHINE LEARNING TECHNIQUES

BY

Abul Hasan Shawn
ID: 213-25-053

This Report Presented in Partial Fulfillment of the Requirements for the Degree of Master of Computer science and Engineering

Supervised By

Ms. Naznin Sultana
Assistant Professor
Department of CSE
Daffodil International University

DAFFODIL INTERNATIONAL UNIVERSITY

DHAKA, BANGLADESH

JANUARY 2023

APPROVAL

This Thesis titled **"LOAN PREDICTION ANALYSIS USING MACHINE LEARNING TECHNIQUES"**, submitted by **Abul Hasan Shawn**, ID No: **213-25-053** to the Department of Computer Science and Engineering, Daffodil International University has been accepted as satisfactory for the partial fulfillment of the requirements for the degree of M.Sc. in Computer Science and Engineering and approved as to its style and contents. The presentation has been held on 17-01-2023.

BOARD OF EXAMINERS

17/01/23

Chairman

Dr. Touhid Bhuiyan, PhD
Professor and Head
Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University

Internal Examiner

Ms. Nazmun Nessa Moon
Associate Professor
Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University

Internal Examiner

Dr. Fizar Ahmed
Associate Professor
Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University

External Examiner

Md. Safaet Hossain
Associate Professor & Head
Department of Computer Science and Engineering
City University

DECLARATION

We hereby declare that, this project has been done by us under the supervision of **Ms. Naznin Sultana**, Assistant Professor, Dept. of CSE, Daffodil International University. We also declare that neither this project nor any part of this project has been submitted elsewhere for the award of any degree or diploma.

Supervised by:

Ms. Naznin Sultana
Assistant Professor
Department of CSE
Daffodil International University

Submitted by:

Shawn

Abul Hasan Shawn
ID: 213-25-053
Department of CSE
Daffodil International University

ACKNOWLEDGEMENT

And first foremost, we offer our heartfelt appreciation and gratitude to Almighty God for His divine gift, which has enabled us to successfully finish the final year proposal.

We are really grateful and wish our profound indebtedness to **Ms. Naznin Sultana, Assistant Professor,** Department of CSE Daffodil International University, Dhaka. Our supervisor has extensive knowledge and a great interest in the subject of Deep Knowledge & keen interest of my supervisor in the field of "Deep Learning, Machine Learning" to carry out this paper. Her unending patience, scholarly guidance, constant encouragement, constant and energetic supervision, constructive criticism, valuable advice, reading many inferior drafts and correcting them at all stages, and reading many inferior drafts and correcting them at all stages enabled us to complete this project.

We would like to express our heartiest gratitude to Mr. Dr. Touhid Bhuiyan, Professor and Head, Department of CSE, for his kind help to finish our project and also to other faculty members and the staff of the CSE department of Daffodil International University.

We would like to thank everyone of our Daffodil International University classmates who participated in this discussion while completing their course work.

ABSTRACT

Loans account for a large portion of bank profits. Despite the fact that many people are looking for loans. Finding a legitimate applicant who will return the loan is difficult. Choosing a real applicant may be difficult if the process is done manually. As a result, we are creating a machine learning-based loan prediction system that will choose the qualified applicants on its own. Both the applicant and the bank staff will benefit from this. There will be a significant reduction in the loan sanctioning period of time. In this research. The majority of the bank's revenue is generated directly from the interest income on loans. Even when the bank authorizes the loan following a lengthy verification and testimony process, there is no guarantee that the chosen hopeful is the best hopeful. When performed manually, this operation requires additional time. We have the ability to foretell whether a specific hopeful is secure or not, and machine literacy has mechanized the entire testifying procedure. Loan Prognostic is extremely beneficial for both banks' clients and potential borrowers. I use some machine learning algorithms techniques to predict the loan data. Those are the Decision Tree, K Nearest Neighbour, SVM, Naive Bayes, and Random Forest Classifier. Nevertheless, I did uncover promising setups for both purposes. I got the best accuracy from Naïve Bayes which was 82.16%.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1

INTRODUCTION

1.1 Introduction

Every retail bank encounters this problem at least once in their lifetime because customer loan projection is typically a lifelong issue. If done correctly, it can save a lot of man-hours when a retail bank closes. If the business decides to partially automate the loan acceptance process, it will use the information customers submit while filling out an online application. Gender, Marital Status, Education, Dependents, Income, Loan Amount, Credit History, and other nuances are among these. They have made it difficult to identify the consumer categories in order to automate this process; these segments are permitted for the total loan amount, so they can plainly target these customers. We need to forecast if a loan will be authorized or not.

The distribution of loans is the primary activity of each and every bank. The primary portion of the bank's resources came directly from the income earned from the advances that the banks gave out. The basic objective of the banking system is to place resources in safe hands wherever they may be. Today, a number of banks and financial institutions grant loans following the second round of verification and validation, but it is still uncertain whether the chosen candidate is the most deserving of the candidates. With the help of this method, we can determine whether a certain candidate is secure or not, and the entire process of attribute validation is automated using machine learning [1][2]. The drawback of this model is that it gives completely different weights to each concern, but in practice, a loan may occasionally be accepted solely on the basis of a single powerful component, which isn't possible using this strategy. Both candidates and bank staff members can benefit from loan prediction.

The main source of revenue for banks is a loan. The majority of the bank's profits are derived directly from the money made from the loans. Even when the bank authorizes the loan following a lengthy verification and testimony process, there is no guarantee that the chosen hopeful is the best hopeful. When performed manually, this operation requires additional time. We have the ability to foretell whether a specific hopeful is secure or not, and machine literacy has mechanized the entire testifying procedure. Loan The use of forecasts by banks and hopefuls alike is really beneficial [3]. Employees of the bank personally verify the applicant's information before awarding loans to those who qualify. It takes a long time to review every applicant's information. The artificial neural network model to forecast a bank's credit risk. To predict credit default, a feed-forward back propagation neural network is used. a technique that produces an ensemble model by combining two or more classifiers for improved prediction [3]. They employed a random forest technique after bagging and boosting procedures. Classifiers work to enhance the performance of the data and increase efficiency. The authors of this paper discuss a variety of ensemble strategies for multiclass and binary classification. COB is the novel method for the ensemble that the authors describe [3]. The novel technique for the ensemble that the authors describe is COB, which provides efficient classification performance but also compromises noise and outlier data. Finally, they came to the conclusion that the ensemble-based algorithm enhances the training data set results.

1.2 Motivation

For financial companies, the loan approval process is crucial. The loan applications were approved or rejected by the system. A key determining factor in a bank's financial results is loan recovery. Predicting whether or not the customer will pay back the loan is exceedingly difficult. The loan approval is predicted using machine learning.

- Loan Applicants will be approved or rejected by the system.
- It'll recover the bank's financial results.
- It'll predict whether the customer will pay back the loan or not.

1.3 Research Question

- Will it recover the financial situation of a bank?
- Is it possible to detect the person whether he will?
- What'll be the Effectiveness of The Loan prediction system?
- Which model is the best?
- Will we be able to find out which person will get the approval for loan?

1.4 Expected Outcome

- Good knowledge of loan prediction analysis.
- Know about machine learning algorithms.
- The banker can predict who can get a loan.

1.5 Report Layout

This report varied in a total of six different chapters. Which are capable of extending the understanding of “Loan prediction” more briefly.

In the first chapter, we’ll mention the introduction, motivation, and research questions and the last one is the expected outcome.

In the second chapter, I’ll briefly talk about some related works, which types of challenges I faced, and the research summary.

In the third chapter, we’ll talk about our research subject and instrumentation, the workflow of the model, and how I got the accuracy of the given models.

In the fourth chapter, I’ll talk about the result that I got, and discussion of the results.

In the fifth chapter, I’ll describe its impact on our society, impact on our environment, and sustainability.

In the sixth chapter, which is our last chapter, I’ll mention the conclusion and our future works.

CHAPTER TWO

BACKGROUND STUDY

2.1 Introduction

Although the banking system offers a wide range of profit-generating products, its credit system is its main source of funding. as the Credit System is capable of interest rates for the loans it credit. For a wide range of challenges, the banking system constantly needs an accurate modeling system. For every bank, one of the most challenging tasks is to predict loan defaulters. However, by predicting the loan defaulters, the banks will undoubtedly be able to cut their losses by decreasing their non-profit assets, allowing the recovery of sanctioned loans to proceed without incurring any losses and acting as a contributing factor to the bank statement. It is highlighting the significance of studying this loan approval forecast.

2.1.1 Machine Learning Approach

Around the game of checkers, one of its own is credited with coining the term "machine learning." The majority of the time, machine learning algorithms are developed utilizing accelerated solution development frameworks like TensorFlow and PyTorch. Node layers, also known as neural networks or artificial neural networks, are made up of an input layer, one or more hidden layers, and an output layer. Deep learning algorithms or deep neural networks are neural networks that include more than three layers. The use of labeled datasets to train algorithms to categorize data is what distinguishes supervised machine learning [4]. The model modifies its weights as input data is fed into it until it is well-fitted. Various scale-up real-world challenges can be solved by companies with the use of supervised learning. An algorithm that is taught without requiring sample data is known as unsupervised machine learning. The lack of sufficient labeled data for a supervised learning system can be resolved via semi-supervised learning [4].

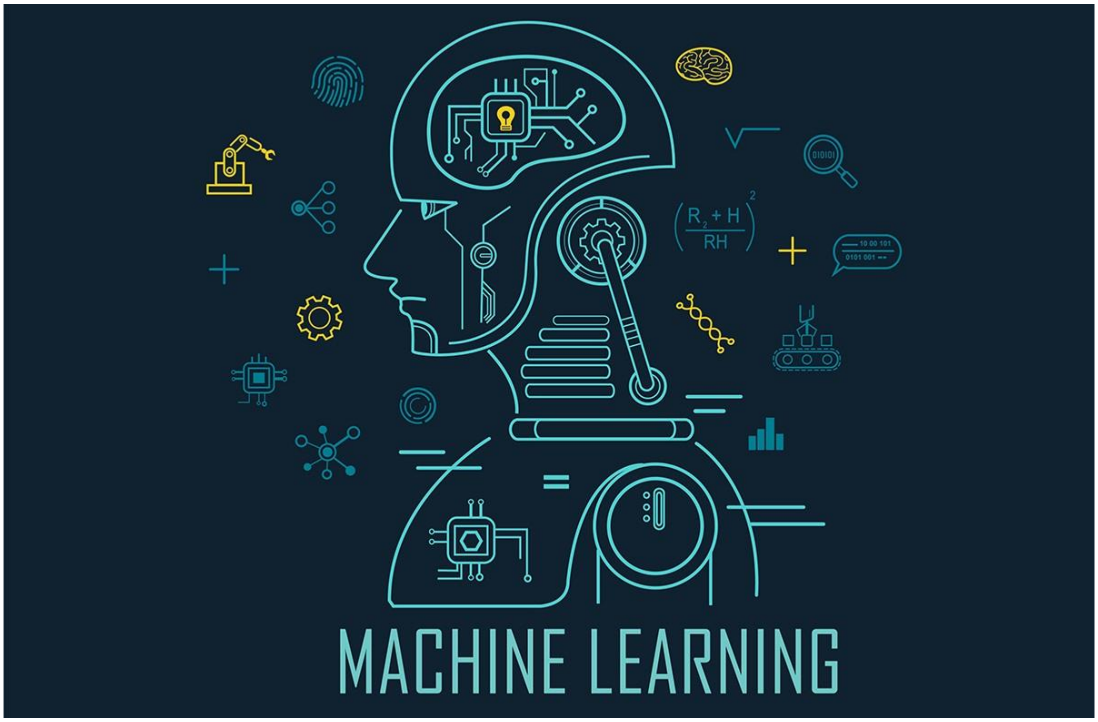

Figure 2.1: Machine Learning [5]

2.1.2 Types of machine learning:

In the field of machine learning (ML), we create algorithms to teach a machine to perform a task without actually performing any calculations on it [5]. Building algorithms that can take input data and apply statistical analysis to predict an output while updating outputs as new data becomes available is the fundamental idea behind machine learning [5]. There are three types of machine learning. Those are given below:

- **Supervised Machine Learning [5]:** A finite collection of data containing the correct responses for each of the input values is given to the algorithm in supervised learning. The machine's job is to accurately analyze the dataset and forecast the correct responses [5]. An illustration of supervised learning As evidenced in the sample above, we first took some data and labeled them as either "Tom" or "Jerry." The training supervised model uses this labeled data; the model is trained using this data. Once it has been trained, we can test our model by using a few test emails to see if it can accurately predict the desired result [5].
- **Unsupervised Machine Learning [6]:** Unsupervised learning is a type of machine learning where the users don't have to watch over the model. Instead, it

enables the model to reason independently and uncover previously hidden patterns and information. It mostly addresses unlabeled data. Unsupervised learning uses data without labels. The program recognizes and learns the patterns present in the dataset [6]. Depending on their density, the algorithm divides the data into different clusters. It allows for the viewing of high-dimensional data. The Principal Component Analysis is an illustration of this type of machine learning technique. K-Means Clustering is an additional unsupervised learning method that separates the data into groups based on the similarity of order. The way that unsupervised learning involves learning.

- **Reinforcement Learning [6]:** Along with supervised learning and unsupervised learning, reinforcement learning is one of the three fundamental machine learning paradigms. One of the most common and rising categories of machine learning algorithms is reinforcement learning. It is utilized in many autonomous systems, including automobiles and commercial robotics [6]. This algorithm's goal is to accomplish a task in a changing environment. Based on the number of prizes that the system offers, it can achieve this aim. The programming of robots to carry out independent tasks makes heavy use of it. Making clever self-driving cars also makes use of it [6].

2.1.3 How Machine Learning works:

Machine learning's overarching objective is to create models that replicate and generalize data. For these models to produce the correct results, they must learn how to discriminate between different items. Simply put, machine learning employs a range of methods, to accomplish a particular objective, these techniques use algorithms [7]. Machine learning's task is to determine that the object being delivered to it is fruit. The clearest explanation of how machine learning functions comes from Interactions' Senior Vice President of Natural Language Research, Jay Wilpon, who uses the example of fruits [7].

2.2 Related work:

Nitesh Pandey, Ramanand Gupta, Sagar Uniyal, and Vishal Kumar (2021) explain for every bank, one of the most challenging tasks is to predict loan defaulters. However, by predicting the loan defaulters, the banks will undoubtedly be able to cut their losses by decreasing their non-profit assets, allowing the recovery of sanctioned loans to proceed without incurring any losses and acting as a contributing factor to the bank statement. This highlights the significance of studying this loan approval forecast. In order to predict this kind of data, machine learning techniques are extremely important and helpful. In this study, four classification-based machine learning algorithms—Logistic Regression, Decision Tree, Support Vector Machine, and Random Forest—are used. The Support Vector Machine approach is the most accurate of these algorithms at predicting loan acceptance with a high degree of accuracy.

Ashwini S. Kadam, et al.'s, (2021) describe with certainty, it can be said that the Naive Bayes model is quite effective and produces better results than other models. It operates properly and satisfies all lenders' requirements. This technology calculates the outcome correctly and precisely. It accurately forecasts whether a loan application or customer will be approved or rejected.

Dr.C K Gomathy et al. (2021) also said about loan prediction. It is safe to say that the product is a very productive member based on a proper study of the member's advantages and limits. This use is properly operating and adhering to all banker requirements. This component is easily pluggable into many different systems. Mathematical instances of software bugs, content violations, and most importantly, the weight of features are resolved in an automated prophecy system, so in the near future, the so-called software might be made more secure, reliable, and dynamic weight conformation. This prophecy module may soon be integrated with the automated processing system module. The system is trained on an older training dataset, but once some time has passed, new testing data should also be included in the training dataset.

The loan prediction analysis is introduced by L. Udaya Bhanu. S. Narayana (2021) utilized the in this study, several algorithms were put into practice to forecast consumer loans. Logistic Regression, Random Forest, KNN, and SVM, as well as decision Tree Classifier, produced the best results. This prediction module may soon be coupled with the machine-driven processing system module.

Sandip Pandit, Pooja Thitme, Mrunal Surve, Priya Shinde, and Swati Sonawane in this study, the main emphasis is on identifying and analyzing the risk associated with a commercial bank loan. They have employed data mining tools to analyze the risk of loaning money. It involves gathering, processing, and synthesizing information from various agencies and assets [12]. They have employed the C4.5 classification algorithm to forecast the risk percentage for a person making loans.

Huakang Li et al. (2021), This paper explains the primary usage of the LSTM-SVM model in predicting user loan risk and elaborates on the present economic climate and conventional risk forecasting techniques. In light of this, a prediction approach based on the LSTM and SVM methods is suggested. The prediction outcomes are compared to those of the conventional algorithm, and the model's viability is confirmed. The LSTM-SVM approach, however, has to be improved in further studies because it does have some limitations.

2.3 Research Summary

According to experimental studies, the Nave Bayes model outperforms other models in terms of loan forecasting. Numerous people are applying for bank loans as a result of the expansion of the banking industry, but due to the bank's limited resources, it can only grant loans to a select group of borrowers. As a result, a typical process involves determining who can receive a loan while still being the safest option for the bank. In order to save a lot of bank resources and work, we are attempting to lessen the risk element involved in choosing a safe individual for this project. This is accomplished by mining the Big Data of the individuals' prior records from when the loan was issued to them and based on these experiences.

2.4 Scope of the Problem:

I’ve reviewed some papers & articles. There they mentioned & applied different approaches. Loans of all kinds, including those for school, shopping, homes, and personal use, are dealt with by finance businesses and banks in our nation. In towns, cities, and villages, there are businesses and banks. These banks/companies want to verify the client details once the consumer applies for a loan to determine whether or not the candidate is eligible for one. The system's primary goal is to determine whether to approve a loan application based on train models.

2.5 Challenges:

I implemented five algorithms. So, it was very challenging for me to learn all the algorithms. I practiced more and more. On the other hand, I had no knowledge about f-1 score, recall, precision. So, I learned how to find out that topic’s results.

CHAPTER 3

RESEARCH METHODOLOGY

3.1 Introduction

In this part, I will quickly describe the steps I took to accomplish our study project. With the help of my project, the banker will be able to analyze which person will get the loan.

3.2. Data Collection Procedure:

Table 3.2.1: Dataset

Variable	Description
Loan_Id	unique loan_id
Gender	male/female
Married	y/n
Dependents	number of dependies
Education	graduate/undergraduate
Self_employeed	y/n
Applicant Income	Applicant’s Income
CoApplicant Income	CoApplicant’s Income
Loan_Amount	Loan amount
Loan_Amount_term	Term of Loan
Credit History	Applicant’s Credit history
Property Area	Urban/Semi-urban/rural
Loan_status	Loan approved (y/n)

3.3 Proposed Methodology:

For the coding part I took some steps:

- Data Collection
- Data Pre-processing
- Model Selection & Evaluation
- Get the best accuracy
- Result
- Testing

3.3.1 Flow Chart of my project:

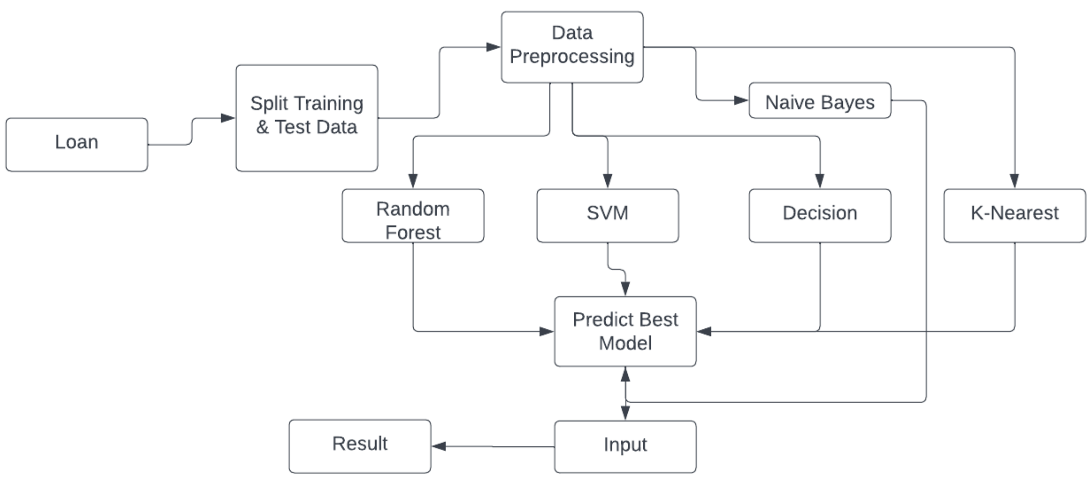

Figure 3.1: Flow chart of my project

3.3.2 Proposed Model:

In this research, we attempt to create a flexible user interface with visual concepts connected by a browser interface. Our aim is to use a machine learning model to classify master card fraud using data obtained from Kaggle as accurately as possible. Once we had done our initial research, we had a tendency to know that the Naïve Bayes model would provide the most accurate results.

- **Data Collection: I** took the data from an online source that was publicly usable. Here they collect the data in a google form: Loan Prediction Problem Dataset | Kaggle.

They arranged 13 questions. After getting the data, they convert it into CSV format. It was very easy for me that they split the data into train & test. So, I didn't have to split them.

	Loan_ID	Gender	Married	Dependents	Education	Self_Employed	ApplicantIncome	CoapplicantIncome	LoanAmount	Loan_Amount_Term	Credit_History	Property_Area
0	LP001002	Male	No	0	Graduate	No	5849	0.0	NaN	360.0	1.0	Urbar
1	LP001003	Male	Yes	1	Graduate	No	4583	1508.0	128.0	360.0	1.0	Rura
2	LP001005	Male	Yes	0	Graduate	Yes	3000	0.0	66.0	360.0	1.0	Urbar
3	LP001006	Male	Yes	0	Not Graduate	No	2583	2358.0	120.0	360.0	1.0	Urbar
4	LP001008	Male	No	0	Graduate	No	6000	0.0	141.0	360.0	1.0	Urbar

Figure 3.2: Head part of my Project

- **Data Pre-processing:** In this part, I cleaned the data. Missing values in the collected data could result in discrepancies. Preprocessing of the data is necessary to improve outcomes and the algorithm's efficiency. I must transform the variables and remove the outliers. To overcome these concerns, we use the chart function.

```
loan_train.isna().sum()

Loan_ID              0
Gender               0
Married              0
Dependents           0
Education            0
Self_Employed        0
ApplicantIncome      0
CoapplicantIncome    0
LoanAmount           0
Loan_Amount_Term     0
Credit_History       0
Property_Area        0
Loan_Status          0
dtype: int64
```

```
loan_test.isna().sum()

Loan_ID               0
Gender               11
Married               0
Dependents           10
Education             0
Self_Employed        23
ApplicantIncome       0
CoapplicantIncome     0
LoanAmount            5
Loan_Amount_Term      6
Credit_History       29
Property_Area         0
dtype: int64
```

Figure 3.3: Train & Test cleaning

- **Correlation:** Based on the association between the traits, it was shown that they were more likely to repay their loans. Individual and significant characteristics can include property kind, educational attainment, loan amount, and initial credit history, which is significant since it is regarded as such by perception. Using the Python platform's core plot and boxplot functions, you may connect the correlation between attributes.

	ApplicantIncome	CoapplicantIncome	LoanAmount	Loan_Amount_Term	Credit_History
ApplicantIncome	1.000000	-0.116605	0.565181	-0.046531	-0.020183
CoapplicantIncome	-0.116605	1.000000	0.189218	-0.059383	0.009391
LoanAmount	0.565181	0.189218	1.000000	0.036960	-0.018454
Loan_Amount_Term	-0.046531	-0.059383	0.036960	1.000000	-0.022967
Credit_History	-0.020183	0.009391	-0.018454	-0.022967	1.000000

Figure 3.4: Correlation

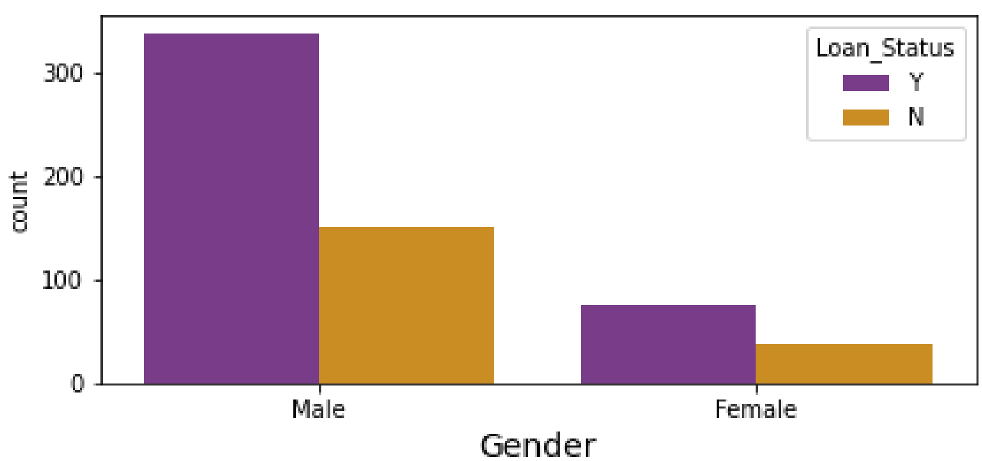

Figure 3.5: Data visualization of Gender

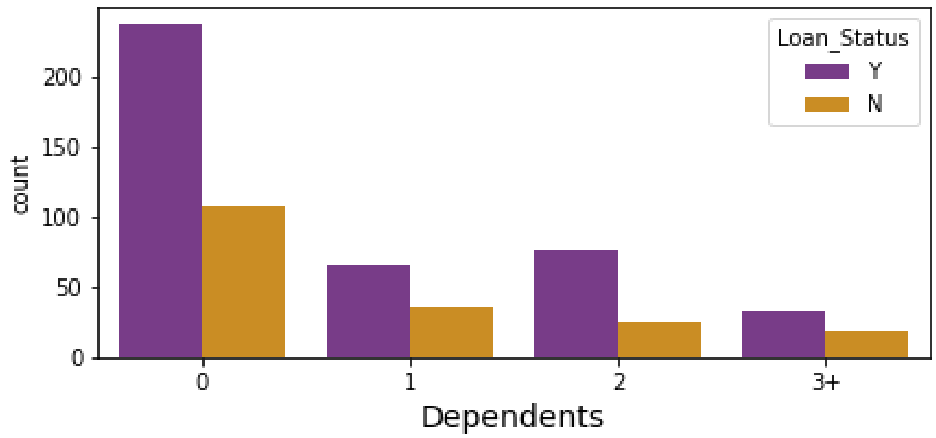

Figure 3.6: Number of Dependencies

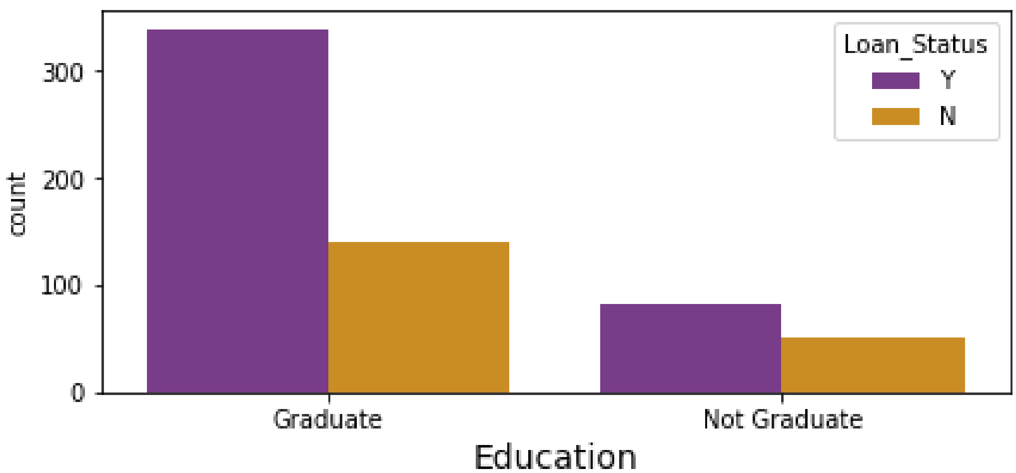

Figure 3.7: Education

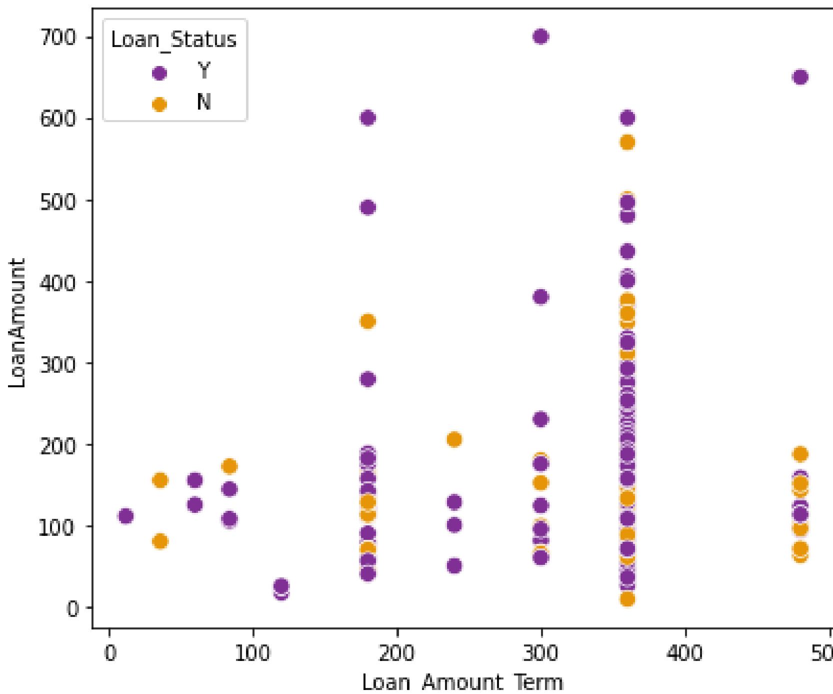

Figure 3.8: Loan Status regardings to loanAmount_Term

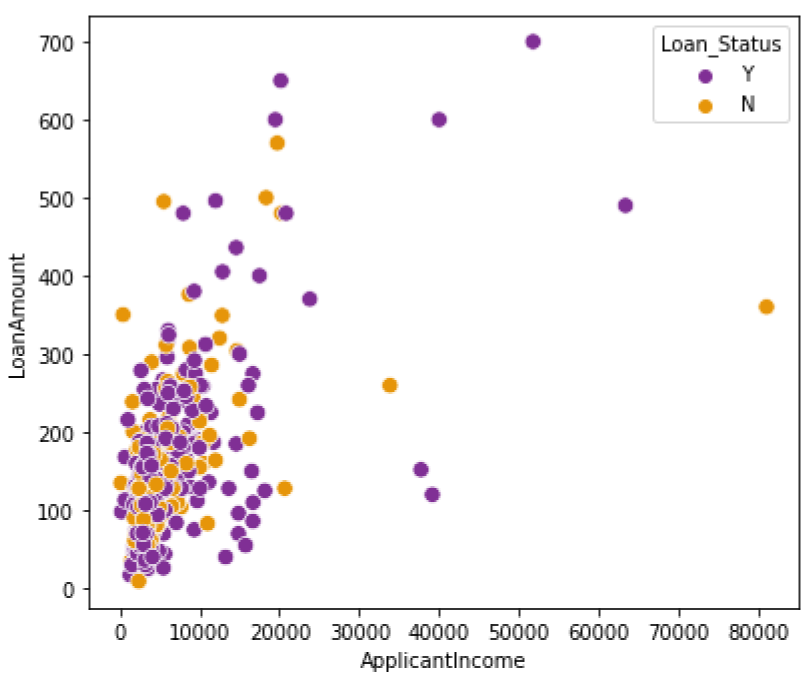

Figure 3.9: Applicant Income

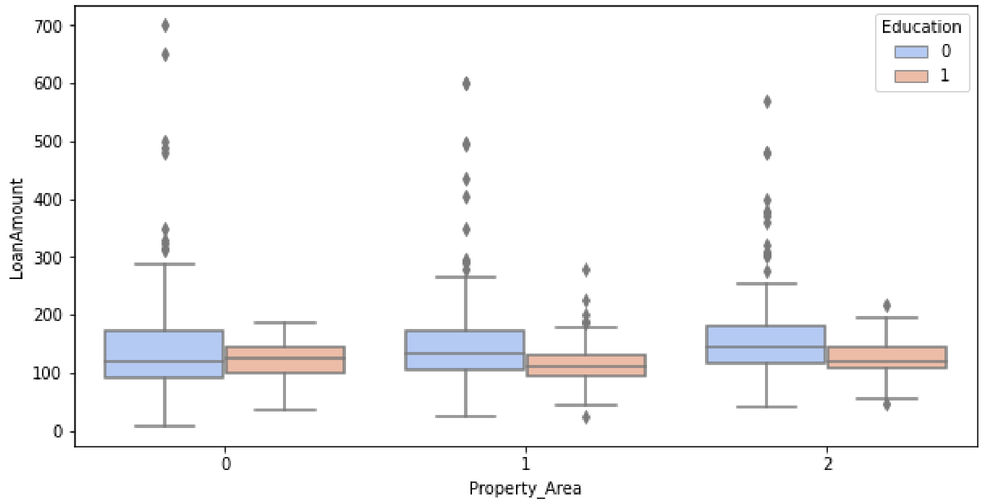

Figure 3.10: Property_Area

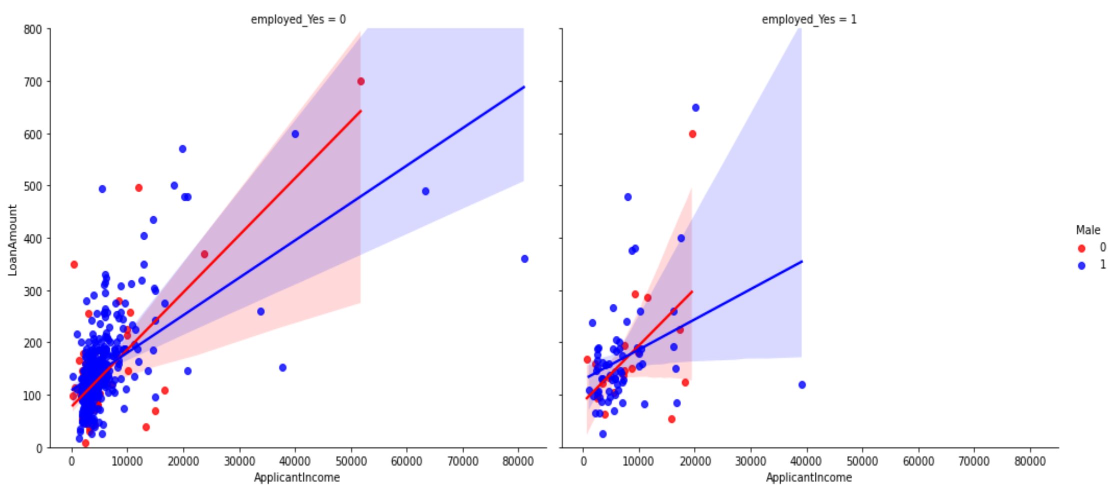

Figure 3.11: Applicant_Income

3.4 Machine Learning Model:

A subtype of artificial intelligence called machine learning teaches machines to think and act like humans without being explicitly taught. We employ supervised techniques in this paper. For the prediction of Android applications, five machine-learning classification models have been applied. The models can be found in free source Python software. Below are brief descriptions of each model.

- **Naive Bayes [13]:** The Naive Bayes technique is typically employed when a huge dataset needs to be predicted. Conditional Probability is utilized. The probability of event A happening given that an earlier event B has already happened is known as conditional probability. The most typical application of this algorithm is the screening of spam emails in your email account. For instance, you recently received new mail. The model employs the Naive Bayes method to predict whether or not the mail received is spam by looking through your previous spam mail records [13].

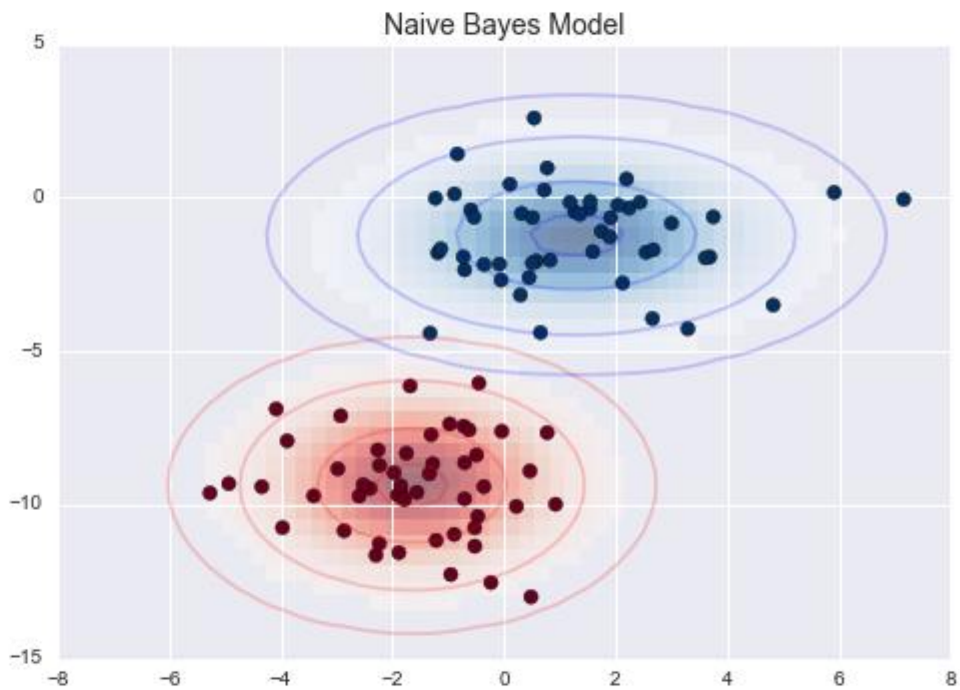

Figure 3.12: Naive Bayes [14]

- **SVM [15]:** Support Vector Machine is a Supervised Machine Learning algorithm that is used for regression and/or classification. Although it is occasionally quite helpful for regression, classification is where it is most often used. In essence, SVM identifies a hyper-plane that establishes a distinction between the various types of data [15]. This hyper-plane is just a line in two-dimensional space. Each dataset item is plotted in an N-dimensional space using SVM, where N is the total number of features and attributes in the dataset. The best hyperplane should then be found to divide the data. You must have realized by now that SVM can only perform binary classification by nature. For multi-class problems, there are numerous techniques to use [15].

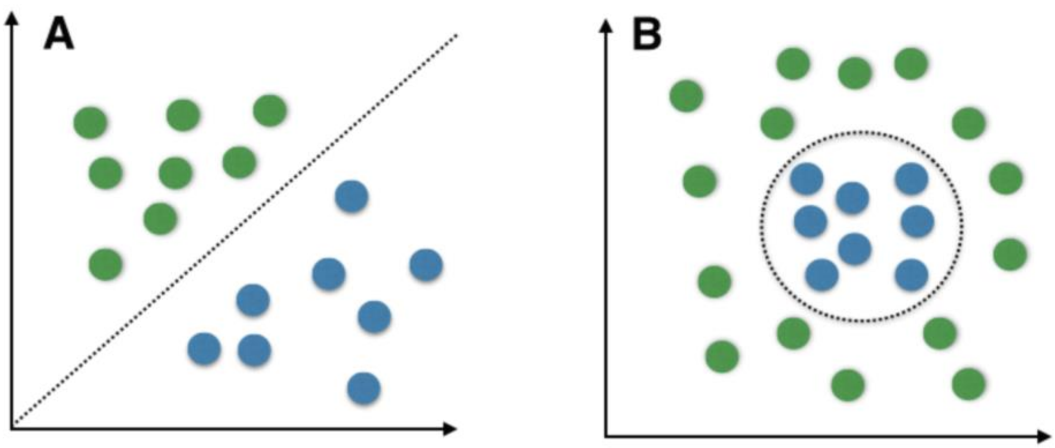

Figure 3.13: SVM [16]

- **Random Forest [17]:** The bagging method is extended by the random forest algorithm, which uses feature randomness in addition to bagging to produce an uncorrelated forest of decision trees. The random subspace method, also known as feature bagging, creates a random subset of features that guarantees a low correlation between decision trees. The main distinction between decision trees and random forests is this. Random forests merely choose a portion of those feature splits, whereas decision trees take into account all possible feature splits [17].

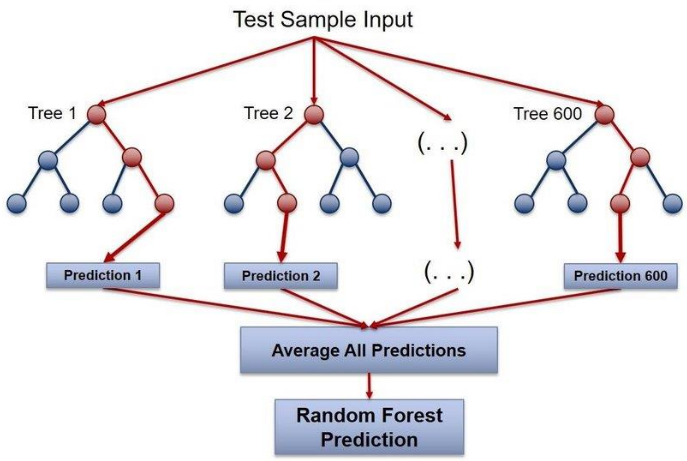

Figure 3.14: Random Forest[18]

- **K-Nearest Neighbour [19]:** A sort of supervised machine learning method is the K-nearest neighbors (KNN) algorithm. It can be applied to situations involving classification and regression. A new data point is classified using the K-NN algorithm depending on how similar it is to previously classified data. The K-NN algorithm makes it simple to determine a dataset's category or class [19]. It utilizes a similarity metric to determine which aspects of the new data set are most similar to the cats and dogs in the previous image. The distance between two points, which we have studied in geometry, is known as the Euclidean distance [19].

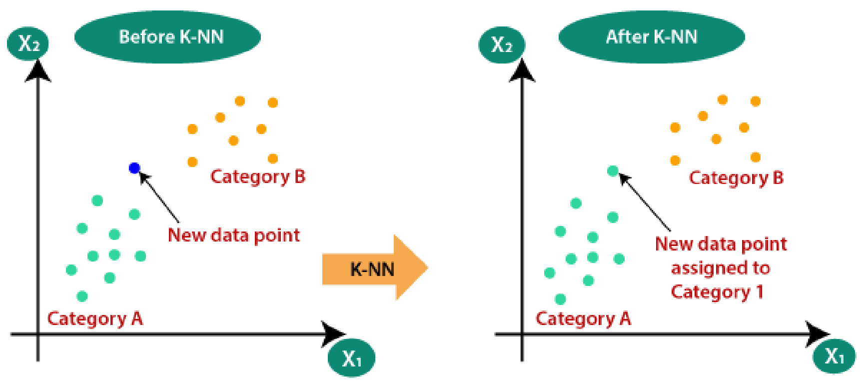

Figure 3.15: K-Nearest Neighbour [20]

- **Decision Tree [13]:** A decision tree is a graph that realizes the problem and makes judgments based on conditions. It shows each potential result of a choice and ultimately forecasts the result. For instance, every eCommerce website that allows you to make purchases will provide you with many choices based on your search criteria. Here, the classification process is carried out using a decision tree method [13].

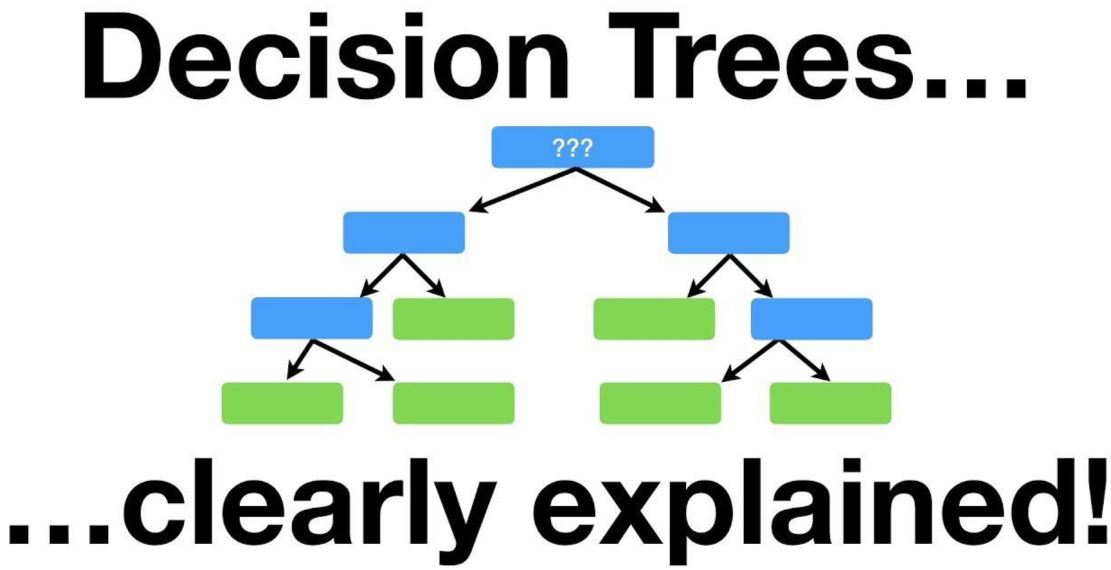

Figure 3.4.4: K-Nearest Neighbour [21]

CHAPTER 4

EXPERIMENTAL RESULTS AND DISCUSSION

4.1 Experimental Setup

I used a Colab notebook for my coding part. My useable language was python. For getting accuracy I uploaded some libraries.

4.2 Result Analysis

The model has to be tested after it has been trained. The model is evaluated using the data that we divided during the test-trained module. Confusion metrics, precision, recall, accuracy, and F1 score techniques are mostly used in utilized to assess the classification issue.

4.2.1 Confusion Matrix:

4.1 True Positive:

Table 4.1: True Positive

Algorithm	TP
Naive Bayes	25
Decision Tree Classifier	28
Random Forest	28
K-Nearest Neighbour	12
SVM	28

4.2 False Positive:

Table 4.2: False Positive

Algorithm	FP
Naive Bayes	25
Decision Tree Classifier	27
Random Forest	27
K-Nearest Neighbour	43
SVM	27

4.3 False Negative:

Table 4.3: False Negative

Algorithm	FN
Naive Bayes	3
Decision Tree Classifier	30
Random Forest	30
K-Nearest Neighbour	29
SVM	30

4.4 True Negative:

Table 4.4: False Negative

Algorithm	TN
Naive Bayes	127
Decision Tree Classifier	100
Random Forest	100
K-Nearest Neighbour	101
SVM	100

4.5 Accuracy:

Table 4.5: Accuracy

Algorithm	Accuracy (%)
Naive Bayes	82
Decision Tree Classifier	69
Random Forest	79
K-Nearest Neighbour	61
SVM	70

4.6 Recall:

Table 4.6: Recall

Algorithm	Recall
Naive Bayes	0.98
Decision Tree Classifier	0.77
Random Forest	0.94
K-Nearest Neighbour	0.78
SVM	1.00

4.7 Precision

Figure 4.7: Precision

Algorithm	Precision
Naive Bayes	0.81
Decision Tree Classifier	0.79
Random Forest	0.80
K-Nearest Neighbour	0.70
SVM	0.70

4.3 Result Discussion

With the help of the Naïve Bayes Algorithm, I got the best accuracy which was 82.16%. With certainty, it can be said that the Naïve Bayes model is quite effective and produces better results than other models. It functions properly and meets all bankers' standards. This technology calculates the outcome correctly and precisely. It accurately forecasts whether a loan application or customer will be approved or rejected.

CHAPTER 5

Impact on Society, Environment and Sustainability

5.1 Impact on Society

Every human feeling may be linked to the words we view on a daily basis on various online platforms in the digital world. In this case, it is critical for these platforms to have a mechanism in place to discern which are genuine emotions and which are pre-programmed aggressiveness. This is why I've decided to focus on one of the most fascinating genres of all time, by doing so, we can expect to create a more definitive and diverse digital era.

5.2 Impact on Environment

Due to the complexity of the network system of openness, sharing of resources, system, linking the variety, the uneven distribution of the terminal, network agnostic, and other barriers, computer networks continue to exhibit their distinctive benefits. Computer's cause. The biggest issue is security, which is one of the numerous issues brought on by the network. Predicting loan is not so easy. Sometimes we see that the person who is badly needed of the loan couldn’t get that. Everybody thinks it is a normal issue. But it is not. So that’s why I decided to work on it.

5.3 Ethical Aspects

Loans account for a large portion of bank profits. Despite the fact that many people are looking for loans. Finding a legitimate applicant who will return the loan is difficult. Choosing a real applicant may be difficult if the process is done manually. As a result, we are creating a machine learning-based loan prediction system that will choose the qualified applicants on its own. Both the applicant and the bank staff will benefit from this. There will be a significant reduction in the loan sanctioning period of time. In this research. The majority of the bank's revenue is generated directly from the interest income on loans.

5.4 Sustainability

- There are over 2.3 billion active internet-based life clients worldwide.
- At least two internet-based life cycles are present in 91 percent of large business brands.
- When they can't access their online life profiles, 65 percent of individuals feel uneasy and uncomfortable.
- It will be a helping hand for the researcher.
- Able to gain more knowledge about loan prediction methods.

CHAPTER 6

SUMMARY, CONCLUSION, RECOMMENDATION, AND IMPLICATION FOR FURTHER STUDY

6.1 Summary of the Study

The purpose of this study was How can we predict the loan. That means whether a person will get the loan or not. This work implements function extraction and data processing for customer basic attribute data and downloads transaction data based on the scenario of a bank credit application. Then, to increase the accuracy of bankruptcy assessment and achieve local optimization, a linear regression model with the penalty and a neural network prediction model are presented. By doing this, the implicit risk prediction and control of customer credit are enhanced, and the risk of bank loan default is significantly decreased. To raise the level of risk management for banks, the most suitable penalty linear regression prediction algorithm is chosen based on the characteristics of the sample data that was collected.

6.2 Conclusion

The majority of the bank's revenue is generated directly from the interest income on loans. Even when the bank authorizes the loan following a lengthy verification and testimony process, there is no guarantee that the chosen hopeful is the best hopeful. When performed manually, this operation requires additional time. We have the ability to foretell whether a specific hopeful is secure or not, and machine literacy has mechanized the entire testifying procedure. Loan Prognostic is extremely beneficial for both banks' clients and potential borrowers. I use some machine learning algorithms techniques to predict the loan data.

The distribution of loans is the primary activity of each and every bank. The primary portion of the bank's resources came directly from the income earned from the advances that the banks gave out. The basic objective of the banking system is to place resources in safe hands wherever they may be. Today, a number of banks and financial institutions grant loans following the second round of verification and validation, but it is still uncertain whether the chosen candidate is the most deserving of the candidates. With the help of this method, we can determine whether a certain candidate is secure or not, and the entire process of attribute validation is automated using machine learning [8][6]. The drawback of this model is that it gives completely different weights to each concern, but in practice, a loan may occasionally be accepted solely on the basis of a single powerful component, which isn't possible using this strategy. Both candidates and bank staff members can benefit from loan prediction.

The main source of revenue for banks is a loan. The majority of the bank's profits are derived directly from the money made from the loans. Even when the bank authorizes the loan following a lengthy verification and testimony process, there is no guarantee that the chosen hopeful is the best hopeful. When performed manually, this operation requires additional time.

6.3 Implication for Further Study

Loan prediction system is a potential work for the modern world. We'll work on more machine learning algorithms as we already worked on 6 model of ML to get even more and fluent output. This project can make an effort to find the suitable customers to give them loan more accurately and we'll working on it in future.

REFERENCES

[1] Loan approval prediction using machine learning algorithms approach – ijirt Available at: https://ijirt.org/master/publishedpaper/IJIRT151769_PAPER.pdf (Accessed: November 30, 2022).

[2] Team, D.F. (2021) Machine learning tutorial - all the essential concepts in single tutorial, Data Flair. Available at: https://data-flair.training/blogs/machine-learning-tutorial/ (Accessed: December 1, 2022).

[3] F. L. Macedo, A. Reverter, and A. Legarra, "Behavior of the linear regression method to estimate bias and accuracies with correct and incorrect genetic evaluation models," *Journal of Dairy Science,* vol. 103, no. 1, pp. 529–544, 2020.

[4] IBM Cloud Education (no date) *What is machine learning? IBM*. Available at: https://www.ibm.com/cloud/learn/machine-learning (Accessed: December 1, 2022).

[5] Behl, A. (2019) *An introduction to machine learning*, *Medium*. Becoming Human: Artificial Intelligence Magazine. Available at: https://becominghuman.ai/an-introduction-to-machine-learning-33a1b5d3a560 (Accessed: December 1, 2022).

[6] Akanksha, Tamara Denning, Vivek Srikumar, Sneha Kumar Kesera "secretsin source code: reducing false positives using ML" software engineering (Microsoft) school of computing, USA (2020).

[7] *The fundamentals of machine learning - interactions* (no date). Available at: https://www.interactions.com/wp-content/uploads/2017/06/machine_learning_wp-5.pdf (Accessed: November 30, 2022).

[8] Mohammad *Ahmad Sheikh, Amit Kumar Goel, Tapas Kumar "An Approach for Prediction of Loan Approval using Machine Learning Algorithm" School Of Computer Science And Engineering Galgotias University Greater Noida,India (2019).*

[9] *Prediction for loan approval using machine learning algorithm* (no date). Available at: https://www.irjet.net/archives/V8/i4/IRJET-V8I4785.pdf (Accessed: November 30, 2022).

[10] *(PDF) the loan prediction using machine learning - researchgate* (no date). Available at: https://www.researchgate.net/publication/357449126_THE_LOAN_PREDICTION_USING_MACHINE_LEARNING (Accessed: November 30, 2022).

[11] Li, X. *et al.* (1970) *Figure 3 from overdue prediction of bank loans based on LSTM-SVM: Semantic scholar*, *undefined*. Available at: https://www.semanticscholar.org/paper/Overdue-Prediction-of-Bank-Loans-Based-on-

LSTM-SVM-Li-Long/490a6f390a2ecb766cc88781bf0b6f76cf0e50b9/figure/2 (Accessed: December 1, 2022).

[12] Survey, M. *et al.* (1970) *Data mining techniques to analyze risk giving loan(bank): Semantic scholar*, *undefined*. Available at: https://www.semanticscholar.org/paper/Data-mining-techniques-to-analyze-risk-giving-Surve-Thitme/c54ae9739803faf76109b78a1bc59e437559a1d9 (Accessed: December 1, 2022).

[13] Kumar Arun, Garg Ishan, Kaur Sanmeet, ― Loan Approval Prediction based on Machine Learning Approach‖, IOSR Journal of Computer Engineering (IOSR-JCE), Vol. 18, Issue 3, pp. 79-81, Ver. I (May-Jun. 2016).

[14] V. C. T. Chan et al., "Designing a Credit Approval System Using Web Services, BPEL, and AJAX," 2009 IEEE International Conference on eBusiness Engineering, Macau, 2009, pp. 287- 294.doi: 10.1109/ICEBE.2009.46

[15] M. Bayraktar, M. S. Aktaş, O. Kalıpsız, O. Susuz and S. Bayracı, "Credit risk analysis with classification Restricted Boltzmann Machine," 2018 26th Signal Processing and Communications Applications Conference (SIU), Izmir, 2018, pp. 1-4.doi: 10.1109/SIU.2018.840 4397

[16] Mohammad Ahmad Sheikh, Amit Kumar Goel,Tapas Kumar. "An Approach for Prediction of Loan Approval using Machine Learning Algorithm", 2020 International Conference on Electronics and Sustainable Communication Systems (ICESC), 2020

[17] R. Samatov, "Application of the linear regression method to determine the effective organization of the transportation," *Acta of Turin Polytechnic University in Tashkent*, vol. 9, no. 3, 4 pages, 2019.

[18] G. Ayoub, T. H. Dang, T. I. Oh, S.-W. Kim, and E. J. Woo, "Feature extraction of upper airway dynamics during sleep apnea using electrical impedance tomography," *Scientific Reports*, vol. 10, no. 1, Article ID 1637, 2020.

[19] Mrunal Surve, Pooja Thitme, Priya Shinde, Swati Sonawane, and SandipPandit. "Data mining techniques to analyze risk giving loan (bank)"Internation Journal of Advance Research and Innovative Ideas in EducationVolume 2 Issue 1 2016 Page 485-490

[20] Ch. Balayesu and S Narayana, "An Improved Algorithm for Efficient Miningof Frequent Item Sets on Large Uncertain Databases" in International Journalof Computer Applications, Volume 73, No. 12 July 2013, Page No. 8-15.

Loan Prediction Report

ORIGINALITY REPORT

19%	11%	8%	%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	dspace.daffodilvarsity.edu.bd:8080 Internet Source	8%
2	doi.org Internet Source	4%
3	www.ncbi.nlm.nih.gov Internet Source	1%
4	technoblender.com Internet Source	1%
5	www.analyticsvidhya.com Internet Source	1%
6	www.machinelearningplus.com Internet Source	1%
7	bmcinfectdis.biomedcentral.com Internet Source	1%
8	machinelearningmastery.com Internet Source	1%
9	journals.plos.org Internet Source	1%