Predictive Modeling for Loan Approval Rates

**A Project Work Synopsis**

*Submitted in the partial fulfillment for the award of the degree of*

**BACHELOR OF ENGINEERING**

**IN**

**COMPUTER SCIENCE WITH SPECIALIZATION IN**

**ARTIFICIAL INTELLIGENCE AND MACHINE LEARNING**

**Submitted by:**

HITESH GARG - 20BCS6614

JAIN KARAN ANAND - 21BCS8804

MITALI GUPTA - 20BCS6890

VATHADA LIKHITH SAI - 20BCS4593

**Under the Supervision of:**

**Prof. Pramod Vishwakarma**

**CHANDIGARH UNIVERSITY, GHARUAN, MOHALI - 140413,**

**PUNJAB**

**January, 2024**

# Abstract

Keywords: Deep Learning, Open AI, Open CV, Artificial Intelligence, Concept Generation.

Technology has boosted the existence of human kind the quality of life they live. Every day we are planning to create something new and different. We have a solution for every other problem we have machines to support our lives and make us somewhat complete in the banking sector candidate gets proofs/ backup before approval of the loan amount. The application approved or not approved depends upon the historical data of the candidate by the system. Every day lots of people applying for the loan in the banking sector but Bank would have limited funds. In this case, the right prediction would be very beneficial using some classes-function algorithm. An example the logistic regression, random forest classifier, support vector machine classifier, etc.

**Table** of Contents

|  |  |
| --- | --- |
| Title Page | i |
| Abstract | ii |
| 1.    Introduction |  |
| 1.1           Problem Definition |  |
| 1.2 Project Overview |  |
| 1.3 Hardware Specification |  |
| 1.4 Software Specification |  |
| 2.    Literature Survey |  |
| 2.1 Existing System |  |
| 2.2 Proposed System |  |
| 2.3 Literature Review Summary |  |
| 3.    Problem Formulation |  |
| 4.    Research Objective |  |
| 5.    Methodologies |  |
| 6.    Experimental Setup |  |
| 7.    Conclusion |  |
| 8.    Tentative Chapter Plan for the proposed work |  |
| 9.    Reference |  |

# 1. INTRODUCTION

## 1.1 Problem Definition

With the increase in banking sector many people are applying for loans in bank. All these loans are not approvable. The main income of bank assets comes from gain earned from loans. The main objective of banks is to invest their assets in safe customers. Today many banks approve loan after many process of verification and validation but still there is no surety that selected customer is safe or not. Therefore it is important to apply various techniques in banking sector for selecting a customer who pays loan on time. In this report we use random forest algorithm for the classification of data. Random forests algorithm builds a model from trained dataset and this model is applied on test data and we get the required output.

**1.2 Problem Overview**

Banks, Housing Finance Companies and some NBFC deal in various types of loans like housing loan, personal loan, business loan etc in all over the part of countries. These companies have existence in Rural, Semi Urban and Urban areas. After applying loan by customer these companies validates the eligibility of customers to get the loan or not. This project provides a solution to automate this process by employing machine learning algorithm. So the customer will fill an online loan application form. This form consist details like Sex, Marital Status, Qualification, Details of Dependents, Annual Income, Amount of Loan, Credit History of Applicant and others.

## 1.3 Hardware Specification

1. PC / Laptop

2. Keyboard

3. Mouse

## 1.4 Software Specification

1. Anaconda Navigator

2. Jupyter Notebook

# 2. LITERATURE SURVEY

## 2.1 Existing System

## Historically, lending risk prediction has used statistical methods, including Linear Discriminant Analysis and Logistic Regression. However, with large credit datasets, ML-driven risk estimation algorithms like k-Nearest Neighbor, Random Forest, and Support Vector Machines are better at capturing complex relationships. Moreover, deep learning methods have gained a particular advantage in modeling non-linear relationships between risk and risk factors for large-scale lending risk and loan prediction datasets. Novel frameworks like DEAL (Deep Ensemble Algorithm), or improvements over existing models of Recurrent Neural Networks (RNN) or Boosted Decision Tree or Autoencoders, give satisfactory accuracy over large datasets and generate features with domain expertise.

## However, more work is available on machine learning models than deep learning architectures since the latter's performances are often specific to the dataset they were designed and tested on. In the figure shown below, a recent paper compares the performance of various machine learning algorithms on the German credit risk dataset. We can see that algorithms like SVM, Random forest, and the Logistic regression model perform better than ELM and ANN. However, decision trees and boosting also give a competitive performance on this dataset.

## 2.2 Proposed System

The proposed system is Loan Approval System software used for approval of loan in banking sector. In this proposed system we have used machine learning algorithm. Machine Learning is process in which a symmetric model is build from the existing dataset; this model is applied for the testing of the new dataset. The system consists of trained dataset and test dataset. The trained dataset is used for construction of model. This model is applied on testing dataset for the required result. We have used Ensemble approach for building of the model. Random forest algorithm uses this ensemble approach and builds a model from the existing training dataset.

## 2.3 Literature Review Summary (Minimum 7 articles should refer)

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Year and**  **Citation** | **Article/ Author** | **Tools/ Software** | **Technique** | **Source** | **Evaluation Parameter** |
| 2011 | K. Hanumantha Rao., G. Srinivas., A. Damodhar., M.Vikas Krishna |  |  | Vol. 2, Issue 3, 2011 | Implementation of Anomaly Detection Technique Using Machine Learning Algorithms |
| 2002 | S.S. Keerthi., E.G. Gilbert |  |  | Vol. 4, Issue 1, pp. 351-360, 2002 | Convergence of a generalize SMO algorithm for SVM classifier design, Machine Learning |
| 2002 | Andy Liaw., Matthew Wiener |  |  | Vol. 2, Issue 3, pp. 9-22 | Classification and Regression by random Forest |
| 2014 | Ekta Gandotra., Divya Bansal., Sanjeev Sofat |  |  | Vol. 05, Issue 02, pp. 56-64 | Malware Analysis and Classification: A Survey, Journal of Information Security |
| 2013 | Aafer Y., Du W., Yin H., Droid |  |  | pp 86-103 | Mining API-Level Features for Robust Malware Detection in Android, Security and privacy in Communication Networks |
| 2021 | J. R. Quinlan |  |  | Vol. 1, No. 1. pp. 81-106. | Induction of Decision Tree, Machine Learning |
| 2022 | Q. Zhang, J. Yang, X. Zhang and T. Cao |  |  | (ICIVC), Xi’an, China, 2022, pp. 848-853, doi: 10.1109/ICIVC55077.2022.9886154. | Generating Adversarial Examples in Audio Classification with Generative Adversarial Network |

.

# 3. PROBLEM FORMULATION

The main purpose of this project is to provide immediate and accurate results for the approval of loan to the eligible customers. In banking sector there will be n number of people who apply loans. It is difficult to check customer’s eligibility through paper work. The system can provide accurate results for the n number of people. In this project we have discussed about credit risk and credit analysis. Banking sectors success mainly depends of credit risk analysis. In this report we have used Random Forest [3] approach to build the model. The use of Random Forest is because Random Forest Approach provides accurate results than the K Nearest Neighbor and Decision Tree. In this project we have used Random Forest approach for building a model. In this report two or more classifiers are combined together and identify a perfect model for loan prediction. Ensemble method compares two or more models and identifies a perfect model from two or more models for better loan prediction which makes banking sector to make a right choice for approval of loan application.

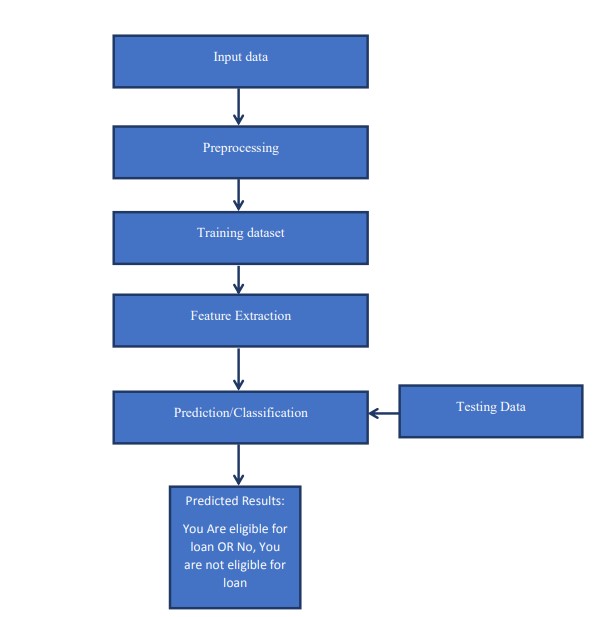
# 4. OBJECTIVES

The main objective of banks is to invest their assets in safe customers. Up to now many banks are processing loans after regress process of verification and validation. But till now no bank can give surety that the customer who is chosen for loan application is safe or not. So to avoid this situation we introduced a system for the approval of bank loans known as Loan Prediction System Using Python. Loan Prediction System is a software which checks the eligibility of a particular customer who is capable of paying loan or not. This system checks various parameters such as customer’s martial status, income, expenditure and various factors. This process is applied for many customers of trained data set. By considering these factors a required model is built. This model is applied on the test data set for getting required output. The output generated will be in the form of yes or no. Yes indicates that a particular customer is capable of paying loan and no indicates that the particular customer is not capable of paying loan. Based on these factors we can approve loans for customers.

# 5. METHODOLOGY

When we submit our training dataset to these classifiers produce output according to their algorithm function. This method is known as Heterogenous classifying or Ensemble learning. We can also take all learner classifier as same algorithm. But when we upload same training dataset these all classifiers produce same output. To avoid this problem we have give different training dataset to these classifiers, So that each classifier gives different outputs. While providing the output these classifiers considering various factors build a model. Each classifier may build a model according to given training dataset. From these different classifiers we combine all these classifiers a build a new classifier model which satisfies all these classifiers. This classifier is considered as Strong Classifier. This is called Strong classifier because it produces accurate output and less error. Random Forest also uses this Ensemble Learning Technique.

# 6.EXPERIMENTAL SETUP



# 7.CONCLUSION

From the project view of analysis this system can be used perfect for detection of clients who are eligible for approval of loan. The software is working perfect and can be used for all banking requirements. This system can be easily uploaded in any operating system. Since the technology is moving towards online, this system has more scope for the upcoming days. This system is more secure and reliable. Since we have used Random Forest Algorithm the system returns very accurate results. There is no issue if there are many no of customers applying for loan. This system accepts data for N no. of customers. In future we can add more algorithms to this system for getting more accurate results.

## REFERENCES

[1] K. Hanumantha Rao., G. Srinivas., A. Damodhar.,

M.Vikas Krishna., Implementation of Anomaly Detection

Technique Using Machine Learning Algorithms:

International Journal of Computer Science and

Telecommunications, Vol. 2, Issue 3, 2011.

[2] S.S. Keerthi., E.G. Gilbert., Convergence of a

generalize SMO algorithm for SVM classifier design,

Machine Learning, Springer, Vol. 4, Issue 1, pp. 351-360,

2002.

[3] Andy Liaw., Matthew Wiener., Classification and

Regression by random Forest, Vol. 2, Issue 3, pp. 9-22,

2002.

[4] Ekta Gandotra., Divya Bansal., Sanjeev Sofat.,

Malware Analysis and Classification: A Survey, Journal

of Information Security, Vol. 05, Issue 02, pp. 56-64,

2014.

[5] Rattle data mining tool, http://rattle.togaware.com/rattle

-download.html.

[6] Aafer Y., Du W., Yin H., Droid APIMiner: Mining

API-Level Features for Robust Malware Detection in

Android, Security and privacy in Communication

Networks, Springer, pp 86-103, 2013.

[7] J. R. Quinlan., Induction of Decision Tree, Machine

Learning, Vol. 1, No. 1. pp. 81-106