A PROJECT REPORT

on

"BANK PERSONAL LOAN MODELLING"

Submitted to

KIIT Deemed to be University

In Partial Fulfilment of the Requirement for the Award of

BACHELOR'S DEGREE IN
COMPUTER SCIENCE AND SYSTEM ENGINEERING

BY

SANTANU GIRI	2128048
SHUBHAM BOSE	2128060
ABHISHEK KUMAR SINGH	2128061
SK ASIF HOSSAIN	2128089

UNDER THE GUIDANCE OF
DR. AMIYA RANJAN PANDA
DR. MANOJ KUMAR MISHRA

SCHOOL OF COMPUTER ENGINEERING

KALINGA INSTITUTE OF INDUSTRIAL TECHNOLOGY

BHUBANESWAR, ODISHA - 751024

April 2024

A PROJECT REPORT

on

"BANK PERSONAL LOAN MODELLING"

Submitted to

KIIT Deemed to be University

In Partial Fulfillment of the Requirement for the Award of

BACHELOR'S DEGREE
IN COMPUTER SCIENCE AND
SYSTEM ENGINEERING

BY

SANTANU GIRI	2128048
SHUBHAM BOSE	2128060
ABHISHEK KUMAR SINGH	2128061
SK ASIF HOSSAIN	2128089

UNDER THE GUIDANCE OF
DR. AMIYA RANJAN PANDA
DR. MANOJ KUMAR MISHRA

SCHOOL OF COMPUTER ENGINEERING

KALINGA INSTITUTE OF INDUSTRIAL TECHNOLOGY

BHUBANESWAR, ODISHA -751024

APRIL 2024

KIIT Deemed to be University

School of Computer Engineering
Bhubaneswar, ODISHA 751024

CERTIFICATE

This is certify that the project entitled

"BANK PERSONAL LOAN MODELLING"

submitted by

SANTANU GIRI	2128048
SHUBHAM BOSE	2128060
ABHISHEK KUMAR SINGH	2128061
SK ASIF HOSSAIN	2128089

is a record of bonafide work carried out by them, in the partial fulfilment of the requirement for the award of Degree of Bachelor of Engineering (Computer Science & System Engineering) at KIIT Deemed to be university, Bhubaneswar. This work is done during year 2023-2024, under our guidance.

Date: 07/04/2024

(Dr. Manoj Kumar Mishra)
(Dr. Amiya Ranjan Panda)
Project Guide

Acknowledgements

We are profoundly grateful to **Dr.Amiya Ranjan Panda** and **Dr. Manoj Kumar Mishra** of KIIT UNIVERSITY

for his expert guidance and continuous encouragement throughout to see that this project rights its target since its commencement to its completion. .....

SANTANU GIRI
SHUBHAM BOSE
ABHISHEK KUMAR SINGH
SK ASIF HOSSAIN

ABSTRACT

The present research is intended to anticipate mathematically the minimal chance of purchasing personal loans from borrowers. For the research to hit the main goals, learning algorithms like KNN, Logistic Regression, Naive Bayes, Decision Tree, Random Forest and SVM were considered. Make stability of model to be based on Testing and Reporting, and use accuracy as a benchmark principle. According to the validity of the testification data employed in modeling the specific problem, random forest classifier shows the highest accuracy whereas support vector machine has the lowest accuracy in comparison with other models. Random Forester's accuracy rate is 98.67%, while K-NN, decision tree, logistic regression and Naive Bayes models' are 90.06%, 97.26%, 89.86% and 89.86% respectively. The last result of SVM performance are taken into account by 89.80%.The subsequent data analysis has graduated the random forest distribution as the most powerful method in correctly predicting the number of people who are going to qualify for credit cards.

Keywords:

KNNs, Logistic Regressions, Naive Bays, Decision Trees, Random Forests as well as SVM.

Contents

List of Figures

\

Chapter 1
Introduction

A loan is actually borrowing the amount from banks or other institutions that can provide credit. Please note that we will return the money we borrowed (principal) plus interest (the bank will receive its money at a certain term) that we all agreed for a given period of time. The most frequent term of repayment is the first month. The fact that banks refuse you loan approval is based on the situation, whether you can afford to compensate the loan. In this zone, the groups consist of such components as our wage, our job experience and our creditworthiness.

The bank offers a variety of loan options, each serving different needs and goals:The bank offers a variety of loan options, each serving different needs and goals:

- Personal Loans – Universal loans for daily personal needs which cover such occasions as jointly entered into business, home repairs, or important events.
- Rental car - The car rental process, however, can become a must if one is in need of a car for wherever the road will take them and their car. The loan is firmly linked to the amount of the car's price.
- Home equity/purchase - purchasing a house and getting a home loan. The house turns into a product and may satisfy the diversity of purposes.
- Students Debt - Funding the student's perception of higher education. The ones which are subsidized (low interest rates) and the other ones which are earned from the very beginning of the loan can be unsubsidized (interest earned from the start).

And the interest rates might be different in different banks or in case of different types of loans which we take. Also we have to bear in mind that the bank will set the important rules for us to follow when we borrow money from the bank.

The project we are undertaking with personal loan company is to prepare a prediction pertaining to clients' application for personal loans. This aids banks in keeping their businessactivities more targeted and lending procedures tighter. Consequently, techniques such as logistic regression can be used to investigate this data and find out the connection between the behavior of the customers and the market sectors they patronize previously. However, the model demonstrates the effectiveness of personal loan product as it is learned from the gathered data that the customers will be satisfied with the personal loan.

Training happens with this model in effect, banks can be able to access and direct their marketing activities at customers who have a higher possibility of being interested in a personal loan. Other than that, the likelihood of the success of their campaigns will also get enhanced with the reduction in operating expenses since the most receptive audience will be targeted. Furthermore, according to the model it is possible to make a loan application process more convenient and transparent. Foreseeing a client's loan approval could quicken up the process and provide a smoother system for the bank and the client on one hand, and on the other hand.

Chapter 2

Basic Concepts/ Literature Review

2.1 Use of ML Models

In this project, we use many learning machines such as Random Forest Classifier, SVM, Naive Bayes, Decision Tree, KNN and Logistic Regression. We will examine the main points of these models as they are used to find the result. The model used is:

- **Random Forest Classifier:** Random Forest Classifier is a learning technique that uses multiple decision trees to predict outcomes. Overfitting can be reduced by splitting the data and features into smaller pieces and training decision trees to combine the results. This method is designed to provide simple and intuitive modeling suitable for complex high-dimensional datasets by considering the prediction of each tree in the prediction process and averaging or sharing them.

- **KNN**: K-Nearest Neighbors (KNN) is a flexible machine learning technique that can be used in both regression and classification techniques. It can predict the new state by considering the nearest neighbors and using the previous value of k during the training process. It determines the average of neighbors for inversion and uses the majority votes of the closest units for distribution. This model can be very expensive when used for large-scale training and its performance may be affected by the k value and measurement distance.

- **Logistic Regression:** Prediction of binary or categorical outcomes as prediction should be based on the statistical method of logistic regression. Increasing the results from 0 to 1 changes the horizontal line equation in the logistic function. The model must calculate the difference between variables or coefficients to determine which predictor variable has the largest effect. The maximum estimate should be used to improve the model and reduce the difference between expected and actual results. Due to its simplicity, logistic regression can be used to predict job response rates, credit risk, and disease outcomes [6] [8].

- **Decision Tree**: For multiple criteria decision making a machine learning technique of the following type is used: a decision tree builds a tree-like model. The page will conduct this by further separating the product into several small groups that will group its most important features, then a final decision has been made. This specific algorithm targets diminishing tree size in favor of accumulating more knowledge/data at the same time that the step is executed to prevent the tree from getting overloaded. Model begins with the top of the tree down to the node that designates for the targeted category with the probabilistic outputs that amount to the complete input features. Design can be said to be a powerful tool to accuse the world of, because it is easy to evaluate and comprehend.

- **Naive Bayes:** The widely used method Naive Bayes enjoys success in the task classification area of machine learning. It is done on the idea of probability perambulating Bayes' formula to estimate the probability of the condition in the defined group. A behavioral model which provides information in a probabilistic language is referred to as a simple probabilistic model. Identification of feature classes markedly different or the ones that require least a priori knowledge [8].

- **SVM**: Artificial intelligence is an all-encompassing term for foreign technology that incorporates various means and methods. They turn to be so great if the tasks are numerous and the tasks distribution becomes the issue. The fact than essentially most machine learning algorithmists are supervised learners and as an SVN it also learns from labeled data is what makes the machine learn from labeled data. This depict that the amplitudes of the data points are split during virus outbreak. [12]. In this case, the planes' sides are found through the gaps (edges) and data points closest to the hyperplane are called support vectors. Such support that these support vectors are able to let you decide what limits of knowledge should be considered by SVM.

2.2 Literature Review

Machine learning helps researchers to reinsure and report concepts for practical research and to testing experiments in the banking industry. The authors of the paper [3] used random forest classifiers, decision tree algorithms, and logistic regression models for predicting (15) what actions the users will take.

The authors of [4] studied the performance of five well-known methods in machine learning for scoring: Examples of classifiers are Bayes theorem model, logistic regression analysis, random forests, decision trees, and k-nearest neighbor classifiers. The least effective as they run their assumptions that random forest was the most good and effective model.

The authors use the numerals and probability point, thus, the logistic regression emerges as a machine learning sort of the model. Employing the logistic regression technique, this paper attempts to draw a clear connection between the activities of lenders on the one hand and the reasons for the rejection of loan applications on the other. Logistic regression turns out to be very convenient for prediction of the outcomes and is really accurate. Incorporated Logistic regression can help in controlling for power points and non linear influence.

Becker, F., and Flordi, H. in[10] developed credit scoring studies and data to develop machine learning models like logistics regression, pruning trees and random forest to examine their usefulness to classify defaulters through the classification exercises. The model is able to classify individuals based on their cobbett coefficient as a better predictor of their behavior thereby making it possible to use more cross validation methods, which will be more suited to the individual and allow the second model which performs best to classify them. Evaluate the credit scoring systems and create data for selecting appropriate machine learning models for credit calls done.

The aim of this paper lies in predicting a loan approval using learning technology [11] In other words, the main target lies in deciding whether to lend money to someone or not.

Chapter 3

Problem Statement / Requirement Specifications

The core purpose for the establishment of this project is to design a machine learning model that can forecast the performance of consumers who go for a personal loan from banks. The model applies different attributes of the client, their financial capacity, credit history and other related parameters to establish whether the applicant will be able to manage to pay back the debt and if they qualify for a loan.

3.1 Project planning-

Characterize the project scope, reason and deliverables. Know the needs of resources, such as materials, instruments and course of action.

Structure your task list and determine see to its assigned priority. Determining who does what and task delegation is also an imperative step towards achieving success.

3.2 Project Evaluation

We pay attention to data collection and allocation as customers' choice is our highest priority.

data (including financial details and credit history) The provided sentence can be humanized as following: data (IE financial details and credit history).Following the search, we

make the decision by the observation of the data so as to reveal the patterns, trends, and some features which may be resulted in credit worthiness. Acknowledge both the business needs and the limitations of a process that evaluates credit approval.

. To measure the efficacy of the model decide on suitable metric like accuracy, precision, recall, and F1 score.

3.3 System Design

3.3.1 Design Limitations

Data availability and quality: Pre-collect all data required for training and testing models and make sure it is sufficient enough for each algorithm.

Evaluation and ethics: Complain with the survival of the laws on the protection of personal information and ethical principles in the lending.

Computing resources: Calculate the number of computing resources that will be used for machine learning training and later for application.

Model Interpretability: To study the impacts of loan decisions, interpretability marks as plastic models versus assembling variables.

3.3.2 System Architecture

The following elements make up the individual loan system diagram that will also include architecture or block diagram.

Data acquisition: Authors too should be at the center of the product development. Get required details about the customers by consulting different sources including credit reporting agencies, data banks, and financial companies alike.

Prerequisites: Materials Data maintenance, building work or data conversion that is prior to modeling have all been included in this item.

Model training: Some include the random forests, logistic regression, decision trees, etc. Examples are also some of the common features of the machine learning algorithm.

Model evaluation: The given topic, therefore will evaluate the effectiveness of the method of training that has been adopted by using applicable indicators such as competition.

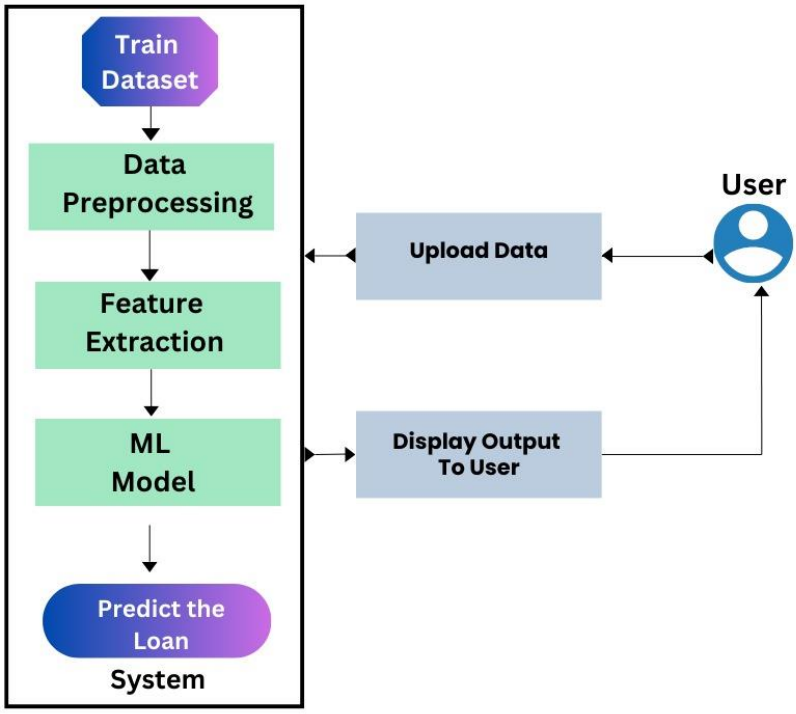
Sample selection: The device is designed for the sample which has the highest score for exporting following the collected information.

Model Deployment: In this section, the installed model goes into the bank's form of loan application and brings up loan estimation of the new goods for the closely using customers.

User Interface: It makes certain part of the business as more transparent.the friendly link to bank officials and customers to quickly access information needed for request and the restriction of information stores to a reasonable estimate of the loan

Monitoring and maintenance: These devices track the current model performance through time and continuously make changes or modify the model along the process to ensure that the model is not deviating off on the path of accuracy and precision.

Below is a figure of the design structure for the business of personal loan.


Train Dataset
Data Preprocessing
Feature Extraction
ML Model
Predict the Loan
System
Upload Data
Display Output To User
User

Chapter 4

Implementation

4.1 Understanding the information: That is the primary step to follow up. It is the very first step. At last loan application candidates must be ready to present information about their work experience. This includes the number of years of work and employment type like permanent, full-time or part-time. They will be able to know the business identity including. Try suggestion: They will be able to determine the identity of their business entities, e.g.It shows incredible properties. The name can be used as the semantic feature, and additionally it can also declare the power that the name provides a supernatural ability to do a specific thing. Above the first row, write a loop that runs through each row, then in each row, check all row elements Secondly, you can create headings using bullets or numbers in order to give special attention to some elements.

Develop a basic profile box to put back the starting row names

In addition, you can use the following to follow, including We have 4999 fonts: they are applying them a credit card in their school, home, they own the age credit, friends, family or card online as well as offline ! From the histogram, it is evident that the distribution of people having green eyes belong to the normal distribution category as a given figure illustrates.

	count	mean	std	min	25%	50%	75%	max
ID	5000.0	2500.500000	1443.520003	1.0	1250.75	2500.5	3750.25	5000.0
Age	5000.0	45.338400	11.463166	23.0	35.00	45.0	55.00	67.0
Experience	5000.0	20.104600	11.467954	-3.0	10.00	20.0	30.00	43.0
Income	5000.0	73.774200	46.033729	8.0	39.00	64.0	98.00	224.0
ZIP Code	5000.0	93152.503000	2121.852197	9307.0	91911.00	93437.0	94608.00	96651.0
Family	5000.0	2.396400	1.147663	1.0	1.00	2.0	3.00	4.0
CCAvg	5000.0	1.937913	1.747666	0.0	0.70	1.5	2.50	10.0
Education	5000.0	1.881000	0.839869	1.0	1.00	2.0	3.00	3.0
Mortgage	5000.0	56.498800	101.713802	0.0	0.00	0.0	101.00	635.0
Personal Loan	5000.0	0.096000	0.294621	0.0	0.00	0.0	0.00	1.0
Securities Account	5000.0	0.104400	0.305809	0.0	0.00	0.0	0.00	1.0
CD Account	5000.0	0.060400	0.238250	0.0	0.00	0.0	0.00	1.0
Online	5000.0	0.596800	0.490589	0.0	0.00	1.0	1.00	1.0
CreditCard	5000.0	0.294000	0.455637	0.0	0.00	0.0	1.00	1.0

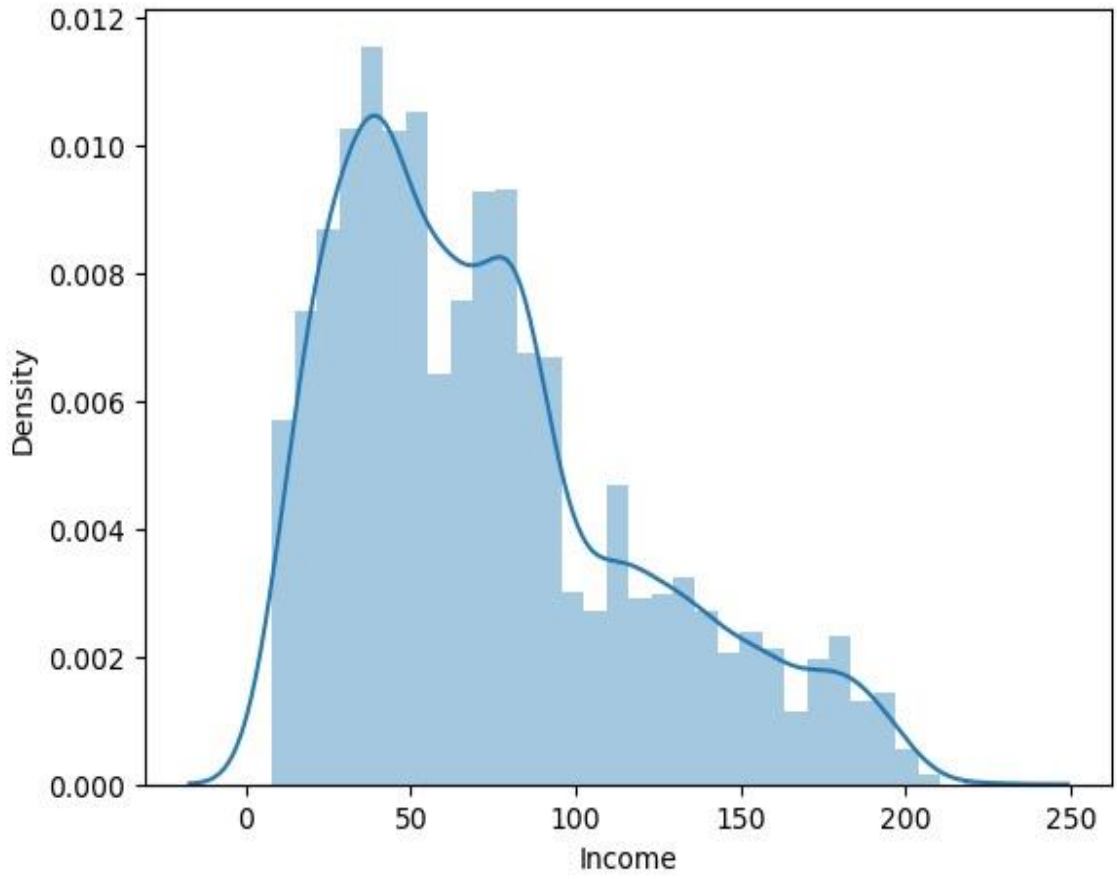
4.2 Data Processing: Data processing is the method used to clean and process the raw data in a manner that it can be understood by ordinary people. It starts from various processes like:
It starts from various processes like:

Invisible Analysis: The key thing in that process is invisible analysis which is characterized by filling the information and context in the details. Missing data; It may be because of many reasons for instance during data collection, when people are asked fewer questions or wrong information is entered during the source of the information or the computer system may crash during data processing, which really affects its validity and reliability.

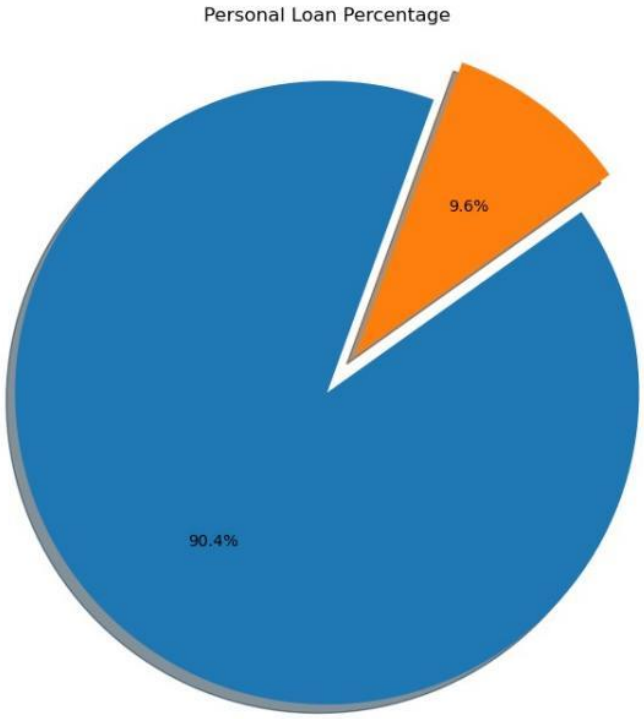
	column	dtypes	nunique	sum_null
0	ID	int64	5000	0
1	Age	int64	45	0
2	Experience	int64	47	0
3	Income	int64	162	0
4	ZIP Code	int64	467	0
5	Family	int64	4	0
6	CCAvg	float64	108	0
7	Education	int64	3	0
8	Mortgage	int64	347	0
9	Personal Loan	int64	2	0
10	Securities Account	int64	2	0
11	CD Account	int64	2	0
12	Online	int64	2	0
13	CreditCard	int64	2	0

The above table represents the missing observation analysis.

- ✧ Outlier Observation analysis: We use the Interquartile range (IQR) to find the outliers in the dataset.



4.3 Univariate Analysis: We see from the pie chart that 9.6% of all the applicants get approved for personal loan.



4.4 Define functions: We introduce the following to set the ML functions which we shall later on use to dissect the outcomes in our model.

Correlation Matrix: The matrix is in the form of rows, also called correlation matrix. It tells how the correlations are formed between some variables. Correlation is a mathematical notion that gives the value and sign of the connection between two variables. The difference factor you are getting is from 1 to 1.

Confusion Matrix: The confusion matrix is a table that widely utilized in the evaluation of machine learning models and distributing the classification activities. It enables you to substantialize the algorithm in an appropriate way by comparing the actual group you defined with the one provided as a reference group. The confusion matrix uses the four main combinations of real and expected classes you commonly see.

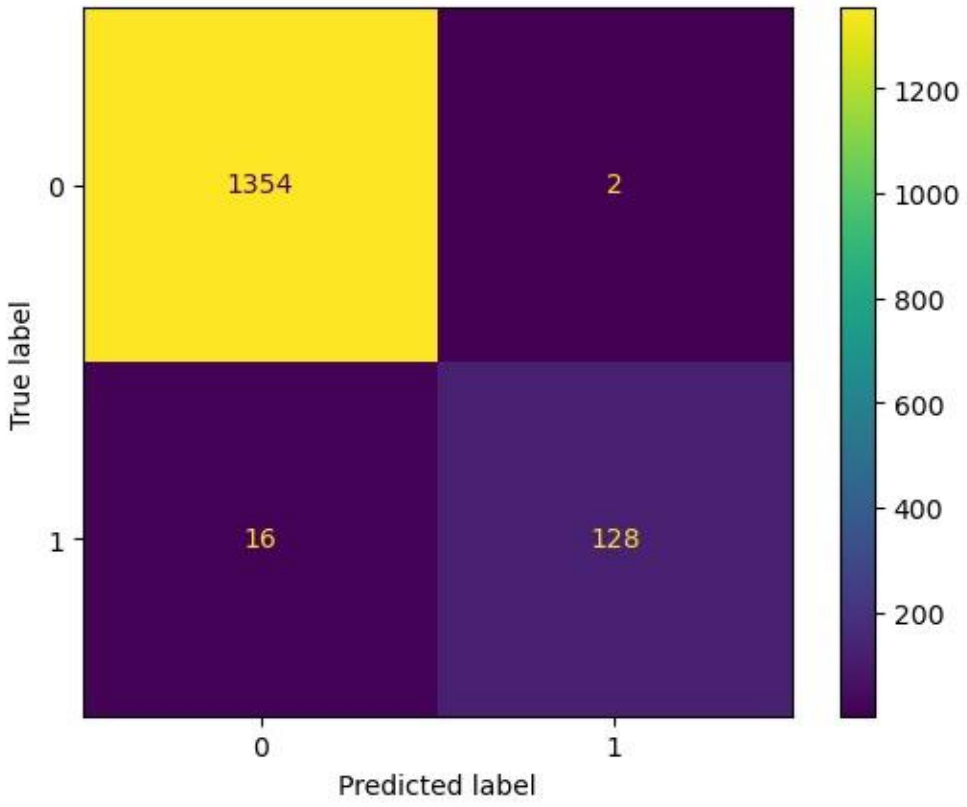
Precision-Recall Curve: It is important to note that Precision-Recall Curve which depicts relationship of precision and recall of distribution model in the binary case is vital for performance assessment as well as annotation task especially when there is no classifier's advantage in a classroom for a given answer.

ROC Score: Besides that, the ROC score also called ROC AUC is a performance measure that is quite practiced to be employed as a base for drawing conclusions about accuracy of the predictive model. This means that the ROC is the graph TPR on the positive axis and FPR on the negative axis plot to show the relationships that exist between TPR and FPR under the specified conditions. On the following block we will compute the area that forms the ROCS curve.

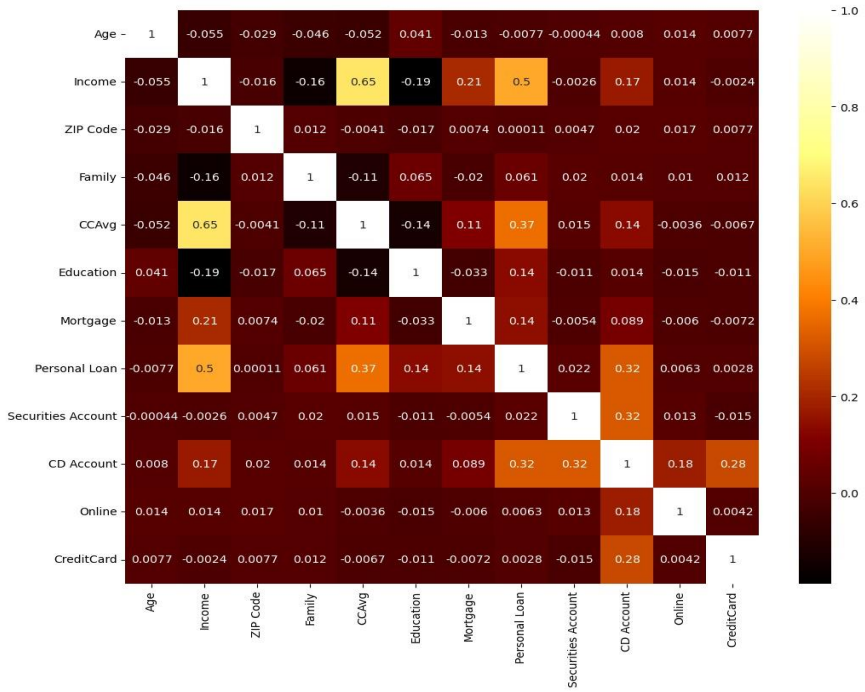
F1 Score: For instance, at the same time, F1 score is one of the predominant ways of summarizing the performance of classification model, produced of various numerical tools, linked with diverse approaches. That is, there is in these two cases combined the measurement of two parameters of an affine transformation with one parameter to give a metric.

The formula for calculating the F1 score is as follows:These are the mathematical steps for calculating F1-score:

However, the F2 index involves double-weighing the precision and recall by multiplying them: F2= (2 * precision * recall)/ (precision + recall).



The above figure represents the confusion matrix of the random forest classifier with a true positive value of 1354.



The above figure represents the correlation matrix of between the different attributes after classification.

Chapter 5
Result Analysis

5.1 Classification results:

Data analysts may use binary classification problems to identify which of the two classes an idea tends to belong. The binary classification problem has four possibilities: T&P = true & positive ; F&P = false & positive ; F&N = false & negative ; F&N = false & negative.

True Positive (TP):The positive class is the target label, which helps the model to learn the various characteristics of the data points that are considered as the target class.

False Positive (FP):The model performs badly as there is an image that triggers to the wrong class.

Negative Instances (TN): The establishing of class belonging to the negatives.

False Negatives (FN): This writing sample failed to predict the unethical behavior that would happen in the experiment.

as one of the key segments of the positive sample.

Evaluation of the four outcomes of the model performance by the binary classification model should take place. Indicators like these direct the model to generate the desired performance metrics which are accuracy, precision, recall, using the model to measure the success of the classification.See performance appraisal standards.

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}}$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}}$$

$$\text{FI score} = 2 \cdot \frac{\text{Precision} - \text{Recall}}{\text{Precision} + \text{Recall}}$$

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}}$$

ROC curve can be a useful tool to graph the performances of binary classification Random Forest machine model and it also can classify distribution. The efficacy was 0.98 for this classification model, it was evident that the model was capable of sorting 98.67% of the data. However, in general, the data may not be able to express the main damage, the data which conflict with each other, and where positive values give different scores than negative values. It is expected that to enable visualize how model´s performance based with an ROC curve, the plot the TPR against FPR as displayed below: Pattery is known not to walk during certain time of the day, as it occurs at intervals..

5.2 Project Result:

The below table shows the accuracy rate ,precision ,recall and f1 score of various machine models for the testing data

Models	Accuracy	Precision	Recall	F1 Score
Logistic Regression	0.89	0.47	0.25	0.33
KNN	0.90	0.40	0.01	0.02
Naive Bayes	0.89	0.48	0.61	0.54
Decision Tree	0.97	0.90	0.80	0.85
Random Forest	0.98	0.98	0.87	0.92
SVM	0.89	0.47	0.39	0.43

Chapter 6

Conclusion and Future Scope

6.1 Conclusion

Modeling complex personal bank reputations in python represents a critical aspect of future smart customer-bank communication. These concepts help the loan approval system of the banks in which with the passage of time may revolutionize the same. The program brings higher accuracy when it identifies the information asked on the credit forms in order to marinate the creditworthiness. Due to this banks' fees for originating new loans have diminished thereby facilitating the process of favorable borrowers screening and these processes have been favourable to the banks'. Pieces of the loaning process, such treasury and data analysis stages, could be completed quicker, which then frees up time and resources for the banks. Model still can deal with the real life use of loans and the nature of borrowers. With all these choices banks can tailor different loan products to different sectors of the market and thus be able to increase loan demand and profits. The model assumes a severe role in contributing to risk management process. Banks, as potential lenders, can appraise the risks and select the loan amount, and can put in place measures to reduce the risk.

6.2 Future Scope

Machine learning algorithm classifiers like Random Forest, SVM, Naive Bayes, Decision Trees, Logistic Regression and KNN can be invaluable cognitive aid for developing the bank's Personal library model with the help of AI in Python. The complete online information include a person's monthly income, credit history, debt-to-income ration, and other required information. Such data will be made available to the lender using statistics like a logistic regression to prove the loan repayment is possible. The situation of the loan provided creates a more precise view of what is feasible regarding their prospects. Thus improving the time efficiency, productivity, profitability and inequity of the personal Bank Credit product, it is possible that the Python Programming Modeling Project in Bank Personal Loan can become the key factor for the development of a better and is fair future of personal credit credit. While these models tend to be effective due to their precision, it is in fact the things we do with models that matter the most. Deciding whether it is right or wrong to use individual information in making the decision and selecting a fair and responsible kind of lending are challenges that need to be overcome to make sure that lending is transparent and accurate.

References

[1] PhilHyo Jin Do ,Ho-Jin Choi, "Sentiment analysis of real-life situations using loca- tion, people and time as contextual features," International Conference on Big Data and Smart Computing (BIGCOMP), pp. 39–42. IEEE, 2015.

[2]Sheikh, Mohammad Ahmad, et al. "An Approach for Prediction of Loan Approval Using Machine Learning Algorithm." 2020 International Conference on Electronics and Sustainable Communication Systems (ICESC), 2020, pp. 490–94. IEEE Xplore, https://doi.org/10.1109/ICESC48915.2020.9155614.

[3] J. Tejaswini, T. M. Kavya, R. D. N. Ramya, P. S. Triveni, and V. R. Maddumala, "Accurate loan approval prediction based on machine learning approach," Journal of Engineering Science, vol. 11, no. 4, pp. 523–532, 2020.

[4] Y. Wang, Y. Zhang, Y. Lu, and X. Yu, "A comparative assessment of credit risk model based on machine learning ——a case study of bank loan data," Procedia Computer Science, vol. 174, pp. 141 – 149,2020. 2019 International Conference on Identification, Information and Knowledge in the Internet of Things.

[5] A. Vaidya, "Predictive and probabilistic approach using logistic regression: Application to prediction of loan approval," in 2017 8th International Conference on Computing, Communication and Networking Technologies (ICCCNT), pp. 1–6, IEEE, 2017.

[6] Yokoyama, Shoichi, and Haruko Sanada. "Logistic regression model for predicting language change." Studies in quantitative linguistics 5 (2009): 176-192.

[7] M. Karim and R. Rahman, "Decision Tree and Naïve Bayes Algorithm for Classification and Generation of Actionable Knowledge for Direct Marketing," Journal of Software Engineering and Applications, Vol. 6 No. 4, 2013, pp. 196-206. doi: 10.4236/jsea.2013.64025.

[8] Maalouf, M. (2011). Logistic regression in data analysis: An overview.International Journal of Data Analysis Techniques and Strategies, 3(3),281–299. https://doi.org/10.1504/IJDATS.2011.0413 35

[9] Chandra Blessie, E., Rekha, R. (2019). Exploring the machine learning algorithm for prediction the loan sanctioning process. International Journal of Innovative Technology and Exploring Engineering, 9(1), 2714–2719. https://doi.org/10.35940/ijitee.A4881.1191 19

[10] Patel, H. H., Prajapati, P. (2018). Study and Analysis of Decision Tree Based Classification Algorithms. International Journal of Computer Sciences and Engineering, 6(10), 74–78.
https://doi.org/10.26438/ijcse/v6i10.7478

[11] Arun, K., Ishan, G., & Sanmeet, K. (2016). Loan approval prediction based on machine learning approach. IOSR J. Comput. Eng, 18(3),1821.

[12]. Boser, B.E., I.M. Guyon, and V.N. Vapnik, A training algorithm for optimal margin classifiers, in Proceedings of the fifth annual workshop on Computational learning theory, 1992. Association for Computing Machinery: Pittsburgh, Pennsylvania, USA: p. 144–152.

[13]Carbo-Valverde, S., Cuadros-Solas, P., & Rodríguez-Fernández, F. (2020). A machine learning approach to the digitalization of bank customers: Evidence from random and causal forests. Plos one, 15(10), e0240362.

[15]Panda, Amiya Ranjan, et al. "Ad Click-Through Rate Prediction: A Comparative Study of Machine Learning Models." 2024 International Conference on Emerging Systems and Intelligent Computing (ESIC), 2024, pp. 679–84. IEEE Xplore,

[15]Panda, Amiya Ranjan, et al. "Ad Click-Through Rate Prediction: A Comparative Study of Machine Learning Models." 2024 International Conference on Emerging Systems and Intelligent Computing (ESIC), 2024, pp. 679–84. IEEE Xplore,

SAMPLE INDIVIDUAL CONTRIBUTION REPORT:

<BANK PERSONAL LOAN MODELLING>

ABHISHEK KUMAR SINGH
2128061

Abstract: This paper targets developing a prediction model of personal loan purchase utilization of various machine learning approaches including KNN, logistic regression, naive Bayes, decision tree, random forest, and SVM. Assure safety by the process of testing and reporting, where the accuracy is the most fundamental aspect. Among these algorithms, Random Forest has the highest accuracy of 98.67%, while SVM ranks the lowest with an accuracy of 89.80%. Data analysis showed that random forest is the best method for credit card eligibility prediction because it performed excellent in predicting the number of people getting an approved loan application.

Individual contribution and findings: To conduct analysis on data processed in machine learning and data, I rely on the Python libraries such as pandas, matplotlib, and numpy among others and then use the precision, recall or F1 score for evaluation, by choosing the model, precision score, recall score or F1 score. Recall, the rate at which all truly relevant items from the dataset are retrieved, is greater than the precision, or the rate at which asked items are identified correctly.

Individual contribution to project report preparation: In the report i have worked on implemeation and result analysis of the topic bank personal loan modelling.

Individual contribution for project presentation and demonstration: In project presentation i have worked on making introduction and role of machine learning in bank personal loan modelling.

Full Signature of Supervisor: Full signature of the student

SHUBHAM BOSE
2128060

Abstract: This paper targets developing a prediction model of personal loan purchase utilization of various machine learning approaches including KNN, logistic regression, naive Bayes, decision tree, random forest, and SVM. Assure safety by the process of testing and reporting, where the accuracy is the most fundamental aspect. Among these algorithms, Random Forest has the highest accuracy of 98.67%, while SVM ranks the lowest with an accuracy of 89.80%. Data analysis showed that random forest is the best method for credit card eligibility prediction because it performed excellent in predicting the number of people getting an approved loan application.

Individual contribution to project report preparation: In project report i had worked on introduction and baic concept/literature review in bank personal loan modelling.

Individual contribution for project presentation and demonstration: In project presentation have worked on basic concepts and methodolgy(data preprocessing,data visulisation,defining function) in project.

Full Signature of Supervisor: Full signature of the student

SANTANU GIRI
2128048

Abstract: This paper targets developing a prediction model of personal loan purchase utilization of various machine learning approaches including KNN, logistic regression, naive Bayes, decision tree, random forest, and SVM. Assure safety by the process of testing and reporting, where the accuracy is the most fundamental aspect. Among these algorithms, Random Forest has the highest accuracy of 98.67%, while SVM ranks the lowest with an accuracy of 89.80%. Data analysis showed that random forest is the best method for credit card eligibility prediction because it performed excellent in predicting the number of people getting an approved loan application.

Individual contribution to project report preparation: In project report i had worked on Problem Statement / Requirement Specifications in bank personal loan modelling.

Individual contribution for project presentation and demonstration: In project presentaion i had worked on result analysis in the project.

Full Signature of Supervisor: Full signature of the student

SK ASIF HOSSAIN
2128089

Abstract: This paper targets developing a prediction model of personal loan purchase utilization of various machine learning approaches including KNN, logistic regression, naive Bayes, decision tree, random forest, and SVM. Assure safety by the process of testing and reporting, where the accuracy is the most fundamental aspect. Among these algorithms, Random Forest has the highest accuracy of 98.67%, while SVM ranks the lowest with an accuracy of 89.80%. Data analysis showed that random forest is the best method for credit card eligibility prediction because it performed excellent in predicting the number of people getting an approved loan application.

Individual contribution to project report preparation: In project report in had worked on conslusion and future scope in the project.

Individual contribution for project presentation and demonstration: In project i have worked on conclusion and future scope in project.

Full Signature of Supervisor: Full signature of the student

PLAGIARISM REPORT

BANK PERSONAL LOAN MODELLING

ORIGINALITY REPORT

19%	21%	11%	15%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	Submitted to KIIT University Student Paper	3%
2	innovationarabia.ae Internet Source	2%
3	Mohammad J. Hamayel, Mobarak A. Abu Mohsen, Mohammed Moreb. "Improvement of personal loans granting methods in banks using machine learning methods and approaches in Palestine", 2021 International Conference on Information Technology (ICIT), 2021 Publication	2%
4	Submitted to Banaras Hindu University Student Paper	1%
5	www.coursehero.com Internet Source	1%
6	www.worldleadershipacademy.live Internet Source	1%
7	scholarworks.rit.edu Internet Source	1%