Innovations

Study of Loan Prediction using Various Machine Learning Models

Monika Kumari[1], Khyati Kocchar[2], Ajay Kumar Yadav[3]

Banasthali Vidyapith, Department of Commerce Banasthali Vidyapith, Rajasthan[1, 2], India

monika_3107@outlook.com, khyatikochhar12@gmail.com

Banasthali Vidyapith, Department of Computer Science[3], Banasthali Vidyapith, Rajasthan, India

ajay.iitdhn@gmail.com

Abstract: *As you can see, it commonly happens that certain borrowers are unable to make loan repayments on time each year. As a result, we have proposed methodology aims to identify the appropriate individual for loan allocation. "Machine Learning Predictive Models" can be used to overcome the existing issues. It solves the problem by analyzing the collected data set and training it using one of the monitored "Machine Learning Algorithms, such as Logistic Regression Algorithm, Random Forest Classifier, Decision Tree Classifier".*

Keywords: *Machine learning; Logistic Regression; Random Forest Classifier; Decision Tree Classifier; Loan; Financial Institution; Capital.*

1. Introduction

The majority of financial organizations today earn from loan sanctions. Even while many banks and financial institutions now grant loans after verifying the repayment process and vetting the applicant, the applicant's deservingness is not yet guaranteed. Because the risk of a loan is substantial and challenging to assess. Credit risk has been calculated using a variety of methods. Additionally, one of the main responsibilities of financial organizations is managing credit risk [1] [2] [3]. Banks are unable to distinguish between applicants who are eligible and those who are not due to the vast amount of data. Given that this problem falls under the category of a classification problem, there is a method called "Machine Learning Techniques" [4] that may be used to answer it with ease.

The algorithms for categorizing data are created with certain likelihood in mind [5]. To date, the most often used reference classifier is a research by [6]. Only a few studies on the most recent developments in predictive learning have been discussed, it has been noted in the literature. The main "Machine Learning" trend right now [4] is to create different algorithms, then enhance weights, employ exploratory search, and create selected multi-classification systems. Artificial intelligence known as "Machine Learning (ML)" creates software programs that more accurately predict events without explicit programming. Machine learning uses past data as an input to forecast future output values. Machine learning is essential because it enables businesses to understand patterns in consumer behavior and business model trends, which in turn supports the creation of new goods.

Today, machine learning is a key component of the operations of many top firms, including "financial institutions, Facebook, Google, and Uber." Machine learning classifiers are used to handle several businesses and other kinds of applications [7]. According to the study [8] "Artificial Intelligence" (AI) and "Machine Learning (ML)" are crucial for the live music industry's long-term growth. Applications of machine learning are crucial in the field of medicine, since many diseases may be anticipated [9]. Many business applications have used "Artificial Intelligence Technology" [10]. The significance of machine learning technology in predictive applications like loan prediction and score prediction has recently been noticed by various financial institutions [11].

Machine learning models have been found to be an essential tool for developing predictive models. According to certain academic research, predictive loan models like "Logistic Regression Classifiers, Random Forest Classifiers, and Decision Tree Classifiers" can be used to determine whether a specific applicant would be approved for a loan or not. The entire functional verification is automated by doing this.

1.1 Motivation

Our objective is to consider a well-defined forecasting model that will help financial institutions reach their objectives in light of the aforementioned problems. Making progress on this important topic is simple with the use of machine learning tools. Technology that can lessen the strain on humans can take the place of people.

1.2 Organization of the paper

The rest part of the article is formatted as: The relevant research about loan prediction models is described in Section 2. Details explanation of the proposed model has been comprised in section 3. The experimental design and outcomes have been explained in sections 4 and 5 respectively. Section 6 concludes with the inference and further work.

2. Background and literature review

Many researches have been carried for loan prediction in the financial and banking sectors. This section briefly presents some of the technology used in the loans prediction model and their findings. The machine learning technology help you learn data based on your own experience predicting the data and making decisions. It brought a revolution in the area of computing as it (information/data) is a very important thing in the world. To analyze the data, machine learning provides several solutions through its algorithms. We have focused to minimize the workload of the bank by providing a model through several machine learning techniques and has been discussed which of the technique can be accurate.

Much research has been done recently using "Machine Learning Technology" in credit modeling emergences several methods of calculating default probabilities, basically: "Support Vector Machine" [15] [16] "Decision Tree" [17] "Random Forest" [18] , and Bagging and Boosting [19] . Most studies emphasize the benefits of using "Machine Learning " systems in credit distress explication. This is because machine learning systems provide better evaluation outcomes than stuffy methods as "logistic regression" [20] - [22].

Wang et al. [23] believed that the DT method scored relatively lower in credit analysis than other methods due to data detonation and overlapping attributes. Therefore, in order to minimize this effect and achieve relatively high classification accuracy, a DT with an ensemble classifier is proposed.
Wang et al. [24] collated the outcomes of three ensemble classifiers “BG, Boosting, and Stacking” based on four algorithms as "LR, DT, Artificial Neural Networks (ANN) and SVM". Marquez et al. [25]. measured the outcomes of seven predictive models using an ensemble classifier and found that “DT (C4.5)” is the best solution for ensemble classifiers, followed by “Neural Networks and LR”. On the other hand, the classifiers "Nearest Neighbor Method" and "Naive Bayes" are even worse.
Conde et al.[26] similituded various “Machine Learning” methods such as "DT, Neural Networks, and Statistical Classifiers" and reached crucial inference: (1) No techniques appear to be completely superior to other methods; (2) The “Machine Learning Technique” seems to be more suitable for multimodal distribution; and (3) Statistical techniques are more effective in the calculation. Barboza et al. [19] used a sample from North American companies and studied the use of the ML algorithm and compared it with stuffy statistical methods as ‘Linear Regression”.
Feng et al. [27] also examined some other techniques in the credit databases and reinforce the use of “Machine Learning” for the imbalanced data found in the database.
Most of the studies only pointed out on credit cards or personal loan businesses. Some studies have analyzed housing loans [28]. Although mortgages are one of the most important asset classes on the balance sheets of some banks, at least in Brazil, default probability models are usually used on a short-term scale and use a limited number of observations in a data set in an academic environment.
Kaarthik, Dharanidharan, Navalarasu, & Sabarinathan [29] proposed a loan sanctions prediction model based on the NB method associated with “K-Nearest Neighbor (KNN)” and “binning” technique. The seven parameters have been taken into account as they are “income, age, occupation, and existing loans and their maturity, amount, and approval status”. The threads include preprocessing (using KNN to process the missing values and using clustering algorithms for data refinement), using NB methods for classification, and frequent updating the data set will lead to appropriate improvements in the loan prediction process. The inference of the experiment is that combining” KNN” and binning algorithm with NB can better predict the loan sanction process.
Goyal, & Kaur [30] proposed a loan prediction technique using multiple “Machine Learning (ML)” techniques. A data set with characteristics namely “gender, marital status, education, number of dependents, employment status, income, applicant's income, loan amount, loan period, credit history, current loan status, and property area”, is used to determine loan eligibility for the approval process. Several “Machine Learning Techniques” have been used in this method includes: “Inear Model, Decision Tree (DT), Neural Network (NN), Random Forest (RF), SVM, Extreme Machine Learning, Model Tree, Multiple Adaptive Regression Spline, Bagged Cart Model, NB, and TGA”. When evaluating these models using the R environment in five executions, TGA gives better loan prediction outcomes in comparison to other techniques.

2.1 Research gap

As we have discussed, the problem occurs in finding the right applicant for a loan. To grant the loan to the right individuals, a lot of manual works are required, including paper work, verifications, signatures, document requirements, etc. We came up with a plan to reduce these costs by applying "Machine Learning" algorithms to anticipate loan defaulters and make it simpler for bankers to select qualified loan.

2.2 Problem identification

Financial institutions occasionally struggle to choose the best applicant for loan approval. When deciding whether to lend to applicants, a financial institution takes into account a number of application-related factors. One applicant's attribute cannot be compared to another applicant's attribute or to another applicant's attribute. The work is really tiresome. We must take into account a number of application characteristics at once while choosing the best applicant. However, there are times when a single characteristic of an applicant is sufficient to evaluate their profile. For instance, if an applicant has a higher monthly salary and needs a smaller loan, the bank will undoubtedly offer them a loan. However, the model might not do so, as would happen, for instance, if the applicant is female, which would increase the likelihood that the loan will be declined. In order to accomplish this, a method known as "Machine Learning" approaches are used, which takes into account all of the applicant's characteristics, compares their data to that of other candidates, and then makes accurate predictions.

2.3 Objectives

The main goal of the suggested approach is for various financial institutions to lend money to individuals. Determination of a borrower's ability to pay back a loan can be exceedingly challenging for lenders. Any bank's financial expansion is possible through loan approval. In general, banks only approve loans after putting them through a number of verification steps; however, it is not always assured that the borrower will be able to pay back the loan. We have suggested a machine learning predictive model that can determine whether or not a loan has been issued to the proper applicant in order to solve this conundrum. The suggested approach can be integrated with the financial organizations and apply loan sanctions to the right candidates.

3. Proposed Model

To solve the existing problems, a loan prediction model has been offered as a solution to the current issues. The applicant who wants to apply for a loan must fill out an "Attribute Relation File Format (ARFF)" form with their personal information. The proposed "Machine Learning (ML)" model that forecasts the qualified loan application uses all of this data. Figure 1 depicts the prediction model.

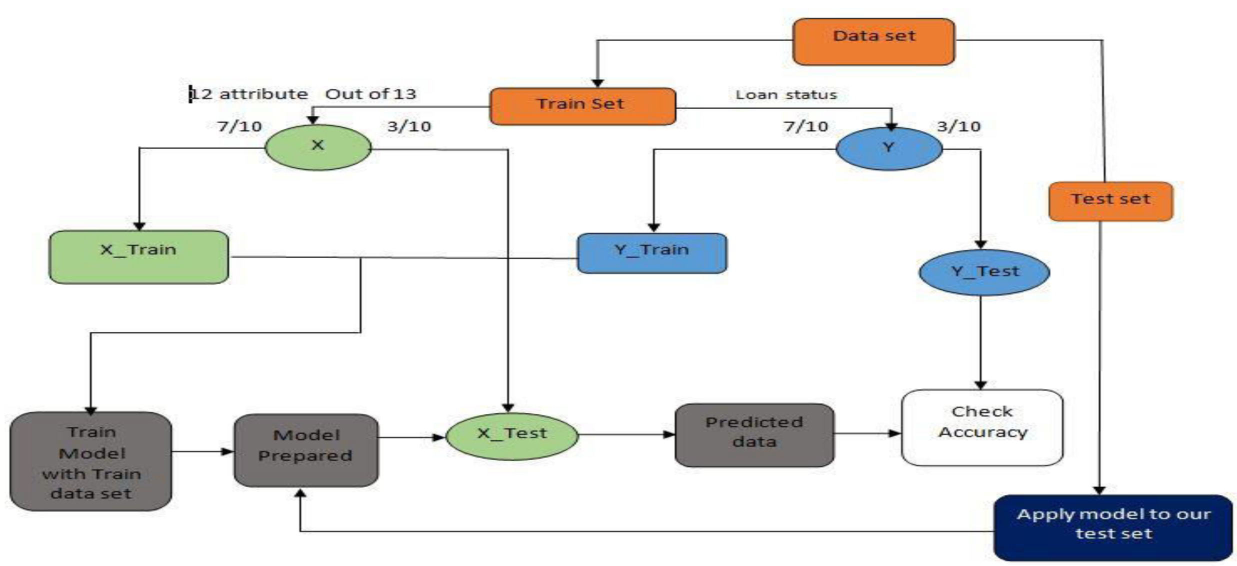


Figure1. Loan prediction model
Source(s): Figure by authors

3.1 Dataset

To train and test the prediction model, the old data set must be gathered. Data was gathered from the well-known "Kaggle" source. "A useful solution for splitting the data set into training and testing portions is a robust and accurate classifier precondition [12]. If there are more data instances, it has been suggested in the literature that they should be separated into portions for training and testing that are 70% and 30%, respectively. The training and testing portions of the data in this study were split 70-30%.

The "Machines Learning" model is now supplied the training dataset, and the model is trained using this dataset [13] [14]. Each new applicant filled out their information while applying, and the system uses it as a test dataset. Following the test, the model infers a conclusion from the training dataset to determine whether the new applicant is qualified for loan sanction. Take into account some of the applicant's characteristics, as listed in the Table.

Table 1: Data set Description **Source(s):** Table by authors

S. No	Attribute Name	Description	Data Type
1.	Loan_id	Unique Loan_id	Integer
2.	Gender	Male or Female	Character
3.	Marriage Status	Yes or No	Character
4.	Dependents	Number of Dependents	Integer
5.	Applicants Education	Graduate or Under Graduate	String
6.	Self –employed	yes or no	character
7.	Applicants Income	Applicants income	integer
8.	Loan Amount	Loan amount in thousands	integer
9	Credit History	Previous history of customers	integer
10.	Property Area	Urban/Semi Arban/Rural	string
11.	Loan amount term	Term of loan in month	integer
12.	Co-applicant Income	Co-applicant Income	integer
13.	Loan_Status	Loan Approved(Y/N)	string

The details will get added into database for further processing and get loan status as shown in Figure 2.

	A	B	C	D	E	F	G	H	I	J	K	L	M
1	Loan_ID	Gender	Married	Dependents	Education	Self_Employ	ApplicantInc	Coapplicant	LoanAmoun	Loan_Amou	Credit_Histc	Property_Ar	Loan_Status
2	LP001002	Male	No	0	Graduate	No	5849	0		360	1	Urban	Y
3	LP001003	Male	Yes	1	Graduate	No	4583	1508	128	360	1	Rural	N
4	LP001005	Male	Yes	0	Graduate	Yes	3000	0	66	360	1	Urban	Y
5	LP001006	Male	Yes	0	Not Gradua	No	2583	2358	120	360	1	Urban	Y
6	LP001008	Male	No	0	Graduate	No	6000	0	141	360	1	Urban	Y
7	LP001011	Male	Yes	2	Graduate	Yes	5417	4196	267	360	1	Urban	Y
8	LP001013	Male	Yes	0	Not Gradua	No	2333	1516	95	360	1	Urban	Y
9	LP001014	Male	Yes	3+	Graduate	No	3036	2504	158	360	0	Semiurban	N
10	LP001018	Male	Yes	2	Graduate	No	4006	1526	168	360	1	Urban	Y
11	LP001020	Male	Yes	1	Graduate	No	12841	10968	349	360	1	Semiurban	N
12	LP001024	Male	Yes	2	Graduate	No	3200	700	70	360	1	Urban	Y
13	LP001027	Male	Yes	2	Graduate		2500	1840	109	360	1	Urban	Y
14	LP001028	Male	Yes	2	Graduate	No	3073	8106	200	360	1	Urban	Y
15	LP001029	Male	No	0	Graduate	No	1853	2840	114	360	1	Rural	N
16	LP001030	Male	Yes	2	Graduate	No	1299	1086	17	120	1	Urban	Y
17	LP001032	Male	No	0	Graduate	No	4950	0	125	360	1	Urban	Y
18	LP001034	Male	No	1	Not Gradua	No	3596	0	100	240		Urban	Y
19	LP001036	Female	No	0	Graduate	No	3510	0	76	360	0	Urban	N
20	LP001038	Male	Yes	0	Not Gradua	No	4887	0	133	360	1	Rural	N
21	LP001041	Male	Yes	0	Graduate		2600	3500	115		1	Urban	Y
22	LP001043	Male	Yes	0	Not Gradua	No	7660	0	104	360	0	Urban	N
23	LP001046	Male	Yes	1	Graduate	No	5955	5625	315	360	1	Urban	Y
24	LP001047	Male	Yes	0	Not Gradua	No	2600	1911	116	360	0	Semiurban	N
25	LP001050		Yes	2	Not Gradua	No	3365	1917	112	360	0	Rural	N
26	LP001052	Male	Yes	1	Graduate		3717	2925	151	360		Semiurban	N
27	LP001066	Male	Yes	0	Graduate	Yes	9560	0	191	360	1	Semiurban	Y
28	LP001068	Male	Yes	0	Graduate	No	2799	2253	122	360	1	Semiurban	Y
29	LP001073	Male	Yes	2	Not Gradua	No	4226	1040	110	360	1	Urban	Y
30	LP001086	Male	No	0	Not Gradua	No	1442	0	35	360	1	Urban	N

Figure2. Details of database

3.2 Prediction Model

In this part, a "Machine Learning Predictive Model" was discussed, which aids bankers in determining who is and is not qualified to use the lending facility. The proposed model had to be trained and tested using a dataset. Kaggale Depositary collected the data, which was stored in an Attribute Relation File Format (ARFF) with 13 attributes of size 981 bytes.

Here, the datasets are split into two parts. The first part is the training dataset and the second is the test dataset. Additionally, the training dataset is split into two sets. For instance, X and Y have 12 and 1 characteristic, respectively. The X dataset is once more split into two sections, X-train and X-test, with data proportions of 70% and 30%, respectively, to train the prediction model. The Y dataset is likewise divided in a similar way. Figure 2 makes it obvious that X-train and Y-train datasets are provided to train the model that will produce the model. The X-test dataset itself has been provided for testing the created model, and the accuracy rate was compared to the Y-test.

"Logistic Regression Classifier, Decision Tree Classifier, and Random Forest Classifier" were used to test the suggested prediction model. The "Logistic Regression Algorithm" outperforms all of these classifiers, with 79.9% accuracy, as seen in Fig. 3.

```
kf= StratifieldKFold( n_splits=5, random_state=1)
for train_index, test_index  in kf.split(X, Y) :
        print ( '\n { } of kfold { }' .format ( i , kf.n_splits) )
        xtr, xvl = x.loc[train_index], x. loc[ test_index]
        ytr, yvl = y[train_index], y[ test_index]
      l_model= LogisticRegression( random_state=1)
      l_model.fit( xtr, ytr)
pred_test= l_model.predict(xvl)
score= accuracy_score (yvl, pred_test)
mean+=score
print (' accuracy_score', score)
pred_l=l_model.predict(test3)
i+=1
print ('\n Mean Validation Accuracy' , mean/ (i-1))
1 of  kfold 5
accuracy_score    0. 8048780487804879
2 of  kfold 5
accuracy_score    0.7642276422764228
3 of  kfold 5
accuracy_score    0. 7804878048780488
4 of  kfold 5
accuracy_score    0.8455284552845529
5 of  kfold 5
accuracy_score    0. 8032786885245902
 Mean Validation Accuracy 0.7996801279488205
```

Figure 3: Logistic Regression using Stratified K Fold

The same procedure was used with the "Decision Tree Classifier" and "Random Forest Classifier," which resulted in mean accuracy rates of 71.0% and 80.4%, respectively, as shown in Figs. 4 and 5.

```
kf= StratifieldKFold( n_splits=S, random_state=1, shuffle=True)
for train_index, test_index  in kf.split(X, Y) :
      print ( '\n { } of kfold { }' .format ( i , kf.n_splits) )
      xtr, xvl = x.loc[train_index], x. loc[ test_index]
      ytr, yvl = y[train_index], y[ test_index]
      d_model= DecisionTree Classifier( random_state=1)
      d_model.fit( xtr, ytr)
    pred_test= d_model.predict(xvl)
    score= accuracy_score(yvl, pred_test)
    mean+=score
    print(' accuracy_score', score)
    i+=1
    pred_d=d_model.predict(test3)
    print ('\n Mean Validation Accuracy' , mean/ (i-1))

    1 of kfold 5
    accuracy_score    0.6991869918699187

    2 of kfold 5
    accuracy_score    0.739837398373987

    3 of kfold 5
    accuracy_score    0.7154471544715447

    4 of kfold 5
    accuracy_score    0.7235772357723578

    5 of kfold 5
    accuracy_score    0.6721311475409836

       Mean Validation Accuracy 0.7100359856057576
```

Figure 4: Decision Tree Classifier using Stratified K Fold

```
kf= StratifieldKFold( n_splits=5, random_state=1, shuffle=True)
for train_index, test_index  in kf.split(X, Y) :
      print ( '\n { } of kfold { }' .format ( i , kf.n_splits) )
      xtr, xvl = x.loc[train_index], x. loc[ test_index]
      ytr, yvl = y[train_index], y[ test_index]
r_model= RandomForest Classifier( random_state=1, max_depth= 1
      r_model.fit( xtr, ytr)
    pred_test= r_model.predict(xvl)
    score= accuracy_score(yvl, pred_test)
    mean+=score
    print(' accuracy_score', score)
    i+=1
    pred_r = r_model.predict(test3)
    print ('\n Mean Validation Accuracy' , mean/ (i-1))

1 of kfold 5
accuracy_score   0.8130081300813008

2 of kfold 5
accuracy_score   0.8455284552845529

3 of kfold 5
accuracy_score   0.7967479674796748

4 of kfold 5
accuracy_score   0.8130081300813008

5 of kfold 5
accuracy_score   0.7540983606557377

 Mean Validation Accuracy 0. 8044782087165135
```

Figure 5: Random Forest Classifier using Stratified K Fold

4. Experimental design

The classification models and filtering methods were carried out using python and the Jupyter notebook soft-ware package. All kinds of calculations have been executed on a such machine configured as "Windows 8 operating system, 8 GB RAM, and an Intel CORE i3" processor. We tested the classification model using Python libraries like "library (caret), library (e1071), and library (C50)".

4.1 Experimental results and Comparison

The proposed model has been tested with three classifiers say "Linear Regression Classifier, Decision Tree Classifier, and Random Forest Classifier". From the experimental outcomes, it has been noticed that the performance of the "Linear Regression" model is outstanding in comparison to the decision tree classifier and "Random Forest Classifier". For example, 80% accuracy.

Let's see the variation of "Loan Status" (target variable) with individual attributes for better understanding.

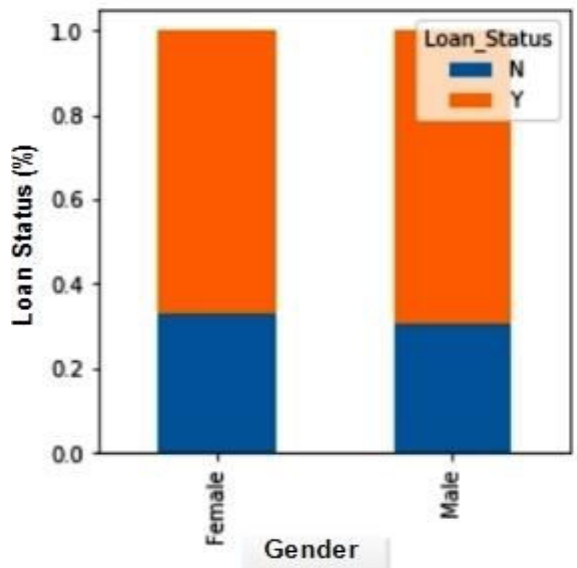
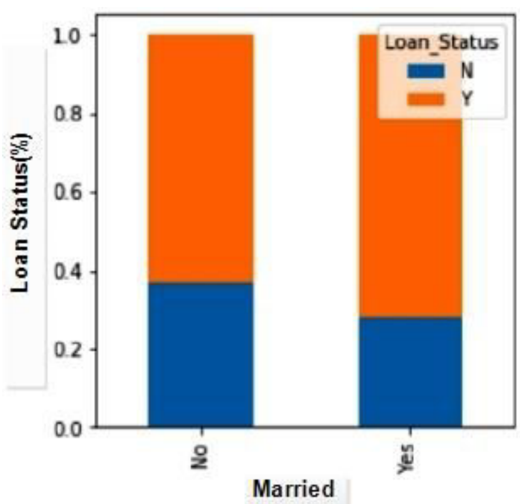


Figure6: Loan Status (%) against Applicants Gender

Figure 6 demonstrates that the proposed model will favor male candidates over female candidates for loan sanctions because it has been presumpted that male candidates are primary applicants and female applicants are co-applicants. The suggested model takes into account a number of additional characteristics of the applicants, such as the highest qualification and income, while allocating the loan amount. Therefore, the preference for the male candidate is marginally higher than for female candidates.



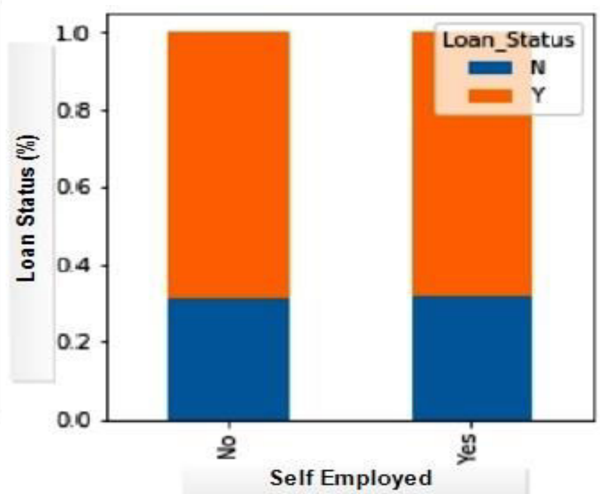


Figure7. Loan Status (%) against married Applicants **Figure8**. Loan Status (%) against Self_Employed

Figure 7 demonstrates that the suggested model gives married applicants a priority over unmarried applicants since it concentrates on other aspects of the applicant at the time of loan approval, such as the number of dependents and their income. Figure 8 show that the suggested model gives all applicants equal weight, regardless of whether they are self-employed or not.

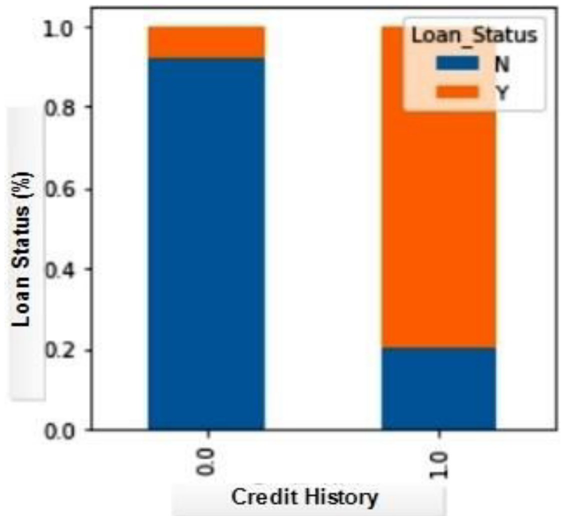


Figure9. Loan Status (%) against Credit History

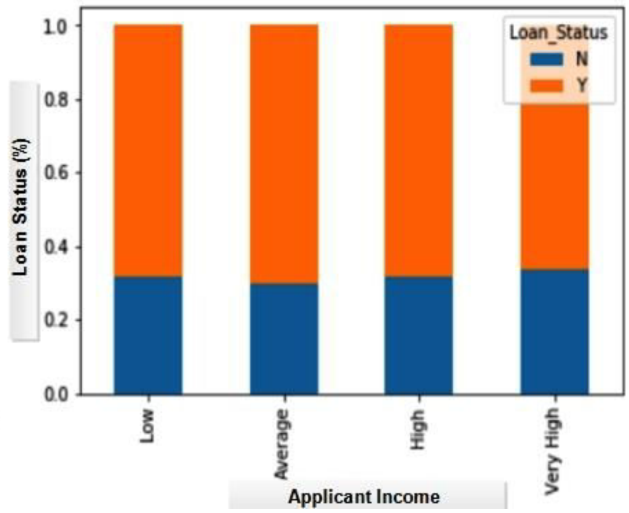


Figure10. Loan Status (%) against Applicant Income

According to Figure 9, the suggested approach will give preference to applicants with strong credit histories because these candidates are more likely to have made their loan payments on time. In a similar way, Fig. 10 demonstrates that the suggested approach favors high-income applicants since they can repay the loan amount on schedule.

5. Conclusion

The proposed model achieved all the requirements of banks for loan sanction. This model can be integrated with several other systems. It can also be used as a bank's decision-making tool to help optimize profit loss due to risk. Model parameters were estimated and their significance tested. With the help of performance metrics, we can conclude that our model is very suitable and also has the correct classification rate. In the near future, the proposed predictive model can be integrated with another automated processing system.

References

1. *Narayana, V. L., Gopi, A. P. & Chaitanya, K,(2019). Avoiding Interoperability and Delay in Healthcare Monitoring System Using Block Chain Technology. Review Intelligence Artificial, 33(1), 45-48.*

2. *Gopi, A. P., Jyothi, R. N. S., Narayana, V. L., & Sandeep, K. S. (2023). Classification of tweets data based on polarity using improved RBF kernel of SVM. International Journal of Information Technology, 15(2), 965-980.*

3. *Arepalli, P. G., Narayana, V. L., Venkatesh, R., & Kumar, N. A. (2019).Certified Node Frequency in Social Network Using Parallel Diffusion Methods. Ingénierie des Systèmes d'Information, 24(1).*

4. *Partalas, I., Tsoumakas, G., & Vlahavas, I. (2010). An ensemble uncertainty aware measure for directed hill climbing ensemble pruning. Machine Learning, 81, 257-282.*

5. *Hand, D.J., & Henley, W. E. (1997). Statistical classification methods in consumer credit scoring: a*

review. Journal of the royal statistical society: series a (statistics in society), 160(3), 523-541.

6. *Baesens, B., Van Gestel, T., Viaene, S., Stepanova, M., Suykens, J., & Vanthienen, J. (2003). Benchmarking state-of-the-art classification algorithms for credit scoring. Journal of the operational research society, 54, 627-635.*
7. *Cubric, M. (2020). Drivers, barriers and social considerations for AI adoption in business and management: A tertiary study. Technology in Society, 62, 101257.*
8. *Naveed, K., Watanabe, C., & Neittaanmäki, P. (2017). Co-evolution between streaming and live music leads a way to the sustainable growth of music industry–Lessons from the US experiences. Technology in Society, 50, 1-19.*
9. *Coccia, M. (2020). Deep learning technology for improving cancer care in society: New directions in cancer imaging driven by artificial intelligence. Technology in Society, 60, 101198.*
10. *Al-Emran, M., Mezhuyev, V., & Kamaludin, A. (2020). Towards a conceptual model for examining the impact of knowledge management factors on mobile learning acceptance. Technology in Society, 61, 101247.*
11. *Wongnaa, C. A., & Babu, S. (2020). Building resilience to shocks of climate change in Ghana's cocoa production and its effect on productivity and incomes. Technology in Society, 62, 101288.*
12. *Patibandla, R. L., Kurra, S. S., Prasad, A., & Veeranjaneyulu, N. (2015). Unstructured Data: Qualitative Analysis. J. of Computation In Biosciences And Engineering, 2(3), 1-4.*
13. *Patibandla, R. L., Kurra, S. S., & Kim, H. J. (2014). Electronic resource management using cloud computing for libraries. International Journal of Applied Engineering Research, 9(23), 18141-18147.*
14. *Shi, J., Zhang, S. Y., & Qiu, L. M. (2013). Credit scoring by feature-weighted support vector machines. Journal of Zhejiang University SCIENCE C, 14(3), 197-204.*
15. *Niklis, D., Doumpos, M., & Zopounidis, C. (2014). Combining market and accounting-based models for credit scoring using a classification scheme based on support vector machines. Applied Mathematics and Computation, 234, 69-81.*
16. *Tsai, C. F., Hsu, Y. F., & Yen, D. C. (2014). A comparative study of classifier ensembles for bankruptcy prediction. Applied Soft Computing, 24, 977-984.*
17. *Kruppa, J., Schwarz, A., Arminger, G., & Ziegler, A. (2013). Consumer credit risk: Individual probability estimates using machine learning. Expert systems with applications, 40(13), 5125-5131.*
18. *Barboza, F., Kimura, H., & Altman, E. (2017). Machine learning models and bankruptcy prediction. Expert Systems with Applications, 83, 405-417.*
19. *Baesens, B., Van Gestel, T., Viaene, S., Stepanova, M., Suykens, J., & Vanthienen, J. (2003). Benchmarking state-of-the-art classification algorithms for credit scoring. Journal of the operational research society, 54, 627-635.*

20. *Tsai, C. F. (2014). Combining cluster analysis with classifier ensembles to predict financial distress. Information Fusion, 16, 46-58*

21. *Chang, Y. C., Chang, K. H., & Wu, G. J. (2018). Application of eXtreme gradient boosting trees in the construction of credit risk assessment models for financial institutions. Applied Soft Computing, 73, 914-920.*

22. *Wang, G., Ma, J., Huang, L., & Xu, K. (2012). Two credit scoring models based on dual strategy ensemble trees. Knowledge-Based Systems, 26, 61-68.*

23. *Wang, G., Hao, J., Ma, J., & Jiang, H. (2011). A comparative assessment of ensemble learning for credit scoring. Expert systems with applications, 38(1), 223-230.*

24. *Marqués, A. I., García, V., & Sánchez, J. S. (2012). Exploring the behaviour of base classifiers in credit scoring ensembles. Expert systems with Applications, 39(11), 10244-10250.*

25. *Conde, P. P., & Carrillo, I. S. (2017, October). Comparison of classifiers based on neural networks and support vector machines. In 2017 5th International Conference in Software Engineering Research and Innovation (CONISOFT) (pp. 107-115). IEEE.*

26. *Feng, X., Xiao, Z., Zhong, B., Qiu, J., & Dong, Y. (2018). Dynamic ensemble classification for credit scoring using soft probability. Applied Soft Computing, 65, 139-151.*

27. *Galindo, J., & Tamayo, P. (2000). Credit risk assessment using statistical and machine learning: basic methodology and risk modeling applications. Computational economics, 15, 107-143.*

28. *Kaarthik, K., Dharanidharan, G., Navalarasu, R. B., & Sabarinathan, G. (2021). Machine Learning Based Loan Prediction System Using Svm and Knn Algorithms. Turkish J. Physiother. Rehabil, 32(2), 3214-3219.*

29. *Goyal, A., & Kaur, R. (2016). A survey on ensemble model for loan prediction. International Journal of Engineering Trends and Applications (IJETA), 3(1), 32-37.*