ISSN NO: 9726-001X

Volume 13 Issue 02 2025

JOURNAL OF CURRENT SCIENCE

# Utilize Machine Learning Approaches to Forecast the Approval of Personal Loans

[1]THOTA HEMA UMA MAHESWARI, [2]VEGESNA VARSHITHA, [3]VALLURI DIVYA TEJA, [4]MOGGA VEERA VENKATA SIVA NARASIMHULU, [5]Dr. B.V. RAM KUMAR

[1234]Student Department of CSE, DNR College of Engineering & Technology, Balusumudi, Bhimavaram, India.

[5] Professor, Department of CSE, DNR College of Engineering & Technology, Balusumudi, Bhimavaram, India.

ABSTRACT

. Although banks make an effort to provide personal loans to dependable borrowers, borrowers do have the option to decline these offers. The prediction of this issue adds more work for banks, but they stand to gain more money if they are successful in determining which clients would accept personal loan offers. Predicting whether or not bank loan proposals will be accepted is, hence, the current focus of this research. This will be accomplished via the use of the Support Vector Machine. Here, support vector machines (SVMs) with four kernels, a grid search technique for improved prediction, and cross validation for much more trustworthy findings were used to forecast outcomes. An accuracy rate of 97.2% was achieved using a poly kernel, whereas an accuracy rate of 83.3% was achieved using a sigmoid kernel, according to the research. Because the dataset is imbalanced, we see lower-than-average accuracy and recall values, such as 0.108 and 0.008, respectively. For every 1 true value, there are 9 negative values, or 9.6% of the total true value. This research suggests that SVC should be used in the banking sector to better anticipate whether customers would accept loan offers. Approval of bank loans Kernel comparison Artificial intelligence Backend vector algorithm A significant portion of a bank's revenue comes from loans.

## Introduction

Lending money is a bank's main activity. Interest on the loan is the primary revenue generator[1, 2]. One the one hand, before approving a loan, financial institutions determine whether the applicant has a history of default or not [3]. Conversely, they do provide personal loans to dependable consumers, but the vast majority of customers, including those in our dataset samples, turn them down [4]. Because of this issue, the banking sector has a significant task: predicting which customers would take the personal loan. A more precise predictive modeling system is needed by the banking sector to address many issues [5]. Bank employees can create such models manually, but it will take a lot of time and effort. When dealing with massive volumes of data, machine learning (ML) methods are becoming quite useful for outcome prediction [5]. This means that the financial sector may benefit from such models by using ML approaches. Following that paradigm, banks may save a ton of time and effort and provide better service to their clients by automating the loan approval process if they can use machine learning to forecast which consumers would accept personal loan offers [6]. The purpose of this research is to identify which customers would accept personal loan offers from banks by using the Support Vector Machine (SVM) technique to solve the classification issue. Boser et al. (1992) published "A Training Algorithm for Optimal Margin Classifiers"[7] which was the first formal usage of support vector machines (SVMs). SVM is an exciting new non-linear, non-parametric classification technique. The combination of non-parametric applied statistics, neural networks, and ML makes it a good fit for binary classification tasks [8]. There is no association between algorithm complexity and sample dimension[9], and SVM's structure offers many computational benefits, such as particular direction at a limited sample. Support vector machines (SVMs) may be useful for bankruptcy analysis when dealing with data that is not normally distributed or has an unclear distribution [8]. SVM techniques are able to resolve a wide variety of optimization problems, including convex problems (such as linear, quadratic, second-order cone, integer, and semi-infinite programming) and more generic and non-convex problems [10]. Utilizing real-life credit card data (245 poor records and 755 good records, with 14 factors) acquired from

ISSN NO: 9726-001X

Volume 13 Issue 02 2025

JOURNAL OF CURRENT SCIENCE

a Chinese commercial bank, Li et al. used SVM to credit evaluation. When it comes to predicting accuracy in the area of credit evaluation, they claim that SVM is superior than the bank's fundamental grade criteria [9]. For the classification challenge, Dall'Asta Rigo used six ML techniques: Support Vector Machines, Linear Regression, Markov Decision Process, Random Forest, XGB, and Stacking, on four real-life credit score datasets: Home Credit, German Credit, Credit Card Default, and Give Me Credit [11]. Xu et al. forecasted borrower repayment variables using four ML methods: RF, XGBT, GBM, and NN. By default, they find that the RF does a good job at classification [12]. When it comes to credit rating evaluations for the US and Taiwan markets, SVM and NN outperform conventional statistical approaches in terms of prediction accuracy, according to Huang et al. [13]. Naïve Bayes (NB) satisfies the demands of bankers, according to Kadam et al., who utilized SVM and NB to forecast loan acceptance [14]. Deep learning techniques, such as Classification Restricted Boltzmann Machines and Multilayer Artificial Neural Networks, were contrasted by Bayraktar et al. with more popular machine learning approaches [15]. To forecast borrowers' creditworthiness, Aphale and Shinde used many ML methods, including Decision Tress, Linear Regression, K-Nearest Neighbor, Discriminant Analysis, Naïve Bayes, and Ensemble Learning [3]. Without exception, every piece of published work on the subject of credit risk management, credit ratings, loan payback, lender decision support, or credit default is devoted to this one central idea. It is the goal of this research, however, to use the Support Vector Machine (SVM) algorithm to forecast whether or not a customer would accept a loan offer from a bank. This study may be the first of its kind to employ support vector machines (SVMs) to forecast whether or not a consumer would accept a bank's loan offer. Since there is currently no published work on the subject, this research will fill a gap in the literature and help advance the loan and banking systems. Table 1. Dataset attributes and statistics details

| Feature Names | Description | Mean | Standard Deviation | Min Value | Max Value | Feature Type |
|---|---|---|---|---|---|---|
| ID | Customer's unique ID number | 2500.50 | 1443.52 | 1 | 5000 | Categorical |
| Age | Customer's age | 45.33 | 11.46 | 23 | 67 | Numeric |
| Experience | Work experience by number of years | 20.10 | 11.46 | -3 | 43 | Numeric |
| Income | 0.1% Percentage of annual income | 73.77 | 46.03 | 8 | 224 | Numeric |
| ZIP Code | ZIP Code of where customer lives | 93152.50 | 2121.85 | 9307 | 96651 | Categorical |
| Family | Family Size | 2.39 | 1.14 | 1 | 4 | Numeric |
| CC Avg | 0.1% Percentage of average credit card spending per month | 1.93 | 1.74 | 0 | 10 | Numeric |
| Education | Level of education (1-Undergraduate, 2-Graduate, 3-Advanced) | 1.88 | 0.83 | 1 | 3 | Categorical |
| Mortgage | If any house mortgage, its value. | 56.49 | 101.71 | 0 | 635 | Numeric |
| Personal Loan | Acceptance of personal loan offer by the customer in the last campaign season | 0.09 | 0.29 | 0 | 1 | Categorical |
| Securities Account | Is there any securities account with the customer? | 0.10 | 0.30 | 0 | 1 | Categorical |
| CD Account | Is there any certificate of deposit account with the customer? | 0.06 | 0.23 | 0 | 1 | Categorical |
| Online | Is this customer using internet banking? | 0.59 | 0.49 | 0 | 1 | Categorical |
| Credit Card | Is this customer using a credit card issued by bank? | 0.29 | 0.45 | 0 | 1 | Categorical |

ISSN NO: 9726-001X

Volume 13 Issue 02 2025

JOURNAL OF CURRENT SCIENCE

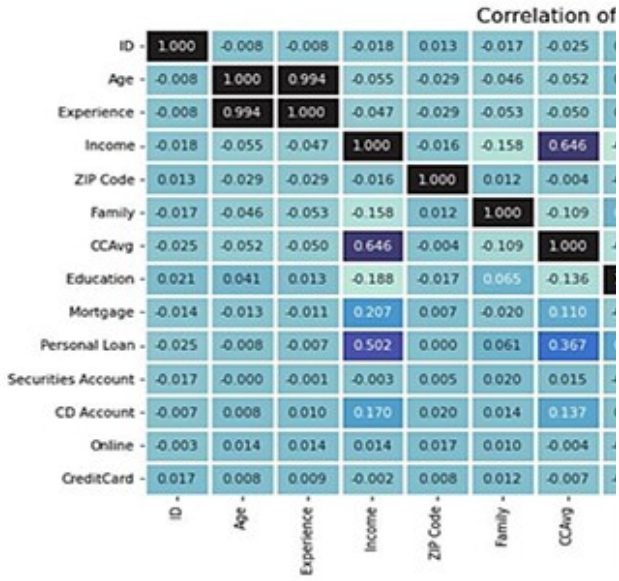


Figure 1. Correlation of attributes

## Material and Methods

### Dataset

Walke collected this publicly accessible dataset from Kaggle and shared it with the world [4]. Included in the "Thera Bank" dataset are 5,000 customer records, together with relationship details such as "Mortgage" and "Securities Account" columns, as well as demographic information such as "Age" and "Income" columns. In addition to feedback from the most recent campaign, such as the Personal Loan section. Just 480 people, or 9.6% of the total, took advantage of this deal [4]. Table 1 displays the characteristics that were considered for this research when the bank loan dataset was examined. Table 1 also displays the feature type, minimum and maximum values, standard deviation, and mean. Nothing is missing or duplicated, and no values are of the string type either. Some ML algorithms struggle with string values, and duplicate or missing data may severely impact prediction outcomes, therefore this is crucial information to have. A label encoder might be used to address the issue if a string value is present. Once these details have been finalized, the columns that are not relevant to this research may be chosen. To begin, you can see how each column impacts the target column, Personal Loan, by checking the correlation matrix in Figure 1. The correlation matrix clearly shows that the goal value is affected by each column. Since each individual's ID is distinct in the dataset and since ZIP Codes reduce prediction accuracy, they were removed. The dataset will be ready for usage with ML algorithms after these details and removals. 2.2 Approaches One must be familiar with the overall procedure for using machine learning algorithms prior to beginning the categorization process. This ML research used grid search and 5-fold cross validation, as shown in Figure 2. Personal loan approval was forecasted using SVM algorithms in this research. The Personal Loan column will serve as our objective, thus we need to extract it to a distinct data frame before making this prediction. The next step is to use train-test-split (TTS), although this strategy is often unreliable for machine learning prediction due to its inconsistent behavior with varying random state values. Thus, we are use Cross Validation to generate both train and test data.

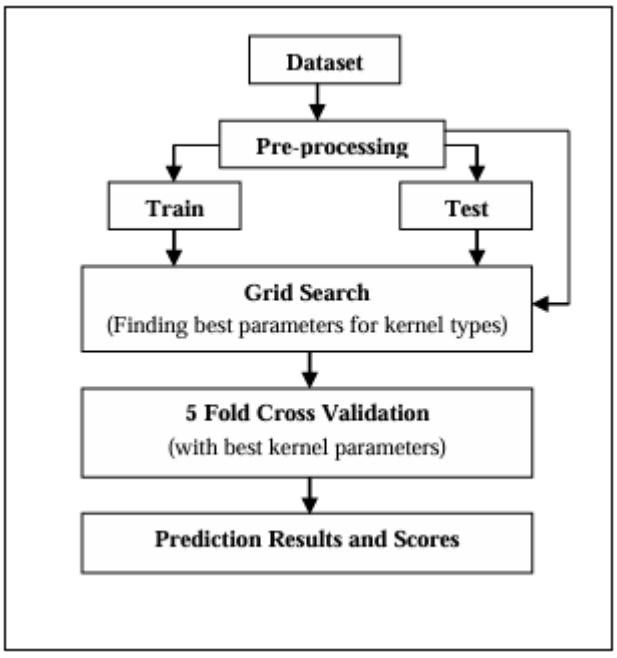


Figure 2. Our study process

### Support Vector Classifier (SVC)

One supervised learning model that has been used for data categorization and prediction is the support vector machine (SVM), which was created by Vladimir Vapnik [16]. The classification issue is the most common use of SVM, however. Using n-dimensional space, the SVM algorithm represents each data item as a point. A certain coordinate is associated with a certain feature's value. The next step in finishing the classification is to find the hyper-plane [17]. This will clearly divide the two groups.

ISSN NO: 9726-001X

Volume 13 Issue 02 2025

JOURNAL OF CURRENT SCIENCE

You may divide the two types of data points along a number of different hyperplanes. The objective is to find a plane that has the greatest margin or range between the two sets of data. The following data points are simpler to identify after increasing the margin distance, which provides reinforcement [18]. Several kernel functions have been implemented in SVM. Nevertheless, the four kernel functions listed below are rather popular: • K(xi,xj) is the linear kernel, which is defined as xiT multiplied by xj'. With γ > 0, the polynomial kernel is defined as K (xi,xj) = (γ xiT ∏ xj + r)d. For any γ greater than 0, the RBF kernel is defined as K (xi,xj) = exp(−γ ‖xi − xj‖2). The sigmoid kernel was defined as: K (xi,xj) = tanh(γ xiT ∗ xj + r). The parameters of the kernel, denoted as γ, are (1), (2), (3), and (4). To determine the optimal settings, this research will use the aforementioned kernel types in conjunction with a grid search technique. Following this decision, we will use cross validation to make predictions and get a number of metrics, including recall, accuracy, precision, and f1 score.

## Experimental Study

and Findings In this study, confusion matrix, accuracy score, precision score, recall score and f1 score metrics will be used to evaluate SVM algorithm. Evaluation Metrics The performance of a model can be explained using evaluation metrics. The ability of evaluation metrics to differentiate between model results is a key feature[20].

3.1.1 Confusion Matrix

| | Predicted:0 | Predicted:1 |
|---|---|---|
| Actual:0 | TN | FP |
| Actual:1 | FN | TP |

Here, N is the projected number of classes, and the resulting matrix is N X N [20]. We will be using a confusion matrix similar to Table 2 for this topic. Section 2. The confusion matrix and its cellular representation Forecast: 0 Real: 0% Projected: 1 tube Real: One match FN 4.1.2 TP Precision Mark Accuracy is defined as the proportion of accurate predictions relative to the total number of forecasts [20]. Equation (5) was used to calculate accuracy.

The formula for accuracy is (TP + TN) divided by (TP + TN + FP + FN). T
The precision score is 5.1.3.
A measure of accuracy is the proportion of positive instances that were correctly identified [20]. The accuracy of a prediction model is shown by its precision score [21]. Equation (6) was used to calculate precision. Preciseness equals to (TP) divided by the sum of (TP) and (FP).

$$\text{Precision} = (\text{TP})/(\text{TP} + \text{FP}) \quad (6)$$

Average Recall (6) Recall is the proportion of correctly recognized true positive cases [20]. The following equation was used to determine recall (7). The formula for recall is (TP) divided by the sum of FP and FN. C F1 Score (3.1.5) (7) The F1-Score is the harmonic mean of the recall and accuracy scores for a classification task [21]. Equation (8) was used to determine F1. F1 =2∗(Precision ∗Recall ) Precision + Recall

$$\text{Recall} = (\text{TP})/(\text{FP} + \text{FN}) \quad (7)$$

F1 Score (3.1.5) (7) The F1-Score is the harmonic mean of the recall and accuracy scores for a classification task [21]. Equation (8) was used to determine F1.

$$\text{F1} = 2 * (\frac{\text{Precision} * \text{Recall}}{\text{Precision} + \text{Recall}}) \quad (8)$$

## Results

Table 3 shows the outcomes of using a support vector machine method with four different kernel types. A mean version of the findings of the 5-fold cross validation is used for confusion matrices and other metrics. Table 4 displays the generalized metric scores. All of our kernels have achieved a successful accuracy score, as shown in Tables 3 and 4, however it is not possible to assess using just this metric. Our accuracy scores, with the exception of the sigmoid kernel, are satisfactory, as they are over 80%. In terms of recall scores, we may declare success for a single kernel, the polynomial. The polynomial and rbf kernels achieve success with respect to F1 scores. The unbalanced dataset causes a small number of positive outcomes, which the support vector machine (SVM) kernels—particularly the sigmoid kernel—fail to categorize, leading to lower scores. According to this study, a polynomial kernel is a better option than a sigmoid kernel when dealing with an imbalanced

ISSN NO: 9726-001X

Volume 13 Issue 02 2025

JOURNAL OF CURRENT SCIENCE

dataset. So yet, no other research has addressed the same questions as this one. The following research are shown in Table 5, all of which pertain to the banking system and the approval of bank loans.

Table 3. Confusion Matrix

| | Actual Value | Predicted Value | |
|---|---|---|---|
| | | 0 | 1 |
| Linear SVC | 0 | 895.2 | 8.8 |
| | 1 | 40.2 | 55.8 |
| Poly SVC | 0 | 892.4 | 11.6 |
| | 1 | 16.4 | 79.6 |
| Sigmoid SVC | 0 | 832.4 | 71.6 |
| | 1 | 95.2 | 0.8 |
| Rbf SVC | 0 | 895.2 | 8.8 |
| | 1 | 23.4 | 72.6 |

Table 4. Metric results

| Kernel Type | Metrics | | | |
|---|---|---|---|---|
| | Accuracy | Precision | Recall | F1 |
| Linear | 0,951 | 0,863 | 0,581 | 0,694 |
| Poly | 0,972 | 0,872 | 0,829 | 0,850 |
| Sigmoid | 0,833 | 0,108 | 0,008 | 0,009 |
| Rbf | 0,967 | 0,893 | 0,757 | 0,818 |

Table 5. Comparison with similar studies in the literature

| Authors of the Article | Highest Score ML Technique | Accuracy |
|---|---|---|
| Sheikh *et al.* [21] | Logistic Regression | 81.1% |
| Vimala and Sharmili [22] | SVM | ~79% |
| Fati [23] | Logistic Regression | 79% |
| Madaan *et al.* [24] | Random Forest | 80% |
| Sreesouthry *et al.* [25] | Logistic Regression | 77% |
| Yaurita and Rustam [26] | SVM (Rbf) | 85% |
| Kumar *et al.* [27] | Decision Tree | 95% |
| Ndayisenga [28] | SVM | 77% |

## Conclusion

Machine learning algorithms have been important in predicting whether or not personal bank loan proposals would be accepted, according to the reviewed literature. Statistical vetting methods (SVMs) are among the most accurate algorithms used in machine learning and statistical analysis [29, 30, 31]. This research made use of a support vector machine technique that makes use of four different kernel types. Based on the studies, a sigmoid kernel produced the lowest results (83%), whereas a polynomial kernel produced the greatest results (97%). Due to the imbalance in our dataset, which results in nine negative values for every true value, certain accuracy and recall numbers are much lower than typical. The usage of an imbalanced dataset may lead to this issue. However, SVMs with a polynomial kernel are an excellent option for predicting loan outcomes, as shown in our work, and support vector machines generally perform well. Different sorts of ML algorithms are employed when we compare with comparable research. In most cases, you should expect an accuracy score of 77% to 85%. Our study's accuracy and other metric scores were higher than previous studies, thus we can state that SVM with a polynomial kernel is effective for banking system classification challenges. At last, a machine learning approach may help banking systems anticipate future profits by gauging the likelihood that customers would accept their loan offers. Declaration Regarding

the study, writing, and publishing of this piece, the writers have not disclosed any possible conflicts of interest.

## References

[1]. Arun, K., G. Ishan, and K. Sanmeet, Loan approval prediction based on machine learning approach. IOSR J. Comput. Eng, 2016. 18(3): p. 18-21.

[2]. Bhandari, M., How to predict loan eligibility using machine learning models. [cited 2022 02 January]; Available from: https://towardsdatascience.com/predict_loan-eligibility-using-machine-learning-models 7a14ef904057.

[3]. Aphale, A.S., and S.R. Shinde, Predict loan approval in banking system machine learning approach for cooperative banks loan approval. International Journal of Engineering Research & Technology, 2020. 9(8): 991-995

[4]. Walke, K. Bank personal loan modelling. [cited 2021 03 October]; Available from: https://www.kaggle.com/krantiswalke/bank-personal-loan modelling.

[5]. Tejaswini, J., 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 Sciences, 2020. 11(4): p. 523-532.