**Loan approval prediction using machine learning algorithm.**

Aniruddh Narlawar**,** Amogh Nimbolkar, Sanket More and Shashank Bhoyar

Department Of Electronics and Telecommunication, Pimpri Chinchwad College Of Engineering, Pune, Maharashtra, India.

[aniruddh.narlawar20@pccoepune.org](mailto:aniruddh.narlawar20@pccoepune.org) , [amogh.nimbolkar20@pccoepune.org](mailto:amogh.nimbolkar20@pccoepune.org), [sanket.more20@pccoepune.org](mailto:sanket.more20@pccoepune.org) [shashank.bhoyar20@pccoepune.org](mailto:shashank.bhoyar20@pccoepune.org)

**Abstract.**

As the banking industry develops, more people take out bank loans, making it challenging to select the best candidate. When selecting the best applicant is done manually, many misunderstandings can occur. Because the technology employs machine learning algorithms, the appropriate candidate is automatically chosen. To ensure that the machine is taught with the use of machine learning algorithms based on these experiences, this is collected by extracting Big Data from the past data of people who have already been granted loans. There are several methods for investigating the drawbacks of credit debt control, according to earlier research from this time period. Technology advancements in areas like machine learning, computer science, and other sciences are helping banks estimate a customer's likelihood of defaulting based on his past behavior, which is a crucial role for technology and which helps bank in many ways .

**Keywords- Machine learning, Loan prediction, Accuracy, banks, KNN, SVM , Logistic Regression**

**Introduction**

Credits are essential to our daily lives. Individuals who don't have a substantial amount of cash can borrow money from banks to establish their own enterprises or for other things. Loans provide banks with enormous revenues. As banks only have a limited supply of products, it is crucial to select a borrower who would repay the loan on time. As so many individuals apply for loans, picking the best one might be challenging. It is the bank's job to choose the best applicant. When selecting the best applicant is handled manually, various misconceptions may occur. One of the key factors influencing financial instability is credit risk. Commercial banks attempt to reduce defaulting risks by evaluating the borrower's ability to repay the loan and requesting collateral prior to the supply of the loan because lending is not only considered by commercial banks as a source of profit but also connected with significant risks. We have created a loan projection as a result. The suitable candidate is chosen automatically by the system using machine learning. Both applicants and bank personnel are supported by this arrangement. This study will be very helpful to financial institutions in determining the creditworthiness of the borrowers and in determining the risk associated with the borrowers by taking into account various Factors.

**Literature Review**

Esha Jain, Gargi Kumar, and Shikhar Srivastava proposed a paper A comparative study of machine

Learning algorithms for loan prediction. The paper approached Predicting whether a loan will be approved or not using machine learning algorithms Overall, the research paper used a supervised machine learning approach for loan prediction and compared the performance of different algorithms.**[2]**

Ritika Gupta and Amit Yadav proposed a paper Loan Approval Prediction using Random Forest Algorithm . The paper consisted To predict whether a loan application will be approved or not using the Random Forest algorithm**. [3]**

Sudarshan D. and Santhosh Kumar G. published a paper A hybrid approach for loan approval prediction using machine learning and sentiment analysis. The problem statement for this paper was To predict whether a loan application will be approved or not using a hybrid approach of machine learning and sentiment analysis . The y used the Random Forest, Naive Bayes, and Sentiment Analysis’**.[6**]

M. Abdullah Al- Farabi, et al .publised a paper A Comparative Study of Machine Learning Algorithms for Loan Prediction .The paper approached A Comparative Study of Machine Learning

Algorithms for Loan Prediction . The study compared different machine learning algorithms&#39; performance in loan prediction and concluded that the Random Forest algorithm outperformed the others**.[9]**

**Methodology**

**Technology used (Machine Learning) :-** A component of artificial intelligence called machine learning enables software applications to predict outcomes more precisely than they could have been programmed to do. Algorithms for machine learning employ real-world data from the past to predict future results. Computers may now gather information without being traditionally programmed thanks to the field of study known as machine learning (ML).One of the numerous interesting approaches that one could come across is machine learning (ML). As the name implies, it endows a computer with the capacity to learn, making it more like a human person. Today will likely see more applications of machine learning than one might anticipate. Some real-world difficulties cannot be resolved using the traditional programming paradigm; in these cases, machine learning is required.

**Algorithms used- 1) Logistic Regression** - One of the well-known ML classifiers, the LR-classifier, is used to categorize learning that is being done under supervision. It is employed for categorical subordinate variable forecasting utilizing a predetermined set of independent features. The sigmoid function, which forms the basis of the technique, gives the concept of logistic regression its name. The possibility to produce distinctive outputs using input is shaped by the LR classifier.

For instance, the most popular two-result logistics return models can provide you with something. To comprehend statistical programmes, one must make an educated assumption as to the relationship between a dependent variable and one or more independent features using two values, such as true/false, yes/no, etc.

**2) SVM -** A deep learning system known as a support vector machine (SVM) uses supervised learning to classify or predict the behavior of groupings of data.

Supervised learning systems in AI and machine learning give input and intended output data that are labelled for classification. A learning foundation for upcoming data processing is provided by the classification. Two data groups are sorted using support vector machines according to similar categorization. The algorithms create hyperplanes (lines) to divide the groups into various configurations.

An SVM creates a learning model that divides up brand-new examples into different groups. SVMs are referred to as non-probabilistic, binary linear classifiers based on these functions. SVMs have the ability to employ techniques like Platt Scaling in probabilistic classification contexts.

3) **KNN-** The k-nearest neighbours algorithm, sometimes referred to as KNN or k-NN, is a supervised learning classifier that employs proximity to produce classifications or predictions about the grouping of a single data point. Although it can be applied to classification or regression issues, it is commonly employed as a classification algorithm because it relies on the idea that comparable points can be discovered close to one another.

In this project we have the selected the dataset from IEEE dataset website . Some of the parameters of this selected datasets are – Gender , Married Status , Education , Self Employed , Past Loan , Property Area , Salary etc . from these various parameters by studing these dataset we compared three classification algorithm SVM , Logistic Regression , KNN. Using various hyper parameters like F1-score , Recall , Accuracy ,Precision we concluded which one will be the best algorithm for such datasets which can give us best possible results .

**Flow Chart**

**DATA COLLECTION**

TEST DATASET

TRAINED DATASET

**DATA PRE-PROCESSING**

HANDLE MISSING VALUES

DATA CLEANING

ENCODING

**MODEL BUILDING**

SVM

LOGISTIC REGRESSION

KNN

**PERFORMANCE PARAMETERS**

RECALL

ACCURACY

F1 SCORE

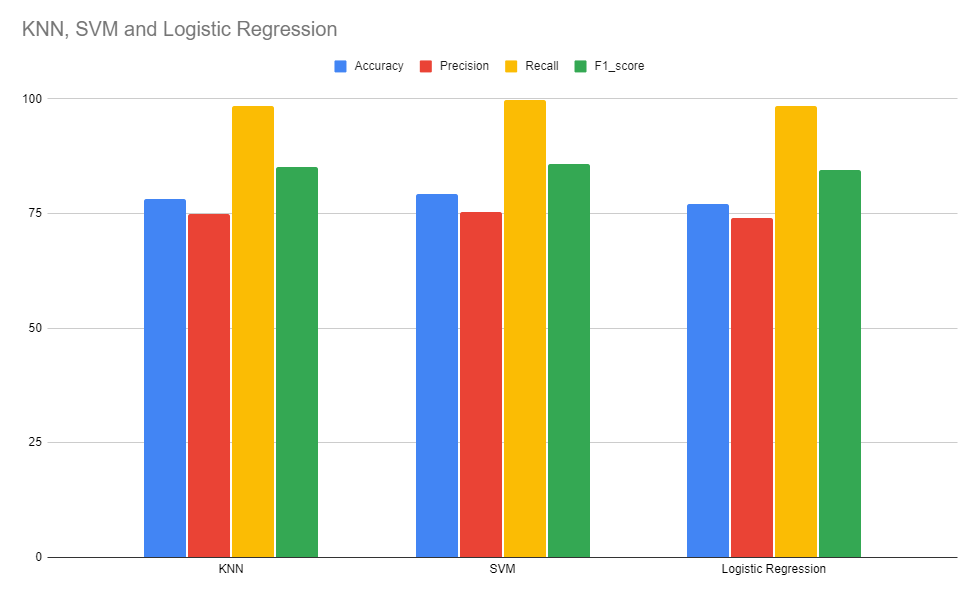
PRECISION

**RESULT ANALYSIS**

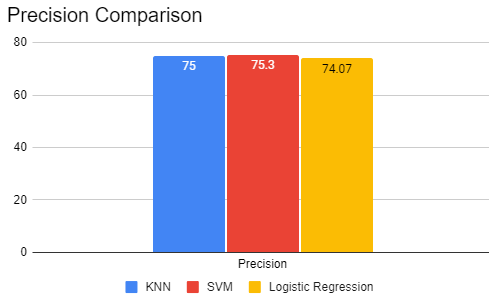
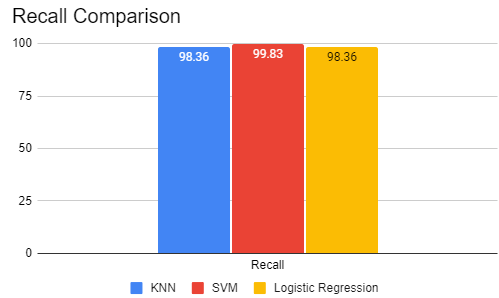
**Result-** We studied different hyper parameters which are most important while comparing different algorithms to evaluate which has the best performance, so here are the results as follows

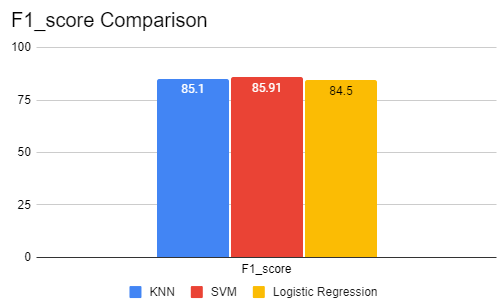
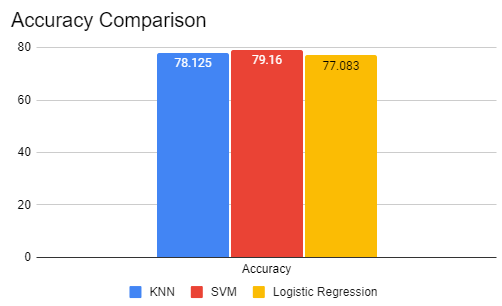
|  |  |  |  |
| --- | --- | --- | --- |
|  | **KNN** | **SVM** | **Logistic Regression** |
| **Accuracy** | **78.125** | **79.160** | **77.083** |
| **Precision** | **75.0** | **75.30** | **74.07** |
| **Recall** | **98.36** | **99.83** | **98.36** |
| **F1\_score** | **85.10** | **85.91** | **84.50** |

**Hyper parameters of algorithms**

****

**Fig 2. Plot of hyper parameters between different algorithms**

** **

** **

**Conclusion:** This paper used various algorithms like K-NearestNeighbour Support Vector Machine and Logistic Regression. We used this algorithms to train the machine. We trained and tested all the algorithms. To conclude we can say that after comparing the performance parameters of algorithms like SVM (Support Vector Machine), KNN(K-Nearest Neighbour) and LR (Logistic Regression).SVM performs the best on our loan predication data set because all the parameters Recall, Precision, Accuracy and F1\_Score outperform when compared to other algorithms when use to train or test the dataset. The Accuracy was 79.16%, Recall 99.83% Precision 75.30% and F1Score of 85.91%Loan approval prediction using machine learning has great potential in the future, especially as the amount of data available for analysis continues to grow there will be tremendous increase in the amount of data, also there will be variety of data available .So we use some modern algorithms which can use a perfect accuracy and other parameters and give the best result possible and make the work easy and reliable

**Literature References**

[1] Esha Jain, Gargi Kumar, and Shikhar Srivastava “Loan Approval Prediction using Machine Learning Techniques” Year-2020

[2] Ritika Gupta and Amit Yadav “Loan Approval Prediction using Machine Learning Techniques “ Year-2020

[3] Sudarshan D. and Santhosh Kumar G. “A hybrid approach for loan approval prediction using machine learning and sentiment analysis” Year-2019

[4] Ekta Gandotra, Divya Bansal, Sanjeev Sofat 2017, ‘Malware Analysis and Classification:

[5] P. Singh and A. Jasuja, “IoT based low-cost distant patient ECG monitoring system,” International Conference on Computing, Communication and Automation, Greater Noida, pp. 1330-1334, 2017.

[6] M. Neyja, S. Mumtaz, K.M.S. Huq, S.A. Busari, J. Rodriguez and Z. Zhou, “AnIoT-Based E-Health Monitoring System Using ECG Signal,” IEEE Global Communications Conference, Singapore, pp. 1-6, 2017.