Loan Prediction Using Machine Learning Algorithms

Sanika Narayanpethkar\textsuperscript{1}, S. Chandana\textsuperscript{2}, M. Rishitha\textsuperscript{3}, Preethi Jeevan\textsuperscript{4}

\textsuperscript{1,2,3}Student, Sreenidhi Institute of Science and Technology
\textsuperscript{4}Associate Professor, Sreenidhi Institute of Science and Technology

ABSTRACT
Loans are a major prerequisite of the present-day world. By this as it were, Banks get a major portion of the entire benefit. It is advantageous for students to oversee their instruction and living costs, and for individuals to purchase any kind of extravagance like houses, cars, etc. But when it comes to choosing whether the applicant’s profile is pertinent to being allowed a credit or not. Banks must see after many aspects. Giving credit is the most commerce of banks. A noteworthy parcel of the bank’s income comes straightforwardly from benefits from credits. Indeed, on the off chance that the bank favors the advance after the confirmation and certification relapse prepare, it cannot be beyond any doubt that the chosen candidate is the genuine candidate. Doing this handle physically requires unused time. Able to anticipate whether a specific candidate is secure, and the whole confirmation handle is robotized with machine fashion. After the presentation of innovation to the world everything began to be automatized. With the introduction of machine learning, the world has seen a modern side of innovation. When the diversion of forecast is played, machine learning gives a few of the most excellent models to test the information for exactness of result. A few of them are broadly utilized. Machine learning calculations are Decision trees, support vector machine, logistic regression, random forest, etc.

I. INTRODUCTION
Nowadays who doesn’t depend on loans. Every person must have taken a loan for some purpose in their life. We take loans for various purposes like home loans, vehicle loans, education loans, etc. Banks provide us with loans based on a certain criterion. The criterion depends on various factors like income, area of living, dependents, job, lifestyle, credit score, etc. Now if we through it manually and check each attribute for every individual it might take up to months to issue the loan. Hence, we use the existing machine learning algorithms to find out which model works best and automatically provide us with a system which tells us whether a certain individual is given the loan or not. Out of all the models used the one with highest accuracy is chosen to be the best fit for prediction problems. We use decision trees to try and predict the accuracy first and then compare the other 5 models to check whether they are better compared to decision trees. This difference is shown in the result after we apply each model to the dataset given. The one which surpasses the accuracy of decision trees is declared best. If not, decision trees is chosen as the best used model.

II. EXISTING SYSTEM
Bank employees manually check the application documents and issue loans to eligible applicants. Viewing all documents requests takes too much time. Therefore, we use an algorithm such as a decision tree to check the correctness of the model. Decision trees help us see how the supervised learning algorithm produces
results. Decision trees in machine learning provide a good way to make decisions once they have identified the problem and all possible outcomes. It allows developers to analyze the consequences of a decision, and as the algorithm accesses more data, it can predict the future outcome of the data.

III. PROPOSED SYSTEM
In the planning process, we use advanced algorithms/models to check the accuracy of the results. We use six different machine learning techniques such as Logistic Regression, Support Vector Machine Models, Random Forests, K-nearest Neighbors and Gaussian Naive Bayes Models. We will use the previous data so the machine can analyze and understand the process. The machine will then check for eligible applicants and provide the results to us. We use the most accurate reporting model to create loan applications.

Advantages
- The lending period will be shortened.
- All processes will be automated to avoid people.
- Credits are granted immediately to eligible applicants.

IV. SYSTEM ARCHITECTURE
Considering each dataset each and each information set has both input as well as yield. Taking models into thought a few information models may require colossal sum of information in arrange to urge exact and exact comes about. Whenever we bolster the information, we’ll discover it categorical, so we ought to utilize a wide extent of pre-processing procedures in arrange to form all the categorical information and convert it into nonstop. We should check the information moreover after the pre-processing and we ought to envision the dataset to induce curious designs. After this the information is put away within the database and it is recovered from the database and sent for the arrangement handle. The strategies that include pre-processing extricates information at that point it changes over the selected set of qualities into persistent traits and makes the information valuable. Now the experiences given from here are utilized develop a model which is scientific by making utilize of machine learning methods.

[Figure 1.1 steps]
V. IMPLEMENTATION

MODULES

THE IMPLEMENTATION CONSISTS OF:

• Import the libraries
• Read the dataset
• Clean the data
• Check the missing values
• Plot the graphs • Data visualization
• Split the data into training and testing
• Apply the algorithms on the trained and tested data
• Check the accuracy

DATASET

The dataset collected for prognosticating advance disappointment clients is anticipated into Preparing set and testing set. By and large 8020 extent is connected to separate the preparing set and testing set. Test set determining is done. There are various attributes which are considered when collecting dataset. These attributes play a vital role in determining the loan given to an individual.

Attributes are

![Figure 1.2 attributes in the dataset](image)

PLOTTING GRAPHS AND VISUALIZATION
Figure 1.3 visualization using seaborn

Figure 1.5 box plot visualization
SPLITTING DATA INTO TRAINING AND TESTING

TRAINING SET: The information collection utilized to prepare our show some time recently making an application is alluded to as the preparing set. The preparation set is regularly physically developed, and your demonstration follows the same rules and definitions as those expressed within the hone set.

TESTING SET: A testing set may be an information set on which you apply your demonstration to see in the event that it is redressed and creating the required comes about. A test set is comparative to a model’s test.
VI. RESULTS
We used 6 algorithms, and each gave a different accuracy:

**Using KNN (K-Nearest Neighbors)**

```python
In [24]: from sklearn.neighbors import KNeighborsClassifier
knn = KNeighborsClassifier()
KNeighborsClassifier Accuracy: 0.8246753246753247
Classification Report:
precision     recall      f1-score  support
    0.0      0.85      0.52            48
    1.0      0.98      0.96            100
accuracy	0.82          154
```

Confusion Matrix:
```
[[ 22  23]
 [  4  92]]
```

Training Accuracy: 0.8246753246753247
Testing Accuracy: 0.8246753246753247

**Using SVM Model**

```python
In [32]: from sklearn.svm import SVC
svc = SVC()
SVC Accuracy: 0.8246753246753247
Classification Report:
precision     recall      f1-score  support
    0.0      1.00      1.00            48
    1.0      0.98      1.00            100
accuracy	0.94          154
```

Confusion Matrix:
```
[[ 22  23]
 [  4  92]]
```

Training Accuracy: 0.8246753246753247
Testing Accuracy: 0.8246753246753247

**Using Decision Tree Model**

```python
In [33]: from sklearn.tree import DecisionTreeClassifier
dt = DecisionTreeClassifier()
DecisionTreeClassifier Accuracy: 0.7507402597402597
Classification Report:
precision     recall      f1-score  support
    0.0      0.63      0.56            48
    1.0      0.81      0.85            100
accuracy	0.76          154
```

Confusion Matrix:
```
[[ 27  23]
 [ 16  98]]
```

Training Accuracy: 0.7507402597402597
Testing Accuracy: 0.7507402597402597

**Using Logistic Regression Model**

```python
In [34]: from sklearn.linear_model import LogisticRegression
lr = LogisticRegression()
LogisticRegression Accuracy: 0.8376923700237002
Classification Report:
precision     recall      f1-score  support
    0.0      1.00      1.00            48
    1.0      1.00      1.00            100
accuracy	0.94          154
```

Confusion Matrix:
```
[[ 22  23]
 [  4  92]]
```

Training Accuracy: 0.8376923700237002
Testing Accuracy: 0.8376923700237002

**Using Gaussian Naive Bayes Model**

```python
In [35]: from sklearn.naive_bayes import GaussianNB
gnb = GaussianNB()
GaussianNB Accuracy: 0.8311688316883168
Classification Report:
precision     recall      f1-score  support
    0.0      0.61      0.40            48
    1.0      0.81      0.80            100
accuracy	0.81          154
```

Confusion Matrix:
```
[[ 22  23]
 [ 1  92]]
```

Training Accuracy: 0.8311688316883168
Testing Accuracy: 0.8311688316883168

**Using Random Forest Model**

```python
In [36]: from sklearn.ensemble import RandomForestClassifier
rf = RandomForestClassifier(n_estimators = 50, max_depth = 10, min_samples_leaf=12)
RandomForestClassifier Accuracy: 0.8766231762317623
Classification Report:
precision     recall      f1-score  support
    0.0      1.00      1.00            48
    1.0      0.96      0.96            100
accuracy	0.98          154
```

Confusion Matrix:
```
[[ 22  23]
 [ 1  92]]
```

Training Accuracy: 0.8766231762317623
Testing Accuracy: 0.8766231762317623

GUI RESULTS
VII. CONCLUSION

With cautious investigation of dynamic substance and members’ limitations, it can be securely done that things are qualified individuals. This app works well and meets all the requirements of an investor. This part can be openly associated in numerous other frameworks. Computer glitches, there are numerical information around the breach points, and the weight of the foremost important highlights has been settled in prophet builder, within the future the program title can be more secure, more solid and more dependable. Energetic overwhelming adaptation. Within the future, this prophet module can be combined with the structure of the computerized process. The framework employs ancient preparing information to prepare future program, so the unused test information ought to be coordinates with the preparing information after a few times.

VIII. REFERENCES

2. Wei Li, Shuai Ding, Yi Chen, and Shanlin Yang, Heterogeneous Ensemble for Default Prediction of Peer-to-Peer Lending in China, Key Laboratory of Process Optimization and Intelligent Decision-Making, Ministry of Education, Hefei University of Technology, Hefei 2009, China
