NAAN MUDHALVAN

PROJECT REPORT

Divya Shree SA

2021503301

LOAN APPROVAL PREDICTION

ABSTRACT

Loan approval is a very important process for banking organizations. The systems approve or reject the loan applications. Recovery of loans is a major contributing parameter in the financial statements of a bank. It is very difficult to predict the possibility of payment of loan by the customer. In recent years many researchers worked on loan approval prediction systems. Machine Learning (ML) techniques are very useful in predicting outcomes for large amount of data. In this project different machine learning algorithms are applied to predict the loan approval of customers. The main focus of this project is to determine whether the loan given to a particular person or an organization shall be approved or not

PROBLEM STATEMENT

The problem statement for loan approval prediction revolves around the challenge of accurately assessing the creditworthiness of loan applicants to determine whether they should be approved or denied. This involves analyzing various factors such as income, credit history, employment status, and debt-to-income ratio to predict the likelihood of loan repayment. The goal is to develop a reliable predictive model that can assist financial institutions in making informed decisions, reducing the risk of default, and ensuring fair and responsible lending practices

NOVELTY

The novelty in this project lies in leveraging advanced machine learning techniques to develop a predictive model for loan approval, integrating not only traditional financial data but also alternative data sources and innovative features such as behavioral patterns, social media activity, and transaction history. By incorporating a holistic approach to credit

risk assessment, the project aims to enhance predictive accuracy, reduce bias, and improve the efficiency and inclusivity of the lending process.

MODELLING

Data Collection: Gather historical loan application data from financial institutions, including information on applicants' demographics, financial history, credit scores, employment details, loan amounts, and approval decisions.

Data Preprocessing: Clean the data by handling missing values, outliers, and inconsistencies. Perform feature engineering to create new features or transform existing ones that may better represent the underlying patterns in the data.

Exploratory Data Analysis (EDA): Conduct exploratory data analysis to understand the distributions, correlations, and trends in the dataset. Identify relevant features and potential relationships between predictors and loan approval decisions.

Feature Selection: Select a subset of informative features that are likely to have a significant impact on loan approval predictions. This may involve techniques such as correlation analysis, feature importance ranking, or domain knowledge expertise.

Model Selection: Choose appropriate machine learning algorithms for loan approval prediction, such as logistic regression, decision trees, random forests, or gradient boosting machines. Consider ensemble methods or neural networks for more complex relationships.

Model Training: Split the dataset into training and testing sets, and train the selected machine learning models on the training data. Tune hyperparameters using techniques like grid search or random search to optimize model performance.

Model Evaluation: Evaluate the trained models using appropriate performance metrics such as accuracy, precision, recall, F1-score, and ROC-AUC curve. Compare the performance of different models to identify the best-performing one.

Model Deployment: Deploy the trained model into a production environment where it can make real-time loan approval predictions. Integrate the model with existing loan approval systems or develop a standalone application for deployment.

Monitoring and Maintenance: Monitor the deployed model's performance over time and update it as needed with new data or retraining. Continuously evaluate the model's accuracy and recalibrate it to ensure reliable predictions and compliance with changing regulatory requirements**.**

PROPOSED WORK

The proposed work involves developing a predictive model for loan approval prediction by leveraging historical loan application data. This process includes data collection, preprocessing, and feature engineering to extract relevant features such as credit score, income, and employment status. We will then select appropriate machine learning algorithms, such as logistic regression or decision trees, to train and evaluate the model's performance. Additionally, fine-tuning and validation techniques will be employed to optimize the model's accuracy and generalization ability. Upon successful validation, the model will be deployed to assist financial institutions in automating the loan approval process, improving efficiency, and minimizing the risk of default.

DATASET

	Loan_ID	Gender	Married	Dependents	Education	Self_Employed	ApplicantIncome	CoapplicantIncome	LoanAmount	Loan_Amount_Term	Credit_History	Property_Area	Loan_Status
0	LP001002	Male	No	0.0	Graduate	No	5849	0.0	NaN	360.0	1.0	Urban	Y
1	LP001003	Male	Yes	1.0	Graduate	No	4583	1508.0	128.0	360.0	1.0	Rural	N
2	LP001005	Male	Yes	0.0	Graduate	Yes	3000	0.0	66.0	360.0	1.0	Urban	Y
3	LP001006	Male	Yes	0.0	Not Graduate	No	2583	2358.0	120.0	360.0	1.0	Urban	Y
4	LP001008	Male	No	0.0	Graduate	No	6000	0.0	141.0	360.0	1.0	Urban	Y

RESULT

```
Accuracy score of  RandomForestClassifier = 98.04469273743017
Accuracy score of  KNeighborsClassifier = 78.49162011173185
Accuracy score of  SVC = 68.71508379888269
Accuracy score of  LogisticRegression = 80.44692737430168
```

```
[ ]  # making predictions on the testing set
     for clf in (rfc, knn, svc,lc):
       clf.fit(X_train, Y_train)
       Y_pred = clf.predict(X_test)
       print("Accuracy score of ",
         clf.__class__.__name__,"=",
         100*metrics.accuracy_score(Y_test,
                       Y_pred))
```

```
Accuracy score of  RandomForestClassifier = 82.5
Accuracy score of  KNeighborsClassifier = 63.74999999999999
Accuracy score of  SVC = 69.16666666666667
Accuracy score of  LogisticRegression = 80.83333333333333
```

CONCLUSION

The System approve or rejects loan applications. Recovery of loans is a major contributing parameter in the financial statements of a bank. It is very difficult to predict the posibility of payment of loan by the customer.

Machine Learning techniques are very useful in predicting outcomes for large amouf data. In our project a total of five machine learning algorithms which includes Logistic Regression, Decision Tree, Random forest classification,Support Vectior Classifier and Gradient Boosting Classifier are applied to predict the loan approval of customers. The experimental results conclude that the accuracy of Logistic Regression Classification machine learning algorithm is better compared to others machine learning approach.