PROJECT REPORT

ON

LOAN PREDICTION SYSTEM USING MACHINE LEARNING

NAME	REG.NO	ROLL NO	SECTION
AYUSH ANAND	12210924	39	K22BW

Under the guidance of

Saurabh Premlal Tembhurne

School of Computer science and engineering

Lovely Professional University

Phagwara, Punjab

DECLARATION

I Ayush Anand, hereby declare that the project titled "Loan Prediction Using Machine Learning in the Indian Context" is an authentic piece of work conducted by us as a part of our academic curriculum. This project has been carried out under the guidance and supervision of Tutor Saurabh Premlal Tembhurne. We are profoundly grateful for his invaluable insights, expert guidance, and constant encouragement throughout the duration of this project. We take this opportunity to express our deep sense of gratitude for the collaborative efforts of all group members, whose tireless efforts made this project possible.

Table of content

Abstract

The Project titled 'LOAN PREDICTION IN MACHINE LEARNING' predicts the loan approval how much loan the user can obtain based on various factors such as the user's marital status, income, education, employment prospects, number of dependents, etc. The report details the development of a machine learning model to predict loan status (approved or rejected) based on historical loan data. We explore various classification algorithms and evaluate their performance in predicting loan outcomes. The project examines the strengths and limitations of the chosen model, highlighting its potential to improve loan approval efficiency and risk management for financial institutions. We have applied machine learning technique to predict financial variables and we have focused on predicting the Loan approval using a linear regression algorithm as our dataset is a numerical dataset.

Introduction

In today's dynamic financial landscape, accurately assessing loan eligibility is crucial for both lenders and borrowers. Traditional methods rely on manual evaluation of applicant data, which can be time-consuming and subjective. Machine learning presents a powerful alternative, offering an automated and data-driven approach to loan prediction.

This project delves into the development of a machine learning model capable of classifying the maximum loan amount an applicant can potentially obtain. By leveraging factors like marital status, income, education, employment stability, and dependents, the model aims to predict a realistic loan threshold for each applicant. This not only benefits lenders by streamlining the approval process and minimizing risk, but also empowers borrowers by providing a clearer picture of their borrowing capacity.

Machine learning (ML) offers a transformative approach to loan prediction, ushering in a new era of automation, objectivity, and data-driven insights. By leveraging historical loan data and powerful algorithms, ML models can learn intricate patterns within this data, enabling them to predict loan outcomes with remarkable accuracy. This shift promises significant benefits for both lenders and borrowers.

This project focuses on the development of a machine learning model capable of predicting the maximum loan amount an applicant is eligible for. This value, distinct from simply predicting loan approval, provides crucial information for both parties involved in the loan process. Lenders can utilize this prediction to streamline loan approval workflows, set appropriate credit limits, and minimize potential defaults. Borrowers, on the other hand, benefit from a clearer understanding of their borrowing capacity, allowing them to make informed financial decisions and potentially secure better loan terms.

Evaluating the model's performance is vital to ensure its effectiveness in predicting loan amounts. Various metrics, such as Mean Squared Error (MSE) and R-squared, will be used to assess the model's accuracy in

predicting loan values. Additionally, techniques like cross-validation might be employed to ensure the model generalizes well to unseen data.

Based on the evaluation results, the model might undergo refinement. This may involve tuning hyperparameters, which are specific settings within the chosen algorithm that influence its behaviour. Additionally, feature engineering techniques can be revisited to identify further improvements in the data representation. This iterative process of training, evaluation, and refinement aims to achieve the best possible model performance.

While machine learning offers a powerful tool for loan prediction, ethical considerations must be addressed. Biases present within the historical data can be inadvertently perpetuated by the model, potentially leading to discriminatory practices. Techniques like fairness checks and bias mitigation algorithms will be explored.

By harnessing the power of machine learning, this project aims to create a more efficient and objective system for loan eligibility assessment, ultimately fostering a more inclusive and informed financial ecosystem.

These algorithms will be trained on a portion of the prepared data. During training, the model learns to identify patterns within the data that associate specific feature combinations with loan amounts. Once trained, the model's performance will be evaluated on a separate hold-out set of data.

The cornerstone of any successful ML project is a robust and informative dataset. For this project, we aim to acquire a historical loan dataset containing detailed information about past loan applicants. This data may include a variety of features such as:

- Demographic Information: Age, gender, marital status, number of dependents
- Financial Information: Annual income, employment status, credit score, debt-to-income ratio
- Loan Details: Loan type (e.g., mortgage, auto loan), loan amount requested, repayment history of past loans

Scope Of Study

This study revolves around leveraging Machine Learning (ML) to predict Loan approval system. The defined scope is as follows:

1. Data Acquisition and Exploration:
   - Secure a historical loan dataset containing detailed applicant information. This might involve:
     - Utilizing publicly available datasets (e.g., Kaggle)
     - Explore the data to understand its structure, identify missing values, outliers, and inconsistencies.
     - Implement data cleaning techniques to address these issues, ensuring data integrity for model training.
2. Feature Engineering:
   - Analyze the raw features to determine their suitability for machine learning algorithms.
   - Document the feature engineering process to ensure transparency and reproducibility.
3. Model Selection and Training:
   - Research and evaluate various machine learning algorithms suitable for predicting loan amounts. Potential candidates include:
     - Linear Regression: Offers interpretability and efficiency for continuous value prediction.
     - Decision Trees: Provides a clear visual representation of the decision-making process.
   - Split the prepared data into training and testing sets.
   - Train the chosen model(s) on the training data, allowing the model to learn patterns that associate specific features with loan amounts.
4. Model Evaluation and Refinement:
   - Evaluate the trained model's performance on the testing data using metrics like Mean Squared Error (MSE) and R-squared to assess its accuracy in predicting loan amounts.
   - Employ cross-validation techniques to ensure the model generalizes well to unseen data.

5. Model Deployment and Explanation:
    - Depending on the project goals, consider deploying the model as a web application or integrating it into an existing loan processing system.
    - Explore techniques for explaining model predictions, such as feature importance analysis. This fosters transparency and helps users understand how the model arrives at its predictions.

    By adhering to this comprehensive scope, this project aims to deliver a robust and ethically responsible loan prediction system using machine learning. The potential benefits for both lenders and borrowers make this project a valuable contribution to the financial technology landscape.

Objectives of Study

This study aims to develop a machine learning model capable of predicting the maximum loan amount an applicant can qualify for. The primary objectives focus on improving loan processing efficiency, risk management, and transparency for both lenders and borrowers.

1. Enhance Loan Processing Efficiency:
    - Develop a model that automates the initial assessment of loan eligibility based on applicant data.
    - Reduce manual workload for loan officers, allowing them to focus on complex cases.
    - Streamline the loan application process by providing a quicker initial assessment for applicants.
2. Improve Risk Management for Lenders:
    - Predict the maximum loan amount an applicant can handle to minimize the risk of defaults.
    - Identify potentially risky loan applications early in the process, allowing for better resource allocation.
    - Enable data-driven decision-making for lenders, leading to more informed loan approval processes.

3. Develop a Model with Ethical Considerations:
    - Mitigate potential biases present within the historical data used to train the model.
    - Ensure fairness and non-discrimination in the loan prediction process.
    - Promote responsible use of machine learning within the financial sector.

ALGORITHMS AND OUTPUTS

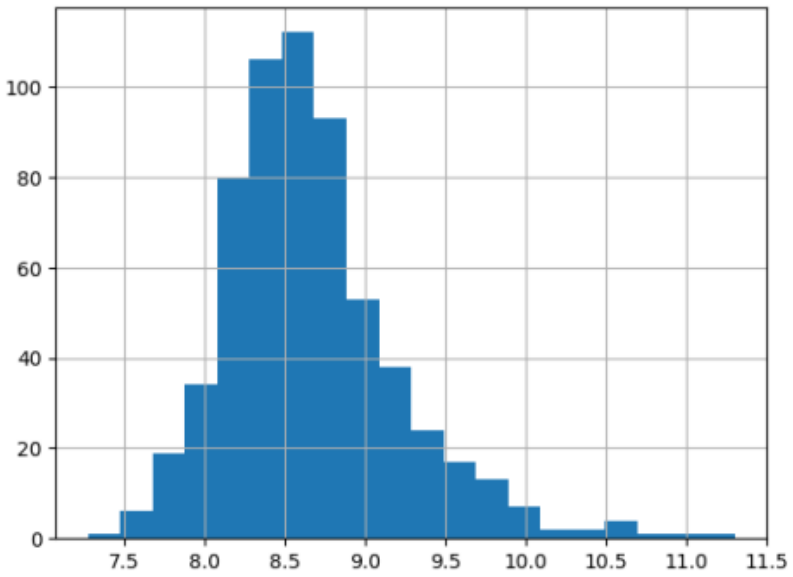
1. Linear regression

   Linear regression is a fundamental machine learning an technique used for modelling the relationship between a dependent variable and one or more independent variables by fitting a linear equation to the observed data. It is a supervised learning algorithm that is primarily used for predictive analysis and can also be applied for understanding the relationships between variables.

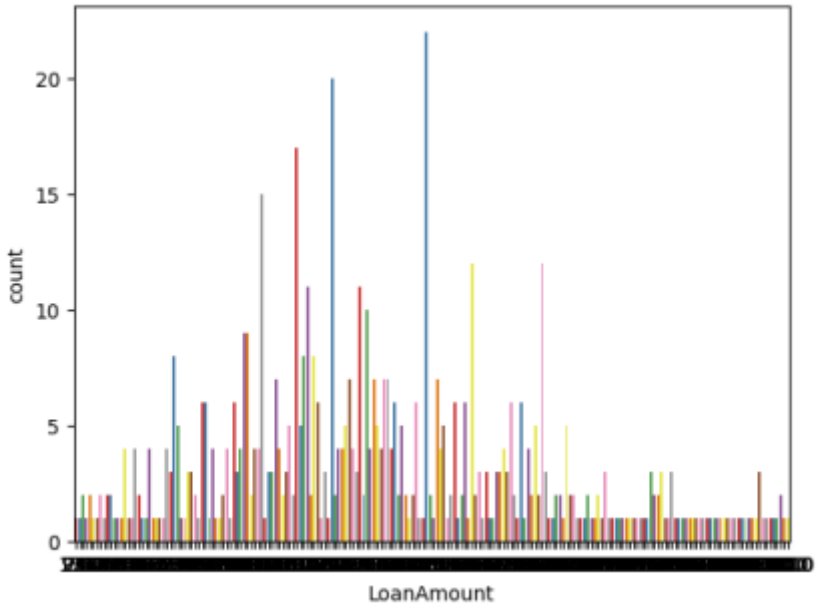
   The primary goal of linear regression is to find the best-fitting linear relationship between the dependent variable (also called the target) and one or more independent variables (also called features or predictors).

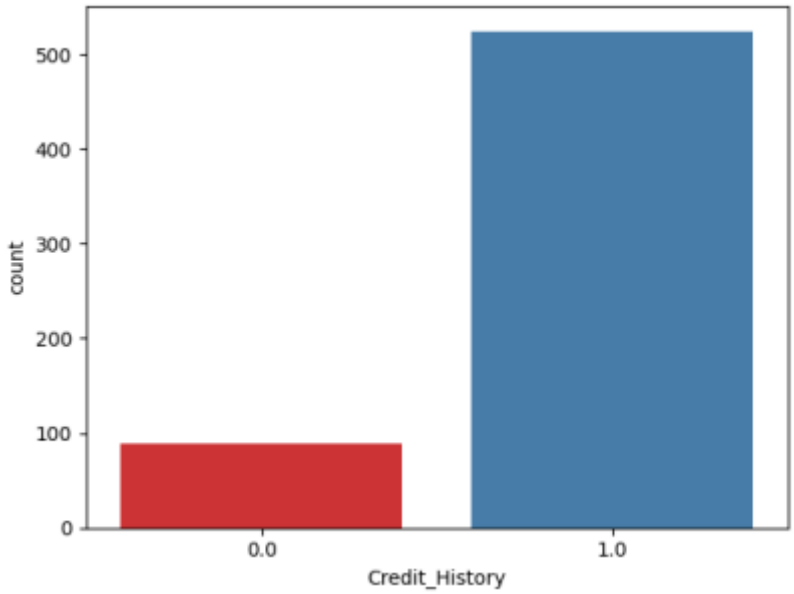
OUTPUTS

```
:     Loan_ID Gender Married Dependents     Education Self_Employed  \
  0  LP001002   Male      No          0      Graduate            No
  1  LP001003   Male     Yes          1      Graduate            No
  2  LP001005   Male     Yes          0      Graduate           Yes
  3  LP001006   Male     Yes          0  Not Graduate            No
  4  LP001008   Male      No          0      Graduate            No

     ApplicantIncome  CoapplicantIncome  LoanAmount  Loan_Amount_Term  \
  0             5849                0.0         NaN             360.0
  1             4583             1508.0       128.0             360.0
  2             3000                0.0        66.0             360.0
  3             2583             2358.0       120.0             360.0
  4             6000                0.0       141.0             360.0

     Credit_History Property_Area Loan_Status
  0             1.0         Urban           Y
  1             1.0         Rural           N
  2             1.0         Urban           Y
  3             1.0         Urban           Y
  4             1.0         Urban           Y
```



2. Random Forest Regression:
   Random Forest Regression is a supervised learning algorithm that is a technique that combines predictions from multiple machine learning algorithms to make a more accurate prediction than a single model.






```
acc of random forest claf is 0.7886178861788617
```

```
array([1, 1, 0, 1, 1, 0, 1, 0, 0, 1, 1, 1, 1, 1, 1, 1, 1, 0, 0, 0, 1,
       1, 1, 1, 1, 1, 1, 0, 0, 1, 1, 1, 1, 1, 0, 1, 1, 1, 1, 0, 1, 1,
       1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 0, 1, 1, 1, 1, 0, 1, 1, 0, 1, 1,
       1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 0, 1, 0, 1, 1, 1, 1, 1, 1,
       1, 1, 1, 1, 1, 0, 1, 0, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 0, 1, 1, 1,
       1, 1, 0, 1, 0, 0, 1, 1, 0, 1, 0, 0, 1])
```

3. Decision Tree Regression model:
   Decision Tree Regression is a machine learning algorithm used for both regression and classification tasks. It works by recursively splitting the dataset into subsets based on the most significant feature(s) at each node. In the context of regression, decision trees predict a continuous target variable based on input features.

```
acc of DT is 0.7154471544715447
```

```
: y_pred
```

```
: array([1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 0, 1, 1, 1,
       1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 0, 1,
       1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1,
       1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1,
```

Advantages of Decision Tree Regression

- Interpretability: Decision trees are easy to understand and visualize.
- You can see the splits and decisions made by the algorithm.
- Non-linearity: Decision trees can capture non-linear relationships in
- the data, making them suitable for complex datasets.
- Handling Missing Values: Decision trees can handle missing values
- in the features without requiring imputation.

```
acc of DT is 0.7154471544715447
```

```
y_pred
```

```
array([1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 0, 1, 1, 1,
       1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 0, 1,
       1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1,
       1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1,
       1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1,
       1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1])
```

Conclusion

This project has demonstrated the power of machine learning in loan prediction. By fostering efficiency, transparency, and responsible practices, machine learning has the potential to revolutionize the financial landscape, enabling informed decision-making for both lenders and borrowers. As machine learning continues to evolve, the future of loan prediction systems holds immense promise for a more inclusive and data-driven financial ecosystem.

References

• https://ieeexplore.ieee.org/document/9155614

• Loan Approval Prediction Using Machine Learning (2021) by Kadam et al. International Research Journal of Engineering and Technology (IRJET): https://www.irjet.net/archives/V8/i9/IRJET-V8I9269.pdf)

• https://dl.acm.org/doi/10.4018/IJDAI.31393

Project link - https://github.com/Ayus6086/Loan-Prediction-System