Loan Approval Prediction using Machine Learning

Short Presentation By Hari Prasad Reddy Sheelam

Student Id : 11601562

CONTEXT

Loans are the primary necessity of the contemporary world. Banks receive the majority of the entire profit only from this. Those who purchase any type of luxury, such as homes, vehicles, etc., and students who balance their living and educational costs will benefit from it.But when it comes to assessing whether the applicant’s profile is suitable to be granted the loan or not. Banks have a lot of things to manage.In order to make their job easier, we will be utilizing Python and machine learning to determine whether or not a candidate's profile is relevant by utilizing important attributes including applicant income, credit history, marital status, and education.

Content

- Data Collection
- Dataset Features
- Data Preprocessing
- Visualization
- Model Training and Evaluation
- Suitable Algorithm and Why ?
- Time and Space Complexity of Algorithm
- Summary

Data Collection

We used Dataset for Loan Approval Prediction Using Machine Learning Algorithm

drive.google.com/file/d/1LIvIdqdHDFEGnfzIgEh4L6GFirzsE3US/view?pli=1

The dataset contains 13 features :

1	Loan	A unique id
2	Gender	Gender of the applicant Male/female
3	Married	Marital Status of the applicant, values will be Yes/ No
4	Dependents	It tells whether the applicant has any dependents or not.
5	Education	It will tell us whether the applicant is Graduated or not.
6	Self_Employed	This defines that the applicant is self-employed i.e. Yes/ No
7	ApplicantIncome	Applicant income
8	CoapplicantIncome	Co-applicant income
9	LoanAmount	Loan amount (in thousands)
10	Loan_Amount_Term	Terms of loan (in months)
11	Credit_History	Credit history of individual’s repayment of their debts
12	Property_Area	Area of property i.e. Rural/Urban/Semi-urban
13	Loan_Status	Status of Loan Approved or not i.e. Y- Yes, N-No

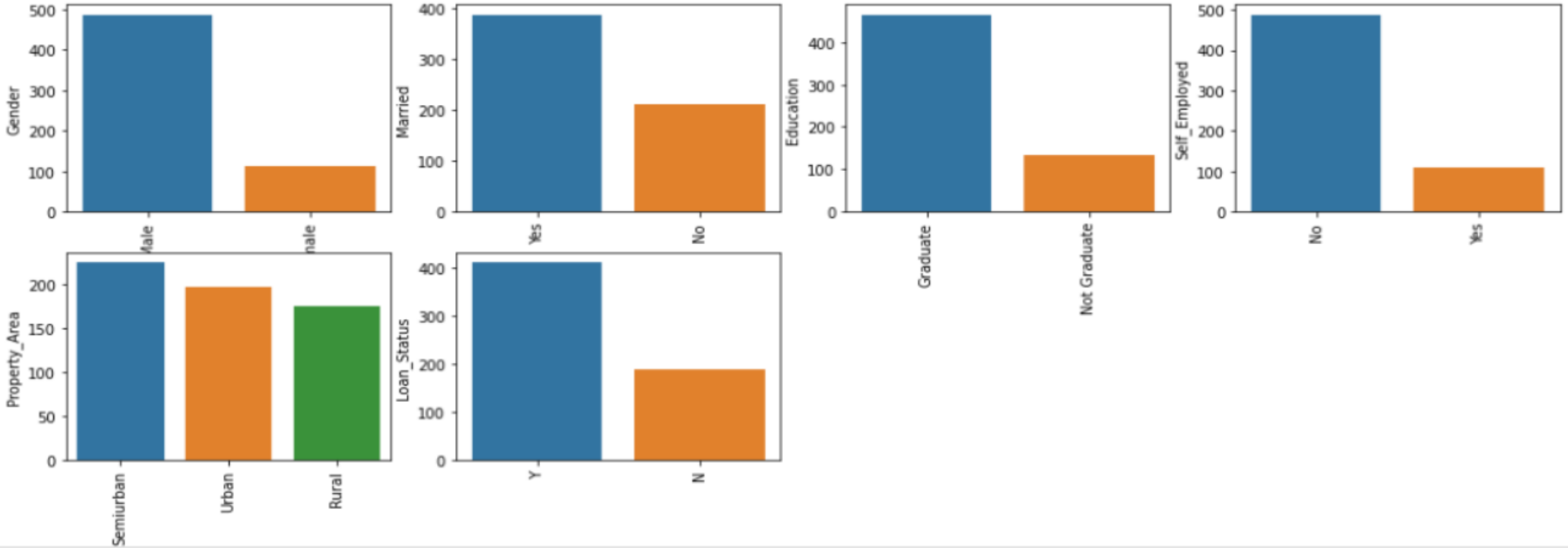
Data Cleaning

Get the number of columns of object datatype.

As Loan_ID is completely unique and not correlated with any of the other column, So we will drop it using .drop() function.

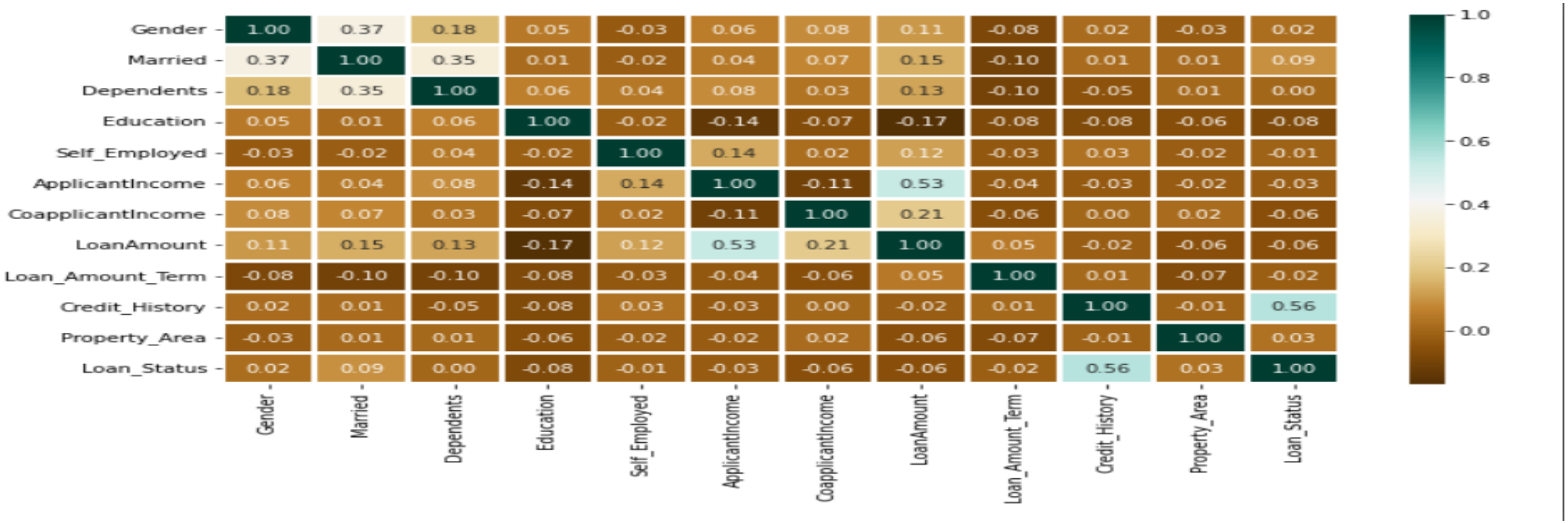
Visualization

- Visualize all the unique values in columns using barplot. This will simply show which value is dominating as per our dataset.



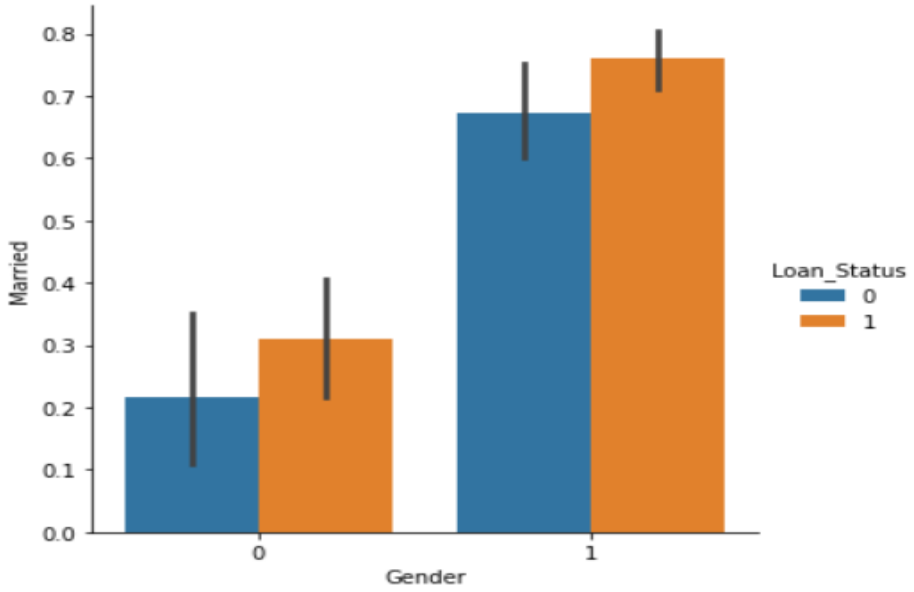
Visualization

As all the categorical values are binary so we can use Label Encoder for all such columns and the values will change into **int** datatype.



Visualization

Now we will use Catplot to visualize the plot for the Gender, and Marital Status of the applicant.



Model Training and Evaluation

As this is a classification problem so we will be using these models :

1) KNeighbours Classifiers
2) Random Forest Classifiers
3) Support Vector Classifiers
4) Logistics Regression

Suitable Algorithm and Why ?

With an accuracy score of 82% for the testing dataset, the Random Forest Classifier provides the best results. Additionally, ensemble learning strategies like bagging and boosting can be applied to obtain far better outcomes.

Time and Space Complexity of Algorithm

Time Complexity:

The time complexity of a machine learning algorithm for loan approval primarily depends on the following factors:

1) Training Phase: The time required to train the model. This phase involves learning patterns from historical loan data. The time complexity often depends on the algorithm used.

For example:

Linear models (e.g., logistic regression): O(n*d), where n is the number of data points and d is the number of features.

Decision Trees: O(n * d * log(n)), where n is the number of data points and d is the number of features.

Random Forest: O(m * n * d * log(n)), where m is the number of trees, n is the number of data points, and d is the number of features.

2) Inference Phase: The time required to make predictions for loan applications once the model is trained. This is typically faster than the training phase. The time complexity depends on the model's structure and the number of features but is often linear in practice.

3) Feature Engineering: Time may be required to preprocess and engineer features, which can vary depending on the complexity of feature extraction and selection.

4) Hyperparameter Tuning: If hyperparameter tuning is performed, it adds an additional layer of complexity in terms of time.

5) Scalability: The time complexity can be influenced by the scalability of the algorithm. Some algorithms may not scale well with large datasets.

Time and Space Complexity of Algorithm

Space Complexity:

The space complexity of a machine learning algorithm for loan approval is primarily determined by:

1) Model Size: The space required to store the trained model, including all its parameters. The space complexity depends on the algorithm, the model architecture, and the number of trees (in the case of ensemble methods like Random Forest).

2) Feature Storage: The space required to store the feature vectors for the dataset. The space complexity depends on the number of data points and the number of features.

3) Intermediate Data: During model training, various intermediate data structures may be used, contributing to space complexity.

4) Inference Data: The space required to store data used for making predictions. For large batch predictions, this can be significant.

It's important to note that modern machine learning libraries and frameworks often optimize memory usage and parallelize operations to make efficient use of computational resources. The actual time and space complexities may vary based on the implementation and hardware used.

In a production environment, the choice of algorithm and the optimization of the workflow play a significant role in ensuring that the system meets real-time or batch processing requirements while remaining resource-efficient. Real-world loan approval systems may also have additional complexities due to data streaming, compliance regulations, and system architecture.

Summary

Finally, by making the process more precise, automated, and scalable and by providing chances for future advancements, the use of a Random Forest Classifier for loan approval prediction offers a data-driven and effective solution that benefits both financial institutions and loan applicants.

THANK YOU