**Name:Salman Pradhan**

**Loan Application Status Prediction**

**This dataset includes details of applicants who have applied for loan. The dataset includes details like credit history, loan amount, their income, dependents etc.**

**Independent Variables:**

* **Loan\_ID**
* **Gender**
* **Married**
* **Dependents**
* **Education**
* **Self\_Employed**
* **ApplicantIncome**
* **CoapplicantIncome**
* **Loan\_Amount**
* **Loan\_Amount\_Term**
* **Credit History**
* **Property\_Area**

**Dependent Variable (Target Variable):**

* **Loan\_Status**

**1.Problem Defination**

We have to build a model that can predict whether the loan of the applicant will be approved or not on the basis of the details provided in the dataset.

**2.Data Analysis**

* **In this Dataset there are 614 rows and 13 columns(a mixture of categorical and numerical columns).Loan stasus is our target variable.**
* **Though Loan status column is categorical column We have to predict our target variable(LoanStatus) by using classification algorithm.**
* **Though there are few null values are present,We have to fill null values of categorical columns by mode() or by forwordfill() or bfill() of that particular column and fill null values of numerical columns by mean() or by ffill() or by bfill() or by interpolate() of that particular column.**
* **Though Our target column has only two outputs,its binary classification problem.**

**3.Exploratory DataAnalysis**

* **In this dataset there are 502 males and 112 females out of which 401 persons are married and 213 persons are unmarried.**
* **Out of 614 peoples 480 persons are graduate and other are not graduate.**
* **People living in semiurban area are applying for loan followed by urban and rural area.**
* **People living in Semiurban gets loan than urban and rural areas.**
* **Credit History column is highly corelated with our target variable.**

**4.Pre-Processing Pipelines**

* **We have to make our dataset ready by converting all categorical variables to numbers by using Encoding Techniques.**
* **We have to remove outliers from numeri columns by using zscore.**
* **Three numerical column has some skewness present,so we remove it by using power\_transform method.**
* **We divide dataset in independent and dependent variable and then divide it with train and test by using train\_test\_split method.**

**5.Building Machine Learning models**

* **We have to use different types of binary classification models(such as LogisticRegression(),SVC(),KNNClassifier(),RandomForestClassifier(),XGBClassifier() etc.) for best accuracy score.**
* **After using Roc-Auc curve we conclude that RandomForestClassifier() is the best model for this dataset.**
* **Later we hypertune RandomForestClassifier() to improve accuracy more.**

**6.Conclusion**

* **We choose RandomForestClassifier() model for this dataset because comparing to other models this model gives us good auc\_score and precision,recall and f1 score.**
* **After hypertuning the model,We get accuracy score of 82.96% .**

**Link of Python code :** **https://github.com/salman972/blogs/blob/main/loanprediction.ipynb**