Title: Loan Approval Prediction

# **Bhargava Gaggainpali DSC680 - Applied Data Science** [bgaggainpali/bgaggainpali\_DSC680 (github.com)](https://github.com/bgaggainpali/bgaggainpali_DSC680)

# Which Domain?

In financial industry, deciding on a loan application, whether to approve or not is a two-edged sword. Bank should not lose business by denying a legitimate customer, who can repay. Also, it should not approve loan to in-eligible customer. Banks are playing important role in challenging times like now, with COVID pandemic across the globe.

# Which Data?

I have identified LoanApprovalPrediction.csv as source for my work, below is the Kaggle link.

<https://www.kaggle.com/premptk/loan-approval-prediction-model/data?select=LoanApprovalPrediction.csv>

Variables:

**Variable Description**

Loan\_ID Unique Loan ID

Gender Male/ Female

Married Applicant married (Y/N)

Dependents Number of dependents

Education Applicant Education (Graduate/ Under Graduate)

Self\_Employed Self employed (Y/N)

ApplicantIncome Applicant income

CoapplicantIncome Coapplicant income

LoanAmount Loan amount in thousands

Loan\_Amount\_Term Term of loan in months

Credit\_History credit history meets guidelines

Property\_Area Urban/ Semi Urban/ Rural

Loan\_Status Loan approved (Y/N)

# Research Questions? Benefits? Why analyze these data?

With the enhancement in the banking sector lots of people are applying for bank loans but the bank has its limited assets which it has to grant to limited people only, so finding out to whom the loan can be granted which will be a safer option for the bank is a typical process. So in this paper we try to predict the customer, whether he is eligible for the loan approval and to reduce this risk factor of bank efforts and assets.

# What Method?

I see this problem as a classification issue, where we should try to understand and able to predict the customers, who have high Loan approval eligibility.

Planning to use supervised machine learning algorithm to work on the classification problem to be trained with algorithms like:

1. Logistic Regression

2. Decision Tree

3. Random Forest

Start with loading data into a data frame and then understand the data, then perform Exploratory Data Analysis (EDA) on the data set. EDA involves doing Univariate and Bivariate Analysis, identify missing values and outliers and fill the gaps with appropriate values. In the next step, building model with starting from logistic regression and observe the accuracy of the model. When the accuracy of the of the model is not high, then planning to use Decision Tree and Random Forest to achieve higher accuracy.

# Potential Issues?

Due to the low volume of the data that is available, I would assume model accuracy would be low. I am planning to use multiple algorithms, at least one would give expected accuracy in the range of 80-85. Also, I might encounter some data type errors while working with variables.

# Concluding Remarks

By taking a sample data set, with set of features, I would like to explore the features to find the pattern which majorly influence the loan approval process and the order of importance. For this to achieve, will try to clean, validate, perform exploratory data analysis and also use the machine learning algorithm to build the predicting model. Based on the data and model efficiency, I would assume the accuracy of the model will be between 80%-85%.

**Reference:**

[4. 18-21.pdf (iosrjournals.org)](https://www.iosrjournals.org/iosr-jce/papers/conf.15013/Volume%203/4.%2018-21.pdf?id=7557) - IOSR Journal of Computer Engineering (IOSR-JCE) - Loan Approval Prediction based on Machine Learning Approach

[An Approach for Prediction of Loan Approval using Machine Learning Algorithm | IEEE Conference Publication | IEEE Xplore](https://ieeexplore.ieee.org/abstract/document/9155614)

[Prediction of Loan Approval using Machine Learning | International Journal of Advanced Science and Technology (sersc.org)](http://sersc.org/journals/index.php/IJAST/article/view/460)

<https://medium.com/@rahulshuklawork/prediction-of-loan-approval-with-machine-learning-539cbd2aad31>

<https://towardsdatascience.com/predict-loan-eligibility-using-machine-learning-models-7a14ef904057>

<https://towardsdatascience.com/top-9-data-science-projects-for-a-beginner-in-2020-26eb7d42b116>

<https://www.kaggle.com/ajaymanwani/loan-approval-prediction/data>

[Bank Loan Approval Prediction With Artificial Neural Nets (coursera.org)](https://www.coursera.org/projects/loan-approval-prediction-using-neural-networks)

<https://towardsdatascience.com/predict-loan-eligibility-using-machine-learning-models-7a14ef904057>

<https://github.com/mridulrb/Predict-loan-eligibility-using-IBM-Watson-Studio/tree/master/Dataset>