Introduction

In the bustling realm of modern finance, where dreams of a better tomorrow are often woven into the tapestry of loans and credit, an invaluable tool lies at the heart of lending institutions' decision-making processes: the loan approval dataset.

This dataset is not merely a collection of numbers and figures; it represents the gateway to financial opportunities for individuals and organizations alike. It holds within its digital grasp the keys to understanding who is deserving of financial support and who may need to journey further to secure their goals.

At its core, the loan approval dataset is a treasure trove of financial records and related information. It shines a spotlight on critical factors, like the CIBIL score, income streams, employment statuses, loan terms, requested amounts, asset valuations, and the loan status. These data points are the building blocks upon which lending institutions make decisions that can shape lives and businesses.

The company seeks your expertise as an AI consultant, to create a predictive model. This model would have to determine whether an individual's loan application would be accepted or rejected.

Dataset Features Explanation

Feature	Explanation
Loan_ID	The loan ID which is irrelevant to the task
Nos_Dep	Number of dependents
Education	Education Level (graduated or not)
Self_employed	Self employed or not
Annual_income	Total Income in a year
Loan_amount	Amount of loan borrowed
Loan_Term	Term to repay loans (in years)
Cibil_score	CIBIL Score

Residential_assets_value	Total residential assets value
Commercial_assets_value	Total commercial assets value
Luxury_assets_value	Total luxury assets value
Bank_asset_value	Total bank asset value
Loan_status	Loan status (1=Accepted, 0= Rejected)

Evaluating Metrics

The metric we are using for this task is the F1-score. F1 score is a machine learning evaluation metric that measures a model's accuracy. It combines the precision and recall scores of a model. The accuracy metric computes how many times a model made a correct prediction across the entire dataset can be seen below.

$$\text{F1 Score} = \frac{TP}{TP + \frac{1}{2}(FP + FN)}$$

Expectations

We would like to see all the process from Data cleaning, Data Analysis, Data modelling, and finally communicating the results summary and presentation which includes recommendation on what to focus on in approving loans.