CIND 820 – Capstone Project

Topic : Financial loan approval prediction

By : Joveria Sadain

The exponential increase in population contributes to a growing demand for housing, which has implications for various aspects of society, including the need for housing loans. In order to fulfil this demand, many individuals rely on housing loans to fulfill their housing needs. The increasing demand for housing loans has a significant impact on financial institutions, such as banks and mortgage lenders. While housing loans are essential for many to achieve homeownership, there can be challenges. High-interest rates, stringent eligibility criteria, and the potential for loan default are factors that borrowers and financial institutions need to navigate. The Dream Housing Finance Company offers a variety of home loan options.

This project aims to apply different supervised machine learning algorithms on applicant’s various factors such as financial attributes, socio-economic indicators and geographical aspects to develop a loan predictive model for Dream Housing Company in US.

The objective of this project aims to develop a classification model for loan approval and identify which features have the most significant impact on loan approval decisions? In addition, the project also aims to identify if there are any geographical disparities in the likelihood of loan approval. Furthermore, what is the optimal loan amount and loan term for maximizing approval rates and can historical credit information effectively predict loan approval outcomes? Lastly, the plan is to analyze how well do different machine learning algorithms perform in predicting loan approval outcomes and can ensemble method improve model accuracy and robustness.

The project will use Python to pre-process data and develop the classifiers. EDA will be performed using ydata profiling or pandas profiling library. Feature selection will be used to identify the most impacting variables in predicting loan approval. The scikit-learn library will then be used in developing logistic regression, decision trees and KNN classifiers will then be developed and analyzed for accuracy. Furthermore, an ensemble method, Random forest can be applied for improving accuracy if needed.

The dataset is a publically available dataset on kaggle.

<https://www.kaggle.com/datasets/rishikeshkonapure/home-loan-approval>