**PROPOSAL FOR FIRST CAPSTONE PROJECT**

**Introduction**:

Credit defaults are a major source of revenue loss for banks. The total amount of bad loans in India was about 8.06 lakh crores in March 2019, (9.3% of the total credit). Even after carrying out an extensive analysis of the borrower profile before lending, it's difficult to ascertain repayment behaviour in the future. Conventional analysis of borrower risk profile depends on assets-liabilities analysis, income-expenditure analysis and inferences from past credit behaviour. In spite of this, the level of bad loans in our country is on the rise. A more robust, mathematical model for adjudication is required.

**Problem at hand:**

Deciding the probability of default by a borrower based on given attributes. This helps extend credit only to deserving customers.

**Value to client:**

Financial institutions could utilise ML models in adjudicating borrowers and improving credit portfolio. This will ensure proper and timely credit delivery to deserving customers, while weeding out untrustworthy and erratic customers. The total NPA amount in India is estimated to be about 6.93 lakh crore (on March 2020). If a model helps reduce this by just 0.1%, NPA amount will decrease by 693000 \* 10000000 \* 0.0001 = 693 crore rupees. It is a relatively small amount when considered on a national level, but continuous improvement may help increase predictive accuracy.

**The stakeholders:**

The top management of financial institutions are the first deciding and reviewing point. On approval, new processes trickle down to branch level.

**Dataset to be used:**

*https://www.kaggle.com/c/home-credit-default-risk/data*

Dataset is provided by “Home Credit Group” for a Kaggle competition (now concluded). Main data is stored in ‘application\_train.csv’, having 122 columns and about 3 lakh instances, and related data is spread across 8 tables.

‘Home Credit Group’ primarily extends loans to customers with little or no history of credit. Models developed using this dataset can be extrapolated to work on Indian loan defaults data.

**Basic Methodology**

1. Cleaning and preprocessing.
2. Carrying out visual EDA for getting a general sense of data and understanding the relationships between various attributes.
3. Ascertaining critical attributes and their effect on target.
4. Applying ML models to predict default probability.

**Deliverables:**

1. Plots showing the relationship between various attributes, and their relevance in final target prediction.
2. Model to predict the probability of borrower default and help the credit appraisal process.