**Group Members:**

1. Anish Shriram Kulkarni
2. Anoushka Gade
3. Satish Dallaram Choudhary
4. Saurabh Suryakant Sathe

**Vehicle loan defaulter risk prediction**

**Description:**Finance Organizations face huge losses due to defaulters. This has led to huge loan rejection rates. This has also adversely affected loan organizations and vehicle manufacturers by significantly reducing business. Hence there is a need for a more robust and efficient risk prediction system which can let the financial institutions know about the possibility of an individual being a defaulter. In this project we aim to accurately predict the probability of borrower defaulting on a vehicle loan in the first EMI (Equated Monthly Instalments) on the due date. Doing so will ensure that clients capable of repayment are not rejected and important determinants can be identified which can be further used for minimizing the default rates.

1. **Propose Methodology:**
   * 1. Dataset will be cleaned for missing values, noise and outliers.
     2. Dimensionality reduction techniques like SVD/PCA will be used to reduce dimensionality and extract attributes contributing to at least 85% of the variance in the dataset.
     3. We will be implementing classification models like Naïve Bayes, random forest algorithm and K- nearest neighbor for classifying the individual. Results from these algorithms would then be used in ensemble fashion to predict results for newer individuals.
     4. Precision, recall, F1 score and weighted f1-score will be calculated to validate the model.
2. **Techniques used:**

PCA/SVD for dimensionality reduction.  
KNN, Naive Bayes and random forest for classification algorithms.

Precision, recall, f1-score, weighted f1-score and accuracy as evaluation metrics.

1. **Datasets:**The data is in CSV format, and have 41 columns and 233154 rows  
   <https://www.kaggle.com/mamtadhaker/lt-vehicle-loan-default-prediction>