Proposal_A: Predict the probability of a Lendingclub loan getting defaulted

Motivation: Credit risk may arise if the consumer fails to make a payment due on a mortgage loan, credit card, line of credit, or other loans.
A credit risk is the risk of default on a debt that may arise from a borrower failing to make required payments.In the first resort, the risk is that of the lender and includes lost principal and interest, disruption to cash flows, and increased collection costs.

Solution: Build a optimized machine learning model that will provide valuable insights to Lendingclub investors for making informed lending decisions. My objective is to optimise resource usage and the learning time whilst decreasing the prediction error. This can be done by either challenging the current code implementation and trying to push it to its limits or by implementing different losses / optimizers (which can be generalized or custom).

Approach: Use logistic regression, multilayer perceptron neural network, and Random forest.

Additional details:
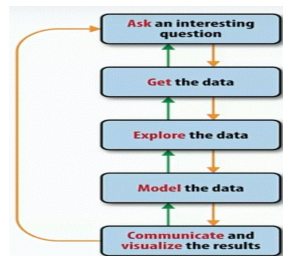
Ask an Interesting Question: What is the risk of a loan getting defaulted?

Get the data: Lendingclub website (https://www.lendingclub.com/info/download-data.action) provides historic datasets(2007-2018) for loan data and decline loan data. I may look into additional financial data as needed.

Explore the data:
- Data Wrangling: Patterns will be analyzed. I will also look for Spreads, Correlation, and Trends.
- Understand the predictors and targets in the dataset: I will check on how predictors are correlated to each other and understand relationship between predictors and target.
- Feature reduction: I will use PCA do eliminate attributes that are not predictors.

Model the data:
- Type of Machine Learning problem: Supervised as well as UnSupervised as both Regression and Classification techniques are involved.
- Build a model: Use logistic regression, multilayer perceptron neural network, and Random forest.
- Fit the model, Validate and optimize the model: I will constantly improve the model.

Communicate and visualize the results:
- Final Deliverable: I will provide an API to predict if a new loan will default. Jupyter notebook will be added to the project to demonstrate the findings and lessons learned.

Resources Needed: Python, Jupyter Notebook.

Proposal_B: Home Credit default risk

Motivation: Many people struggle to get loans due to insufficient or non-existent credit histories. Home Credit strives to broaden financial inclusion for the unbanked population by providing a positive and safe borrowing experience.

Solution: This project will build optimized Machine learning models that will help "Home Credit" to offer a positive loan experience to the underserved population.

Approach:
Use historical loan application data to predict whether or not an applicant will be able to repay a loan. This is a standard supervised classification task.

Additional Details:
Ask an Interesting Question: Should the loan be issued to the Customer?
Get the data: Kaggle provides datasets for credit applications, credit bureuau, credit balances, installment payments and more. The total size is 688 MB and available for download at https://www.kaggle.com/c/home-credit-default-risk/data.

Proposal_C: Credit card fraud detection

Motivation: Determine if a given transaction is fraud.

Solution: This project will build optimized Machine learning models to detect Fraud.

Approach: Random Forest, AdaBoostClassifier, XGBoost, LightGBM

Additional Details:
Ask an Interesting Question: Is the current transaction classified as fraud or legitimate?
Get the data:
Datasets are available at https://www.kaggle.com/mlg-ulb/creditcardfraud