Summary Report
Data Science Technocolabs Internship

Abhishek Tyagi
Team A
Batch - 20 June

Title- Optimizing Lending Club's Financial

Aim- The main purpose of this project is to Analyze given data and design a machine learning model to predict loan repayment Ability of the client, whether or not the client will be able to pay off the loan.

Abstract- We analyzed LendingClub's dataset of roughly 2.2M loans between 2008–18. We chose to only analyze loans that were paid off in full, charged off or defaulted in this case. There are over 400 borrower characteristics at time of application and loan characteristics at time of issuance.

Methodology -

1) Different models predict default probabilities.
2) The relative value provided by the predictions from these models was then evaluated using EMP estimation. EMP is a metric of comparison between classifiers. It can be interpreted as an upper-bound on the additional profit gained by using the classifier versus performing no classification.

Overview -

EDA- Exploratory data analysis
Building Logistic regression model
Building neural network
Deploying the ML model

Dataset - Dataset contains the details of all the accepted as well as rejected clients during the period 2008-2018.

EDA (Exploratory data analysis)-

1) Eda started with data cleaning in which we first tried to analyze given data and then step by step proceeded with cleaning the unrequired rows and columns.
2) Changed data types of various rows as per our requirement.
3) Dropped unnecessary columns and rows.
4) Replaced all the NaN values to sensible entries.
5) By using the techniques of data visualization various plots such as barplots, scatter plots and line plots were used to showcase trends and data.

Logistic regression model -

1) We use stochastic gradient descent and minimize cross entropy loss.
2) All the objects present in the data were converted into labels using label encoder.
3) Training model was trained using sklearn.
4) Model was evaluated and an accuracy score of 92% was obtained.

Neural Network model -

1) We use a fully connected, 5-layer network with hidden layers of shape (89,89,45,20,2) and ReLU activation.
2) All the objects present in the data were converted into labels using label encoder.
3) Training model was trained using sklearn.
4) Model was evaluated and an accuracy score of 99% was obtained.

Model Deployment -

1) Once both models were successfully implemented the next task was to deploy the models using flask and streamlit.
2) Flask Implementation- Contained three files ‘model.py’,’index.html’,’app.py’.
3) Model was deployed on the heroku platform successfully.
4) Steamlit implementation - Contained two files ‘app.py’ and ‘model.py’.
5) The model was deployed on stramlit platform.
6) Deployed models used thirteen key features from the database table and gave out predictions about whether or not the loan was fully charged off.

Result - All the steps were successfully implemented and the whole process gave out an effective and efficient result of loan repayment ability.