Natural Language Processing with Deep Learning

Lending club defaulter prediction

DESCRIPTION DATA SUBMISSION

Problem statement:

In this project we are working on the `Lending club` dataset. Lending Club is a peer to peer lending company based in the United States, in which investors provide funds for potential borrowers and investors earn a profit depending on the risk they take (the borrowers credit score).

From the given set of data we want to predict `loan_status` of the borrower. We have to predict the loan staus based on the features like `Loan amount` , `payment plan` , `grade` , `verification status` , `recoveries` etc. The loan status having the various categories like `Fully paid` , `charged off` , `late` , `Issued` , `In a grace period` etc.

About the dataset

A zipped file containing following items is given:

- `train.csv`

The data file `train.csv` contains the `621165` loan status entries with the `80` features including the target feature.

- `test.csv`

The datafile `test.csv` contains the `266214` loan status entries with the `79` features excluding the target feature.

- `sample_submission.csv`

Explained under the `submission` sub-heading

- `LCDataDictionary.csv`

The file contains data dictionary(Dictionary explaining what each feature of the dataset means) of the lending club dataset

- `challenge_entity_student_template.ipynb`

A template notebook explaining the task breakdown to solve the given problem statement (***Learners***