**Data606\_Gsahaw\_Hailemichael**

**Related Work**

There have been many studies on classification models predicting LendingClub loan default. According to the data, examining borrowers’ characteristics and their influence on borrowers’ default is important. The study shows in both paper are in consensus that the Credit Grade assigned by Lending Club is the best predictor for borrowers’ default. Moreover, Revolving Credit Line Utilization is another variable influencing the default rate mentioned in all. Findings of other default determinants vary. The first factor is the selection of variables potentially having an impact on borrowers’ default. For instance, both references [1 & 2] found out that the FICO score has an influence on default. In the research paper, the dataset used also another factor creating differences between the findings for instance, the timeframe. Reference [1] used only 36-month loans. Reference. [2] Used both, 36- and 60-month loans also classification of loan status or type of loan length might be the cause. Borrower characteristics can be used to predict a loan’s default chances. However, since default determinants depend on the loan’s risk class, caution is warranted. What seems to be a good predictor of loan default based on overall data may not be reliable in the highest loan risk class.

**Data Cleaning and Exploratory Analysis**

In order to achieve our goal, it is important to identifying the relevant features from a set of data and removing the irrelevant or less important features, that does not contribute much to our target variable. Such as dropped the columns, which had more than 50% of missing data and imputed the remaining attributes of missing data using median for continuous variables and mode for categorical variables. Also removed the columns that has no direct relation to our analysis such as ID’s and other unique identifiers, Removing Redundant Attributes that carried same or similar information values for employment length – such as 10+ years,, We labeled our data based on the features provided in data. “Default", "Charged Off", "Late (31-120 days)", "Late (16-30 days)", "Does not meet the credit policy. Status:Charged Off“ were considered as defaulted loans.

References,

1. Emekter, R.; Tu, Y.; Jirasakuldech, B.; Lu, M. Evaluating credit risk and loan performance in online Peer-to-Peer (P2P) lending. Appl. Econ. 2015, 47, 54–70.
2. Carmichael, D. Modeling Default for Peer-to-Peer Loans. 2014. Available online: http://ssrn.com/abstract= 2529240 (accessed on 31 August 2018).
3. Determinants of Borrowers’ Default in P2P Lending under Consideration of the Loan Risk Class by Michal Polena and Toblas Regner (2018) <https://www.mdpi.com/2073-4336/9/4/82>