Workflow for Phase 4 – FP\_GROUPN\_9

There are 7 different sources of data:

*application\_train/application\_test:* The data for training and testing with information about each loan application at Home Credit. Each loan has its row as feature SK\_ID\_CURR as an identifier or a unique key. The TARGET of training application data has two values indicating 0: indicated the loan was repaid or 1: the loan was not repaid.

*bureau:* data from other financial organizations about the client's prior credit. Each previous credit has its own row in the bureau.

*bureau\_balance:* monthly information about past credit history in the bureau. A previous credit can include numerous rows, one for each month of the credit period.

*previous\_application:* Past loan applications made by customers with loans at Home Credit are included in the application data. The application data allows for many prior loans for each current loan. The feature SK ID PREV serves to distinguish each previous application, which contains one row.

*POS\_CASH\_BALANCE:* monthly information on prior point-of-sale or cash loans that customers have taken out through Home Credit. A single previous loan can have numerous rows, each representing a month from a previous point of sale or cash loan.

*credit\_card\_balance:* data about prior credit cards that Home Credit customers have had on a monthly basis. Every row represents a month's worth of credit card debt, and a single credit card may have several rows.

*installments\_payment :* history of payments for prior loans with Home Credit. Every made payment has its own row, and every missed payment has its own row.

Tasks to be tackled:

1. Neural Network Implementation

2. Modeling Pipelines using Loss Functions

3. Checking for Leakage in the pipeline

**Workflow:**

1. We make sure to use the

given dataset as much as

possible by tweaking it

wherever necessary.

2. To tweak the data, we need

to visualize the dataset and

understand it. Therefore, we

perform EDA.

3. Once we gather the required

data, we plan to performed

feature engineering in Phase 3 and

selecting the essential

attributes that contribute to

the accuracy of the data and put it

into the dataframe called final\_features

which is the dataset we will be used to train

our model with.

4. We will be designing neural network

architecture.We plan to perform experiment

with atleast two different network architectures

by playing around with activation functions and

hidden layers and report the findings.

5. Later we plan to check for any leakage

present in our pipeline

6. We record the results

generated.