Question 1 In you own data set, is the task supervised / or unsupervised learning? Is it classification or regression? Provide the target variables and corresponding loss functions

Answer: The Data set tasks are supervised as data set is 122 features in application csv.

TARGET
NAME_CONTRACT_TYPE
CODE_GENDER
FLAG_OWN_CAR
FLAG_OWN_REALTY
CNT_CHILDREN
AMT_INCOME_TOTAL
AMT_CREDIT
AMT_ANNUITY
AMT_GOODS_PRICE
NAME_TYPE_SUITE
NAME_INCOME_TYPE
NAME_EDUCATION_TYPE
NAME_FAMILY_STATUS
NAME_HOUSING_TYPE
REGION_POPULATION_RELATIVE
DAYS_BIRTH
DAYS_EMPLOYED
DAYS_REGISTRATION
DAYS_ID_PUBLISH
OWN_CAR_AGE
OCCUPATION_TYPE
CNT_FAM_MEMBERS
REGION_RATING_CLIENT
REGION_RATING_CLIENT_W_CITY

These are the features that can be important I found that can help to understand the people about their credit history and reasons of rejection of their loan or credit they applied for.
The loan data can be classification as we need binary answers about credit approved or not.

Question 2: What is your strategy to split your total train data when you are training models? In your setup, how many models you will have eventually? How can you use these trained models to predict for new input?
Answer: The Dataset can be divided in 3 parts (Training dataset, Test dataset and Validation dataset). Training dataset will train the classification model. Test dataset is to test the trained model and validation will further test after a great model is trained. Thus, it will revalidate trained model for perfection. We could provide same input parameter as we did previously with features required to train model and result can be predicted.

Question 3: imbalanced data set in your own data set, if it is the classification task, please analyze the ratio between two classes (1 and 0)? If the ratio is very imbalanced (imbalanced data means the amount of 1 or 0 is extremely small comparing the other class), what is the possible methodologies to train models?
Answer: In imbalanced data set we can analyze data with some of the different techniques

- Using Right evaluation metrics
- Resampling Dataset

- Using K-Fold Cross validation
- Ensemble Different resampled data
- Resample with different ratios
- Cluster the abundant class

Resource: https://www.kdnuggets.com/2017/06/7-techniques-handle-imbalanced-data.html