EE5183 FinTech Final Project Report- Project Churn Prediction (Group 5)

T. Prayoonkittikul, B. Vatcharakomonphan, S. Suttawuttkwong, T. Pornkraisri, A. Pongsawang

***Abstract*— The telecommunication market is one of the toughest markets as customers are always free to choose from plenty of providers even within one product category. The fundamental method to reduce costs and generating revenue is to focus on retaining existing customers rather than acquiring a new client because the latter one is costs more. It will be tough and waste of the resources to ingratiate all of the customers. Thus, it would be better if we know which customer is going to leave our service. That when Churn prediction come in to play, it is an implementation to detect when customers want to leave a company so that the company can provide special services and focuses on earning these customer loyalty and satisfaction.**

I. INTRODUCTION

Let's begin with what is consumer churn and their importance. Customer churn is whilst a patron or subscriber forestall his or her relationship with a business enterprise. Usually, businesses understand a client as churned whilst a selected amount of time has exceeded on the grounds that the patron's final interplay with the employer's service. The value of churn consists of the advertising fee that organizations ought to replace those customers and the revenue of the gap when there's no replacement. Therefore, lowering this value is one of the key business desires for many telecom businesses.

Having the capability to predict that a specific patron has the capacity to churn and making the business enterprise have a while to do something can clearly save cost and help boom revenue as the fee can be from direct loss of sales resulting from a client unsubscribing whilst the client's spending to date cannot cowl the acquiring cost. Also, the fee of acquiring new customers is normally more luxurious than retaining the modern-day ones.

There are many strategies used within the beyond for predicting consumer churn for example gadget studying algorithms which include Logistic Regression, Decision trees, Random Forest, SupportVectorMachine, NaïveBayes, NeuralNetworkand many more. Another interesting approach to expect churn for telecom organization is by the use of a gadget learning technique referred to as Deep mastering.

It is a multiple layer of illustration with a couple of nodes in every layer connecting and shape an artificial neural network which is a network that imitates of the way the human brain works. By the use of this method, we are able to build version that suits our information the usage of diverse layers of concepts to be able to help boom the overall performance of the model built.

II. MATERIAL AND METHOD

We obtained an open-source dataset from Kaggle. perform an Exploratory Data Analysis (EDA) so, our data set quite organized. The only thing we have to do is remove some rows that have null value and cast '*TotalCharges'* column to float. Then split the data to training set 80 percent and test set 20 percent with *train_test_split* method. Then use this particular to train all of the models with 10 k-folds cross validation.

A. Baseline model

Logistic regression – use l2 penalty and C value is 1.
SVM – C value is 100, gamma is 0.0001 and kernel is rbf
Random forest – n estimators is 300, max_depth is and use gini criterion

B. DNN model

```
Model: "sequential"
_________________________________________________________________
Layer (type)                 Output Shape              Param #
=================================================================
dense (Dense)                (None, 46)                2162
_________________________________________________________________
dense_1 (Dense)              (None, 32)                1504
_________________________________________________________________
dense_2 (Dense)              (None, 32)                1056
_________________________________________________________________
dense_3 (Dense)              (None, 32)                1056
_________________________________________________________________
dense_4 (Dense)              (None, 1)                 33
=================================================================
Total params: 5,811
Trainable params: 5,811
Non-trainable params: 0
_________________________________________________________________
```

For our project, we use 3 common machine learning approaches which we think work the best for classification problems as the baseline which consists of Support Vector Machine, Logistic regression, and Random Forest.

The approach we choose is to develop a deep learning model. By doing grid-search, we acquired the best settings for our model. The model consists of 5 layers (1 input layer, 1 output layer, and 3 hidden layers). Each hidden layer has 32 nodes. The input layer and all the hidden layers use the rectified linear activation function. The output layer uses sigmoid function as the activation function. We choose Adam optimizer as our optimization algorithm due to its computational efficiency. The lost function for the model is Binary Crossentropy. The

learning rate is 0.003, the batch size is 64, and the number of epochs is 200. We also add kernel regularizer L2=0.1.

III. Evaluation

	Predicted Churn (Sensitivity)	Predicted Non-Churn (Specificity)	Accuracy
Logistic Regression	63%	85%	79.6%
SVM	65%	83%	79.3%
Random Forest	67%	84%	80.4%
DNN	66.9%	84.5%	80.5%

IV. Discussion

We try to build a DNN model that outperform the baseline models. As the result shown in the table, the overall accuracy of our model only reaches the same level as the baseline models even though we have conducted grid search multiple times with many hyperparameters.

Here what we have found from grid search:

- There has to be at least 3 Hidden layers
- Dropout doesn't help increase the accuracy or reduce the overfitting in our case
- L2 regularizers help stabilize the model since every time we perform the 10 k-folds cross validation, the accuracy of the training set was 80%

We hope that our model will perform better if we acquire some more data so we plot accuracy vs data size graph. From the graphs, we have two assumptions since the trend of the blue line doesn't show the obvious increase in terms of accuracy. First, the DNN model is still overfitting because the blue line graph is relatively constant as the accuracy when there is less data compared with when there is more data are nearly the same. The second assumption is that our data is too small to fully capture the increasing trend.

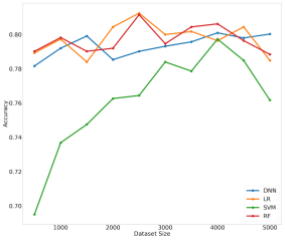

Fig.1

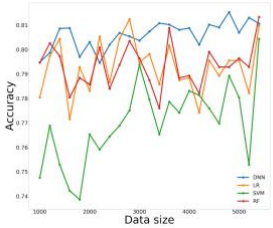

Fig.2

Fig.1 and Fig.2 represent increase 500 data units per iteration and 200 data units per iteration, respectively.

For work distribution, Satsawant Suttawuttiwong doing on exploratory data analysis (EDA). I and Anon Pongsawang responsible for training the DNN model and baseline models. We try to tune the parameter via by gridsearch and find the way to reduce the overfitting of the DNN model. Boonsita Vatcharakomonphan is a project manager that take care overall of our project progress and also working on poster. Thunyasorn Pornkraisri is responsible for researching the related paper work, provide more information and insights for us and also working on the Powerpoint.