Good morning everyone, thank you for the opportunity. I am Depus, hereby bringing the presentation of predicting customer churn of Telco System. With the goal is to find best model for predicting Churn customer within the actual churn customer. The dataset is obtained from Kaggle.

Here is the outline for today’s presentation: first we will go to exploratory data analysis, data preprocessing, model evaluation, model inference, and then the last is conclusion and future improvement.

From the exploratory data analysis, it is known that 7 from 10 customer of Telco is retained customer. It is a good insight of the company’s condition. But, this also indicating that before the data could be used in the dataset, this data should be handled for data imbalance. Therefore, smote will be used to ensure that the model have enough data to learn about churn customer characteristic.

Here are the features that doesn’t affect customer churn: Gender, Phone Service and Multiple lines. It is indicated from there are no significance difference in the proportion of one or more category in the feature. It means that there is no difference whether the customer is in one or another category of the feature.

Ok, next we move on to the features that affect the target: as can be seen in all graphs, there is significance difference between one or more category within a feature. Those features are:…………….

Here are the steps done in the data preprocessing:

Splitting dataset, Feature Selection, Missing value handling in Total Charge by Trimming, Outliers checking which is found that there is no outliers in Monthly Charges & Total Charges, feature Scaling and Encoding, and then the last data imbalance handling both for train & validation dataset.

Now, let’s proceed to model evaluation:

The first model is sequential model

Parameters used in the model are stated in the slide and graph of loss and accuracy is in the right showing the result of model training. From the graph it can be seen that there is no exploding gradient and there is reduction in loss and improvement in the accuracy. The result is good but accuracy score showing that the model is overfit and furthermore classification report shows an underfit.

The second model is functional model where the result of training shows no exploding gradient and improving performance. Both test and train evaluation result show a good fit.

The third model is improvement model with improvement is indicated with green, where loss and accuracy graph show a slight gradient exploding but still acceptable. Both test and train evaluation result show a good fit and better performance than the functional. Therefore, this model will be used for model inference and deployment.

Then we proceed to model inference, the result of model inference is 5 misprediction out of 30 data, and the accuracy is good which is 83.3%.

At last, let’s continue to the model deployment:

For the overall conclusion, the model used 16 out of 20 categories, improved model has best performance with goal of higher recall, and the accuracy obtained from inference data is 83.3%. For future improvement we can try another parameter tuning, using functional model and adding features that excluded in previous model.