Hi, this is Chuanhai Zhu. My capstone project at data incubator is the risk prediction for this is alice. This is alice is a startup company at Boston, which makes the 30-day Metrocard more available by letting people pay in weekly installments. They send the payment history to the credit bureaus so people can build credit. The goal of this project is develop a predictive model help this is alice predict the potential default risk of their customers based on their communication patterns.

In order to achieve this goal, over 1000 customers information were extracted from the sql database, and converted into pandas dataframe. Then, different machine learning algorithms were used to developed models, the models’ performance were visuized and compared, the model with highest accuracy were delivered to this is Alice.

Based on the charge-off and payment failed history, the customers were classified into three groups, which are safe, votaile, and risk. Then, based on customer’s text messages with This is Alice, three models were developed, using logistic, svm and random forest. Then, their prediction results of testing set were compared. Based on the confusion matrix here, random forest has the best performance, which has over 96% accuracy. As a result, this model was the final model for this is Alice.