Machine Learning Model for prediction of Customer Attrition

By Sattam Ghosal

Problem Statement

- Given a csv dataset containing different feature details like Customer_ID, Weeks, Contract_Renewal, Data_Plan, Data_Usage, Calls_To_Customer_Care, DayMins, DayCalls, MonthlyCharge, OverageFee, RoamMins, Customer_Attrition, Inserted_Date predict the Customer Attrition from the given data.

Steps

- Preprocessing
- Filling the missing data
- Fitting a model
- Hyperparameter tuning
- Prediction

Preprocessing

- The raw data needs to be preprocessed
  - Scaling is applied to non-catagorical data
  - Label Encoding and One Hot Encoding are added for categorical features

Filling the missing values

- The missing values are filled using RandomForestClassifier(Categorical Features) and RandomForestRegressor(Normal variables)

- Random forests were used because it was outperforming the classical Machine Learning Models and also prevents the model from overfitting

- ANNs were not used as I had used them in the future for actual prediction.

Fitting A Model

- An artificial neural network was used for the final model.

    – It consisted of 5 Dense Layers(including 1 input and one output layer)

    – Each Layer was followed by a batch norm and one dropout layer

Hyperparameter Tuning

- The final model has the following specifications:
  - Dense Layer(4)(followed by BatchNorm layer and Dropout layer)
  - For Dropout layer p=.2-for all
  - Optimizer=adam
  - Final layer activation=sigmoid
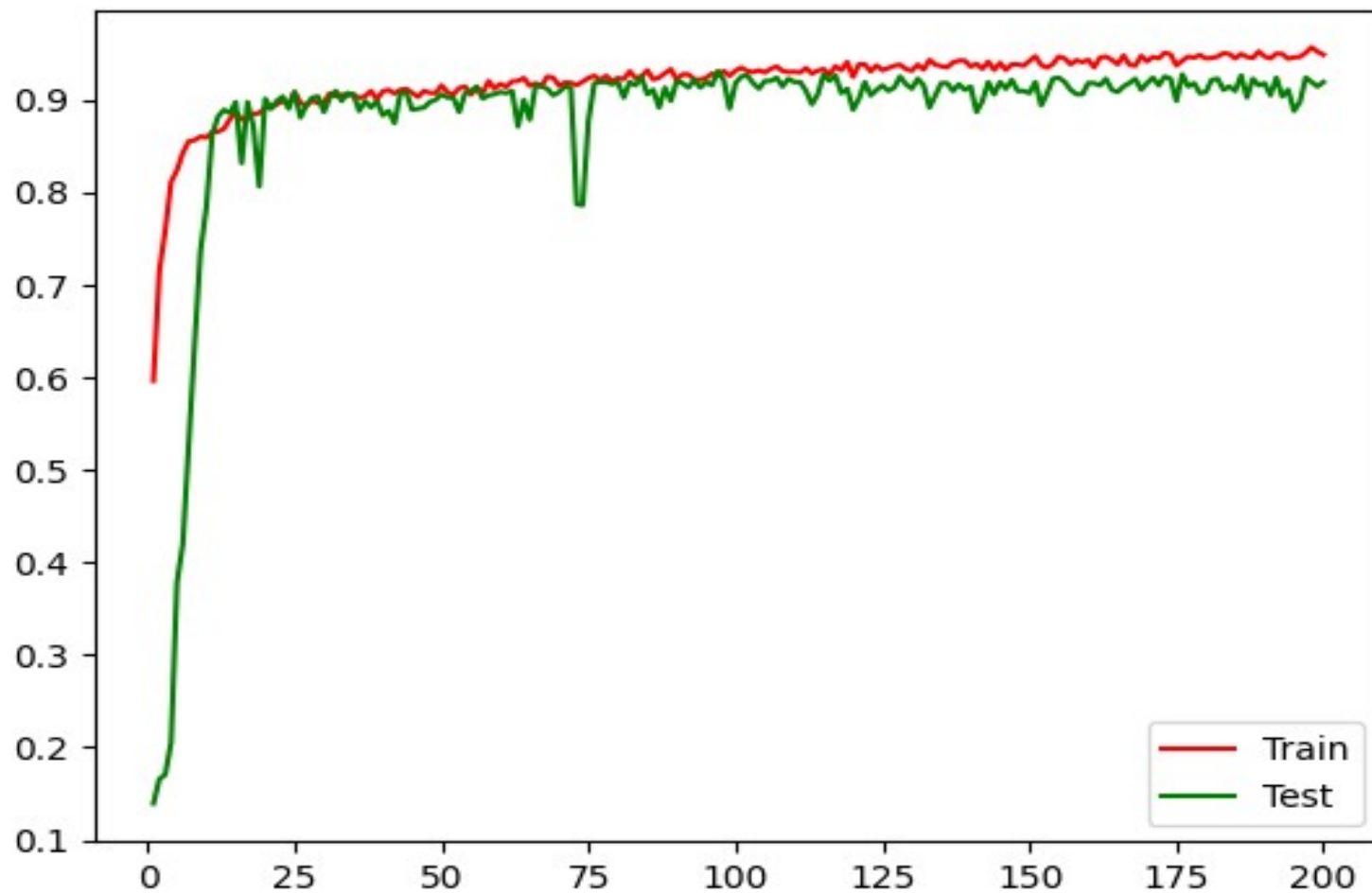
Accuracy

- Training accuracy: 94%

- Validation accuracy:92%

- Confusion Matrix:

	Predicted(Yes)	Predicted (No)
Actual(Yes)	*699*	*19*
Actual(No)	*46*	*69*

Thank You

- *Thank You*