Competition 2016

Predict a purchased policy based on transaction history

INFO 7309 Machine Learning for Business Intelligence

Team: Data Wizards

Leela Gangadhar Vallabhaneni

Nuhiya Rafeeq

Sumit Deshmukh

Department of Information Systems,

Northeastern University

Date: 03/30/2016

Objective:

Develop a model for an Insurance Company which would accurately predict the policy number that a customer is likely to purchase and its price, given transaction and purchase history of old customers.

Approach:

Calculate Correlation Co-efficient and Information Gain for the attributes in train.csv. Since there are four policy options available to the customer, the model can be trained by applying the Classification Technique for the Target Variable ‘policy’. As the price of the policy depends on the customer characteristics and the policy selected, Linear Regression can be applied to estimate the policy ‘price’ to quote.

Information to be used:

Train.csv will be used to train the model. The trained model will be used on test\_session\_history.csv to test the accuracy of the model. Although Cross-validation is used for limited datasets, it may be considered to obtain a model with higher accuracy if Classification Techniques do not yield the desired accuracy.

Performance Measure:

(rough)We will perform sensitivity analysis and calculate accuracy for model. Success criteria: we would calculate the absolute errors between predicted price and actual price on test\_session\_history.csv. We will define the tolerance value, any predicted value for ‘price’ which lies within the tolerance range is considered to be successful prediction

Success criteria:

(rough) Decision tree, logistic regression or SVM whichever gives the highest performance, will be used to select… depending on hiheger accuracy is picked. For regression, we will use linear regression or other techniques to check model accuracy.