Department of Industrial and Systems Engineering

ISE 560, Stochastic Models in Industrial Engineering

Course Project Proposal under the Guidance of **Professor Dr. Julie Ivy**

Using Sentiment & Telemetry Data to Predict Customer Satisfaction

INTRODUCTION

The Lenovo Group is one of the largest technology companies in the world. For a technology company like Lenovo, early and efficient NPS Scores of its products plays a key role in providing best quality Products and thus, driving customer satisfaction. The Project requires us predict the customer satisfaction based on the CID Sentiment/ Star Rating given to different Lenovo products on different web platforms like Amazon, Best Buy and others based on different factors which governed the customer sentiment and Star rating to provide the predicted NPS score which Lenovo gets from its NPS Survey after a duration of 6 months to two year after selling a product.

The telemetry data that Lenovo receives for a window of 28 days from Microsoft on its different hardware components can also be used to predict the NPS Score and Star Rating which the Customers will give based on the optimal functioning of the hardware components, thus helping retain the customer count and getting a Positive NPS Score.

The scope of the project is restricted to obtaining the NPS Score of the future NPS Survey based on the current Web Sentiment and telemetry data and then take suitable actions based on the Markov Decision Process to retain the customer loyalty and increase in customer count.

OVERVIEW OF THE PROBLEM

Lenovo uses customers' responses to the NPS survey as their primary measurement of customer satisfaction The problem that Lenovo faces to get the NPS Score is a very time taking process as the surveys are sent to customer in the first 30 days of their purchase but it takes time to collect them and get enough responses to be relevant. Unfortunately, feedback from the NPS survey results arrives rather late (approx. six months after product release) in the product life cycle and that to with not a 100% Accuracy. The Formulas for PSI, Star Rating and pNPS industry standard calculations is not upto the mark and in time which ultimately affect the actions that needs to be taken to resolve a particular issue faced by the customers and thus affecting sales and brand value.

One is not able to find the right Actions in terms of Cost efficiency and correct prediction of NPS Score for the future so Lenovo would ideally like to be able to predict NPS score based on the customer sentiment/star rating and the telemetry data.

OBJECTIVE OF THE PROJECT

The objective of the Project is to perform the following given below –
* To predict an analytical model using CID sentiment/stars data to predict pNPS Survey scores.
* To find how the Telemetry data influence pNPS Survey Scores and sentiment/star data

- To predict the NPS score by testing the predicted predictive analytical model on 5 products which have been eliminated at first place from our CID sentiment data and the influence of telemetry data on these 5 products.
- Based on our results, provide actions for Lenovo to focus on to improve pNPS, to focus on to improve total pNPS for all products and what should Lenovo focus on to improve pNPS for bottom 3 Customer and bottom 3 Commercial products.
- To find what impact do the products stars have on the analysis.
- To propose and recommend any changes based upon our results to Lenovo on the data collection methods.

METHODOLGY AND STRUCTURE OF THE MODEL-

The Project will be divided into three broad phases – Data Preparation and Analysis, Predictive Modeling on the basis of Web Sentiment and Telemetry Data and Stochastic Optimization using Markov Chain and Markov Decision Process.

DATA PREPARATION & ANALYSIS

Lenovo provided us with 3 files containing Web Sentiment Data from Customers across the globe, Telemetry data from Microsoft and actual NPS scores from their customers from their primary Survey. We will be sorting the 142224 rows in the Web Sentiment Data based on Product Type in order to get 402 rows (402 products given in the whole data after predictive analysis) corresponding to similar products on the NPS data file in order to obtain data frame to carry out the predictive analytical modeling. Similarly, Telemetry Data is prepared according to the Lenovo Product Classes on Lenovo's Product Class Star Rating in Web Sentiment Data and actual NPS data file. The data cleaning will be done using SAS and Excel to clean and sort the data and then export to a csv file format on which regression can be done. We will also be sorting data based on Consumer, Commercial and SMB products to do regression on individual categories.

PREDICTIVE ANALYTICAL MODELING

After Data Preparation, we will be doing predictive analytics on a regression model method –multinomial logistic regression and linear regression to predict customer satisfaction using customer sentiment data for overall Lenovo products and we will also be doing these regression methods on Product Categories data frames of Consumer, Commercial and SMB Products as well as Principal Component Analysis and Linear regression model on telemetry data. Exploring alternative metrics such as Star Rating , Ownership Period, Location, Product Category (Commercial, Consumer and SMB) and Actual Lenovo products in each category as model of measure for quantifying customer satisfaction in addition to pNPS and sentiment data.

We will be finding important factors that influence the customer satisfaction in each predictor category and how these factors evolve over time.

Our results will include probabilities of the Negative, Neutral and Positive Sentiments for overall Lenovo Products as well the probabilities for Sentiments in each Products Category (Consumer, Commercial and SMB) The results will be able to tell us what are prediction of each product and the Product Categories along with the Location that influence the most. The telemetry models will help and tell if telemetry data will influence the Star Rating and NPS Scores or not. Thus, the Models will help in us in finding optimal results for intervening to influence customer satisfaction based on learning from sentiment. And also provide Probabilities for our transition states.

STOCHASTIC OPTIMIZATION USING MARKOV CHAIN AND MARKOV DECISION PROCESS

In this Phase, the project team is developing a stochastic model of the evolution of sentiment and satisfaction. We will be able to explain the evolution of customer's perceptions of various Lenovo products following release over time and provide certain actions that must be taken by Lenovo to improve their NPS scores. We have defined 3 States in Sentiment form which are Positive Sentiment, Neutral Sentiment and Negative Sentiment from our Predictive Model. We will derive probabilities for the Positive, Neutral and Negative Sentiment from regression model and thus we are able to create First State Transition Matrix. Actions chosen are to do nothing and do something to improve NPS and Rewards are our positive NPS Scores which needs to be achieved.

4. Identify states that can be used to develop a Markov chain including the time period associated with a transition. Define states, actions, rewards/costs and how you will use the data to calculate. (This will be used to solve a Markov decision process).

The states that can be used to develop a Markov Chain are Positive Sentiment, Neutral Sentiment and Negative Sentiment associated for a duration of 28 as per the telemetry data and 6 months corresponding to the Actual NPS scores.

States are defined as the particular condition of a customer sentiment that customer has at a specific time period. Actions defined here is to do nothing or do something which can bring out change in the NPS scores for a Product. Reward is defined as the positive NPS which Lenovo needs to attain in order to conclude positive customer satisfaction for their products.

The Markov Chain will have Probabilities for each of the customer sentiment and then based on our actions, we will attain different Markov Chain based on which rewards will be calculated and the best rewards (positive NPS Score) will give the best decisions that needs to be taken by Lenovo.