- **METHODOLGY AND STRUCTURE OF THE MODEL-**

The Project is divided into three broad phases –

1. Data Preparation and Analysis
2. Predictive Analytical Modeling
3. Stochastic Optimization using Markov Chain and Markov Decision Process.

## 1. 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.

Data Validation and Cleaning was done on the 142223 rows in the Web Sentiment and 48921 rows in the NPS dataset & the data was divided based on product type (Consumer & Commercial Products) to get 142 unique products. For each category, the unique products were filtered out to match with the web sentiment & the NPS Survey data. There were 74 consumer & 68 commercial products identified after initial filtering process. The unique products are shown in Appendix A.

The unique products were defined based on their Star-Sentiment evolution. There were 15 states defined for the analysis with 5 Star ratings (ranging from 1 to 5) and 3 Sentiment evolutions (positive, neutral and negative). The evolution of the star-rating and sentiment for each product was observed on a weekly basis. The figure below explains the evolution of the star-sentiment for the 5 products (these are the same 5 products whose NPS was supposed to be calculated).

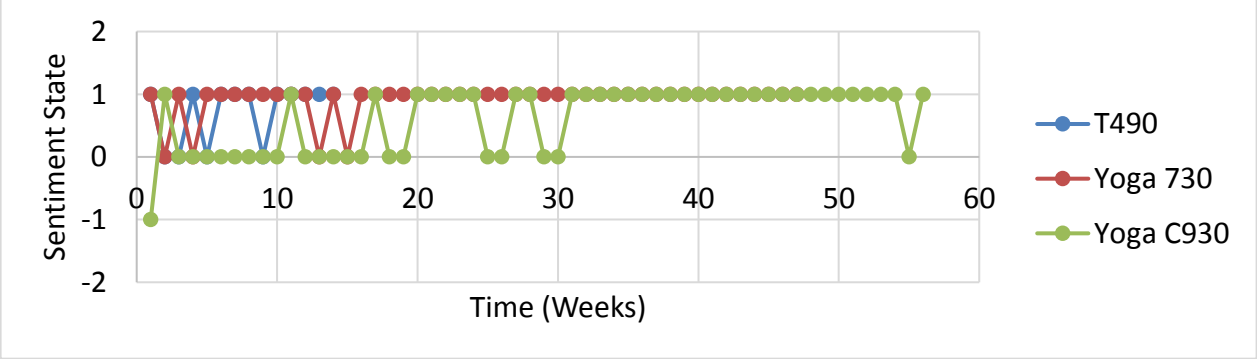


Figure 1: Sentiment Analysis of products based on weekly epochs

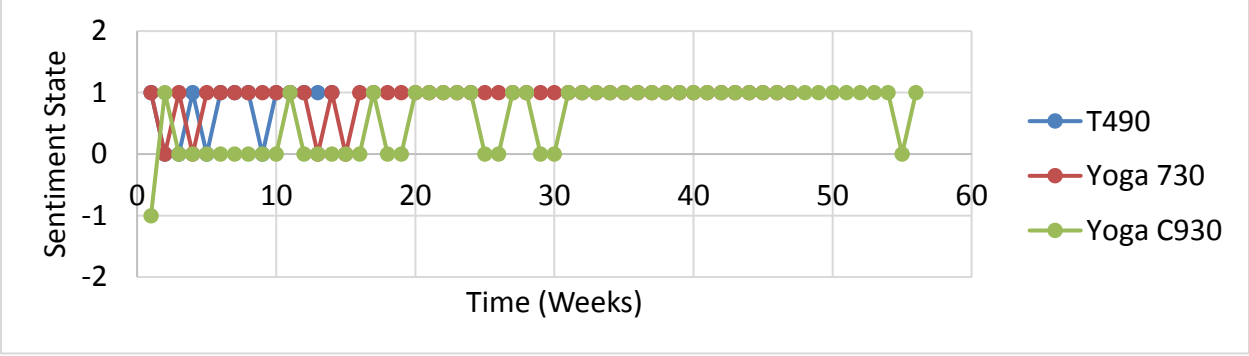


Figure 2: Star Analysis of products based on weekly epochs

## 2. PREDICTIVE ANALYTICAL MODELING

NPS survey and Web Sentiment for each product category was obtained after data filtration & data cleaning and exported to R Studio software for analysis. Analytical Code was developed to match each product in both the datasets (Web Sentiment & NPS Survey). Moreover, calculation of the product sentiment through PSI and overall Star Rating for each Product in each category was done and were used as significant factors to formulate the NPS Score & two Multiple Linear regression model with PSI and Star Rating (Consumer and Commercial Models) were selected among Liner Regression Model with separate PSI and Star Rating, Multiple Regression Model with PSI & Star Rating and Multiple Regression Model with Interaction to predict the NPS Score of the required five products.

Likewise, the NPS Score was predicted for the required five products based on this model.

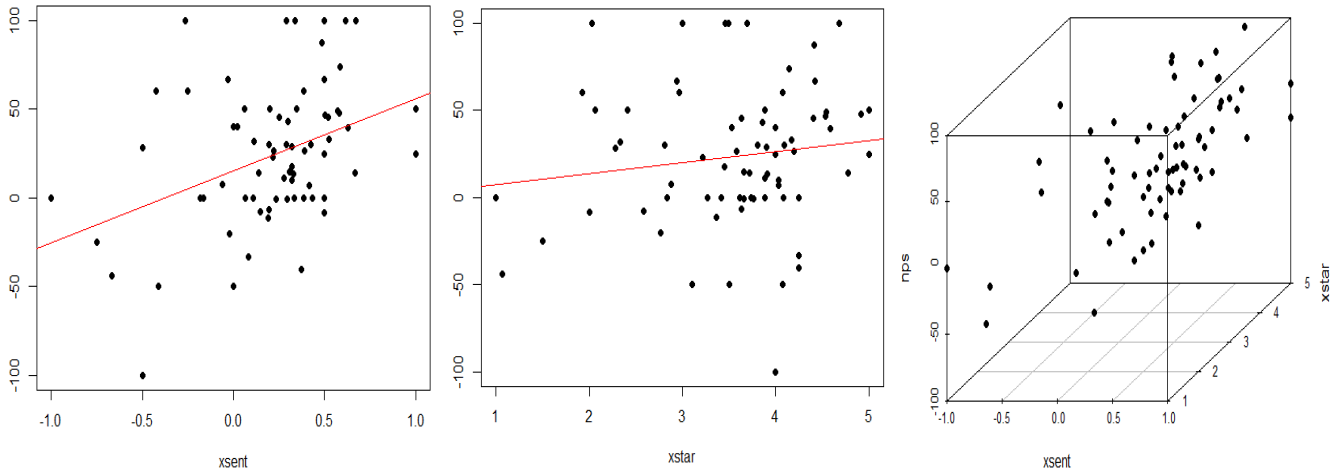


Figure 3: Consumer products NPS variance compared with the average sentiment & star score

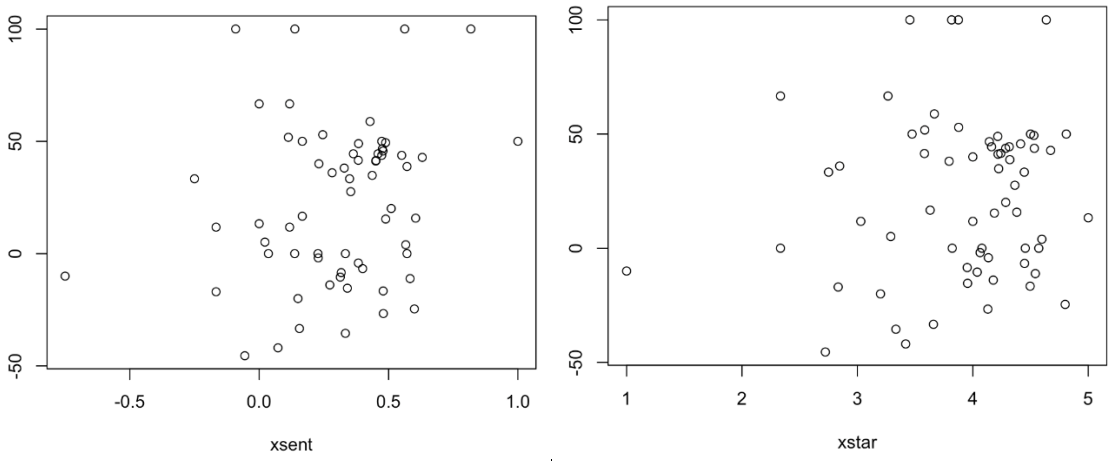


**Figure 4: Commercial products pNPS variance compared with the average sentiment & star score**

The models gave us the regression fit line with $R^2$ value of 24.25% for consumer products and 4.5% for the commercial products. The more information on the model is given in Appendix B. The models were finally used to predict the product NPS values of the required 5 products.

## 3. 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.

### 3.3 Markov Decision Process

The objective of formulating a Markov Decision model was to understand the evolution of customer sentiment and need to take appropriate action. Markov Decision process is a mathematical framework for modeling decisions where outcomes are probabilistic rather than deterministic but can be influenced by the decision maker. This process also suggested best actions to take based on the optimal policy defined.

The components of Markov Chains we used for our project are –

- State Space (S): The objective was to assess the evolution of sentiment over time and hence the state space comprised of 3 customer sentiments states- Positive, Negative and Neutral
- Action Space (A): This includes the set of possible actions Lenovo could take to impact its NPS score. There are 2 possible actions – 'Do Nothing' and 'Intervene'
- Decision Epoch (T): The epoch refers to the time interval over which decision is made. Since, the data was available for more than a year, we divided our epochs on a weekly basis.
- Transition Probability (P): This represents the probability of each possible state of the model in the next time period.
- Rewards (R): This considers the immediate value of taking an action at each of the given state.

The Markovian transition probabilities were built by observing the change in sentiments for the given 5 products on weekly epochs. The transition probabilities for each product is shown in Appendix C. This transition probability was assigned random rewards as and when Lenovo intervenes. The objective function is formulated based on the transition and reward matrix and solved on Excel to predict the improved NPS. Thus, this Markovian Decision Process helped improve the NPS score of the 5 given products by allocating rewards for each intervention and associating a probability of change in state for these interventions. This helped in analyzing how timely intervention of Lenovo would help improve the NPS score of their products.

Assumptions:

1. The probabilities for the ‘Do Nothing’ and ‘Intervene’ actions are kept same for the analysis.
2. The discount factor is assumed to be 0.9 which is used for analysis in Microsoft Excel.
3. Any intervention action by Lenovo invoked the same cost irrespective of the state the system is in.