Using customer sentiment to predict customer satisfaction

Lenovo™

NC STATE UNIVERSITY

- Addagarla Suraj
- Bendale Priyanka
- Narayanan Satyajit
- Sinha Mallika

Contents

- Problem Definition
- Markov Decision Process (MDP)
- Variable Selection
- Prediction
- Appendix

Problem definition

Defining the problem to get to the right key questions

Current State

- Lenovo currently uses customers' responses to the Net Promoter Score (NPS) survey as their primary measurement of customer satisfaction
- Feedback from the NPS survey results arrives late in the product life cycle

Gap

- Lack of understanding of customer's concerns and satisfaction with the product

Future State

- Timely addressing customers' concerns
- Taking corrective actions on problems
- Identifying pain points of customers

Adopting a holistic approach to understand customer’s concerns and satisfaction

Predict product-NPS (pNPS) using the sentiment data available

Evaluate the use of ‘evolution of sentiment’ (EOS) in predicting pNPS

Recognize major drivers of customer satisfaction

Markov Decision Process (MDP)

Markov Decision Process (MDP)

Assumptions:

- Data available corresponds to ‘Do Nothing’ scenario
- Better improvement in sentiment with intervention
- Easier to convert negative sentiment into neutral sentiment than for negative sentiment to positive sentiment

Markov Decision Process (MDP)

Elements of MDP:

State Space (S):

Positive, Neutral and Negative (sentiments)

Action Space (A):

Do Nothing; Intervention

Decision Epoch (T):

Monthly

Transition Probability (P)

Reward Space (R)

Markov Decision Process (MDP)

Transition probabilities:

P (Do Nothing)	Positive	Neutral	Negative
Positive	0.25	0.25	0.5
Neutral	0.4	0.2	0.4
Negative	0.167	0.334	0.5

P (Intervene)	Positive	Neutral	Negative
Positive	0.25 + 0.5a	0.25 + 0.5b	0.5(1 - a - b)
Neutral	0.4 + 0.4a	0.2 + 0.4b	0.4(1 - a - b)
Negative	0.167 + 0.5a	0.334 + 0.5b	0.5(1 - a - b)

P (Intervene for a = 5%, b = 10%)	Positive	Neutral	Negative
Positive	0.275	0.3	0.425
Neutral	0.42	0.24	0.34
Negative	0.192	0.384	0.425

Markov Decision Process (MDP)

Reward Space (R):

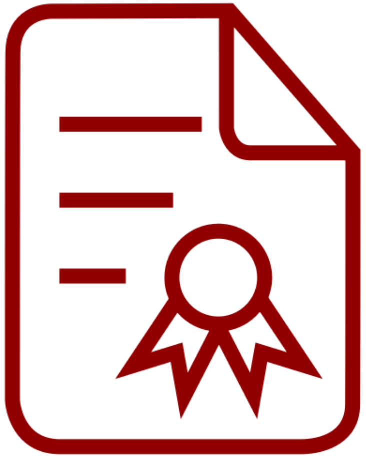
State (S)	Action (A)	Customer Satisfaction (CS)		Monetary Costs (MC)		Net Reward R (S, A)
		Ranking	R_{CS}	Ranking	R_{MC}	
Positive	Intervene	1	10	1	(10)	0
Neutral	Intervene	2	6.67	1	(10)	(3.34)
Positive	Do Nothing	2	6.67	2	0	6.67
Negative	Intervene	3	3.34	1	(10)	(6.67)
Neutral	Do Nothing	3	3.34	2	0	3.34
Negative	Do Nothing	4	0	2	0	0

Note: Values in bracket denote negative values

Markov Decision Process (MDP)

Optimal Policy

- Policy Iteration algorithm was employed to obtain an optimal policy for each state (over an infinite horizon)
- Optimal policy of ‘Do Nothing’ was arrived at for all the three states
- Policy remains constant for increasing values of (a, b)



Variable Selection

Selecting the variables to be used as predictors of pNPS

Taxonomy sentiments

- Assumption: The frequency of occurrence of a sentiment is calculated for each taxonomy is a predictor of pNPS
- Why at a taxonomy level?
  - Can help us understand the reason for a survey score
  - To capture the magnitude of sentiment across every taxonomy
- To reduce dimensionality, similar taxonomies were identified and grouped based on:
  - Class
  - Correlation between them

Selecting the variables to be used as predictors of pNPS

Evolution of Sentiment (EOS)

- Assumption: The EOS calculated at each month is a predictor of pNPS
- Sentiment distribution in the first month was taken as the initial sentiment distribution (q)
- Distribution in the following months (q_n) was taken as per Chapman-Kolmogorov equation ($q_n = q*P_n$)
- Note: the distribution of sentiment for the first month should be known to calculate EOS values

Selecting the variables to be used as predictors of pNPS

Calculating pNPS value to be used as a dependent variable

- The Product Net Promoter Score (pNPS) value for a month is calculated as a cumulative subtraction of promoter % and detractor % till that month
- We have assumed that the sentiments will reflect on the pNPS after a lag of 5 months

Level of Data for analysis

- We have taken Series-Month combination as level of data for the analysis

Prediction

Direct application of Multiple Linear Regression and its failure

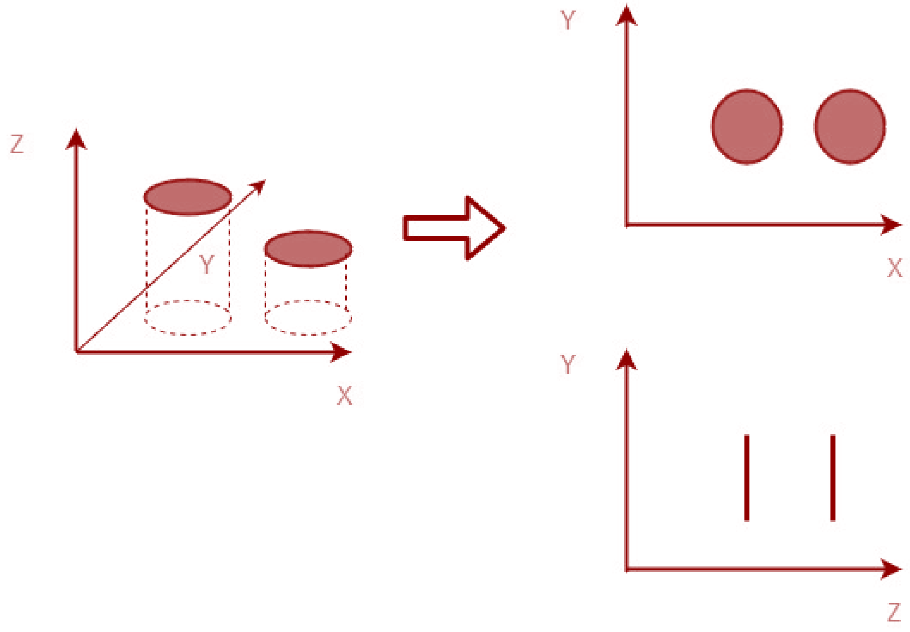
Why it failed:

- Couldn’t handle strongly correlated Taxonomy-Sentiment despite grouping of taxonomies
- Couldn’t handle missing values
- Variable selection to gain independence between columns is not always obvious and may have lead to suboptimal predictions

Principal Component Analysis (PCA) helps mitigate most of MLR's problems

Why PCA?

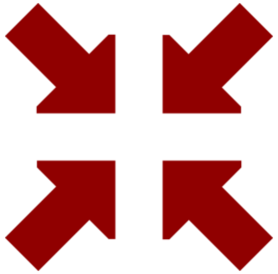
- Summarizes different characteristics of dataset into fewer predictors which explains data
- Removes dependency (multi-collinearity) between variables creating a set of uncorrelated variables
- The scores in the components can be calculated even if there are missing data
- Each Component explains a percentage of the total variance in the data



Principal Components Regression is used to predict pNPS

PCA on original data

- Build a PCA model on the original data to obtain components



Dimension reduction

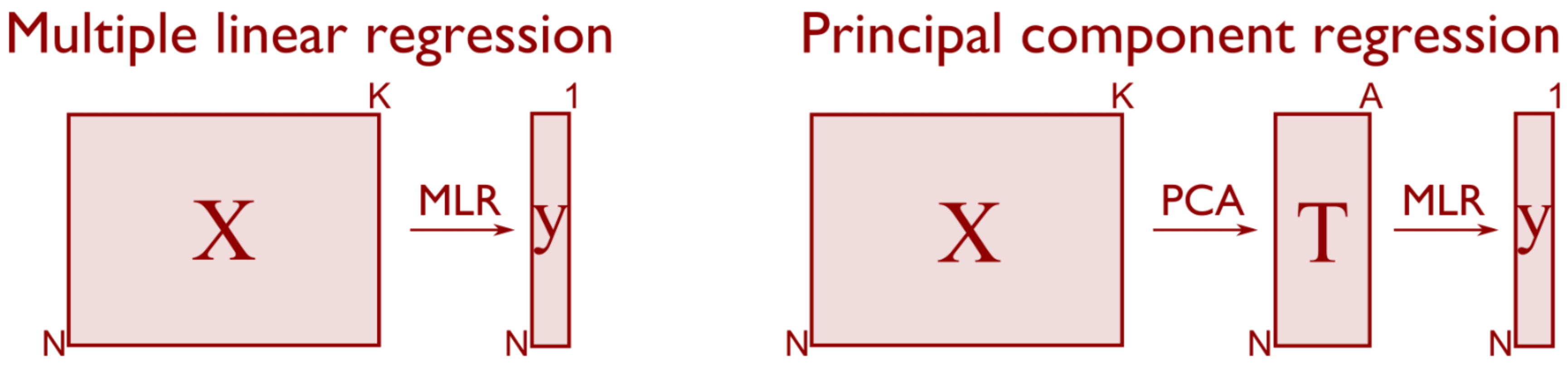
- Fit components which explain maximum variance and by iteration

Regression on components

- Use the columns of components from PCA as data for multiple linear regression model

Principal Components Regression is used to predict pNPS

- The main idea with principal component regression is to replace the K columns in X with their uncorrelated A score vectors from PCA



Example of Principal Components Regression for one Series-Month combination

- Example: Yoga Series
- In the Month of May 2017, X values are:

Tax #	Variable	Sentiment	Frequency of sentiments
1	ACCESSORIES	NEGATIVE	84
2		NEUTRAL	92
3		POSITIVE	50
4	CLIENT OS	NEGATIVE	18
5		NEUTRAL	35
6		POSITIVE	14
…	…	…	…
…	…	…	…
…	…	…	…
94	GENERAL COMMENT	NEGATIVE	22
95		NEUTRAL	25
96		POSITIVE	122

Example of Principal Components Regression for one Series-Month combination

- After PCA, the T values we get for the month of May 2017 are:

#	Component	Frequency of sentiments
1	C1	23.163
2	C2	-0.789
3	C3	7.458
…	…	…
…	…	…
…	…	…
29	NEGATIVE	-0.309
30	NEUTRAL	0.038
31	POSITIVE	-0.139

Example of Principal Components Regression for one Series-Month combination

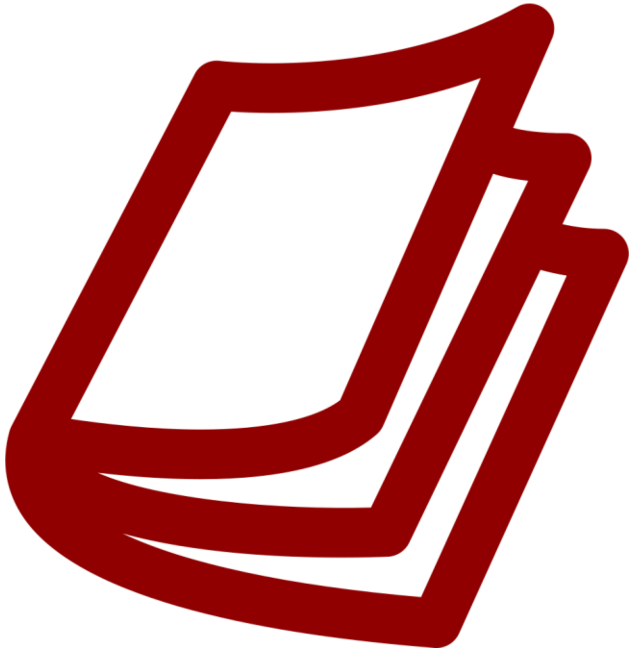
- Substituting the T values of the variables onto the regression equation of *Consumer* Products:

$pNPS_{Consumer}$ = 30.685 + (0.385)*C1 + (-1.379)*C2 +…..+ (-1.33)*C30 + (-2.022)*C31

- We get, $pNPS_{Consumer}$ for YOGA series for the month of May 2017 as 44.144.
- This means, based on May 2017 data, we expect a pNPS for the month of October 2017 (5 month lag) to be 44.144
- From the survey data, we observe that the actual pNPS value for the month of October 2017 is 43.778

Summary

- EOS obtained from MDP was used as a variable to account for the effect of sentiment evolution in prediction of pNPS
- 2 final equations was obtained based on regression analysis, each for commercial and consumer product type to predict pNPS
- Based on the equations obtained, pNPS was calculated for test products



Predictions for Consumer and Commercial products

- Predictions were made for the required products based on equations for Consumer and Customer

- Consumer:

Product	Month of prediction	Predicted pNPS
IDEAPAD 120S 11	May-18	27.312
	Jun-18	27.029
	Jul-18	26.803
YOGA 920	May-18	*33.571*
	Jun-18	25.237
	Jul-18	34.280

- Commercial:

Product	Month of prediction	Predicted pNPS
T480	Mar-18	60.100
	Apr-18	59.920
	May-18	54.533
X1 CARBON 2018	Mar-18	*80.983*
	Apr-18	21.681
	May-18	36.506

Appendix

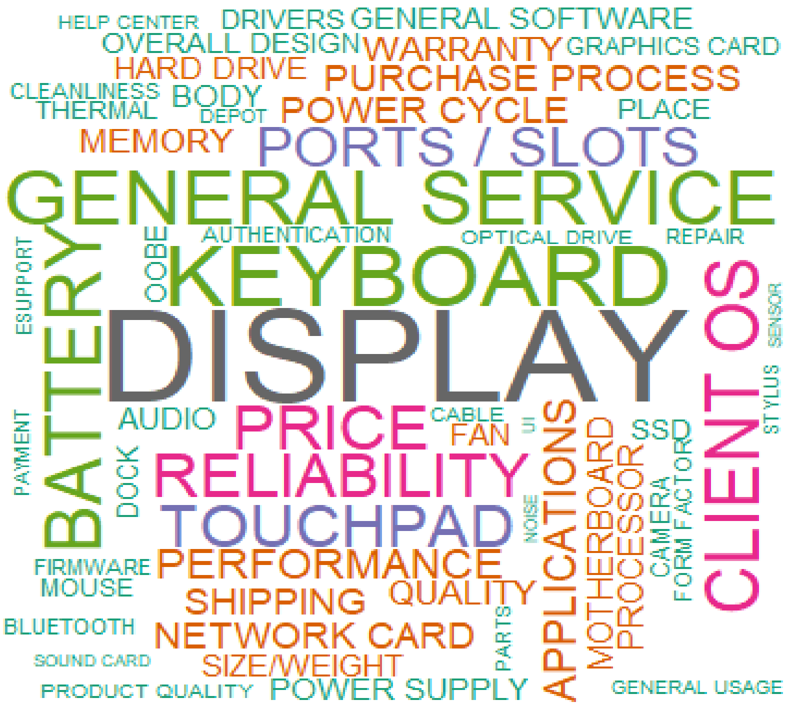
Defining the problem to get to the right key questions

Net Promoter Score (NPS)

- It is the key metric of a company's or a product's performance in the market
- Customers are asked to rate their willingness to suggest the product to a friend on a scale of 0 to 10

Taxonomy Sentiment Analysis

- Consumer Negative Sentiment

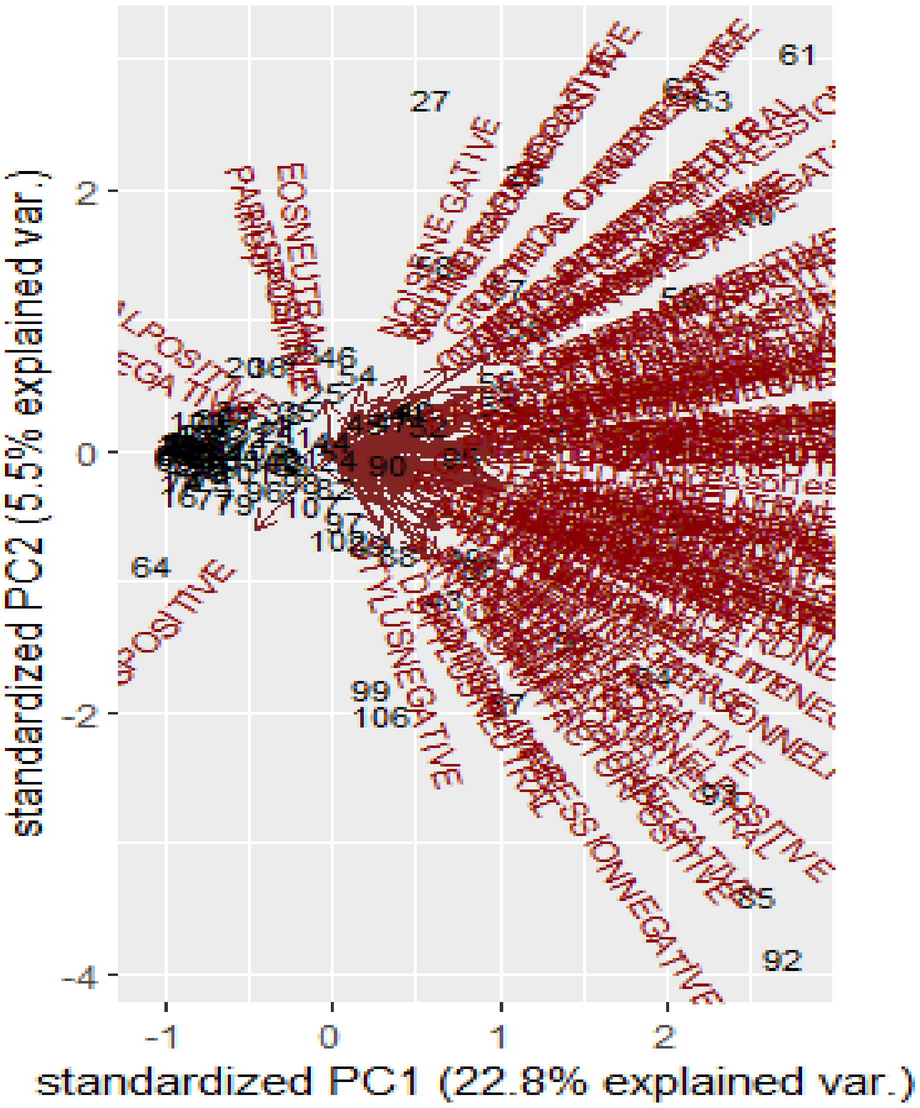


- Consumer Positive Sentiment

PCA For Commercial

- First principle component has large positive association with:
  - Help-positive
  - Parts-positive
  - Sensor-Negative
  - Thermal-Positive
  - EOS-Positive
  - EOS-Neutral

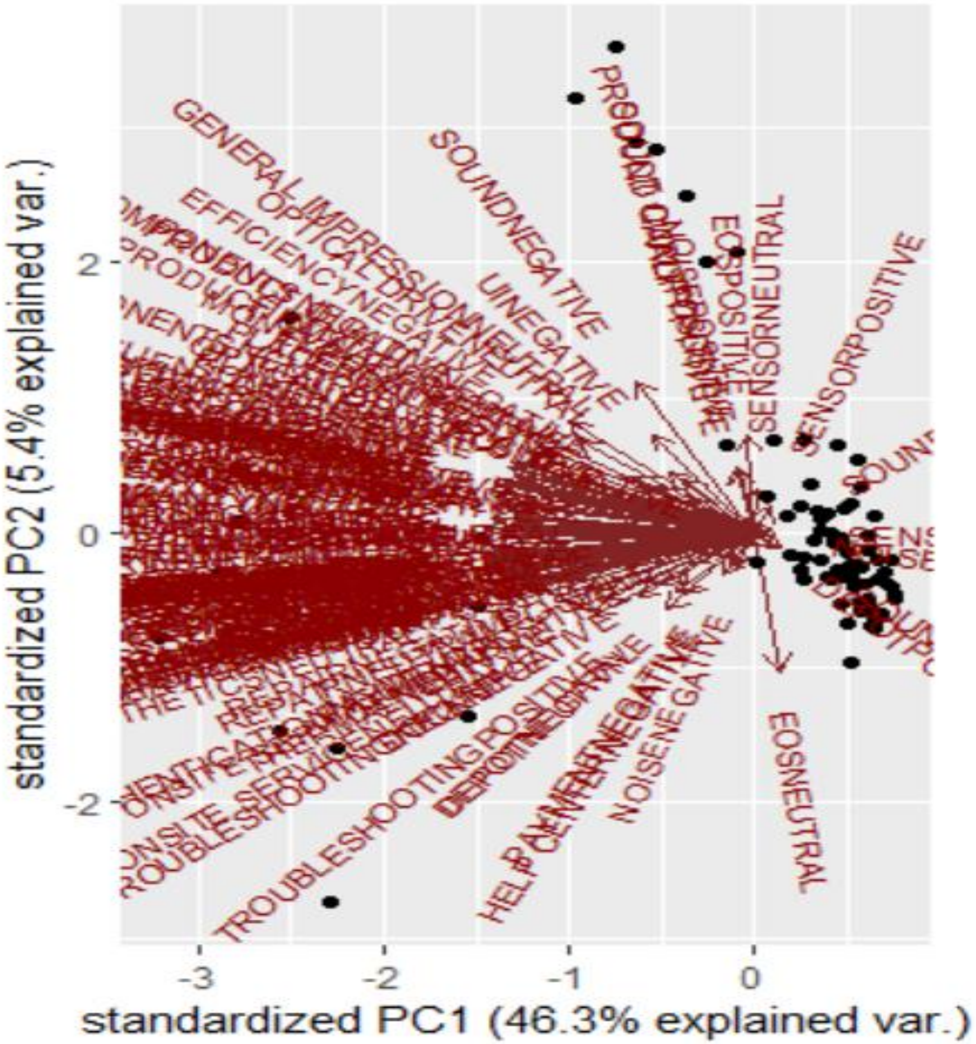
- Second principle component has negative association with:
  - Money-Positive
  - External – Positive
  - General Impression- Positive
  - Optical Driven-Negative
  - Noise-Negative



PCA For Consumer

- First principle component has large positive association with:
  - Indicators-Neutral
  - Efficiency-Positive
  - Accessories-Neutral
  - Accessories-Negative
  - Indicators-Positive
  - Power-Neutral

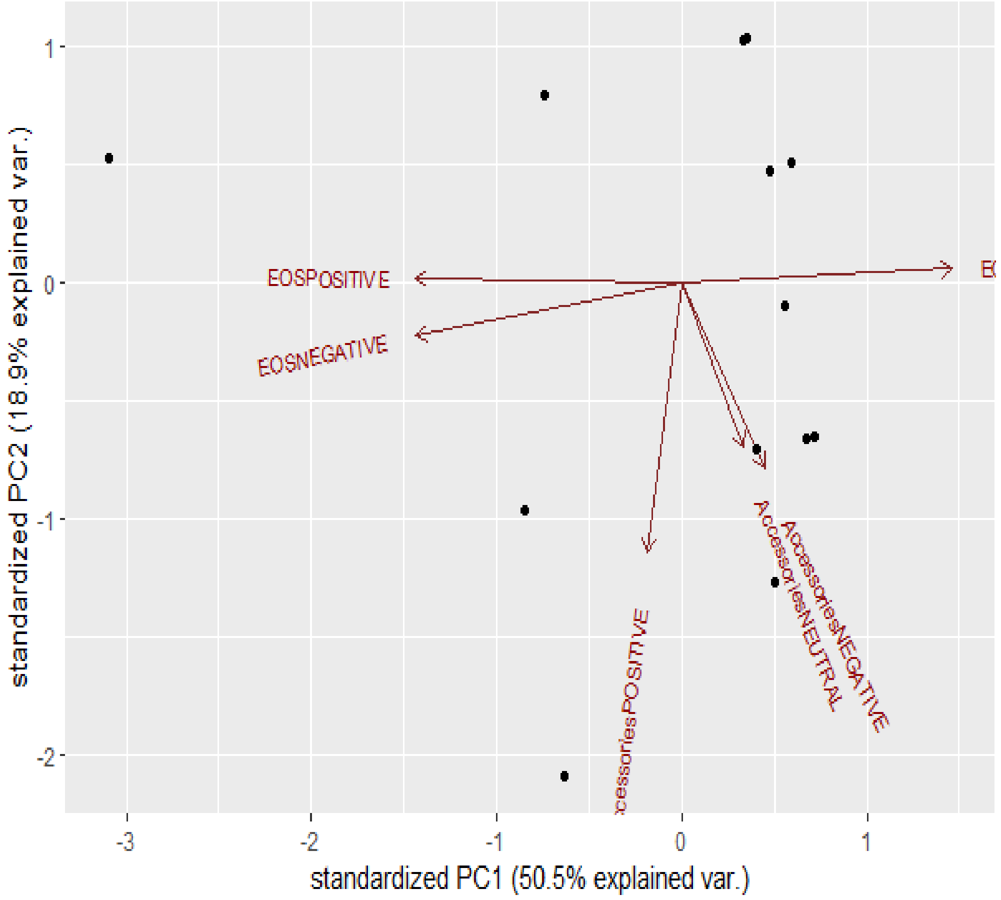
- Second principle component has negative association with:
  - EOS-Neutral
  - Onsite Service- Neutral
  - Depot-Neutral
  - Authentication-Neutral
  - Aesthetic-Neutral



Principal Component Analysis (PCA) helps mitigate most of MLR's problems

PCA Interpretation

- For interpretation, magnitude and direction of the coefficients of original value are examined
- The larger the value, the more important the corresponding variable is in calculating the component
- The level at which correlation value is important is decided



Evolution of the model

#	Test Case	Pass/Fail – (Reason)
1	Apply regression across 184 variables/Predictors (taxonomy-sentiment level) vs pNPS	Fail - very low value of Coefficient of Determination; Multicollinearity between variables
2	Eliminated variables based on the number of 0 responses (142 variables) in order to reduce dimensionality	Fail – very low value of Coefficient of Determination; Multicollinearity between variables
3	Based on the median of frequency of responses, considered first 50% of variables with variables with Highest response being higher limit; this reduced dimensionality to 71 variables	Fail – better value of coefficient of Determination; Multicollinearity between variables
4	Clubbed variables on the basis of their correlation value and performed principle component analysis to reduce the dimensionality	Pass- No Multicollinearity (the components are orthogonal), reduced dimensionality (from 138 variables to 31)
5	Performed Multilinear Regression Analysis on 31 variables as predictors for determining pNPS	Pass

Principal Components Regression

- For Consumer Data, X is given by:

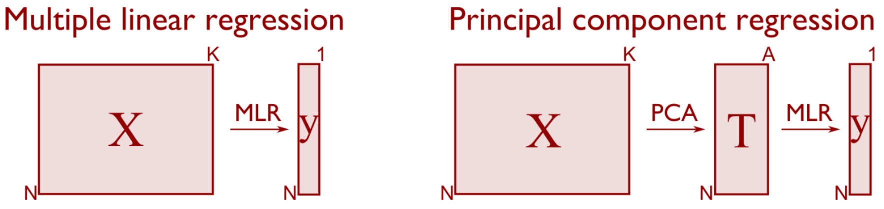
			1	2	3		94	95	K = 96
			ACCESSORIES			...	CLIENT OS		
	Series	Month	NEGATIVE	NEUTRAL	POSITIVE	...	NEGATIVE	NEUTRAL	POSITIVE
1	A SERIES	May-17	5	15	2	...	5	10	3
2	A SERIES	Jun-17	0	7	4	...	1	5	0
3	A SERIES	Jul-17	8	6	1	...	4	0	2
4	A SERIES	Aug-17	1	9	2	...	0	2	3
5	A SERIES	Sep-17	5	3	3	...	4	2	1
	...	...	...	...	...	...	...	...	...
82	IDEAPAD 100 SERIES	Aug-17	18	21	8	...	8	27	11
83	IDEAPAD 100 SERIES	Sep-17	16	12	6	...	13	10	8
N = 84	IDEAPAD 100 SERIES	Oct-17	14	16	13	...	10	24	11

Principal Components Regression

- The K variables are converted to A = 31 components after PCA and we obtain T:

	Series	Month	C 1	C 2	C 3	...	C 29	C 30	C 31
1	A SERIES	May-17	-0.881	1.882	-0.819	...	1.077	-0.382	-0.119
2	A SERIES	Jun-17	-4.191	0.561	-0.318	...	-0.129	-0.379	0.981
3	A SERIES	Jul-17	-3.759	-0.107	0.046	...	0.384	-0.4	0.916
4	A SERIES	Aug-17	-3.266	-0.013	-0.884	...	-0.269	0.749	0.105
5	A SERIES	Sep-17	-3.082	0.385	0.294	...	-0.693	0.425	-1.213
	...	...	...	...	...	...	...	...	...
82	IDEAPAD 100 SERIES	Aug-17	2.922	6.78	-1.98	...	-0.438	1.809	-0.826
83	IDEAPAD 100 SERIES	Sep-17	2.093	5.478	-1.602	...	2.609	0.676	0.072
84	IDEAPAD 100 SERIES	Oct-17	4.219	7.737	-3.215	...	0.383	0.217	0.098

Principal Components Regression

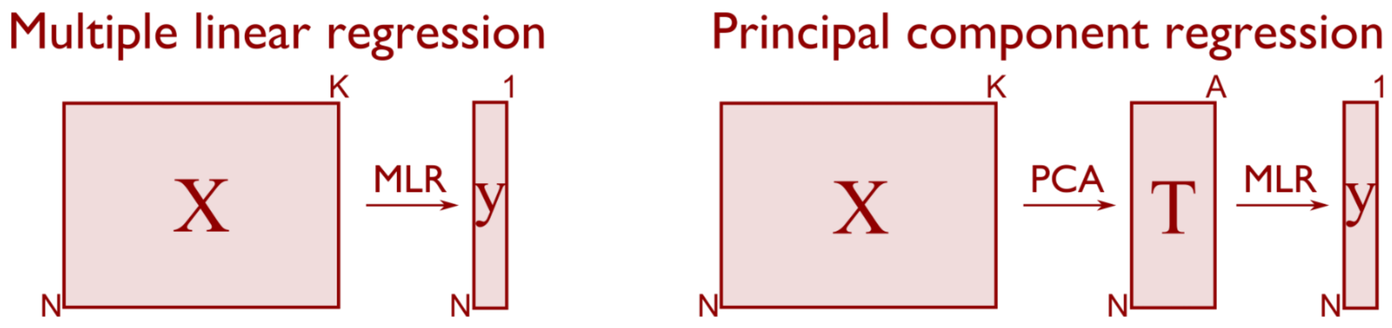


- The K variables are converted to A = 31 components after PCA and we obtain T:

	Series	Month	C 1	C 2	C 3	...	C 29	C 30	C 31
1	A SERIES	May-17	-0.881	1.882	-0.819	...	1.077	-0.382	-0.119
2	A SERIES	Jun-17	-4.191	0.561	-0.318	...	-0.129	-0.379	0.981
3	A SERIES	Jul-17	-3.759	-0.107	0.046	...	0.384	-0.4	0.916
4	A SERIES	Aug-17	-3.266	-0.013	-0.884	...	-0.269	0.749	0.105
5	A SERIES	Sep-17	-3.082	0.385	0.294	...	-0.693	0.425	-1.213
	...	...	...	...	...	...	...	...	...
82	IDEAPAD 100 SERIES	Aug-17	2.922	6.78	-1.98	...	-0.438	1.809	-0.826
83	IDEAPAD 100 SERIES	Sep-17	2.093	5.478	-1.602	...	2.609	0.676	0.072
84	IDEAPAD 100 SERIES	Oct-17	4.219	7.737	-3.215	...	0.383	0.217	0.098

Month	pNPS
Oct-17	43.75
Nov-17	44.89
Dec-17	33.33
Jan-18	33.33
Feb-18	32.91
...	...
Jan-18	15.36
Feb-18	14.47
Mar-18	14.24

Principal Components Regression



- The main idea with principal component regression is to replace the K columns in X with their uncorrelated A score vectors from PCA
- We replace the N×K matrix of raw data with a smaller N×A matrix of data that summarizes the original X matrix.
- Then we relate these A scores to the y variable. Mathematically it is a two-step process:
- T=XP (from the PCA model)
- yˆ=Tb (and can be solved as b = (T′T)−1T′y)

Markov Decision Process (MDP)

Markov Decision Process:

- Aids in decision making in case of probabilistic processes
- Suggests optimal actions at various points in time or for any state (infinite horizon situation)
- Provides expected rewards at each stage which can then be used to select among alternatives
- Characterized by five parameters: {P, A, R, T, S}

Sentiment aggregation across Comments and Months

2									
3	**Comment Level:**								
4		Positive	Neutral	Negative	Sum		Positive	Neutral	**Negative**
5	Comment	9	3	8	20		0.45/0.75	0.15/0.167	0.4/0.0834
6	Sentiment Distribution	0.45	0.15	0.4			0.6	0.9	**4.8**
7									
8	Total	9000	2000	1000	12000				
9	Distribution	0.75	0.167	0.0834					
10									
11									
12	**Month Level:**								
13		Positive	Neutral	Negative	Sum		Positive	Neutral	**Negative**
14	Month	90	30	80	200		0.45/0.75	0.15/0.167	0.4/0.0834
15	Sentiment Distribution	0.45	0.15	0.4			0.6	0.9	**4.8**
16									
17	Total	900	200	100	1200				
18	Distribution	0.75	0.167	0.0834					
19									
20									

P-Matrices (Consumer Products)

P (Do Nothing)	Positive	Neutral	Negative
Positive	0.25	0.25	0.5
Neutral	0.142	0.714	0.142
Negative	0.25	0.5	0.25

P (Intervene)	Positive	Neutral	Negative
Positive	0.25 + 0.5a	0.25 + 0.5b	0.5(1 - a - b)
Neutral	0.142 + 0.142a	0.714 + 0.142b	0.142(1 - a - b)
Negative	0.25 + 0.25a	0.5 + 0.25b	0.25(1 - a - b)

P (Intervene) a = 5%, b = 10%	Positive	Neutral	Negative
Positive	0.275	0.3	0.425
Neutral	0.149	0.728	0.12
Negative	0.263	0.525	0.213