SENTIMENTAL ANALYSIS ON AMAZON PRODUCT REVIEWS

Gayathri Shanmuga Sundaram, Manoj Arasada, Sumanth Jinka, Vinaya Rajaram Nayak

INRODUCTION:

The industries are growing their online operations at an exponential rate due to the rising demand for online buying. There are many brands on the market, but some of them are prominent and control a sizeable portion of the sector. One of the largest e-commerce sites is Amazon, where many sellers offer their goods. Therefore, online commerce is essential for boosting product sales and influencing consumer purchasing behavior. Consumers use the reviews offered on these e-commerce platforms as a resource to help them make wise purchases. Reviewers have a variety of alternatives for submitting their reviews on retail websites like Amazon.com. For instance, the customer can rate the product numerically (1–5) or by writing remarks. Considering the sheer number of goods produced by numerous brands, it is imperative that buyers receive reviews that are pertinent to their needs. Studies on the efficacy of online consumer reviews and the numbers show that people are becoming increasingly dependent on this valuable form of social proof. Here are some of the takeaways from the 2016 study:

- **84** percent of shoppers trust online reviews as much as personal recommendations.
- **9** out of **10** consumers read less than 10 reviews before forming an opinion about a business.
- **More than half** of people will visit a company's website after reading positive reviews.
- **74** percent of people say that positive reviews make them trust local businesses more.
- **58** percent of shoppers believe the star rating of a business is one of the most important factors.

As the data clearly shows, consumers connect with review and ratings. It's the equivalent of online word of mouth and goes a long way towards helping brands establish trust within their target demographics. The quantity of reviews connected to a product or brand is growing at an alarming rate, which is comparable to managing big data. Reviews are sentiment-oriented when they are divided into those with positive and those with negative customer sentiment. This produces superior judgment. Future customers as well as the product sellers can benefit from the segmentation of reviews based on their attitude by being able to examine both positive and negative comments constructively and make better selections based on their needs.

In this paper, we focus on main three research questions that are:

1. What will the future sales of specific products look like? Will it rise or fall?
2. Who are the major competitors in a particular product category?
3. Is the customer content? How can the customer experience be improved?

Future Sale Prediction

The proposed research question is resolved by conducting nine different processes as shown in Fig. 2

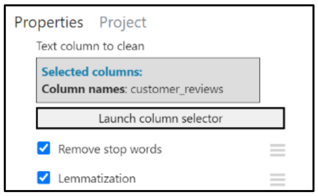
Data Collection

The first part of the project includes data collection and pre-processing of data. A large sample of online reviews is collected about the e-commerce giant Amazon.com from snap.standford.edu. The data set consists of over 91000 reviews for approximately 10000 toys, and it is converted to CSV format. It includes mainly nine features as explained in the Table 1 in which Product recommend will be the target variable.

Table 1: Dataset Features

Feature	Data Type	Description
Uniq_id	ID	Product ID
Product_name	Text	Name of the product
Manufacturer	Categorical	Manufacturing brand
Price	Numerical	Price of the toy in dollars
Number_available_in_stock	Numerical	Available stock count of the product
Number_of_reviews	Numerical	Total number of reviews
Number_of_answered_questions	Numerical	Number of questions on the product answered by the users
Average_review_rating	Categorical	Average user rating between 1 to 5
Customer_reviews	Text	Product review provided by the user
customer_questions_and_answers	Text	Questions on product and answers.
Sellers	Text	Other sellers and price
Product_recommend	Categorical	Shows if the product is recommended or not

Data Preprocessing



As preprocessing is crucial in opinion mining and sentiment analysis. Tokenization, stop word removal, stemming, punctuation mark removal, etc., were done during preprocessing using Azure ML studio preprocess text task.

Dataset Limitation and Resolution

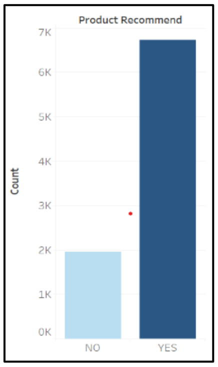


Figure 1

As the Figure 1 indicates that there are more than twice as many products recommended as negative ones. This implies that the target variable has an unbalanced distribution of classes. Hence, using machine learning models for classification would yield biased results and inaccurate predictions. Data balance is used to prevent this situation. Under sampling and oversampling are two techniques that can be used to convert unbalanced data into balanced data.

The under sampling method has been applied to handle the imbalanced amazon data. To balance the data set, under sampling involves reducing observations from the dominant class in this case its products with recommendations. The resulting balanced data has about equal amounts of both products recommended and not recommended.

Feature Selection

To train and test the dataset on two classifiers, a random sample of reviews is selected, after an imbalanced proportion of classes in the dataset has been removed. To extract pertinent features from the preprocessed dataset, feature selection is conducted. Only six of the features mentioned in the above table have been taken into account in the provided data set.

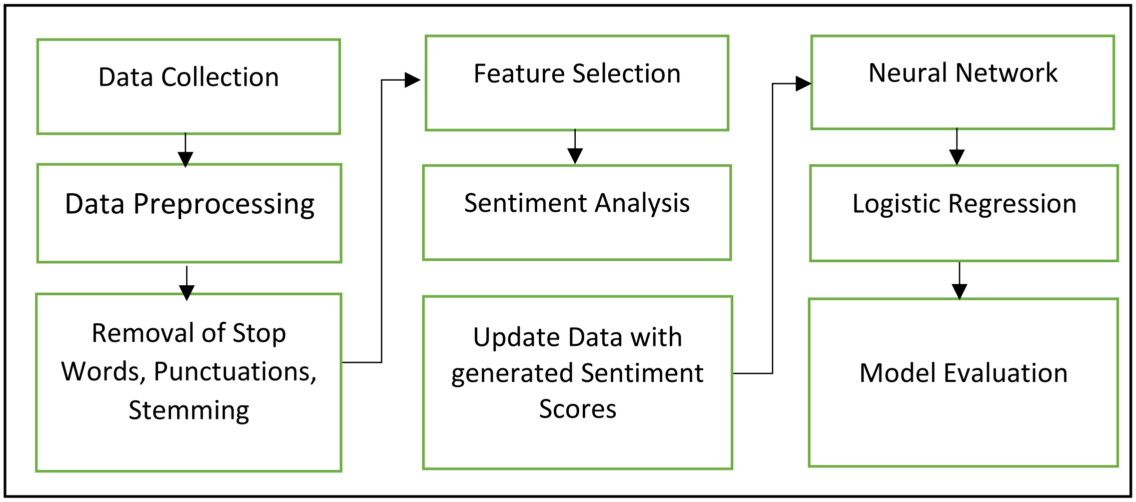


Figure 2: Process Workflow

Sentimental Analysis

Sentiment Analysis or Opinion Mining is the process of analyzing a text and interpreting the sentiment behind it. It majorly relies on an AI engine that is powered by both natural language processing and machine learning. Through machine learning and text analysis, it's possible to classify the statement as positive, negative, or neutral. As people express their views and opinions more openly than ever before, sentimental analysis has rapidly become the essential tool to monitor and understand the sentiment behind all types of data. The sentiment orientation of the reviews is established in the next phase using Python.

```python
pip install vaderSentiment
import numpy as np
import pandas as pd
average_list=[]
data_score=pd.DataFrame()
data = pd.read_excel('C:/Users/gayat/Downloads/DM Final Project/amazon_co-ecommerce_sample.xlsx')
```

```python
from vaderSentiment.vaderSentiment import SentimentIntensityAnalyzer
sentiment = SentimentIntensityAnalyzer()
for i in range (0,len(data)):
        review=data['customer_reviews'].iloc[i]
        review_score=sentiment.polarity_scores(str(review))
        average_list.append(review_score['compound'])
```

A rule-based sentiment analyzer called VADER (Valence Aware Dictionary and Sentiment Reasoner) has been trained on social media text. It's implemented in Python for the amazon dataset. A Vader Sentiment object returns a dictionary of sentiment scores for the text to be analyzed. If the compound value is more than or equal to 0.05, it's positive. Likewise, if the compound value of the text is less than or equal to -0.05, it's considered as negative. Else, it the sentence is considered as neutral.

```python
review1="The service here is great"
review_score_1=sentiment.polarity_scores(review1)
review_score_1
```

```
{'neg': 0.0, 'neu': 0.494, 'pos': 0.506, 'compound': 0.6249}
```

Dataset Update

The dataset is added with a new column with the compound sentiment score correlating each review in order to do supervised learning.

```python
data['compound']=average_list
```

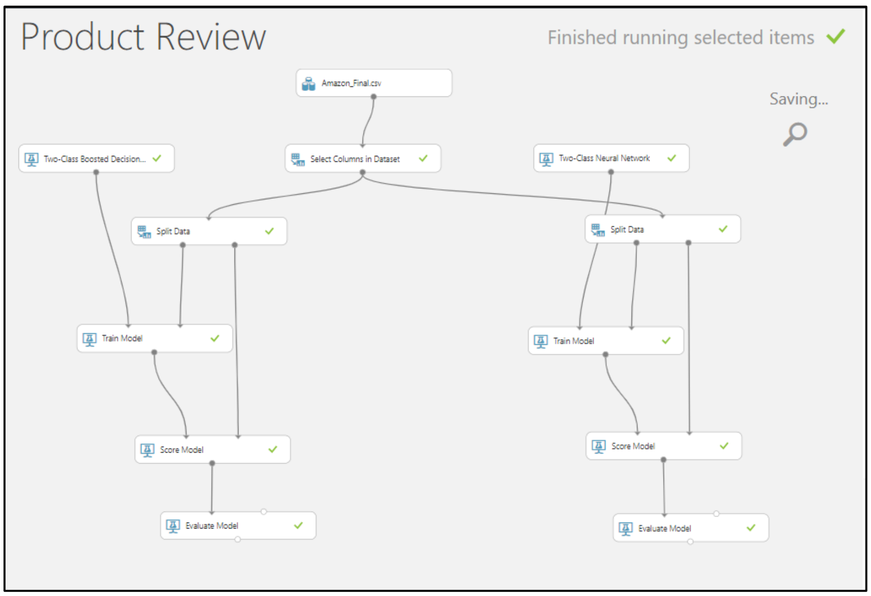
```python
data.to_excel('compoundData.xlsx')
```

Classification is the process of categorizing reviews into Positive and Negative categories based on their sentiment. The reviews dataset is then updated with the Compound sentiment score calculated using Vader sentiment dictionary for each positive, negative, and neutral polarity. The acquired subsets are combined with the original dataset in a CSV file.

A	B	C	D	E	F	G	H	I
uniq_id	product_name	price	number_avail	number_of_rev	number_of_answ	average_review_ra	Sentimental_Score	Product_Recommend
eac7efa5dbd3d667f26eb3d3ab504464	Hornby 2014 Cata	$3.42	5	15	1	4.9	0.9977	YES
b17540ef7e86e461d37f3ae58b7b72ac	FunkyBuys ® Larg	$16.99		2	1	4.5	0.8439	YES
348f344247b0c1a935b1223072ef9d8a	CLASSIC TOY TRAI	$9.99	2	17	2	3.9	0.9934	YES
e12b92dbb8eaee78b22965d2a9bbbd9f	HORNBY Coach R	$39.99		1	2	5	0.9393	YES
e33a9adeed5f36840ccc227db4682a36	Hornby 00 Gauge	$32.19		3	2	4.7	0.9947	YES
cb34f0a84102c1ebc3ef6892d7444d36	20pcs Model Gar	$6.99		2	1	5	0.6486	YES
f74b562470571dfb689324adf236f82c	Hornby 00 Gauge	$24.99		2	1	4.5	0.902	NO
87bbb472ef9d90dcef140a551665c929	Hornby Santa s Ex	$69.93	3	36	7	4.3	0.9956	NO
7e2aa2b4596a39ba852449718413d7cc	Hornby Gauge W	$235.58	4	1	1	5	0.7506	NO
5afbaf65680c9f378af5b3a3ae22427e	Learning Curve Chuggington		1	8	1	4.8	0.9858	NO
5c76389a8c302c6d7d6e179393031b97	Hornby Gauge Ra	$27.49	6	1	1	5	0.8748	NO
878048c41f3c249badb3704e160b4c6e	Kato (USA) 176-1	$273.60		1	1	5	0.7146	NO
f910c6542ededa5abf81787c0fd87c99	Bachmann 37-66	$9.60	2	1	1	5	0.4927	NO

Model Building

The classification models selected for categorization of text are: Neural Network, and Logistic Regression.



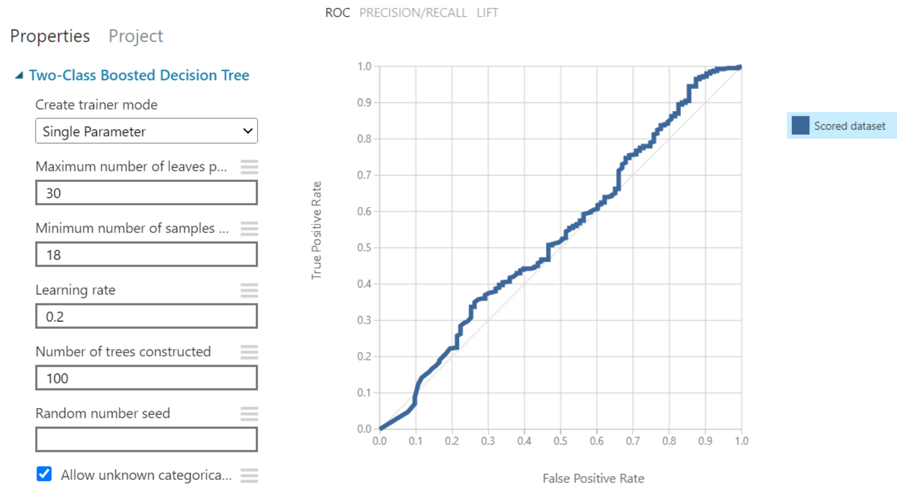
Decision Tree

A hierarchical tree encompassing decision nodes with attributes and edges represents attributes values. Building classification rules for new instances of the data is made possible by this representation of the data as a tree.

Following are the different parameters selected for two class boosted decision tree and corresponding results:

Model 1:

This model is parameterized with 30 maximum number of leaves and 18 minimum number of samples, above output shows the AUC is 0.535 with 344 true positives and false positives 83.

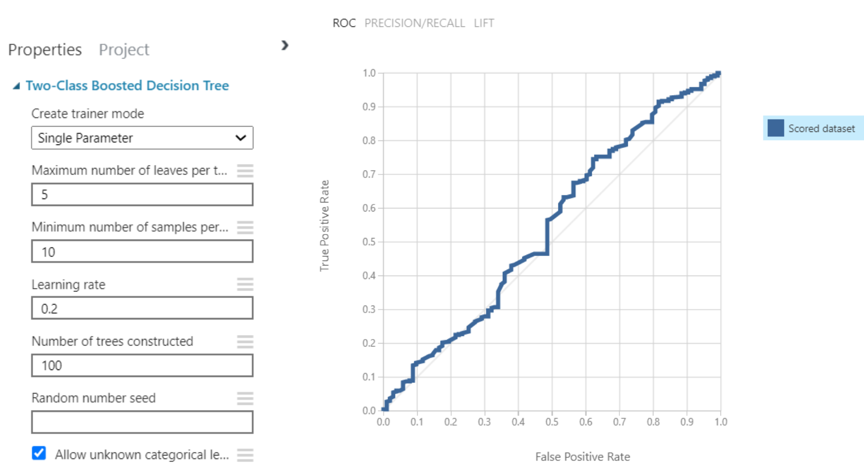


True Positive	False Negative	Accuracy	Precision	Threshold	AUC
344	56	0.724	0.806	0.5	0.535
False Positive	**True Negative**	**Recall**	**F1 Score**		
83	20	0.860	0.832		
Positive Label	**Negative Label**				
YES	NO				

Score Bin	Positive Examples	Negative Examples	Fraction Above Threshold	Accuracy	F1 Score	Precision	Recall	Negative Precision	Negative Recall	Cumulative AUC
(0.900,1.000]	203	48	0.499	0.513	0.624	0.809	0.507	0.218	0.534	0.123
(0.800,0.900]	57	19	0.650	0.588	0.715	0.795	0.650	0.205	0.350	0.229
(0.700,0.800]	39	3	0.734	0.660	0.778	0.810	0.748	0.246	0.320	0.249
(0.600,0.700]	25	8	0.799	0.694	0.808	0.806	0.810	0.248	0.243	0.309
(0.500,0.600]	20	5	0.849	0.724	0.832	0.806	0.860	0.263	0.194	0.350
(0.400,0.500]	13	2	0.879	0.746	0.848	0.808	0.892	0.295	0.175	0.366
(0.300,0.400]	10	3	0.905	0.759	0.858	0.807	0.917	0.313	0.146	0.393
(0.200,0.300]	19	3	0.948	0.791	0.880	0.809	0.965	0.462	0.117	0.420
(0.100,0.200]	8	4	0.972	0.799	0.886	0.806	0.985	0.571	0.078	0.458
(0.000,0.100]	6	8	1.000	0.795	0.886	0.795	1.000	1.000	0.000	0.535

Model 2:

This model is parameterized with 5 maximum number of leaves and 10 minimum number of samples, above output shows the AUC is 0.502 with 346 true positives and false positives 16.



True Positive	False Negative	Accuracy	Precision	Threshold	AUC
379	21	0.773	0.803	0.5	0.543

False Positive	True Negative	Recall	F1 Score
93	10	0.948	0.869

Positive Label	Negative Label
YES	NO

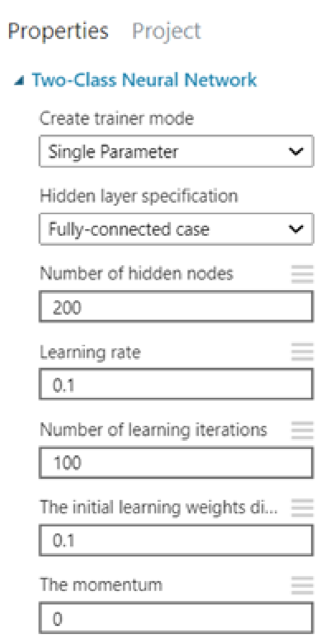
Score Bin	Positive Examples	Negative Examples	Fraction Above Threshold	Accuracy	F1 Score	Precision	Recall	Negative Precision	Negative Recall	Cumulative AUC
(0.900,1.000]	79	18	0.193	0.326	0.318	0.814	0.198	0.209	0.825	0.019
(0.800,0.900]	157	36	0.577	0.567	0.684	0.814	0.590	0.230	0.476	0.146
(0.700,0.800]	83	20	0.781	0.692	0.805	0.812	0.797	0.264	0.282	0.285
(0.600,0.700]	47	11	0.897	0.763	0.860	0.812	0.915	0.346	0.175	0.376
(0.500,0.600]	13	8	0.938	0.773	0.869	0.803	0.948	0.323	0.097	0.448
(0.400,0.500]	12	5	0.972	0.787	0.880	0.800	0.978	0.357	0.049	0.495
(0.300,0.400]	5	3	0.988	0.791	0.883	0.797	0.990	0.333	0.019	0.523
(0.200,0.300]	3	1	0.996	0.795	0.886	0.796	0.998	0.500	0.010	0.533
(0.100,0.200]	1	0	0.998	0.797	0.887	0.797	1.000	1.000	0.010	0.533
(0.000,0.100]	0	1	1.000	0.795	0.886	0.795	1.000	1.000	0.000	0.543

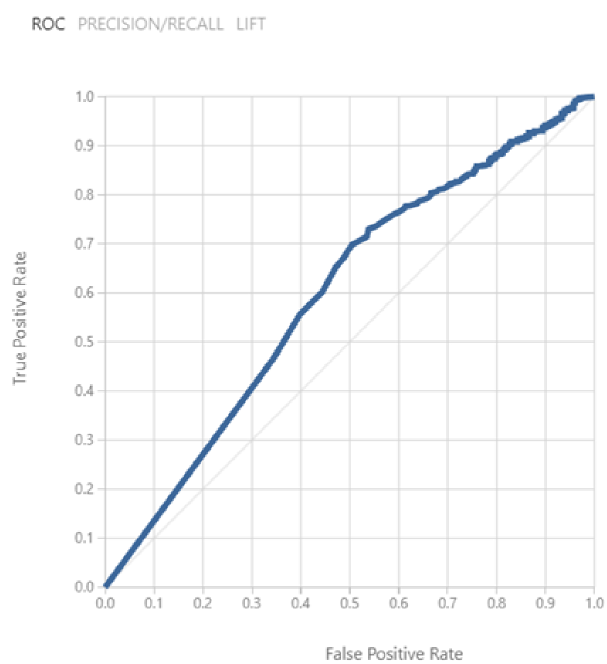
As per the above-mentioned values, model 2 is better suitable to achieve the required outcome for our problem as it has high area under the curve and greater number of true positives and false positives.

Neural Networks

A neural network is a streamlined representation of how the human brain functions. It simulates a huge number of connected processing units that mimic abstract representations of neurons in order to function. Layers of arrangement make up the processing units. A neural network typically consists of three parts: an input layer with units that represent the input fields; one or more hidden layers; and an output layer with a unit or units that represent the target field (s). The connections between the units have different connection strengths (or weights). The first layer receives input data, and each neuron in the subsequent layer receives values propagated from the previous layer's neuron. The output layer eventually delivers a result.

Following are the parameters selected for two class neural networks and corresponding results:





True Positive	False Negative	Accuracy	Precision	Threshold	AUC
1121	74	0.757	0.791	0.5	0.570

False Positive	True Negative	Recall	F1 Score
296	33	0.938	0.858

Positive Label	Negative Label
YES	NO

Score Bin	Positive Examples	Negative Examples	Fraction Above Threshold	Accuracy	F1 Score	Precision	Recall	Negative Precision	Negative Recall	Cumulative AUC
(0.900,1.000]	1046	262	0.858	0.730	0.836	0.800	0.875	0.310	0.204	0.379
(0.800,0.900]	39	12	0.892	0.748	0.850	0.798	0.908	0.333	0.167	0.411
(0.700,0.800]	13	11	0.907	0.749	0.852	0.794	0.919	0.312	0.134	0.442
(0.600,0.700]	14	9	0.923	0.753	0.855	0.791	0.931	0.297	0.106	0.467
(0.500,0.600]	9	2	0.930	0.757	0.858	0.791	0.938	0.308	0.100	0.473
(0.400,0.500]	7	3	0.936	0.760	0.860	0.790	0.944	0.309	0.091	0.481
(0.300,0.400]	8	4	0.944	0.762	0.863	0.789	0.951	0.306	0.079	0.493
(0.200,0.300]	6	4	0.951	0.764	0.864	0.788	0.956	0.293	0.067	0.504
(0.100,0.200]	15	2	0.962	0.772	0.870	0.789	0.968	0.345	0.061	0.510
(0.000,0.100]	38	20	1.000	0.784	0.879	0.784	1.000	1.000	0.000	0.570

By comparing both the models, we derive at a conclusion that two-class neural network is better since it identifies more true positives and false negatives.

Our assumption here is that if the product recommendation could be predicted efficiently, on seeing the product being recommended by Amazon there is high chance of customers buying it resulting in boosting of the sales.

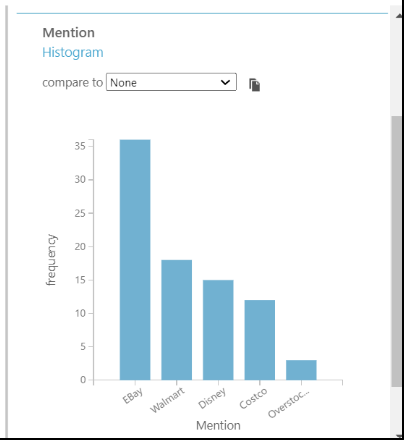
Competitor Identification

Here the dataset has been preprocessed and missing values are removed using Azure ML tasks. The customer review attributes of the dataset is provided as the input to Named Entity Recognition task to identify the competitors who sell the similar product.

Competitors Identification › Named Entity Recognition › Entities

rows 84 columns 5

Article	Mention	Offset	Length	Type
799	Overstock.com	86	13	ORG
799	Disney	149	6	ORG
802	Disney	30	6	ORG
805	EBay	146	4	ORG
808	Disney	30	6	ORG
816	EBay	31	4	ORG
816	Costco	86	6	ORG
816	Walmart	143	7	ORG
816	Walmart	201	7	ORG
817	EBay	31	4	ORG
817	Walmart	86	7	ORG
825	EBay	30	4	ORG
853	Disney	30	6	ORG
897	Disney	30	6	ORG



The visualization helps in identifying the top 5 competitors for this product category.

Customer Experience

Since Amazon is a vast ecommerce website, we have gathered a customer reviews data set from Kaggle.com using which we will answer our third research question which focuses on consumer experience. The below is the table of our data set contents

Feature	Description
customer id	This includes the unique ids of the customers

review_id	This includes the unique ids of the reviews
review_score	This has a rating from 1 to 5.
review_creation_date	Date of creation of reviews
review_answer_timestamp	Time of answering the reviews
compound	This is a sentiment score based on the reviews
Satisfaction (target)	This has the values 0 or 1

Our target field is Satisfaction which considers the customer rating and sentiment score. The rating above 3 is taken as 1 and below the rating 3 is taken as 0 which means satisfied or not respectively.

Classification:

The above problem can be formulated as a binary classification problem. Based on the overall reviews and ratings given by the customer on we will predict if the customer is satisfied with the experience.

Classifiers:

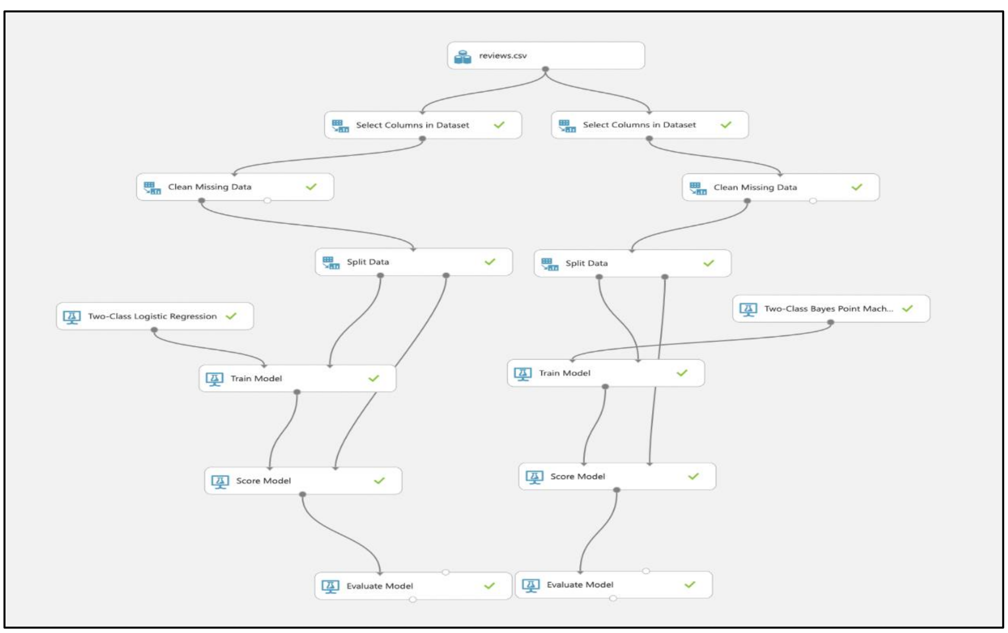
The classification models selected for categorization are:

1. Two class logistic regression
2. Two class Bayes Point Machine

Model Evaluation:

The analysis of two class logistic regression and two class Bayes Point Machine was based on terms of different parameters like, Area under the curve, accuracy, precision, the number of true positives, false positives, true negatives, and false negatives.

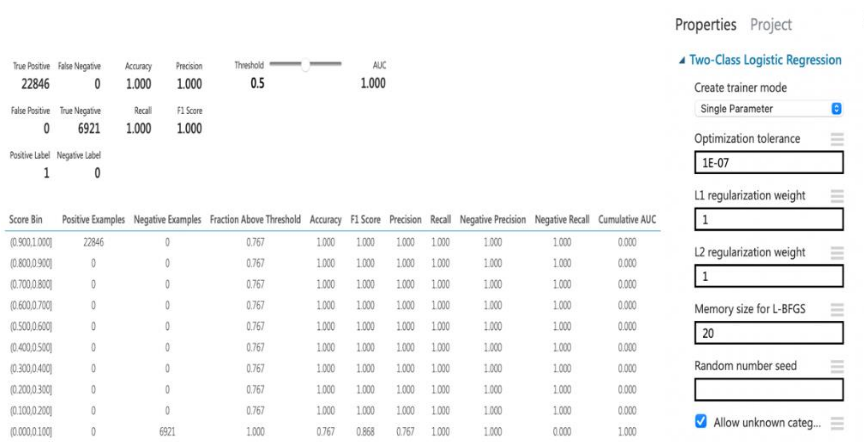
The screenshot below shows the implementation of these two algorithms using different parameters



Two class logistic regression:

A well-known statistical technique for estimating the likelihood of an outcome, logistic regression is particularly well-liked for classification tasks. The technique fits data to a logistic function to estimate the likelihood that an event will occur.

Following are the different parameters selected for two class boosted decision tree and corresponding results.



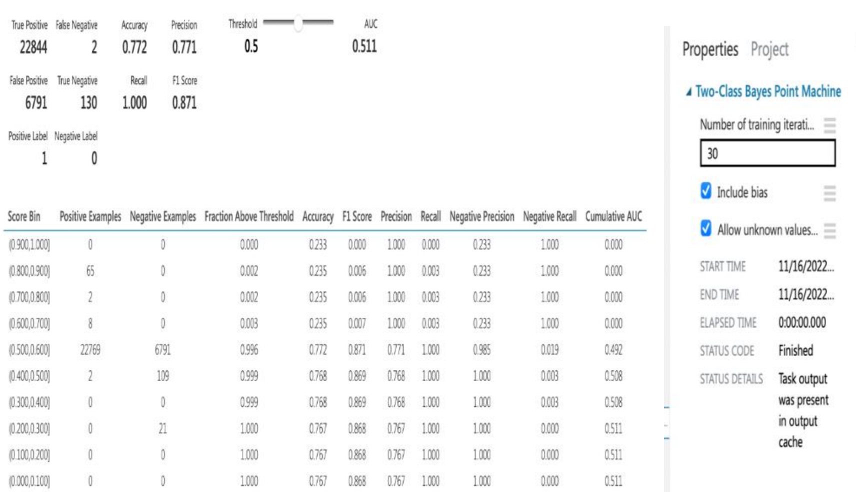
Score Bin	Positive Examples	Negative Examples	Fraction Above Threshold	Accuracy	F1 Score	Precision	Recall	Negative Precision	Negative Recall	Cumulative AUC
(0.900,1.000]	22846	0	0.767	1.000	1.000	1.000	1.000	1.000	1.000	0.000
(0.800,0.900]	0	0	0.767	1.000	1.000	1.000	1.000	1.000	1.000	0.000
(0.700,0.800]	0	0	0.767	1.000	1.000	1.000	1.000	1.000	1.000	0.000
(0.600,0.700]	0	0	0.767	1.000	1.000	1.000	1.000	1.000	1.000	0.000
(0.500,0.600]	0	0	0.767	1.000	1.000	1.000	1.000	1.000	1.000	0.000
(0.400,0.500]	0	0	0.767	1.000	1.000	1.000	1.000	1.000	1.000	0.000
(0.300,0.400]	0	0	0.767	1.000	1.000	1.000	1.000	1.000	1.000	0.000
(0.200,0.300]	0	0	0.767	1.000	1.000	1.000	1.000	1.000	1.000	0.000
(0.100,0.200]	0	0	0.767	1.000	1.000	1.000	1.000	1.000	1.000	0.000
(0.000,0.100]	0	6921	1.000	0.767	0.868	0.767	1.000	1.000	0.000	1.000

Here we can observe that the model identified 22846 true positive values and 6921 true negative and the accuracy is 1.00.

Two class Bayes Point Machine:

A statistical classifier that maps input feature vectors to output class labels. For a set of training data D, each row is represented by an n-dimensional feature vector, X =x1, x2, …, xn. There are K classes, K1, K2, ..., Km in the output class label. For every tuple X, the classifier will predict 2 as given by Eq. 2 that X belongs to Ki if and only if: P (Ki|X) > P (Kj |X), where i, j □ [1,m] and i □ j. P(Ki|X)=□n a=1P(xk|Ki)

Following are the different parameters selected for two class boosted decision tree and corresponding results:



Score Bin	Positive Examples	Negative Examples	Fraction Above Threshold	Accuracy	F1 Score	Precision	Recall	Negative Precision	Negative Recall	Cumulative AUC
(0.900,1.000]	0	0	0.000	0.233	0.000	1.000	0.000	0.233	1.000	0.000
(0.800,0.900]	65	0	0.002	0.235	0.006	1.000	0.003	0.233	1.000	0.000
(0.700,0.800]	2	0	0.002	0.235	0.006	1.000	0.003	0.233	1.000	0.000
(0.600,0.700]	8	0	0.003	0.235	0.007	1.000	0.003	0.233	1.000	0.000
(0.500,0.600]	22769	6791	0.996	0.772	0.871	0.771	1.000	0.985	0.019	0.492
(0.400,0.500]	2	109	0.999	0.768	0.869	0.768	1.000	1.000	0.003	0.508
(0.300,0.400]	0	0	0.999	0.768	0.869	0.768	1.000	1.000	0.003	0.508
(0.200,0.300]	0	21	1.000	0.767	0.868	0.767	1.000	1.000	0.000	0.511
(0.100,0.200]	0	0	1.000	0.767	0.868	0.767	1.000	1.000	0.000	0.511
(0.000,0.100]	0	0	1.000	0.767	0.868	0.767	1.000	1.000	0.000	0.511

Here we can observe that the model identified 22844 true positive values and 130 true negative 2 false negative and 6791 false positives.

By comparing both the models, we derive at a conclusion that two class Bayes Point Machine is better since it identifies more false positives and false negatives.

Conclusion:

From our analysis, we found the major competitors, performed sentimental analysis on the reviews and predicted if products could be recommended to customers based on the sentiment score and identified if the customer is satisfied by the experience.

Future Work:

1. **Competitive analysis**: Now that we identified the competitors our future goal is to perform sentimental analysis on the product reviews of the competitors

2. **Product Development**: Reviews assist you in directing the development of new items in a way that makes sense for your customer rather than generating them at random. They are more inclined to buy from you, repurchase your goods, and refer you to their friends when you design things around their needs.

3. **Detecting product bugs**: Reviews are a crucial tool for letting you know if there are any problems with your goods from customers. It's crucial that you consider and act upon these reviews. As a result, you may provide your customers a product that works properly. You may avoid any customer annoyance and ensure that your consumers like using your items by swiftly acting on their reviews.