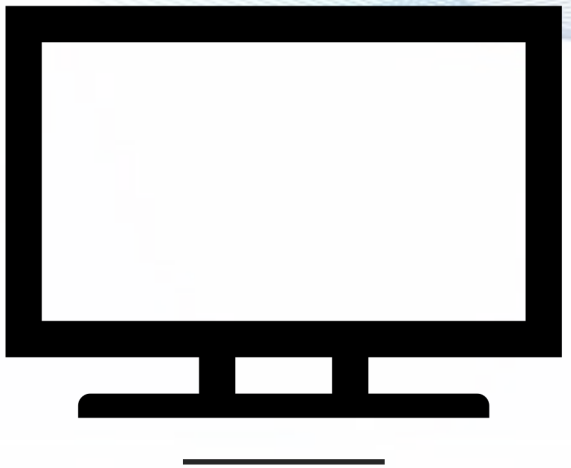
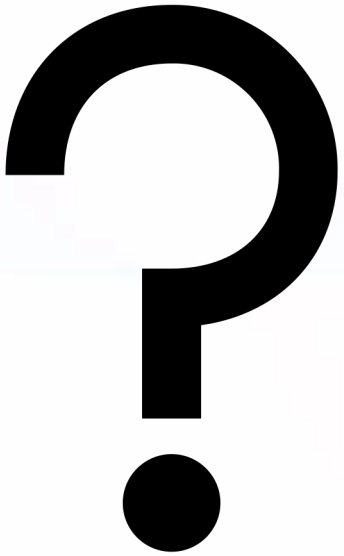
Capstone Project On Sentiment Analysis

By Kanwal Arora
September 2020



Problem Statement

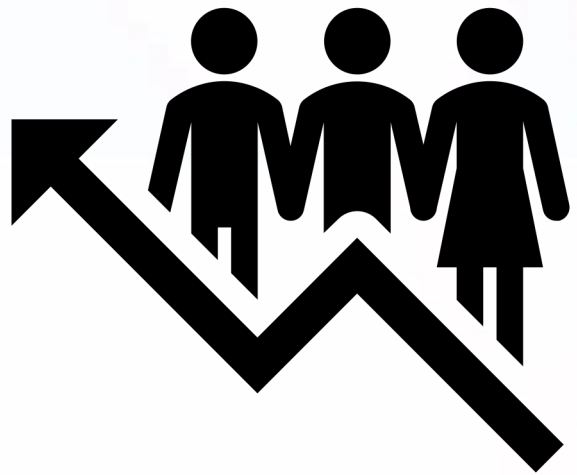
In this section I am going to shed a light on main issues to get the better understanding of how a reviews put impact on the sale of any product or any department and how we can maximize profit by understating the nature of review. We will discuss some problems and solution in next few slides.



Business Values

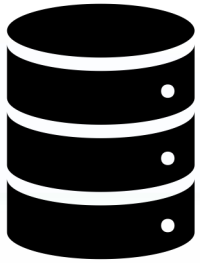
Here we will touch some most crucial aspects to grow in this industry.

Some major aspects to understand.



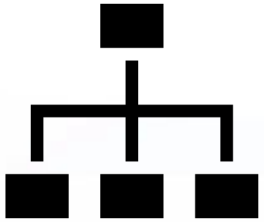
- Understand the sentiment of review.
- Will predict the how department or product will perform in future
- Most Important is sentiment of the audience.
- We will touch the surface of other important aspects.

Methodology



Load & analyze the data

- Kaggle Amazon Review Database
- IMDB Movie Review Database

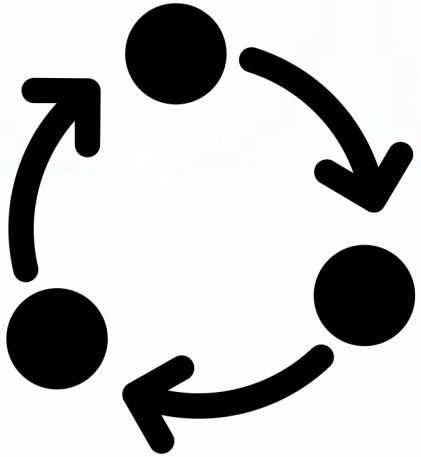


Organizing Data Using

Train model using different Machine Learning and Deep Learning Models for highest accuracy

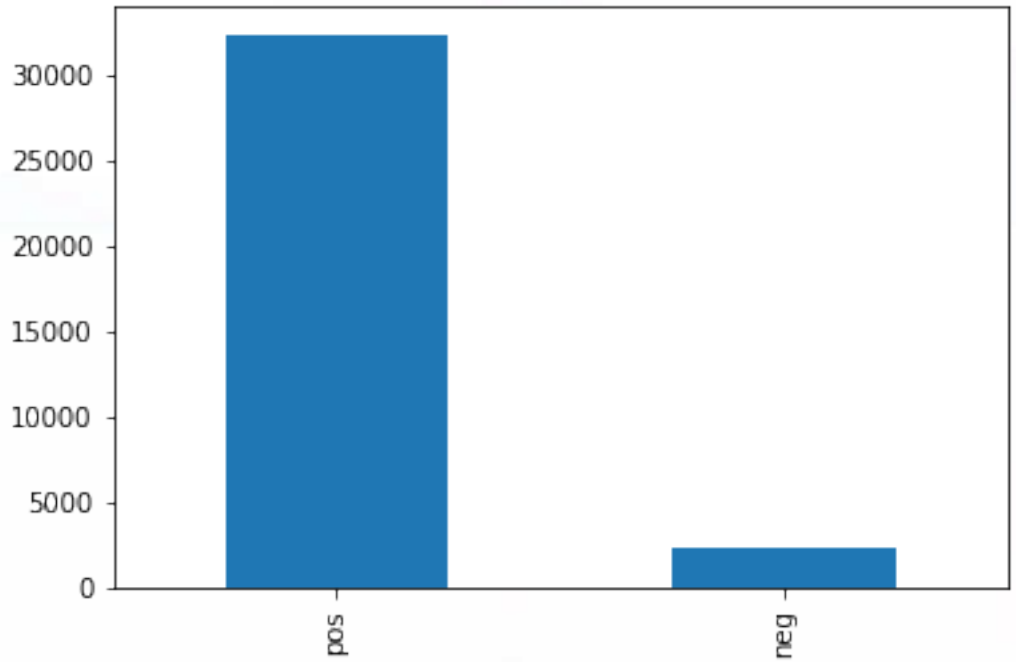
Use Stats Tools

We will extract the solutions using averages, totals, and will use graphs to paint a better picture.



Methodology

- Data Cleaning
- Feature Engineering
- Logistic Regression
- Bernouli NB
- Multinomial NB
- Naïve Bayes
- TFIDF
- NLTK

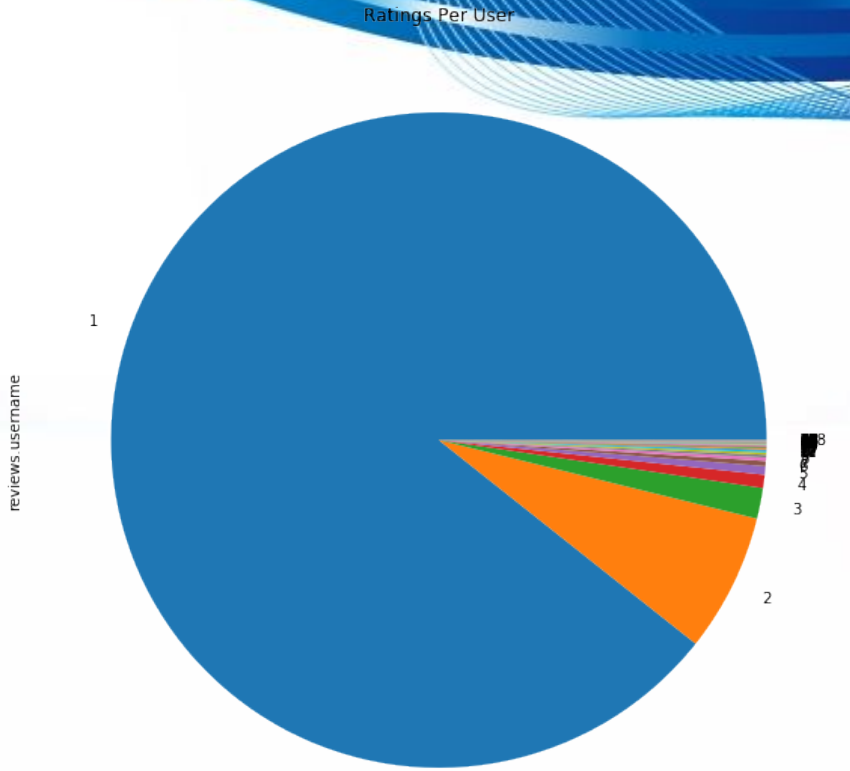


Amazon Review Sentiment Analysis:

Finding 1 -

We have learned from this section:

1. **Only 0.55 % of the users are bulk users**
2. **Around 9 % of the ratings have been submitted by just 0.55% users - Does it seem odd to you ?**



Amazon Review Sentiment Analysis:

Multinomial NB:-

	Precision	Recall	F1-score	Support
positive	0.00	0.00	0.00	464
negative	0.93	1.0	0.97	6461
accuracy			0.93	6925
macro avg	0.47	0.50	0.48	6925
weighted avg	0.87	0.93	0.90	6925

Amazon Review Sentiment Analysis:

Bernoulli NB:-

	Precision	Recall	F1-score	Support
positive	0.33	0.17	0.23	464
negative	0.94	0.97	0.96	6461
accuracy			0.92	6925
macro avg	0.63	0.57	0.59	6925
weighted avg	0.90	0.92	0.91	6925

Amazon Review Sentiment Analysis:

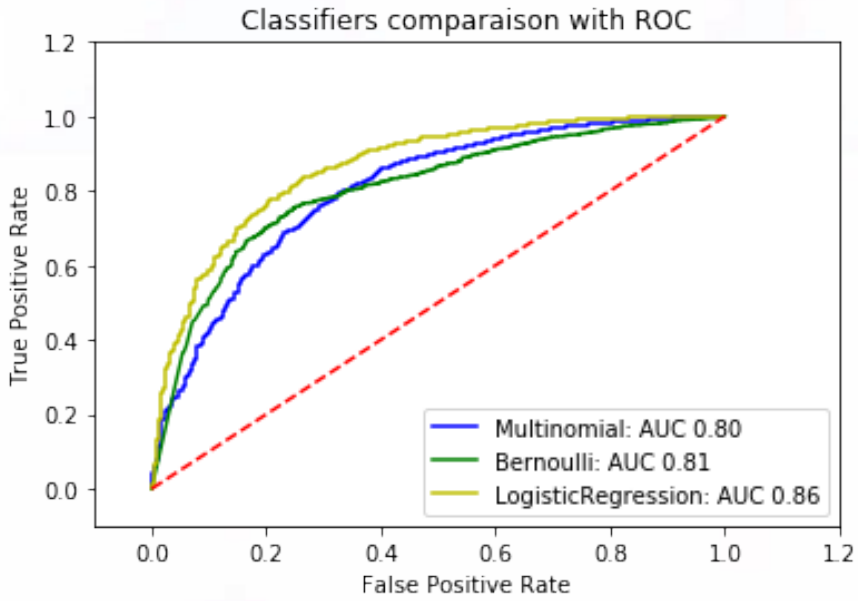
Logistic Regression:-

	Precision	Recall	F1-score	Support
positive	0.56	0.33	0.41	464
negative	0.95	0.98	0.97	6461
accuracy			0.94	6925
macro avg	0.75	0.65	0.69	6925
weighted avg	0.93	0.94	0.93	6925

Conclusion

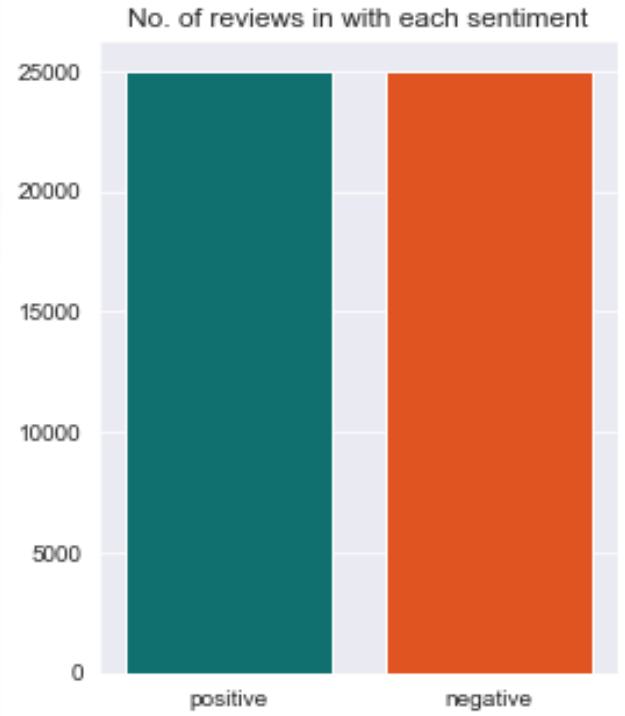
Now we have better understanding of the model which can give us higher area under curve to predict the sentiment of a particular product on Amazon . Concluding with following points.

- Logistic Regression Model is giving us highest area under curve among all trained models with 86%.
- Multinomial is with least AOC of 80%.
- Bernoulli is with 81 %.



IMDB Review Sentiment Analysis:

Numbers of Positive & Negative Reviews-

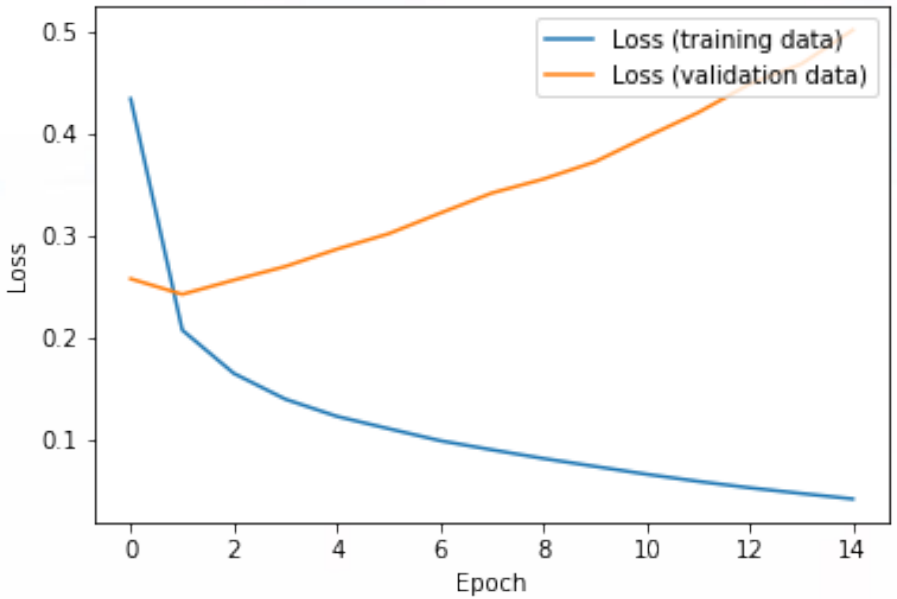


We have learned from this section:

1. **We have equal number of reviews 25000 each of positive and negative**

IMDB Review Sentiment Analysis:

Artificial Neural Network Loss-



We have learned from this section:

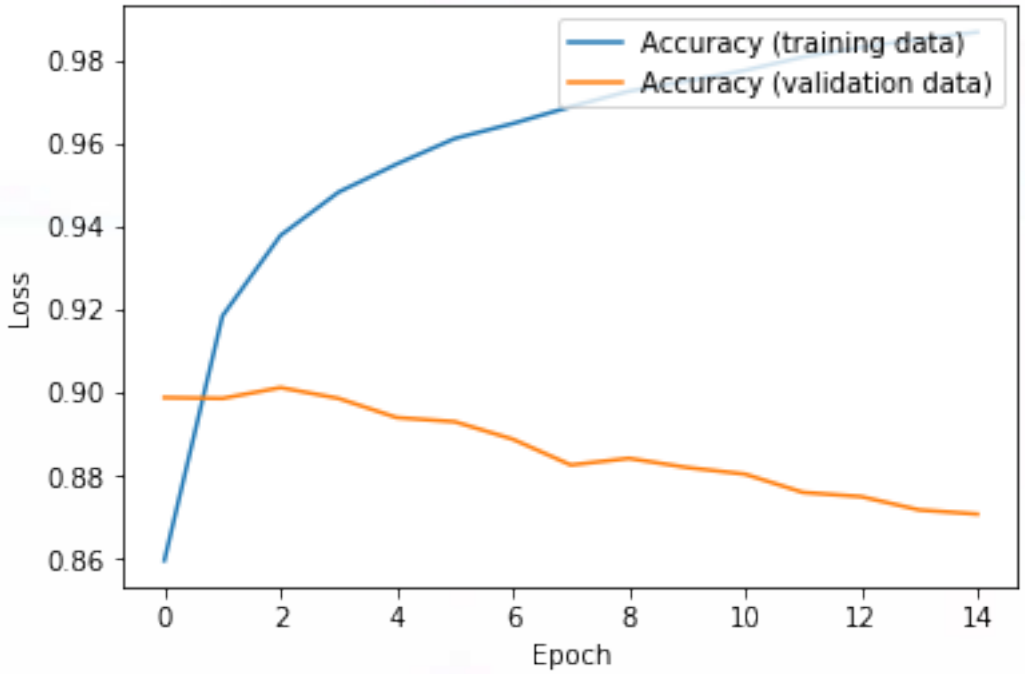
1. **In this picture we can see loss has been reduced to 0.043 and validation loss touched 0.5017**

IMDB Review Sentiment Analysis:

Artificial Neural Network Accuracy-

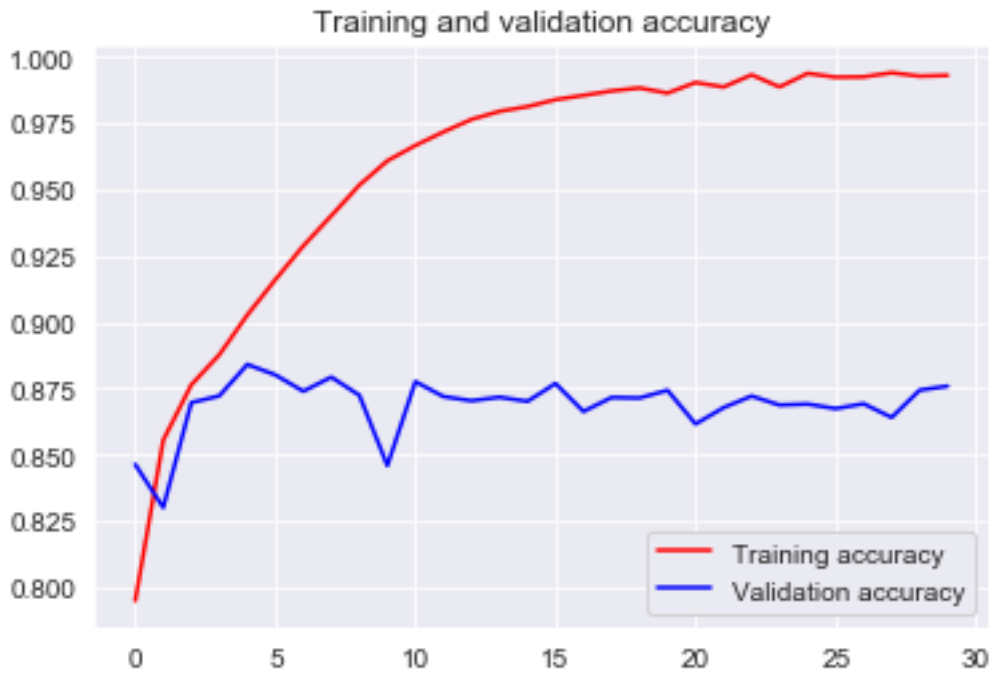
We have learned from this section:

1. **In this picture we can see max accuracy is 0.9866 and validation accuracy reached to 0.8708**



IMDB Review Sentiment Analysis:

Deep Neural Network Accuracy-

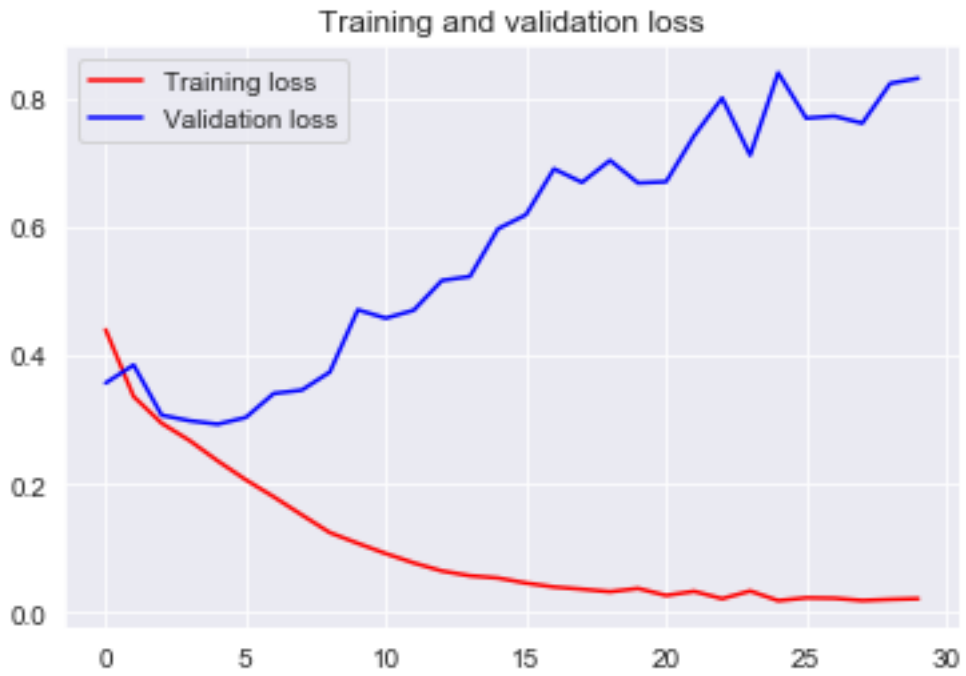


We have learned from this section:

1. **In this picture we can see max accuracy is 0.9930 and validation accuracy reached to 0.8758**

IMDB Review Sentiment Analysis:

Deep Neural Network Loss-



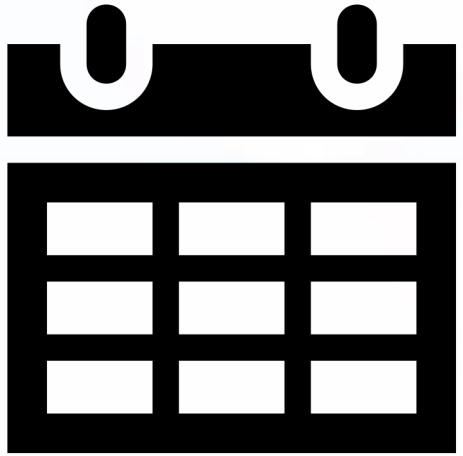
We have learned from this section:

1. **In this picture we can see loss has been reduced to 0.0214 and validation loss touched 0.8758**

Future Work

Plans to dig deeper to get precise information

- Understand the legitimacy of review whether its (Fake or Real)



Thank You

Capstone Project On Sentiment Analysis

By Kanwal Arora

September 2020