

# A Lexicon based Unsupervised Model to Evaluate Product Ratings Vs Reviews

**Taiba Naz**  
**Mahapara G.**  
**Ramamani V.**

# Topics Covered

- ☐ *Abstract*
- ☐ *Problem Statement*
- ☐ *Literature Survey*
- ☐ *Approach Layout*
- ☐ *Research Methodology*
- ☐ *Data Understanding*
- ☐ *Lexicon-based Approach*
- ☐ *ML(KNN) Model*
- ☐ *Result and Conclusion*

# *Abstract*

- Now in this advancing World E-commerce shoppers would like to look at Customer Reviews as a reliable source of information. These E-commerce websites provide features for Customers to write the **Product Reviews and Score the Product from 1 to 5 or it's commonly referred to as star rating.**
- But sometimes it is seen that there are **Inconsistencies between the actual Sentiment of the Review which differ from the Star Ratings given.** Because of that it is necessary to Validate the Star Rating versus the Reviews.
- **2000 Product Reviews from Amazon** has been taken and this methodology has been applied in this study, to validate the results.

# ***Problem Statement***

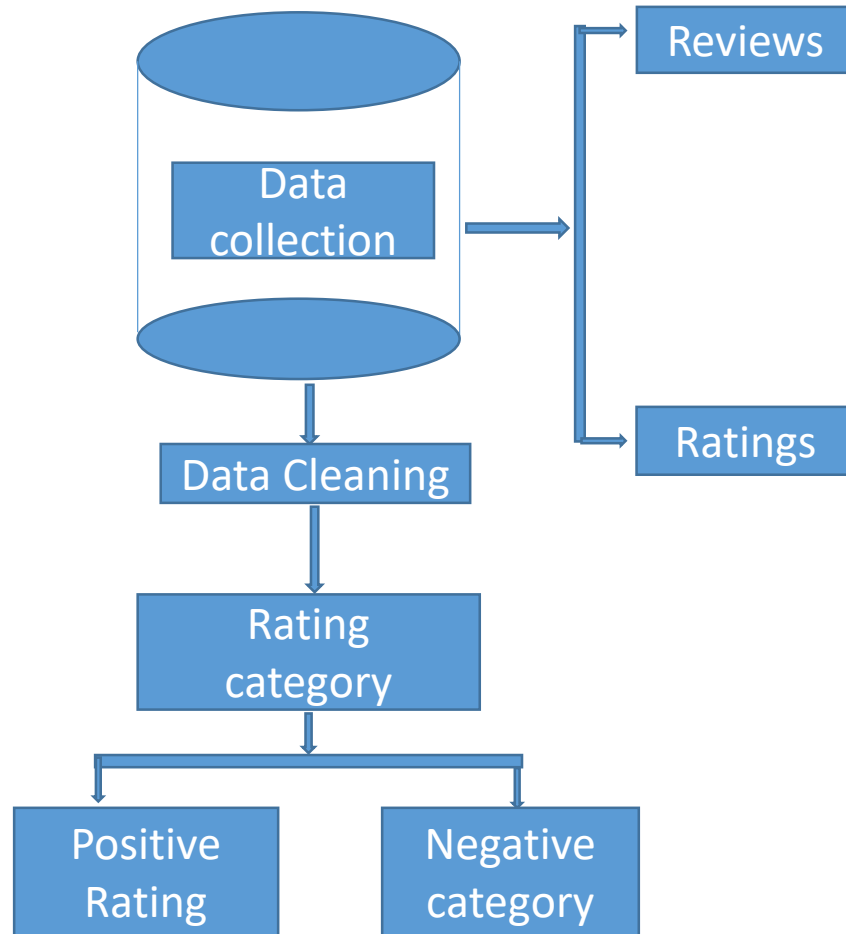
To Find the **inconsistencies** between the **Actual Sentiment** of the Product **Reviews** and the Star **ratings** given by the Customers for the “One Plus 7pro” Mobile Phones.

# Literature Survey

AUTHORS	WORK DONE
Vyas et.al	They describe how both <b>Machine learning and lexicon-based approaches</b> can be effectively used for sentiment analysis.
Shah et.al	They used <b>n-gram classification along with POS tagging for customer reviews.</b>
Kumar et.al	used the <b>lexicon-based approach to derive star rating for the reviews.</b>
Jagdale et.al	used <b>machine learning techniques to classify reviews that are positive or negative.</b>

**From the literature review, it has been found that no work is available in evaluating consistency of reviews w.r.t ratings.**

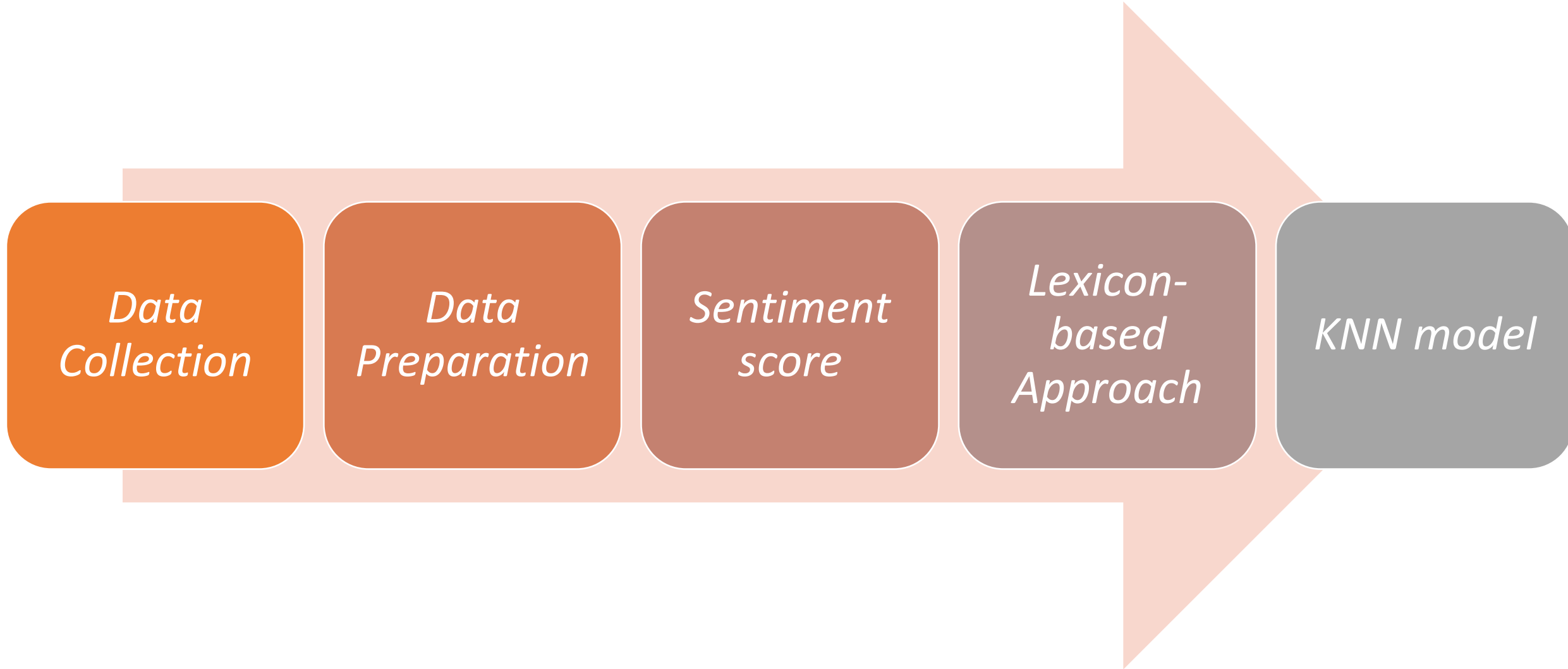
# Approach Layout



**Data Source:** Amazon Shopping site

- Data have been scrapped using Parse hub tool to get the desired dataset.
- Data is then cleaned to get the proper product reviews of the customers who have purchased history of the product.
- The then reviews have been classified into Positive and Negative reviews based on their sentiments.

# *Research Methodology*

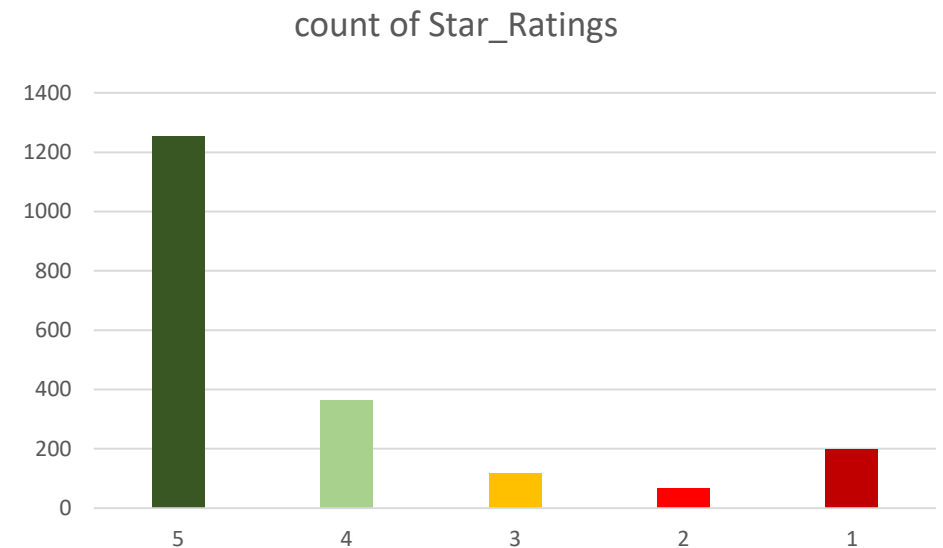


- For our work we collected the Reviews for “**One Plus 7pro**” Mobile Phone from Amazon. A total of **2000** reviews along with the respective Star Ratings given by the customers who have bought this Phone through were collected.

## *Dataset Used for the Analysis*

Reviews	Star_Ratings
Best Display in Market!	5
Awesome purchase	5
Picture Quality and Slow Motion Video Recording Blanking problem	2
Overpriced to the cheap quality	1
Unhappy with one plus	3
Pay more for more happiness	4
Just OK , Consider as 3rd option to buy	3
Go for it	5
Battery life..?	1
Not receive complete product	1
Wow. Excellent Mobile	5

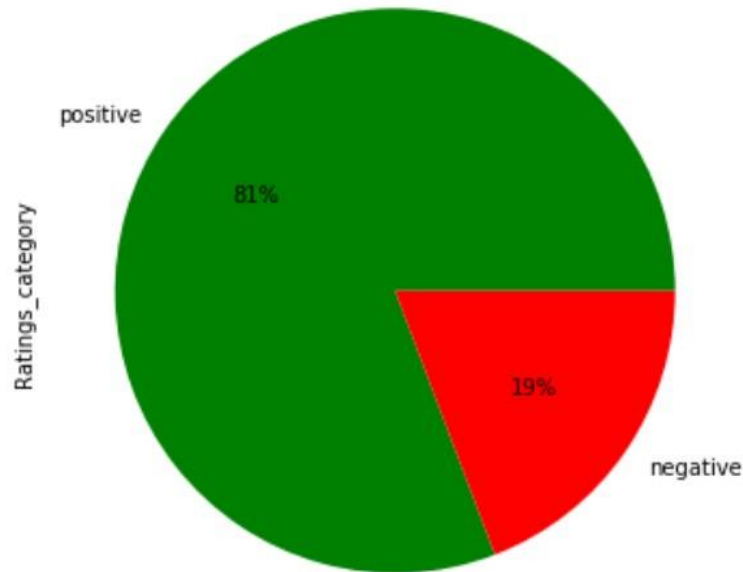
## *Star Ratings based on the product Reviews*





# Data Preparation

From The Frequency of star rating figure we got to know that Over 81% reviews were positive (4 or 5 out of 5 stars).The Reviews which received the Star Rating of 4 or 5,has been categorized as the” **Positive rating**”.



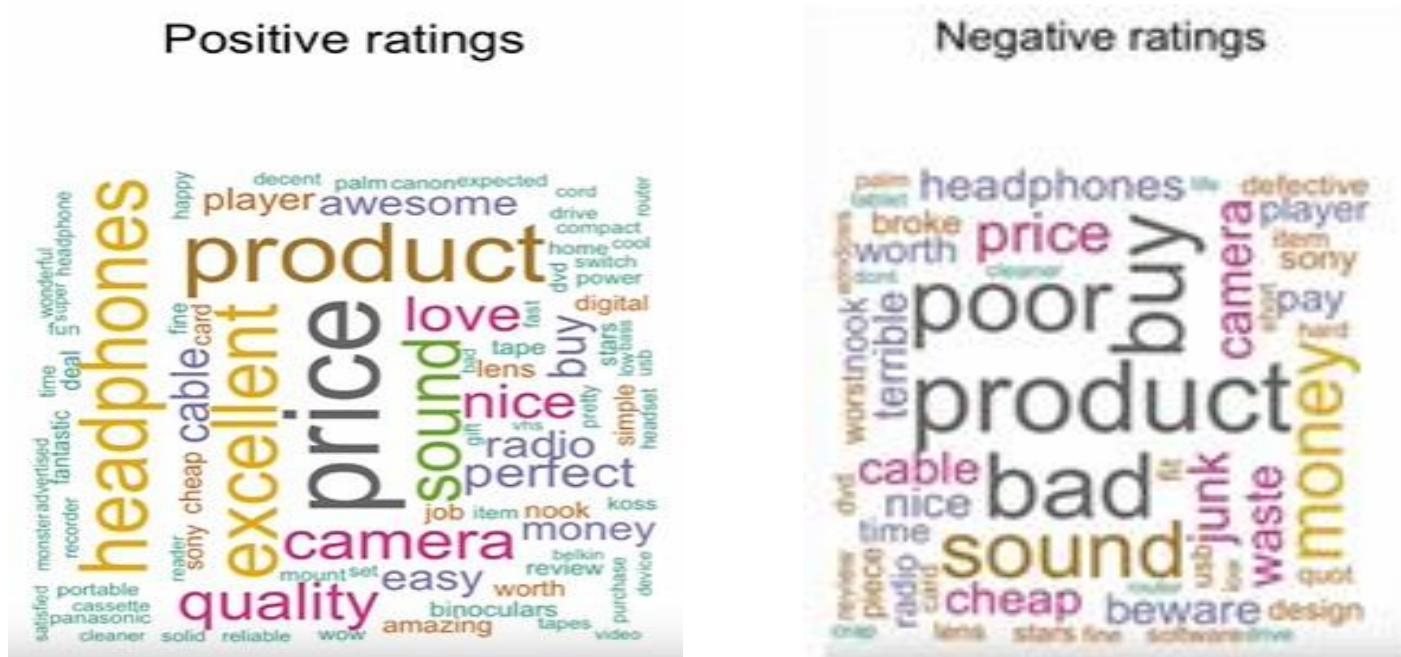
**Figure 1:** Percentage of Star Rating

Ratings_Category	Actual_Count
Positive	1616
Negative	384

**Table 1.** Count of Star-ratings category

# Data Pre-processing

We performed pre-processing steps before applying Sentiment analysis methods. We categorized the reviews as “Positive” whose star ratings are “>3” and “Negative” for those whose star ratings are “<3”.



**Figure 2:** Word cloud of Positive and Negative words

Reviews	Star_Ratings	Rating Category
Best Display in Market!	5	Positive
Awesome purchase	5	Positive
Overpriced to the cheap quality	2	Negative

**Table 2.** Reviews along with the Star-ratings category

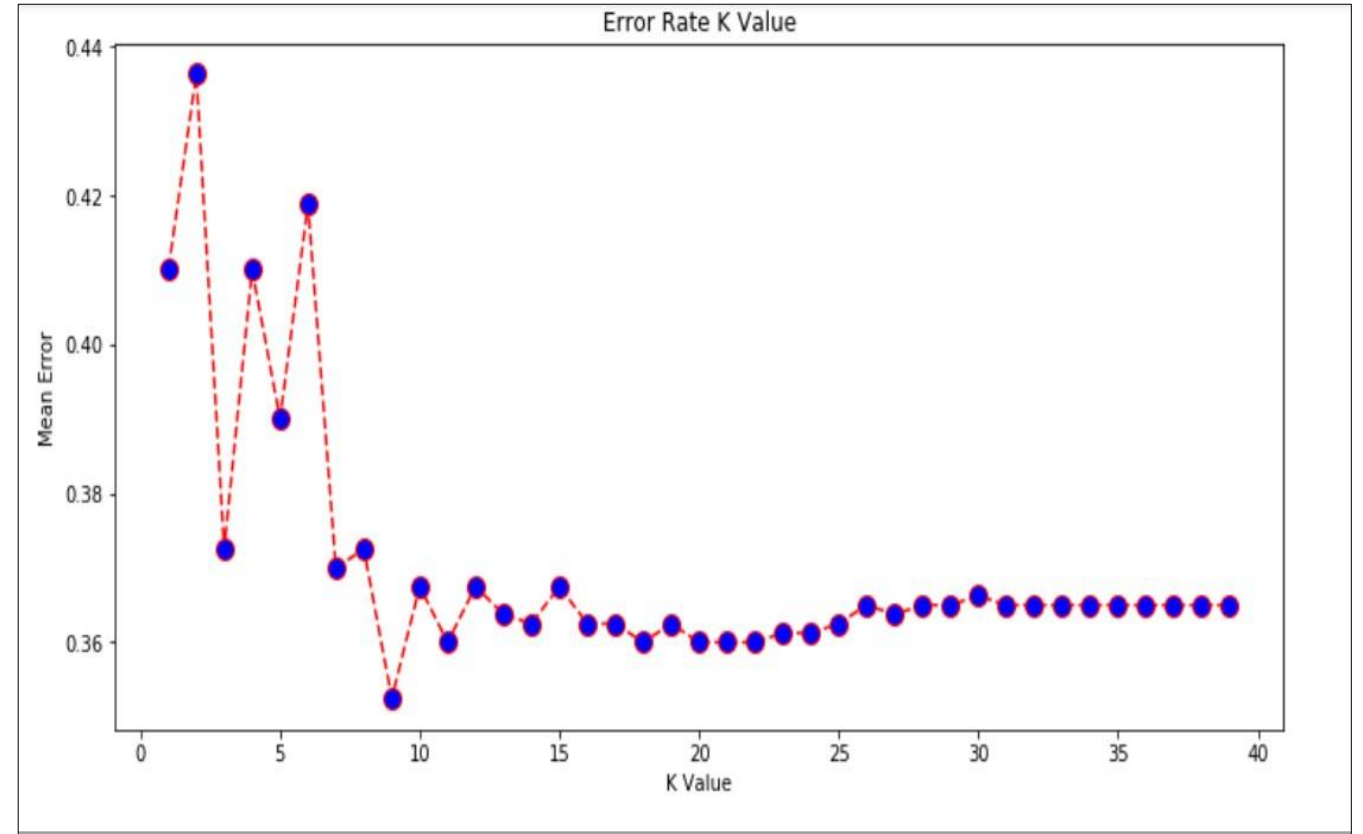
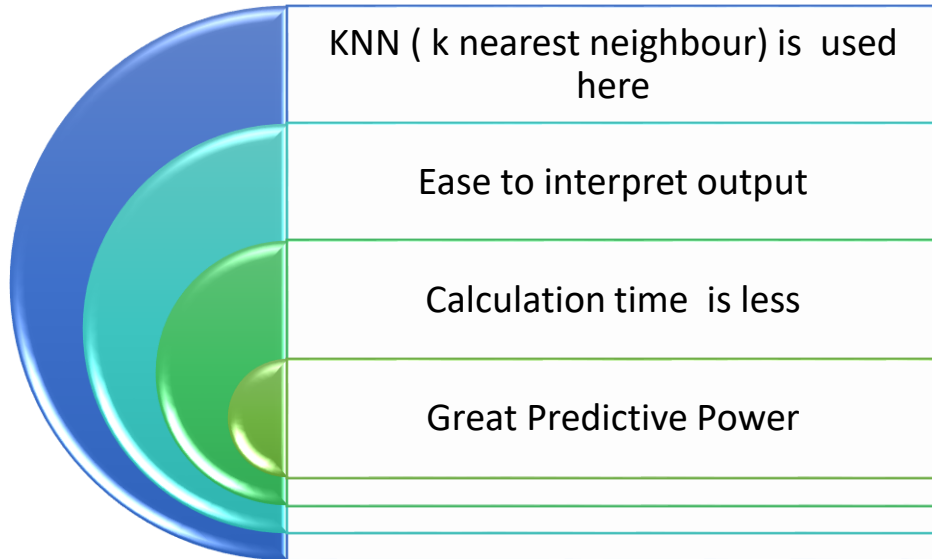
# *Lexicon-based Approach*

- Lexicon-based approach is used here for Sentiment classification.
- AFINN- 111 is a database of words with polarity attached to each word.
- The Valence is numbered between minus five (negative) and plus five (positive).
- Sentiment score was calculated for each review with the help of this.

Reviews	Star_Ratings	Sentiment_score
Best Display in Market!	5	3
Awesome purchase	5	4
Overpriced to the cheap quality	2	0

**Table 3:**Product reviews with their respective star ratings, sentiment score

# *KNN Model*



**Error Rate at Different K values**

# Result and Conclusion

- Sentiment analysis or opinion mining is a field of study that analyzes people's sentiments, attitudes, or emotions towards certain entities. This paper tackles a fundamental problem of sentiment analysis, sentiment polarity categorization. Online product reviews from Amazon.com are selected as data used for this study. A sentiment polarity categorization has been done along with detailed descriptions of each step.
- Experiments have been done both With the unsupervised Lexicon based Approach where we achieved the accuracy of 62%, and with the machine learning based supervised KNN algorithm where the accuracy achieved is 66%.

Model	Accuracy
Lexicon <u>based model</u>	62%
KNN model	66%

**Table 3.** *The results Obtained from the Lexicon based KNN Algorithm*

- From Machine learning model also the desired accuracy is not satisfactory, so here we can conclude that the Reviews are not consistent w.r.t the star ratings.
- So for future work would like to can go for **Feature Engineering** and can use **Deep Learning Model** to see if the Accuracy of the model improves.

