A
Project Report
On

ONLINE GROCERY SHOPPING IN INDIA - SENTIMENT ANALYSIS

By

Protyush Jana

Roll No: 30018021007

Reg No: 213001818010007

Submitted in Partial Fulfillment of the Requirement for the Degree of

MASTER OF SCIENCE IN APPLIED STATISTICS

DEPARTMENT OF APPLIED STATISTICS

MAULANA ABUL KALAM AZAD UNIVERSITY OF TECHNOLOGY (WBUT)

WEST BENGAL, NADIA- 741249

NOVEMBER, 2022

ONLINE GROCERY SHOPPING IN INDIA - SENTIMENT ANALYSIS

Project work submitted by

Protyush Jana

Roll No: 30018021007

Reg No: 213001818010007

Under the Supervision of

Mr. Taranga Mukherjee

Assistant Professor

Department of Applied Statistics

Maulana Abul Kalam Azad University of Technology (WBUT)

West Bengal, Nadia 741249

DEPARTMENT OF APPLIED STATISTICS

MAULANA ABUL KALAM AZAD UNIVERSITY OF TECHNOLOGY (WBUT)

WEST BENGAL, NADIA- 741249

NOVEMBER, 2022

MAULANA ABUL KALAM AZAD UNIVERSITY OF TECHNOLOGY (WBUT)

WEST BENGAL, NADIA- 741249

CERTIFICATE

I hereby forward this project thesis entitled "ONLINE GROCERY SHOPPING IN INDIA - SENTIMENT ANALYSIS" by *Protyush Jana* (Roll No: 30018021007; Reg No: 213001818010007) of 2021-23 in partial fulfillment of the requirement for the degree of MASTER IN APPLIED STATISTICS AND ANALYTICS of the DEPARTMENT OF APPLIED STATISTICS, MAULANA ABUL KALAM AZAD UNIVERSITY OF TECHNOLOGY (WBUT), WEST BENGAL, NADIA- 741249.

This project thesis has been completed under my guidance in the Department of Statistics, Maulana Abul Kalam Azad University of Technology (WBUT), West Bengal.

_____

TARANGA MUKHERJEE
Supervisor
Assistant Professor
Department of Applied Statistics
Maulana Abul Kalam Azad University of Technology, WB

Countersigned:

_____

ANWESHA SENGUPTA
Professor & Head of the Department
Department of Applied Statistics
Maulana Abul Kalam Azad University of Technology
West Bengal, Nadia 741249

ACKNOWLEDGEMENT

Acknowledgement is not simply a ritual. It is an expression of heartfelt gratitude and indebtedness to all those who have been associated with the development of the thesis.

I would like to express my profound and deep sense of gratitude to my guide *Prof. (Mr.) Taranga Mukherjee* for his unending help, guidance, and suggestions without which this thesis would not have been a reality. He has acted as a friend, philosopher, and guide to me. I owe great indebtedness for his untiring effort thoughtful the period of my project work.

I express my sincere thanks to *Prof. Anwesha Sengupta*, Head of the Department of Applied Science, Maulana Abul Kalam Azad University of Technology, Kalyani, West Bengal, India, for his cooperation and extended during the period of my project work.

I am great fully indebted to *Prof. Prasanta Narayan Dutta*, Course Coordinator, and *Prof. Sukhendu Samajdar*, Director, Department of Applied Science, Maulana Abul Kalam Azad University of Technology, Kalyani, West Bengal, India, for their great inspiration and encouragement for my work and for providing me with a favourable environment in the form of infrastructural facilities for the research work.

I am greatly indebted to all the faculty members, who often took pains and stood by me in adverse circumstances. Without their encouragement and inspiration, it was not possible for me to complete this project.

Finally, my earnest thanks go to my friends who were always beside me when I needed them without any excuses and made these three years worthwhile.

_____

(Protyush Jana)

CONTENTS

ABSTRACT

India is facing a COVID-19 lockdown that relies heavily on digitalization. When the lockdown began, people began to run out of groceries. Consumers experienced many problems while receiving groceries from grocery stores during the pandemic, causing abnormally high sales for online grocery stores. When products cannot be delivered on time, maintaining consumer satisfaction levels becomes very difficult. At that time, consumers posted many online reviews on different digital platforms. Online grocery stores can easily deal with huge online reviews and segment consumers based on their positive and negative comments. This paper presents a new lexicon-based sentiment analysis algorithm that has been designed with the main focus on people's opinions regarding Indian online grocery shopping companies' services and how they can perform better. This study focused on analyzing those reviews by Lexicon methods. By using this method, we identified which app people prefer the most and which facts the company should keep in mind to better their service.

INTRODUCTION

Grocery delivery apps are transforming how Indian businesses operate. Grocery app development services in India have helped grocery delivery companies to revolutionize the way people shop for their groceries. We've seen significant growth in the number of grocery stores that offer online shopping, and this is only going to increase as more Indians adopt smartphones over the next few years.

The recent covid-19 pandemic has driven increased demand for these services. Whether healthcare, production, or management, every sector strives to improve online shopping experiences by providing better customer service. Among the leading grocery apps available in India are Big Basket, Blinkit, etc. It is becoming more and more popular because it saves people time and effort by providing them with a comfortable facility.

It's no longer clothes or books or makeup items that people want to buy online. Even when it comes to buying everyday groceries, people are increasingly turning online, amid the pandemic. A recent study

indicates that online grocery shopping is likely to grow more than two-fold as more shoppers are now ordering items online.

Multiple studies show that online grocery shopping has surged in the last few months. While studies indicate a 76% growth in the online grocery market as compared to last year, another study by Local Circles found that 80% of people ordered groceries online for convenience and not price. Due to the pandemic and lockdown, while sales on online grocery sites such as Grofers and BigBasket increased, aggregators like Zomato, Swiggy, and JioMart also started delivering groceries.

Amazon and Flipkart have also seen an increase in grocery orders. As reported in Economic Times, BigBasket says that since March, its new customer acquisition grew three-fold in terms of orders, compared to pre-COVID, while retention of new customer cohorts has increased by 60%.

What is Sentiment Analysis?

Sentiment Analysis is a Natural Language Processing task; it refers to finding patterns in data and inferring the emotion of the given piece of information which could be classified into one of these categories: Negative, Neutral, and Positive.

As, for the first time in India, millions of people from different communities' order groceries from the online delivery app. So, just like it is new for Indian customers to use these apps, it's also new for the app company to build, grow, manage orders, maintain good service, and satisfy customers. Here, in this project we picked up the top three online grocery apps in India i.e., BigBasket, Blinkit, and Jio Mart, and analyze customers' sentiments over these three apps using their reviews. We manage to find out which app is better than the other two and why, also it will help new app makers to look up at the glitch side of these apps that they can rectify.

The online grocery delivery app is used by millions of people from different communities in India for the first time. So, just like it is a first for Indian customers to use these apps, it's also a first for the app company to build, grow, manage orders, maintain high service, and satisfy customers. Here, in this project we picked up the top three online grocery apps in India i.e., BigBasket, Blinkit, and Jio Mart. We analyzed customers' sentiments about these three apps based on their reviews. Besides finding out which app does better than the other two and why we can also look for glitches in these apps to be able to rectify them in the future.

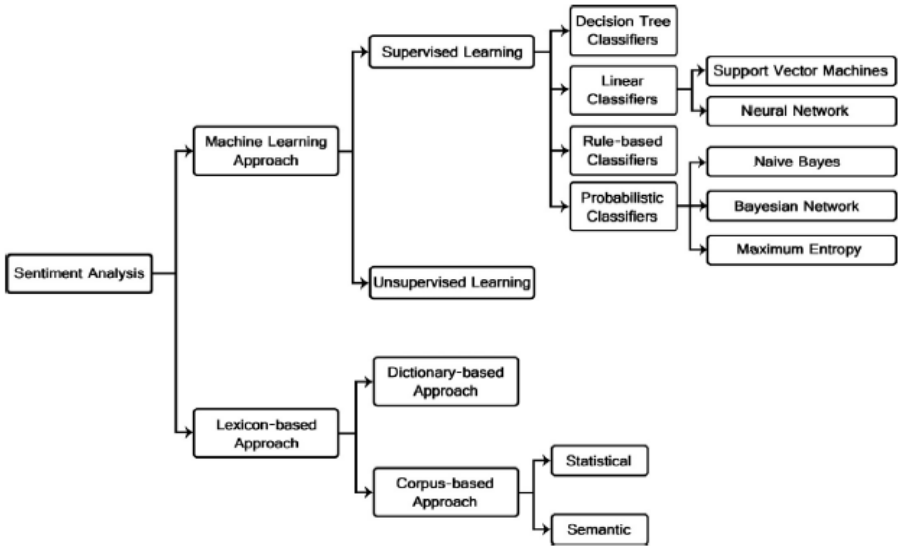
METHODOLOGY

To draw the Conclusion, the steps that we have done in our project are described below.

In words, first we collected the data; we cleaned the data by dropping the null values and by eliminating the numerical values and special characters. After that we analyze the sentiment of the reviews by following methods.

Sentiment analysis is used to determine whether a given text contains negative, positive, or neutral emotions. It's a form of text analytics that uses natural language processing (NLP) and machine learning. Sentiment analysis is also known as "opinion mining" or "emotion artificial intelligence".

Now, Sentiment Analysis can be done in different ways:



Sentiment classification techniques

Here the method we have used in our project is Lexicon-based Sentiment Analysis. Now, generally speaking, in lexicon-based approaches, a piece of the text message is represented as a bag of words.

Following this representation of the message, sentiment values from the dictionary are assigned to all positive and negative words or phrases within the message. A combining function, such as sum or average, is applied in order to make the final prediction regarding the overall sentiment for the message. Apart from a sentiment value, the aspect of the local context of a word is usually taken into consideration, such as negation or intensification.

The lexicon-based approach depends on finding the opinion lexicon which is used to analyze the text. There are two methods in this approach. The dictionary-based approach depends on finding opinion-seed words and then searching the dictionary for their synonyms and antonyms. The corpus-based approach begins with a seed list of opinion words and then finds other opinion words in a large corpus to help in finding opinion words with context-specific orientations. This could be done by using statistical or semantic methods.

We tried to analyze by using the user rating and four types of dictionary-based approaches; i.e., TextBlob, SentiWordNet and Vader.

Our project work is done by NLP. Natural language processing (NLP) is an area of artificial intelligence (AI) that focuses on assisting computers in understanding how people write and communicate. This is a tough process because of the large amount of unstructured data involved. NLTK is a leading platform for building Python programs to work with human language data. It provides easy-to-use interfaces to over 50 corpora and lexical resources such as WordNet, along with a suite of text processing libraries for classification, tokenization, stemming, tagging, parsing, and semantic reasoning, wrappers for industrial-strength NLP libraries, and an active discussion forum. NLTK has been called “a wonderful tool for teaching and working in, computational linguistics using Python,” and “an amazing library to play with natural language.”

Now, before using the Textblob method one has to make sure all the works are in lower/ upper case, remove stopwords, tag parts of speech and lemmatization. Parts of speech tagging simply refers to assigning parts of speech to individual words in a sentence, which means that, unlike phrase matching, which is performed at the sentence or multi-word level, parts of speech tagging are performed at the token level. From a very small age, we have been made accustomed to identifying part of speech tags. For example, reading a sentence and being able to identify what words act as nouns, pronouns, verbs, adverbs, and so on. All these are referred to as part of speech tags. There are several Lemmas for the same word at times. Lemmatization is a text normalization technique used in Natural Language

Processing (NLP), that switches any kind of word to its base root mode. Lemmatization is responsible for grouping different inflected forms of words into the root form, having the same meaning. It takes into account the context when converting a word to its meaning basic form, known as a Lemma.

❖ **TextBlob:**

A very useful NLP library comes prepackaged with its own sentiment analysis functionality.It is also based on NLTK. The sentiment property of the API/library returns polarity and subjectivity.

Polarity range: -1.0 to 1.0

Subjectivity range: 0.0 – 1.0 (0.0 is very objective and 1.0 is very subjective)

Example:

Problem instance: "Very bad delivery. Customer care no response. Goods not delivery 1-week delivery date."

Polarity: - 0.7

Subjectivity: 0.666667

- Polarity: as mentioned earlier, it is a measurement of how positive or negative the given problem instance is. In other words, it is related to the emotion of the given text.
- Subjectivity: It refers to opinions or views (can be allegations, expressions or speculations) that need to be analyzed in the given context of the problem statement. The more subjective the instance is, the less objective it becomes and vice versa. A subjective instance (e.g., a sentence) may or may not carry any emotion. Examples: "Sam likes watching football", "Sam is driving to work".

❖ **SentiWordNet:**

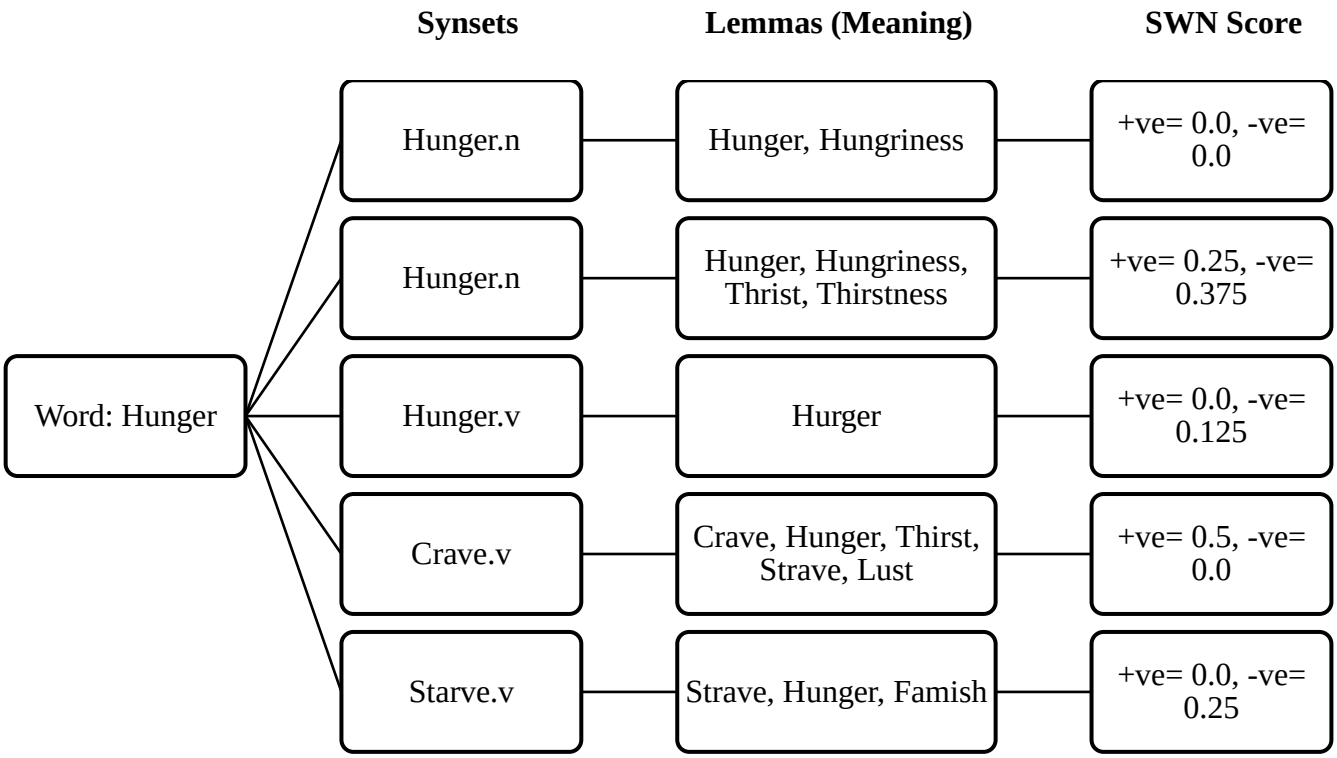
This is also built into NLTK. It is used for opinion mining. This helps in deducing the polarity information from the given problem instance. SWN extends wordnet which is a lexical database of words (the relationship between words, hence the term net), developed at Princeton and is a part of NLTK corpus. Here I'd focus primarily on synsets (Synset — "synonym set" — a collection of synonymous words), which are the logical groupings of nouns, verbs, adjectives and adverbs, into sets or collections of

cognitive synonyms, hence the term synset. NLTK module includes the English WordNet with 155 287 words and 117 659 synonym sets that are logically related to each other.

Example:

Let's take the word 'hunger'.

- The wordnet part: Get a list of all the synsets against hunger
- The Sentiwordnet (swn) part: Get the polarity for all those synsets
- Use one or maybe a combination of all of them to determine the polarity score for that word (hunger in ours). The following diagram shows this flow:



- I am showing lemmas (meanings) only for illustrative purposes here, I find it helpful to understand the synsets I am working on.
- Here is how to interpret the tags assigned to the given synsets:

hunger: the word we need the polarity of

n: part of speech (n = noun)

The following should guide in understanding the given grouping/synset of a word:

n	NOUN
v	VERB
a	ADJECTIVE
s	ADJECTIVE SATELLITE
r	ADVERB

❖ **Vader:**

Vader stands for Valence Aware Dictionary and Sentiment Reasoner. Widely used in analyzing sentiment on social media text because it has been specifically attuned to analyze sentiments expressed in social media. It now comes out of the box in the Natural language toolkit (NLTK). VADER is sensitive to both polarity and intensity. It gives result of four dictionary keys: Positive, Negative, Neutral and Compound. The sum of positive, negative, neutral is 1 or near to 1. We use Compound Score to find the sentiment.

The *compound score* is computed by summing the valence scores of each word in the lexicon, adjusted according to the rules, and then normalized to be between -1 (most extreme negative) and +1 (most extreme positive). This is the most useful metric if you want a single unidimensional measure of sentiment for a given sentence.

As explained in the paper, researchers used the below normalization,

$$Compound\ Score = \frac{x}{\sqrt{x^2 + \alpha}}$$

where x = sum of valence scores of constituent words,

and α = Normalization constant (default value is 15).

Here is how to read the measurements:

- -1: extremely negative
- 1: extremely positive
- 0: Neutral or N/A

It is a score assigned to the word under consideration by means of observation and experiences rather than pure logic.

Example:

Consider the words 'terrible', 'hopeless', and 'miserable'. Any self-aware Human would easily gauge the sentiment of these words as Negative.

While on the other side, words like 'marvelous', 'worthy', and 'adequate' are signifying positive sentiment.

According to the academic paper on VADER, the Valence score is measured on a scale from -1 to +1, where -1 stands for the most ‘Negative’ sentiment and +1 for the most ‘Positive’ sentiment. Intuitively one can guess that midpoint 0 represents a ‘Neutral’ Sentiment, and this is how it is defined actually too.

ANALYSIS

➢ DATA COLLECTION :

The data of customer reviews of three top Indian online grocery home delivery apps, BigBasket, Blinkit and Jio Mart; have been collected from the Indian Google play website. We scrapped the data from Google Play (https://play.google.com/store/apps/details?id=com.bigbasket.mobileapp &hl=en, https://play.google.com/store/apps/details?id=com.grofers.customerapp&hl=en, https://play.google.com/store/apps/details?id=com.jpl.jiomart&hl=en) in Python by using packages like- 'google_play_scraper', 'transformers', 'plotly-express'. The review data of BigBasket, Blinkit and Jio Mart contain 235687, 267824, and 204424 Reviews till 19th October 2022.

➢ DATA DESCRIPTION :

BigBasket, Blinkit and Jio Mart contain 235687, 267824, 204424 Reviews respectively. Each of the data contains Review Id, User Name, User Image, Content, Score, Thumbs Up Count, Review Created Version, At, Reply Content, and Replied At.

In this column, 'Content' describes the user's review, 'Score' describes their rating on a scale of 0 to 5, 'Thumbs Up Count' describes how many people liked the review, 'Reply Content' indicates the company's reply on that particular review of that user, and 'At' and 'Reply At' indicate when the user made the comment and when the company replied.

	reviewId	userName	userImage	content	score	thumbsUpCount	reviewCreatedVersion	at	replyContent	repliedAt
0	f793d90d-73dc-4810-9f1c-6f2d6cd4658c	Tanu Sharma	https://play-lh.googleusercontent.com/a/ALm5wu...	Nice but price higher than jio mart	5	0	None	2022-10-19 08:19:18	Your happiness and ease is our prime importanc...	2022-10-19 08:35:29
1	f8a9f795-01a2-405d-8be2-acedce6ad56d	Mahesh Mahi	https://play-lh.googleusercontent.com/a/ALm5wu...	good fresh products... Well service... On time...	5	0	7.6.0	2022-10-19 08:04:47	Thank you for taking the time to write about y...	2022-10-19 08:20:25
2	d8d1bed2-4551-4d97-9fec-9fa7464e2e7f	Ramesh Datla	https://play-lh.googleusercontent.com/a/ALm5wu...	Fast and easy using and ordering. Will wait an...	4	0	7.6.2	2022-10-19 08:04:16	We greatly appreciate your business.... Thank yo...	2022-10-19 08:07:07
3	bee55523-b64c-4420-8889-f08e3bb896a2	swechchha sarojkar	https://play-lh.googleusercontent.com/a-/ACNPE...	Fake one. If u cancelled an order believe me t...	1	0	7.6.2	2022-10-19 07:37:46	Hi, Please accept our sincere apology. We are ...	2022-10-19 07:55:20
4	c7a61b2a-e044-4c2e-9917-fc4a008cefb2	sam shinchan	https://play-lh.googleusercontent.com/a-/ACNPE...	Didn't enjoy shipping nothing I'm really sugge...	1	0	7.6.2	2022-10-19 07:36:21	Hi sam, Please accept our sincere apology. We ...	2022-10-19 07:42:29

BigBasket data description

We have similar data for Blinkit and Jio Mart.

- **DATA CLEANING :**

In the first step, we try to eliminate missing values from the data. After removing the missing values, we drop the columns that we do not use and remove special characters, emojis and numerical values from each review using Regular Expressions. We then got the cleaned data.

	userName	content	score	Cleaned Reviews
0	Vivek Lakum	❤️❤️❤️❤️	5	
1	Liyakath ali	good app	5	good app
2	jagjeet singh	Worst experience on the name jio company....i ...	1	Worst experience on the name jio company i ord...
3	Pintu Das	Very bad delivery. Customer care no response. ...	1	Very bad delivery Customer care no response Go...
4	Chirag Puvar	Nice	5	Nice

Jio Mart Cleaned Data

- **DATA ANALYSIS :**

In this project, we used Python to draw conclusions. We tried to analyze the sentiment by one (A) rating-based method and other three Lexicon Based methods, i.e., (B) TextBlob, (C) SentiWordNet, (D) Vader.
Those are described below:

A. **By User Rating**: At first, we analyze sentiment by using users' Ratings. Ratings allow customers to share their experience with a product or service and give it an overall star rating. Here we are analysis that how many ratings gave to the product. With the proliferation of the internet, digital communication, and hundreds of review sites, ratings have become a critical aspect of today's world. Here, using the user rating we tried to figure out which app is better by calculating how many good ratings are there. We divided the rating into three groups. If the rating is greater than or equal to 4, we take that as a positive review (+1); if less than or equal to 2 the negative (-1) and others as neutral (0).

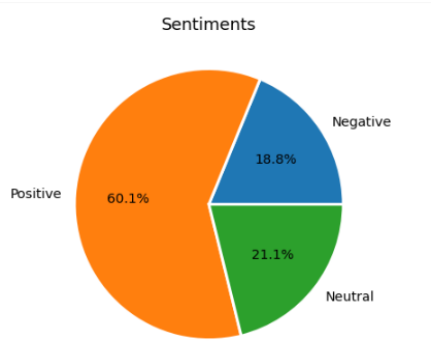
	Cleaned Reviews	cleaned Rating
0	Nice Stuff	4
1	Wow fast delivery and good items	5
2	Such a helpful app	4
3	on time delivery	5
4	Good service	5

Before transformation

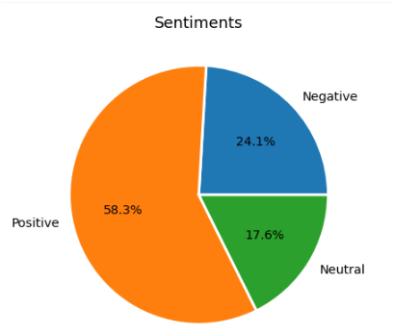
	Cleaned Reviews	rating_transform
0	Nice Stuff	0
1	Wow fast delivery and good items	1
2	Such a helpful app	0
3	on time delivery	1
4	Good service	1

After transformation

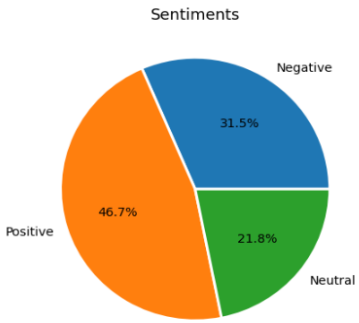
By categorizing these we described the result in the form of a Pie Chart.



BigBasket



Blinkit



Jio Mart

From this Pie chart, we can clearly see that 60.1% of people gave a higher rating to BigBasket, 58.3% of people to Blinkit and 46.7% of people to Jio mart. So, by user rating, we can clearly conclude that Millions of people prefer BigBasket.

B. **Using Textblob:** At first, we used the Parts of Speech tagger to find the parts of speech of each of the words that are used in lemmatization to find the root word.

	Cleaned Reviews	rating_transform	POS tagged	Lemma
0	Nice Stuff	0	[(Nice, n), (Stuff, n)]	Nice Stuff
1	Wow fast delivery and good items	1	[(Wow, n), (fast, a), (delivery, n), (good, a)...	Wow fast delivery good item
2	Such a helpful app	0	[(helpful, a), (app, n)]	helpful app
3	on time delivery	1	[(time, n), (delivery, n)]	time delivery
4	Good service	1	[(Good, a), (service, n)]	Good service

Blinkit after POS tag and Lemmatization

After finding the lemma, we calculated the subjectivity and polarity score (positive, negative or neutral) that describes the pie chart.

	Cleaned Reviews	Lemma	Subjectivity	Polarity	Analysis
0	Nice but price higher than jio mart	Nice price high jio mart	0.770000	0.380000	Neutral
1	good fresh products Well service On time deliv...	good fresh product Well service time deliver...	0.575000	0.600000	Positive
2	Fast and easy using and ordering Will wait and...	Fast easy use order wait see delivery qualit...	0.716667	0.316667	Neutral
3	Fake one If u cancelled an order believe me th...	Fake one u cancel order believe refund yr am...	1.000000	-0.500000	Negative
4	Didn t enjoy shipping nothing I m really sugge...	enjoy ship nothing really suggest people dow...	0.200000	0.100000	Neutral

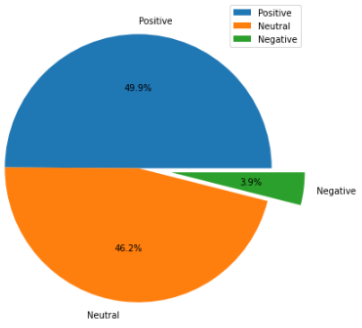
BigBasket

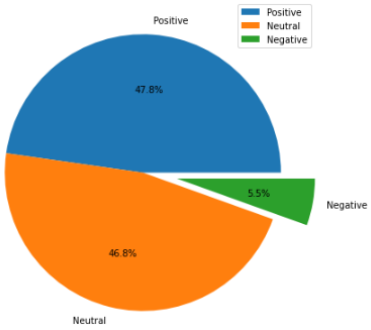
	Cleaned Reviews	Lemma	Subjectivity	Polarity	Analysis
0	Nice Stuff	Nice Stuff	1.000000	0.600000	Positive
1	Wow fast delivery and good items	Wow fast delivery good item	0.733333	0.333333	Neutral
2	Such a helpful app	helpful app	0.000000	0.000000	Neutral
3	on time delivery	time delivery	0.000000	0.000000	Neutral
4	Good service	Good service	0.600000	0.700000	Positive

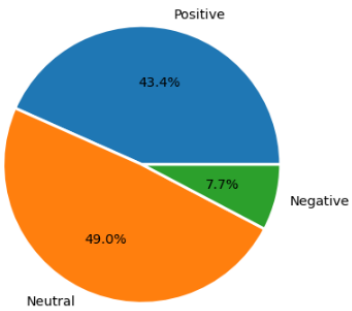
Blinkit

	Cleaned Reviews	Lemma	Subjectivity	Polarity	Analysis
1	good app	good app	0.600000	0.7	Positive
2	Worst experience on the name jio company i ord...	Worst experience name jio company order cook...	1.000000	-0.5	Negative
3	Very bad delivery Customer care no response Go...	bad delivery Customer care response Goods de...	0.666667	-0.7	Negative
4	Nice	Nice	1.000000	0.6	Positive
5	good	good	0.600000	0.7	Positive

Jio Mart







BigBasket Blinkit Jio Mart

From this Pie chart, we can clearly see that 49.9% of people gave a higher rating to BigBasket, 47.8% of people to Blinkit and 43.4% of people to Jio mart. So, by TextBlob Method, we can clearly conclude that Millions of people prefer BigBasket.

C. SentiWordNet: In this part, we have an inbuilt dictionary that contains a collection of synonymous words.

	Cleaned Reviews	Lemma	Subjectivity	Polarity	Analysis	Vader Sentiment	Vader Analysis	SWN analysis
0	Nice but price higher than jio mart	Nice price high jio mart	0.770000	0.380000	Neutral	0.4215	Neutral	Neutral
1	good fresh products Well service On time deliv...	good fresh product Well service time deliver...	0.575000	0.600000	Positive	0.9217	Positive	Positive
2	Fast and easy using and ordering Will wait and...	Fast easy use order wait see delivery qualit...	0.716667	0.316667	Neutral	0.4404	Neutral	Positive
3	Fake one If u cancelled an order believe me th...	Fake one u cancel order believe refund yr am...	1.000000	-0.500000	Negative	-0.8020	Negative	Negative
4	Didn t enjoy shipping nothing I m really sugge...	enjoy ship nothing really suggest people dow...	0.200000	0.100000	Neutral	0.7003	Positive	Positive

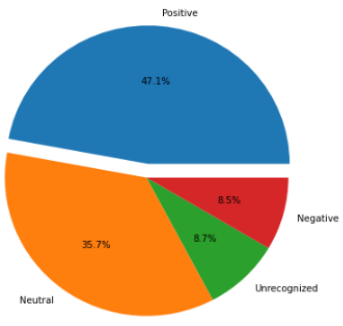
BigBasket

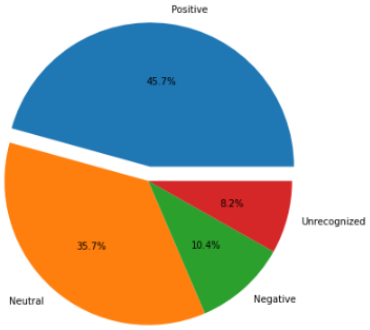
	Cleaned Reviews	Lemma	Subjectivity	Polarity	Analysis	Vader Sentiment	Vader Analysis	SWN analysis
0	Nice Stuff	Nice Stuff	1.000000	0.600000	Positive	0.4215	Neutral	Neutral
1	Wow fast delivery and good items	Wow fast delivery good item	0.733333	0.333333	Neutral	0.7717	Positive	Positive
2	Such a helpful app	helpful app	0.000000	0.000000	Neutral	0.4215	Neutral	Neutral
3	on time delivery	time delivery	0.000000	0.000000	Neutral	0.0000	Neutral	Neutral
4	Good service	Good service	0.600000	0.700000	Positive	0.4404	Neutral	Positive

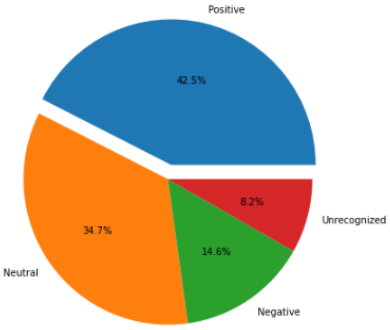
Blinkit

	Cleaned Reviews	Lemma	Subjectivity	Polarity	Analysis	Vader Sentiment	Vader Analysis	SWN analysis
0	Nice Stuff	Nice Stuff	1.000000	0.600000	Positive	0.4215	Neutral	Neutral
1	Wow fast delivery and good items	Wow fast delivery good item	0.733333	0.333333	Neutral	0.7717	Positive	Positive
2	Such a helpful app	helpful app	0.000000	0.000000	Neutral	0.4215	Neutral	Neutral
3	on time delivery	time delivery	0.000000	0.000000	Neutral	0.0000	Neutral	Neutral
4	Good service	Good service	0.600000	0.700000	Positive	0.4404	Neutral	Positive

Jio Mart







BigBasket Blinkit Jio Mart

From this Pie chart, we can clearly see that 47.1% of people gave a higher rating to BigBasket, 45.7% of people to Blinkit and 42.5% of people to Jio mart. So, by SentiWordNet Method, we can clearly conclude that Millions of people prefer BigBasket.

D. **Vader:** By calculating the compound score, it gives us the sentiment of each review as positive, negative or neutral.

	Cleaned Reviews	Lemma	Subjectivity	Polarity	Analysis	Vader Sentiment	Vader Analysis
0	Nice but price higher than jio mart	Nice price high jio mart	0.770000	0.380000	Neutral	0.4215	Neutral
1	good fresh products Well service On time deliv...	good fresh product Well service time deliver...	0.575000	0.600000	Positive	0.9217	Positive
2	Fast and easy using and ordering Will wait and...	Fast easy use order wait see delivery qualit...	0.716667	0.316667	Neutral	0.4404	Neutral
3	Fake one If u cancelled an order believe me th...	Fake one u cancel order believe refund yr am...	1.000000	-0.500000	Negative	-0.8020	Negative
4	Didn t enjoy shipping nothing I m really sugge...	enjoy ship nothing really suggest people dow...	0.200000	0.100000	Neutral	0.7003	Positive

BigBasket

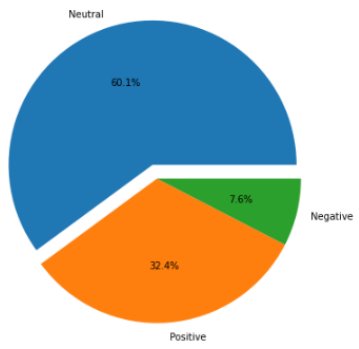
	Cleaned Reviews	Lemma	Subjectivity	Polarity	Analysis	Vader Sentiment	Vader Analysis
0	Nice Stuff	Nice Stuff	1.000000	0.600000	Positive	0.4215	Neutral
1	Wow fast delivery and good items	Wow fast delivery good item	0.733333	0.333333	Neutral	0.7717	Positive
2	Such a helpful app	helpful app	0.000000	0.000000	Neutral	0.4215	Neutral
3	on time delivery	time delivery	0.000000	0.000000	Neutral	0.0000	Neutral
4	Good service	Good service	0.600000	0.700000	Positive	0.4404	Neutral

Blinkit

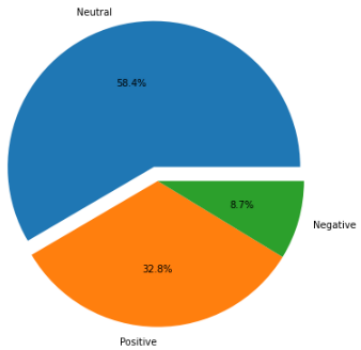
	Cleaned Reviews	Lemma	Subjectivity	Polarity	Analysis	Vader Sentiment	Vader Analysis
1	good app	good app	0.600000	0.7	Positive	0.4404	Neutral
2	Worst experience on the name jio company i ord...	Worst experience name jio company order cook...	1.000000	-0.5	Negative	-0.6808	Negative
3	Very bad delivery Customer care no response Go...	bad delivery Customer care response Goods de...	0.666667	-0.7	Negative	-0.0772	Neutral
4	Nice	Nice	1.000000	0.6	Positive	0.4215	Neutral
5	good	good	0.600000	0.7	Positive	0.4404	Neutral

Jio Mart

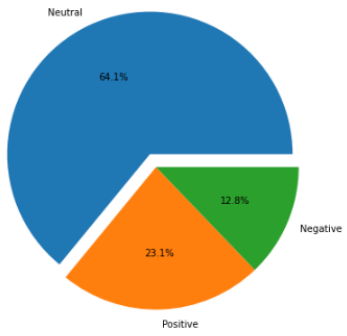
Each of the results describes by a pie chart.



BigBasket



Blinkit



Jio Mart

From this Pie chart, we can clearly see that 32.4% of people gave a higher rating to BigBasket, 32.8% of people to Blinkit and 23.1% of people to Jio mart. So, by Vader Method, we can clearly conclude that Millions of people prefer Blinkit.

Table:

Application Name	Sentiment	Rating Based	TextBlob	SentiWordNet	Vader
Big Basket	Positive	60.10%	49.90%	47.10%	32.40%
	Neutral	21.10%	46.20%	35.70%	60.10%
	Negative	18.8	3.90%	8.50%	7.50%
	Unrecognised			8.70%	
Blinkit	Positive	58.30%	47.60%	45.70%	32.80%
	Neutral	24.10%	5.50%	35.70%	8.70%
	Negative	17.60%	46.80%	10.40%	58.40%
	Unrecognised			8.20%	
Jio Mart	Positive	46.70%	43.40%	42.50%	23.10%
	Neutral	31.50%	7.70%	14.60%	12.80%
	Negative	21.80%	49.00%	34.70%	64.10%
	Unrecognised			8.20%	

Model Comparison:

Here, we want to identify which model is better for analyzing sentiments and after detecting the model we can find which app people prefer for online grocery shopping in India Nowadays.

For identifying the best model, we found the accuracy of three Lexicon-based models, i.e., TextBlob, SentiWordNet, and Vader by comparing them with the ratings of the people.

To find the Accuracy, we created the confusion matrix for three models where the Actual value is the percentage of users who gave positive and negative sentiments. As we used a 2×2 confusion matrix, we calculated predicted positive and actual positive values by adding positive and neutral.

Actual Class= Percentage of Positive and Negative ratings given by User.

Predicted Class= Lexicon Methods (TextBlob, SentiWordNet, and Vader)

Now, we have chosen Big Basket data to measure the accuracy of the models, as from four methods we have seen people prefer Big Basket over the other two apps.

Accuracy Measure for Text Blob:

	Positive	Negative
Rating-Based Percentage Score	81.2	18.8
Percentage Score of Text Blob	96.1	3.9

Confusion Matrix:

	Actual Positive	Actual Negative
Predicted Positive	66.3	14.9
Predicted Negative	14.9	3.9

$$\text{Accuracy} = \frac{66.3+3.9}{100} \times 100 = 70.2\%$$

Similarly, we calculated accuracy for SentiWordNet and Vader methods by comparing the Rating-Based Percentage Score; got 79.4% and 77.4% respectively.

As SentiWordNet has higher accuracy we prefer this method to conclude our result on 'which app is better?'

Now, in the SentiWordNet method, 47.1% of people gave a higher rating to BigBasket, 45.7% of people to Blinkit and 42.5% of people to Jio mart. So, we can clearly conclude that BigBasket is better than Blinkit and Jio Mart.

CONCLUSION

Based on these problems faced by the customers, grocery apps like BigBasket, Blinkit and Jio Mart have come into existence. It is one of the easiest ways of buying all your home needs. So, most people prefer the online grocery delivery process.

Bigbasket has a great website and user interface when compared to Jio Mart. It offers a user-friendly interface that makes shopping easier. In Jiomart only one or fewer banks will offer cashback or offers when we shop using their cards. In Bigbasket you can find various banks that provide offers on various products and it has more product categories like exotic fruits and vegetables, Fresh Range and many more. Blinkit has fewer food product categories when compared to Big Basket.

Like these with many pros and cons, we detected which online grocery app is better. By the Sentiwordnet method, BigBasket is a better app based on accuracy which, in theory, proves that Vader is a better method as Vader reads the whole sentence and tries to identify the sentiment of the text.

So, in conclusion, SentiWordNet performs well than Vader and TextBlob sentiment analysis tools and BigBasket is a better app than Blinkit and Jio Mart in many ways.

As we found out which app is better, we also tried to find out why the other two apps are not good as compared to BigBasket. If we look at the average reply time of BigBasket to their customer, we can clearly see it is 131 minutes whereas; for Blinkit it is 485 minutes and Jio Mart only replies to those who give bad ratings (like 1 or 2). From this, we can inspire future online grocery app makers, they can fix this type of glitch into good customer service. Also, by predicting which are the negative reviews, they can easily go through the comments and make their app better.

REFERENCE

[1] Vikas Malik and Amit Kumar. “Sentiment Analysis of Twitter Data Using Naive Bayes Algorithm”, International Journal on Recent and Innovation Trends in Computing and Communication, Vol. 6, No. 4, 2018.

[2] Z. Nanli, Z. Ping, L. Weiguo, and C. Meng, “Sentiment analysis: A literature review”, Proceedings of the International Symposium on Management of Technology (ISMOT), Hangzhou, IEEE, pp. 572-576, 2012.

[3] J.W. Pennebaker, R.L. Boyd, K. Jordan, and K. Blackburn, “The development and psychometric properties of LIWC2015”, Austin, TX: University of Texas at Austin, 2015

[4] Wordnet.com, “WordNet, a Lexical database for English”, [online] Available: http://wordnet.princeton.edu/

[5] S.Feng, L. Wang, W. Xu, D. Wang, G. Yu,"Unsupervised learning Chinese sentiment lexicon from massive microblog data",Advanced Data Mining and Applications, Springer-Verlag Berlin Heidelberg 2012.

[6] Wilson, J. Wiebe, and P. Hoffman. 2005. Recognizing contextual polarity in phrase level sentiment analysis.ACL.

[7] Ratnaparkhi Adwait. A maximum entropy model for part-of speech tagging. In: Proceedings of the conference on empirical methods in natural language processing, April 1996

[8] Aggarwal Charu C, Zhai Cheng Xiang. Mining Text Data. Springer New York Dordrecht Heidelberg London: © Springer Science+Business Media, LLC’12; 2012.

[9] Chenlo J, Hogenboom A, Losada D. Sentiment-based ranking of blog posts using rhetorical structure theory. In: Presented at the 18th international conference on applications of Natural Language to Information Systems (NLDB’13); 2013.

[10] Asian Journal of Computer Science and Technology ISSN: 2249-0701 Vol.8 No.S2, 2019, pp. 1-6 © The Research Publication, www.trp.org.in

[11] Sentiment Analysis and Opinion Mining April 22, 2012 Bing Lientiment Analysis