Opinion Mining and Sentiment Analysis for Amazon Product Reviews using Lexicon and Rule-Based Approach and Testing on Machine Learning Algorithms

Sai Kiran Chintalapudi[#1], Harshavardhan Metla[#2], Keerthi Shrikar[#3] , Nalluri Rahul[#4]

[#]*School of Computer Science and Engineering,*
Vellore Institute of Technology, Vellore, India

Abstract— *As people are free to say their opinions on anything using various social networking sites like Twitter, Facebook, Discussion forums, and blogs. Particularly Microblogging and text messaging have emerged and become dominated tool over the web. Amazon Reviews are often used to share opinions and sentiments about the products they purchased the products. The availability of social content generated on sites such as Amazon creates new opportunities to study public opinion about the entity. This analysis we took amazon data for sentiment classification. The Sentiment analysis is done on a per-review basis. The words in each review are compared with those in other reviews that have been previously labeled as "positive", "negative" or "neutral". After looking at these words, the algorithm then judges whether the text in the Tweet is positive or negative or neutral based on the likelihood for each possibility. The overall objective of this paper is to determine the sentiment of the text, whether it is positive, negative or neutral, which is extended to strength of polarity also this approach is used to obtain the significant features and to analyzing the overall sentiment for each object by using training data and testing data using Machine Learning Algorithms: Naïve Bayes Algorithm and Support Vector Machines(SVM) Algorithm.*

Keywords— *Web sentiment analysis, Opinion mining, Vader tool, Support Vector Machines, Naïve Bayes Algorithm, Recommendation system.*

I.   Introduction

In the decision-making process each and every piece of information are very important. After arriving internet world user doesn't bother about other opinions from individuals newspaper, surveys, opinion pools, consultants because web analytics introduce new system called opinion mining, which is find out the opinions and experience of other people over the internet using digital social media network like Facebook , reviews, forums, blogs, Twitter, micro-blogs, etc.,

Indeed, according to surveys about 6 in 10 (60%) online shoppers say user generated customer product reviews have a significant or good impact on their buying behavior. Also Data from the 2011 Social Shopping Study indicates that 50% of consumers spend 75% or more of their total shopping time conducting online product research, with 15% spending 90% or more of their shopping time in this manner.

Another surveys by Deloitte Consumer Products Group found that almost two-thirds (62%) of consumers read consumer written product reviews online. In fact, a recent study by Deloitte found that "82% of purchase decisions have been directly influenced by reviews". The objective of this paper is to throw lime light on determine the sentiment of the text, whether it is positive or negative, which is extended to strength of polarity. With the explosion of Web 2.0 platforms such as blogs, discussion forums, peer-to-peer networks, and various other types of social media. Consumers have at their disposal a soapbox of unprecedented reach and power by which to share their brand experiences and opinions, positive or negative, regarding any product or service.

As major companies are increasingly coming to realize, these consumer voices can wield enormous influence in shaping the opinions of other consumer and, ultimately, their brand loyalties, their purchase decisions, and their own brand advocacy. Companies can respond to the consumer insights they generate through social media monitoring and analysis by modifying their marketing message, brand positing, product development, and other activities accordingly.

II.   Literature survey

A. Literature Survey

Sentiment analysis, also called Opinion mining, is the field of study that analyzes people's opinions, sentiments, evaluations, appraisals, and emotions towards entities such as products, services, organizations, individuals, issues, events, topics and their attributes. In general opinion cannot structure a problem but it can subjective and in case opinion gathered from many

The basic components of an opinion are:

Opinion holder: it is the person that gives a specific opinion on an object.

Object: it is entity on which an opinion is expressed by user.

Opinion: it is a view, sentiment, or appraisal of an object done by user.

There are two other types of opinions: *Regular and Comparative.*
Regular opinion is expressions on some target entities, which can be classified into direct and indirect opinion. Comparative opinion is Comparisons of more than one entity.
An opinion is a quintuple:
 (ej, ajk, soijkl, hi, tl), Where ,

ej, is a target entity / Named Entity Extraction ajk is an aspect/feature of the entity ej / Information Extraction soijkl is the sentiment value of the opinion from the opinion holder hi on feature ajk of entity ej at time tl. soijkl is +ve, -ve, or neu, or more granular ratings. / soijkl is Sentiment identification. hi is an opinion holder. / information / Data extraction tl is the time when the opinion is expressed.

Sentiment analysis – otherwise known as opinion mining – is a much bandied about but often misunderstood term.

Sentiment Analysis is the process of determining whether a piece of writing is positive, negative or neutral. It's also known as opinion mining, deriving the opinion or attitude of a speaker. A common use case for this technology is to discover how people feel about a particular topic.

In essence, it is the process of determining the emotional tone behind a series of words, used to gain an understanding of the the attitudes, opinions and emotions expressed within an online mention. [4]

The human language is complex. Teaching a machine to analyse the various grammatical nuances, cultural variations, slang and misspellings that occur in online mentions is a difficult process. Teaching a machine to understand how context can affect tone is even more difficult.

The best businesses understand sentiment of their customers – what people are saying, how they're saying it, and what they mean. Sentiment Analysis is the domain of understanding these emotions with software, and it's a must-understand for developers and business leaders in a modern workplace.

Analysis is used in business intelligence to understand the subjective reasons why consumers are or are not responding to something (e.x. why are consumers buying a product? What do they think of the user experience? Did customer service support meet their expectations?). Sentiment analysis can also be used in the areas of political science, sociology, and psychology to analyze trends, ideological bias, opinions, gauge reactions, etc. [5]

An important part of our information-gathering behavior has always been to find out what other people think. With the growing availability and popularity of opinion-rich resources such as online review sites and personal blogs, new opportunities and challenges arise as people can, and do, actively use information technologies to seek out and understand the opinions of others. The sudden eruption of activity in the area of opinion mining and sentiment analysis, which deals with the computational treatment of opinion, sentiment, and subjectivity in text, has thus occurred at least in part as a direct response to the surge of interest in new systems that deal directly with opinions as a first-class object. [6]

III.   Methodology

A. Motivation and Objective

A number of approaches are used to identify the significant features of opinion mining and to determine the sentiment of the text, whether it is positive, negative or neutral, which is extended to strength of polarity. The aim of this approach is used to obtain the significant features and to analyzing the overall sentiment for each object by Naïve Bayes Classification and Support Vector Machines(SVM) which are machine learning algorithms.

B. Data Set

The data set is obtained by creating a function to crawl the reviews into a text file. Then after noise removal, we obtain the training set by using Vader sentiment analysis tool. We have obtained reviews from about 150 Amazon Review pages. There are 1150 instances of reviews.

C. Steps for Sentiment analysis

In opinion mining are various types of sentiment analysis as: word level , feature-level, entity level, sentence-level, document-level.

Data set are collected from Amazon by web crawling, in this step will be explained very clearly here.

The PAPER has 4 main stages -

- Collect the data of reviews for Amazon products.

- Pre-process and data cleaning.

- Calculate the sentiments of the product reviews

- Apply Machine Learning Algorithms to get the metrics

We take all the words and phrases that imply positive or negative sentiment and apply rules that consider how context might affect the tone of the content. Carefully crafted rules help our software know the first sentence below is positive and the second is negative.

Sentiment analysis tools can be configured to determine sentiment on a range of levels. We'll score sentiment on a document level (does this express a general positive or negative tone), but we'll also score the sentiment of individual words or phrases in the document.

Less precise sentiment analysis tools lump together the sentiment expressed at individual entities into a general document sentiment score. [3]

Fig. 1.   Appendix

	label	post
0	neutral	got a defected piece
1	pos	the phone is as good as it looks the feel the ...
2	pos	still super happy
3	neutral	as good as should be prompt service
4	neutral	thats a classic piece
5	pos	one of the best phones from camera to the desi...
6	neutral	good
7	neutral	waao product
8	pos	best phone ever
9	pos	initially i was a bit hesitant in buying onlin...
10	pos	my first apple device and its simple awesome
11	pos	got it in best deal
12	pos	loving the product
13	pos	one of the best deal and the phone is fabulous
14	pos	fantastic product and excellent delivery perfo...
15	neutral	cant expect anything better except the ear phones
16	pos	migrated to ios really loved it
17	neutral	nice product and timely delivery
18	pos	wonderful and a superb example of uniqueness
19	pos	awesome m loving it

IV.   DATA HANDLING

A. Data Collection

In September 2016, iPhone 7 was released. We were interested in finding out people reactions to the new iPhone. We wanted to write a simple Python script to scrape and parse online reviews and run a sentiment analysis on the collected reviews. There are many applications for automated opinion mining where companies are interested in finding out their customers reactions to new products release.

For scraping reviews we used Python urllib module. For parsing pages contents and grabbing the required HTML elements we have used a Python library called BeautifulSoup.

B. Data Cleaning

This stage comprises of removing the punctautions, emojis and transformation of upper case to lower case.

Fig. 2.   RawData

Fig. 4.   Output

class	fscore	precision	recall
neg	0.514286	0.428571	0.642857
neutral	0.637931	0.532374	0.795699
pos	0.569948	0.785714	0.447154

Fig. 5.   Classification Report Naives Baiyes

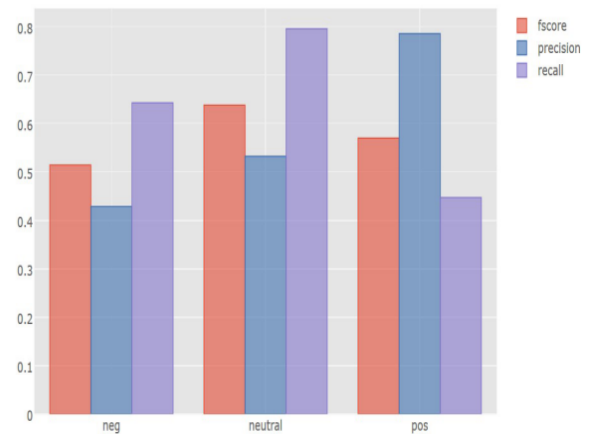## V.   RESULTS AND DISCUSSIONS

Fig. 3.   Opinions
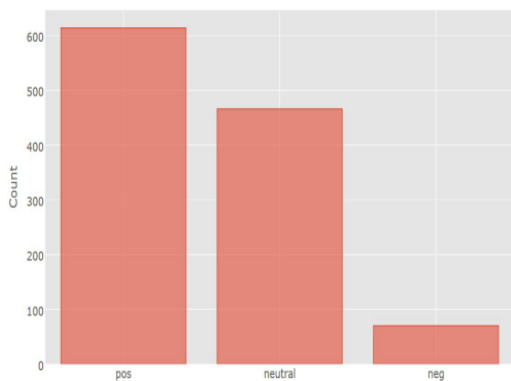
Fig. 6. Output

```
Training time: 0.648 s
Predicting time: 0.172 s
Average Accuracy: 83.0434782609 %
Classification Report:
```

	fscore	precision	recall
class			
neg	0.352941	1.000000	0.214286
neutral	0.813725	0.747748	0.892473
pos	0.878661	0.905172	0.853659

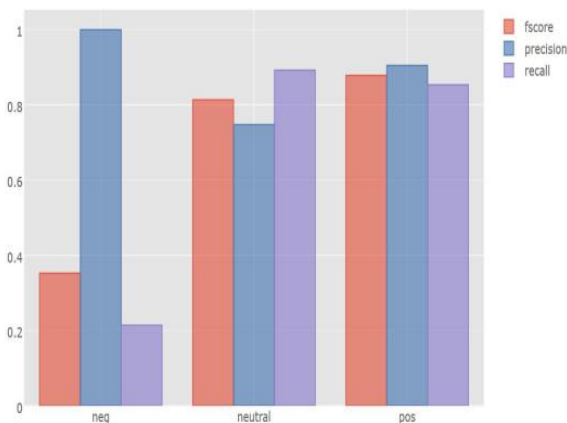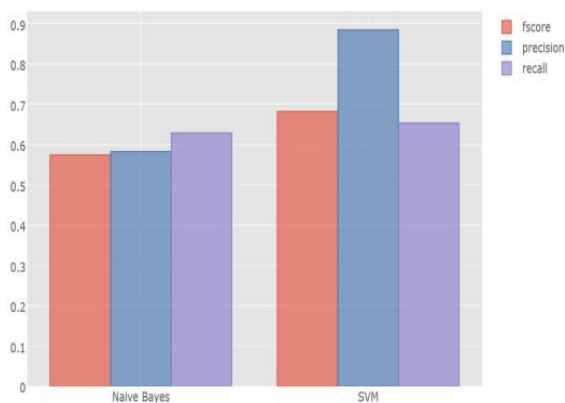Fig. 7. Classification Report Support Vector Machines

Fig. 8. Result

VI. CONCLUSION

This research introduces the theoretical basic of opinion mining. The proposed approach determines the sentiment of the text, whether it is positive, negative or neutral, which is extended to strength of polarity and also which was obtain the significant features and to Analyzing the overall sentiment for each object by computing the weighted average for all the sentiments in the textual data. Hence it enables the company officials to determine the opinion about the product in the market and its demand in the market.

Sentiment analysis provides solutions directly to businesses, as well as offering APIs for integration into our client's own products.

VII. FUTURE SCOPE

While it's difficult to speculate how a relatively immature system might evolve in the the future, there is a general assumption that sentiment analysis needs to move beyond a one-dimensional positive to negative scale.

Most of the current thinking in sentiment analysis happens in a categorical framework: sentiment is analyzed as belonging to a certain bucket, to a certain degree.

To address the context issue, a lot of research surrounding sentiment analysis has focused on feature engineering. Creating inputs to a model that recognize context, tone, and previous indications of sentiment can help increase accuracy and get a better overall sense. [5]

For the future, to truly understand and capture the broad range of emotions that humans express as written word, we need a more sophisticated multidimensional scale.

We will see a shift in perception of the reliability of sentiment analysis. Users will become more comfortable with the idea that the automatic analysis of individual text material is hard to match human performance.

The insight that can be gained from large datasets (millions of Reviews) will overshadow the concerns about reliability at a granular level (a single Review).

Instead, the focus will be on how to make results interpretable and actionable. In the meantime, we'll be ensuring we are working at making sentiment analysis as accurate and easy to understand as possible. [4]

REFERENCES

Journal Articles:

[1] Sentiment Analysis : A Literature Review, 8-9 Nov. 2012, Management of Technology (ISMOT), 2012 International Symposium

[2]   A Literature review on Sentiment Analysis - Vol- 3 issue- 5, 2017 IJARIIT

Online Sources:

[3]   What is Sentiment Analysis? Lexalytics

[4]   Understanding Sentimental Analysis: What it is and why is it used - Kristian Bannister, Brandwatch.

[5]   Introduction to Sentiment Analysis - Algorithmia Blog.

[6]   Opinion Mining and Sentiment Analysis - cornell.edu.