Product Analyst Advisor

Parth Kanani[1], Pradeep Ramola[2]

[1]Parth Kanani, Undergraduate student, Information Technology Department, K. J. Somaiya Institute of Technology, Mumbai, Maharashtra, India

[2]Pradeep Ramola, Undergraduate student, Information Technology Department, K. J. Somaiya Institute of Technology, Mumbai, Maharashtra, India

---***--

Abstract - *We live in a digital world where data is the new oil. One can do wonders if he processes the data well and extracts the useful facts out of it which could be used as useful information. We human beings are sensitive and have our own views, opinion, judgment, and emotions. We do express our emotions, views, opinions, and judgments through text messages. Wouldn't it be wonderful if we could extract the emotions out of mere texts? And then use if for our purpose. Yes, it's very possible thanks to Sentimental Analysis and NLP we can extract emotions out of texts. Our project is based on extracting emotions and opinions out of the text field and uses it forgiving us some insight into the product. Suppose if one must buy a mobile phone, he has to do a lot of research from price range to features and its durability and then select the best mobile in the range which fits maximum criteria. Yes, it could be possible with the help of our project product analyst (mobile) advisor. What we will be trying to do is extract reviews from different sites about the product (mobile) which are looking for and then process that data and transform it so it could be used for our analysis which in short will be helpful to display the result of all reviews and make it easy for the user to select the best choice available*

Keywords: Sentiment Analysis, Product Analyst Advisor, Natural Language Processing (NLP), Opinion Mining

1.INTRODUCTION

Product Analyst Advisor is a software framework that uses state-of-the-art natural language processing techniques like Sentiment Analysis, for getting the overall view of the customers about the product. Sentimental analysis is a process that uses natural language processing, statistics, and text analysis to analyze the sentiment of netizens. The best in the business, business giants understand the sentiment of their customers well, what people are saying about their product or brand, how they're saying about it, and what they really mean. Sentimental Analysis is one such domain which can help us to understand the emotions of humans with ease using software techniques, and it's a must-understand for developers and business leaders in a modern workplace in this digital era. Sentiment analysis is used to extract useful information from different platforms with the help of natural language

processing (NLP), which could be useful for developers developing their product or businessmen to invest in the field where there is maximum interaction of people in such a way that they gain maximum profit. Sentimental analysis is used to find out what kind of attitude a speaker has or a writer's perspective depending on the topic or the overall emotional polarity of a document. The point of view of a reviewer can be his or her opinion, or the intended emotional communication which could help companies to target what they should produce or manufacture. In this world, data is critical and plays a vital role in making any decision plus if we must choose between two products, we have to compare them along with features to price range and how they differ from each other and what would be the best product based on the data of comparison wouldn't it make life easy? If the comparison is accurate based on the provided data. In this digital world we have lots of data but how we use it and process it for our benefits still depends on us with the help of sentimental analysis the expressions, judgment, and opinion can be extracted from the text and make it as real as possible as it is fact coming from humans. Product Analyst Advisor can have a wide range of applications and use from normal users to big industrial companies can use it for their benefit and make maximum profit out of it .Such an approach will be more user-centric of what users want and the related products will be developed in the future because as the demand of those products will increase manufacturing of those items will increase which will result in more supply of that product by increasing the competition among companies to provide the best feature what customers are looking for in what price range there would be maximum profit for them**.**

2. ACTUAL SURVEY:

A. Problem Discussed

Emotional analysis on social media is more difficult than different text types due to limitations such as abbreviations, short text for words, and references to existing content or concepts. Social media provides visual data without text, for example connected media, user responses, and interpersonal relationships. This can often be referred to as public content [6]. Recent works have successfully benefited from the integration of text and social context of emotional analysis activities. However,

these activities are generally restricted to certain aspects of the public context, and there have never been any attempts to analyze and use the public context systematically.

B. Traditional Methods and their disadvantages

The ancient methods use the Bag Of Words model, in which the document is placed on a feature map, separated by machine learning techniques. Although the Bag Of Words approach is straightforward and expensive, a lot of respectable information from the original language is lost. Therefore, various styles of selection are exploited, such as high-n-grams. Another feature of the peace that will be used is the part of marking the Speech, which is often used throughout the analysis of grammar. This feature is known as the above forms. Lexical resources are a very useful field in the field of emotional analysis, as they provide a tool that provides direct emotional information [12]. Typically, a dictionary tends to recognize the similarity of words and their occasional dictionary correlations of text. Resources like these have problems with vocabulary usage and domain change. This workplace unit is restricted only to certain aspects of the public domain, and there have never been any attempts to analyze and use the public domain systematically.

C. Problem in existing systems:

The existing system [14] consists of approaches which mainly includes traditional machine learning based algorithms like SVM, Naive Bayes etc. for performing sentiment analysis on the given text. However, the accuracy of such models is not more than 75%.

Another problem of the existing system is that they mostly work on static data, that means all the analysis and process would be applied to the data that is previously extracted,

but for a project like analyzing and comparing two products needs real time data in order to examine the current trend of the market and analyze it in real sense.

D. Modern Approaches

Apart from this, many modern techniques use the concept of language distribution, usually through the use of word embedding. The metaphor of language similarity is calculated between text and vocabulary words translated in [1]. To undermine this metric, this paper suggests a model of emotional separation using live language similarities on the side of embedding presentations [13]. Since there is a certain problem about efficiency between idea, feeling and feeling, they print the idea as a concept to change to reflect the situation in that business. The word unit that feels is used in many ideas for the preparation of ideas. Positive repetition words are used to specify other

appropriate regions, while negative verbal sense is used to specify other unwanted regions. There is also a unit of similar test expressions and speech statistics called unit lexicon [3]. Public context, and there have never been any attempts to analyze and use social context systematically.

E. Deep learning in Sentiment Analysis

In-depth Emotional Analysis Techniques turned out to be unusually fashionable [2]. It offers automatic feature rendering, high display power and intelligent presentation compared to traditional feature-based techniques. These long-term approaches can produce solid foundations, and their potential strength is a unit often used in conjunction with in-depth learning methods [2]. Throughout this paper, the authors explored the spice in the application of in-depth reading strategies and even ancient facial expressions based on hand-drawn choices. At the end of the day, web-based mining life. Twitter, Facebook bowed to the hot seat. Twitter is one of the leading microblogging management systems that allows consumers to put their ideas on web-based business sites, sports, legal issues, gift day developments, animated images, etc. watchman's view. It is difficult to accept a feeling related to an object or event by collecting and writing physical microblog assignments [4]. And, these days, today's standard shortcut unit is used specifically for informative content as an easy way to create a longer message. These methods eliminate these cut pattern names. To overcome the obstacle, a marking model is developed [4], which includes the use of abbreviations and emoji. With the highest objective of this activity, we have compiled a summary of abbreviated words and emoji.

F. Machine learning in Sentiment Analysis

Each reading machine and Lexon-based approach is presented by Rathor [5]. This paper focuses on taking the go-ahead for the implementation of 3 AI (Naive mathematician (NB), Support Vector Machines (SVM), and Entropy (ME) formats for online review using misused learning methods. the basic premise of this sophisticated customer research test is that the placement of customer evaluation as negative, positive or bipartisan can be a repetitive and exaggerated task to be performed. The results showed that machine learning statistics work respectfully for weighted unigrams and SVM showed outstanding accuracy. this does not help consumers who have to disclose pre-purchase items but by participating in organizations that have to look for open reactions to their products. Over time, the principles of online life, for example, Twitter, Facebook, and web page reviews convey an incredible amount of printed data. The written information complements the basic source of understanding customer perceptions regarding the integration of ideas, objects, or events. Trends provided by customers are within the framework of a good, bad, or

honest limit. The information written on online health goals sets the stage for dynamic work on the phone and the plans of individual phone manufacturers. The process of making a machine the most common proof of experimentation during an object is a sensory analysis. The net is very powerful, and the test of every customer is on the web. There are completely different ongoing studies and proposed models for the analysis of such emotions.

This survey paper talks concerning this takes and shot at opinion mining and conclusion characterization of consumer criticism and audits on the online, and assesses the varied ways utilized for the procedure. This look on the zones caused by the assessed papers, brings up the territories that are pleasantly secured by completely different specialists and regions that are neglected in conclusion mining and slant order that are open for future analysis works.

3. ARCHITECTURE

The basic architecture involves the mechanism of two fundamental processes which is required in order to serve the purpose of effective sentimental analysis and there by doing the effective analysis on texts for efficient results in terms of good product insights. This product insights can later be use by companies or business for any improvement in their product or by consumer for better choice of product according to their requirement.

Initially consumers or any business organization can search for their respective product, then the request would be forwarded to Live Data Web Scraping module, which would then get the essentials information like reviews, comments etc. using some third-party APIs. From the extracted data reviews, comments would then be fed to Deep learning algorithm module which would do essential analysis on the comments and reviews and base on various text-analysis final results in terms of product insights would be showcase to the end user.

There can be two types of analysis can be done based on user query, if the user choose for comparing any two product, then analysis would be done by comparing various aspects of the product and its commonality features and then revert the result back in terms of insights like comparing cost, features like camera quality in case of a mobile product etc., and another type of search query that user can avail is just getting insights about a particular product. The mechanism mentioned above is demonstrated below in figure here: -
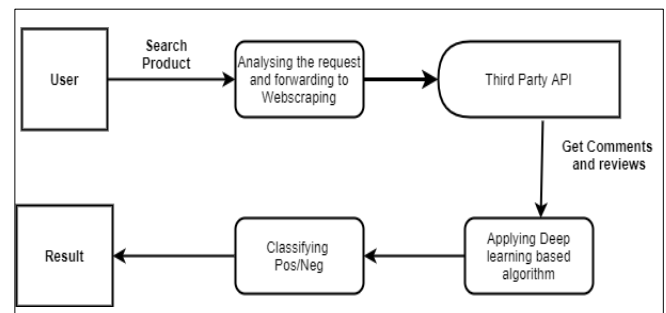

Fig 1. Process flow for sentiment analysis

4. METHODOLOGY

The selection of proper methodology plays an important role in the success of any project. Various techniques and methods chosen in the project are selected based on various research papers chosen as references.

The Product Analyst Advisor (PAA) is a web-based application which uses various python-based frameworks like NumPy, ploty, etc. for analysis and visualization of analysis done. It uses Django as a backend framework and ReactJS as an frontend framework. For analysis on data, Google colab and Jupiter notebooks are used.

For Sentimental analysis, Deep Learning transformer-based BERT (Bidirectional Encoding Representational Transformer) model is used. BERT model is Google Brain's state of the art algorithm which is used to work extensively now a days on natural language processing, in which it gives about 98% of accuracy on sentiment analysis, which helps to make the product more reliable in terms of getting good product insights.

Above all one thing which is mandatory without which project won't work is the required data, we need. Concerning this paper for a product like product analyst advisor, we need data that is related to the list of products, their reviews on various e-commerce websites. One of the popular ways to extract information from the web are:

(a) Web Scraping

(b) Web Crawler

(c) From third party APIs

A. Web Scraping:

Web Scraping is typically the extraction of data from any of the websites available on the internet. This information is extracted/collected from various data and store in form of an excel file, which we are going to use for giving predictive analysis on the product.

In general, there are two ways of implementing web scraping:

Manually or automatically, in manual web scraping is done when the user would provide the website URL to the model from which it could extract the required data, whereas in automatic web scraping it is much more related to web crawler, but it has some limitation i.e.

extracting required information from the website URL mention in the code. Advance web scraper would extract the whole website including the HTML, CSS part of it. And then segregate and get the required information like reviews, comments, etc.

B. Web Crawler:

Web Crawler is another way of extracting information from the web. However, the difference lies in the scalability of the information it extracts. It means the web crawler crawls all the websites available on the web and searches for the data which is related to what is been given as input data. It is getting us the indexes of related data available on the web. Such a scale is not required by applications like product analyst advisor which focuses on data based on a particular product.

C. From Third-Party APIs:

This is another approach of extracting data from the web, where all the hard work is taken care of by a third party. But with this comfort also comes the price. Most of the available API providers on the web are having charges for their APIs, but another reason for not using Third-party APIs is customizability. That product like the product analyst advisor needs to access varied types of data. Few Third-party APIs available on the web are PriceJSON, ScrapperBox, etc.

Given the requirements of a project like a product analyst advisor, web scraping can be the best available option for using it as a tool for extracting particular data from various e-commerce websites. And use for further analytical steps.

For the classification task, data is required divided into two parts: a test set and a training set. Training set is used to learn to divide, and the other set further used to evaluate the performance of classifiers. The data set contains the following details either attribute from the 'open phone' category of Amzon.com:

1. Product Title
2. Brand
3. Price
4. Rating
5. Text of the review
6. Number of persons who initiate the appraisal helpful.

5. SENTIMENT CLASSIFICATION TECHNIQUE:

Sentimental analysis is one of the major areas of NLP (Natural Language Processing) which involves the process to identify and categorize the opinions expressed in a piece of text, especially to determine the writer's attitude towards a particular topic, product, etc., in the form of positive, negative, or neutral.

Sometimes referred to as opinion mining, although the emphasis, in this case, is on extraction. Four reasons for using deep learning are:

1)   Deep learning algorithms can perform complex calculations, any traditional approach like the Random tree forest and Naive Bayes method which is used as a convenient way for performing the sentimental analysis can be performed using the deep learning algorithm but vice-versa is not possible.

2)   Deep learning algorithms are adaptive.

The deep learning algorithm we have used here are written in order of their computational accuracy:

A. Logistic Regression Based Sentiment Analysis: -

It is the model that is used to determine output or result when there is one or more than one independent variable. The output value can be in form of 0 or 1 i.e., in binary form. Accuracy with Logistic Regression we are getting is 74%.

B. LSTM (Long Short-Term Memory): -

LSTM is an advanced version of RNN (Recurrent Neural Network), this deep learning algorithm is widely used in the industry. It is based on the principle of remembering the features of the text once trained and accordingly updating its memory.

LSTM overcomes the problem with the traditional approach of deep learning which was, traditional algorithm use to perform sentiment analysis on a word by word basis which was not giving higher accuracy as word meaning can change depending on its neighboring words this care is taken by RNN algorithm of the neural network but it lacks in performance when the sentence is too big. LSTM solves this problem by looping over each word and combining the sentiments till it completes the whole sentence thus by getting the intuition behind the overall sentence fed to the inputs of the neurons.

The accuracy of the LSTM model is about 86%.

C. Google Brain's BERT model: -

BERT stands for (Bidirectional Encoding Representational Transformer was proposed by researchers at Google brain

in 2018. Although the main aim of that was to improve the understanding of the meaning of queries related to Google Search, BERT becomes one of the most efficient and complete solutions for various natural language processing tasks. BERT has the capability of having generated the state-of-the-art results on Sentence pair classification task, question-answer task, and Sentiment analysis, etc.

The accuracy of the BERT model is about 98% which is best among various algorithms used for performing sentiment analysis.
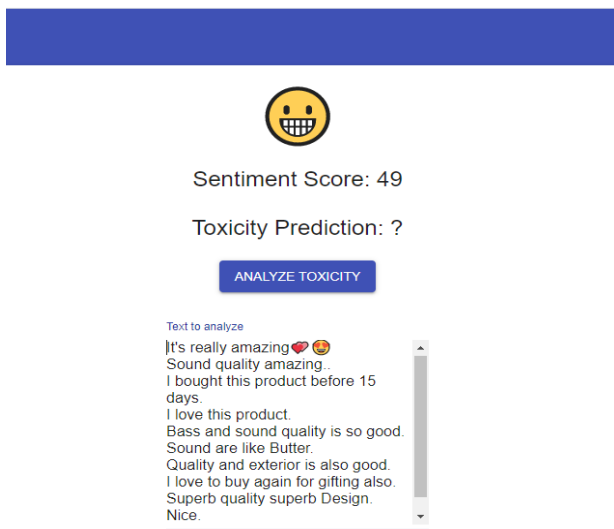
Fig 2. Sentiment Analysis of live reviews

The above image (Fig 2.) is a demonstration of sentiment analysis using the BERT model, wherein a random comment/review is inserted in the textbox, for every positive or negative comment the algorithm will return its respective emoji representation and the sentiment score. The sentiment score is based on the Bag of words concept, which returns a positive number i.e number > 0, if sentiment of the comment is positive and it returns a negative number i.e number<0, if the sentiment of the comment is negative, otherwise it returns 0, as for neutral statement.

6. IMPLEMENTATION

The beginning phase in the opinion mining is to separate and choose content highlights. A portion of the present highlights are:

i. Terms nearness and recurrence: This gives various events of a word in an info content.

ii. Part of Speech (POS): labeling the word in given content and discovering descriptors, as they are significant pointers of assessments.

iii. Feeling words and expressions: these are words generally used to express sentiments including fortunate or unfortunate, as or loathe. While a few expressions express opinions lacking utilizing conclusion words.

iv. Refutations: the presence of negative words may change the conclusion significance like not great is equal to awful.

Feature Selection strategies are often isolated into vocabulary- based techniques that require human explanation, and measurable strategies which are programmed techniques which are utilized as often as possible. Dictionary based methodologies typically start with a touch arrangement of 'seed' words. At that time, they bootstrap this set through synonym location or on-line assets to urge a much bigger dictionary. While factual methodologies are completely programmed. The feature selection techniques treat the reports either as gathering of words (Bag of Words (BOWs)), or as a string which holds the succession of words within the archive. Bag of Words is regularly utilized in view of it's a basic methodology for the order procedure. The most widely recognized component choice advance is the expulsion of stop-words and stemming. Three of the foremost as often as possible utilized factual strategies in highlight determination are examined within the following are and their related articles.

By filtering out the specific content of our need using algorithms on dataset we can perform different operations to get what we are looking for exactly:

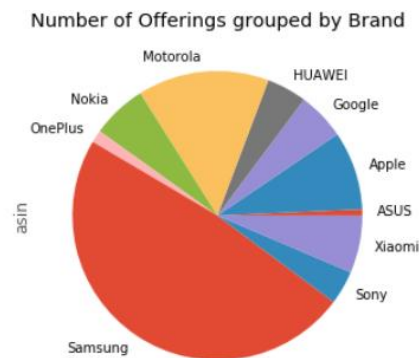### A) Number of offerings in the market

Fig 3. Number of offerings in the market

This feature of the project helps us to understand the market distribution. Which brand has most numbers, which brand has larger market share. This feature tells us the market influenced by the brand. Brands represented by different colors in the pie chart shows the market capture by each brand and its influence present in the market till which extent, who has the major offerings

present in the market. As in the figure we can say nearly half of the market is captured by Samsung which is represented by red color. All the analysis are based on the number of reviews.

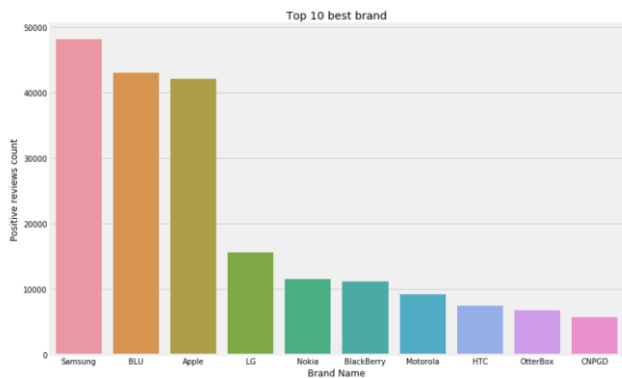## B) Top ten best Brands in the market

Fig 4. Top ten best Brand

With This feature we can find out the top 10 best products present in the market. Based on that a company can make a similar product which will be popular. It will help customer to buy from the best product present in the market.

C) Monthly number of reviews per Brand

Fig 5. Monthly number of reviews per Brand

Monthly reviews will help to analyze the activity of brands in the market. Eventually which will make it easy to understand the sentiments among people of the product by brand present in the market activity?

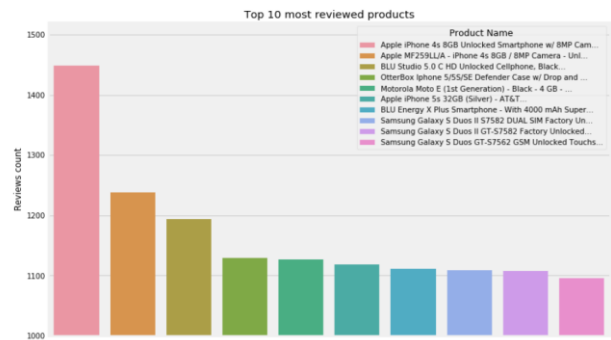## D) Top 10 most reviewed product-brands in the market

Fig 6. Top 10 most reviewed product-brands

Most reviewed product-brands help to understand which brand is more popular and active among masses. Business giants can study on such products why they are most reviewed, what they are reviewed for.
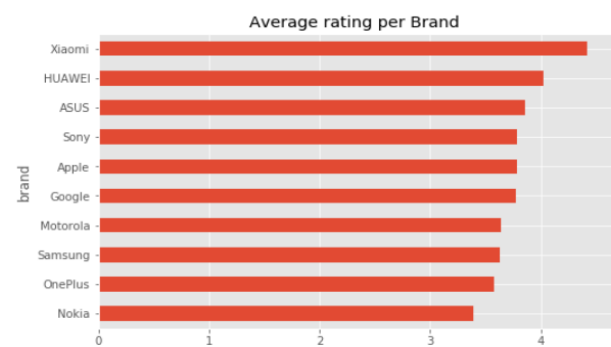
E) Average rating per Brand

Fig 7. Average rating per Brand

Average rating helps to realize about the product, its importance where does it stand out in the market and the opinion of public about the product.
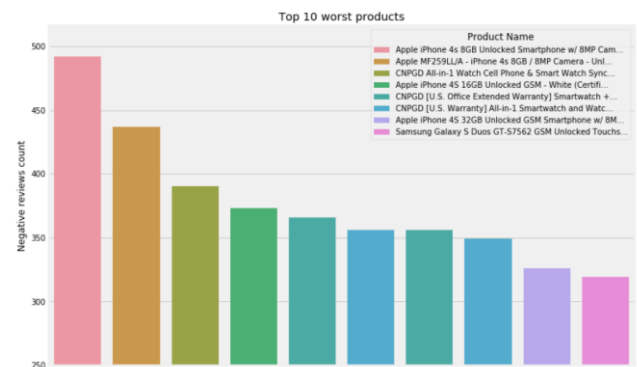
F) Top 10 worst product as per reviews

Fig 8. Top 10 worst product

By knowing which are the top worst product can help the brands to understand where they lacked and comeback stronger

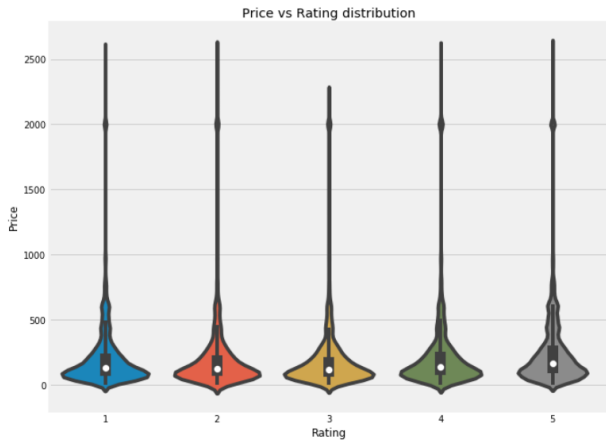## G) Price vs Rating distribution

Fig 9. Price vs Rating distribution

This price vs rating distribution helps to identify the price range which has the greatest number of users or buyers. It can help brands to target specific groups. It shows what the greatest number of people can go for or afford.

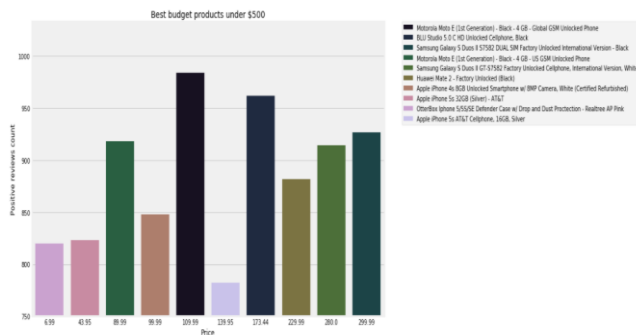## H) Best product as per price filter

Fig 10. Best product as per price filter

Filtering the product based on price range and reviews count which are positive can help a user to quickly find the best product within that price range.

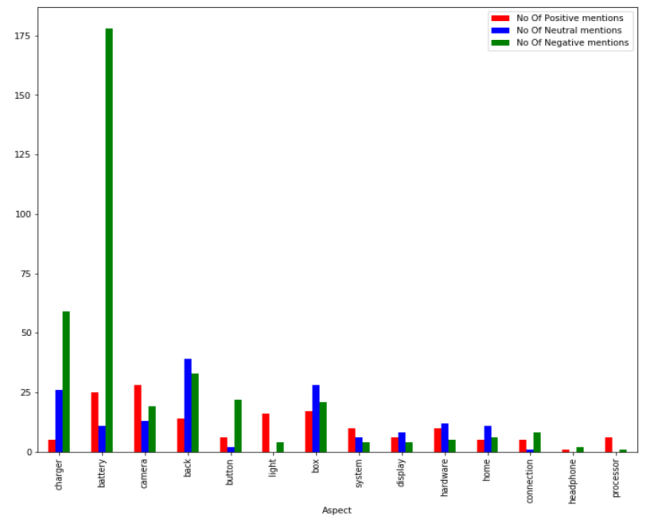## I) Features opinion mining of a product

Fig 11. Features Opinion Mining of a product

With feature mining of a product users can get to know more about the product and its features what are they good for and what they lack. Each feature shows the strength and weakness of the product. Based on the need of the user he/she can go for the appropriate feature they are looking for.

7. DISCUSSION AND ANALYSIS:

Data unit is an important factor in emotional analysis. IMDB and Amazon.com are sources of information recognized by the research information unit. IMDB can be a source for Flick surveys and amazon.com can be a source for price research for a variety of items. These sources are used in emotional analysis and emotional planning activities. It appears that twitter was used individually now and was redesigned within the most up-to-date year. Twitter is a well-respected social media website wherever its conclusions affect the conclusions of certain people, and its length is greater than one hundred and forty characters. The unit of machine learning calculations undoubtedly seeks to address the problem of its direct separation of the senses and the ability to use the preparatory data that provides the advantage of spatial flexibility. The Lexicon is primarily based on the calculator as often as possible to deal with general emotional analysis problems to the best of their knowledge. They should implement basic savings and computer savings. The number and level of articles using machine readings and therefore Lexicon-based statistical estimate is strong over the years. The final work over the next few years shows that analysts are using a dictionary-based approach even more now and once again. This is often because it describes many ways of analyzing emotions despite having many advanced features. Methods of machine learning is currently an open method of pursuit.

8. OUTCOMES:

The Product Analyst advisor achieved its goal. different features of the project were implemented and the top 10 best products present in the market was achieved in Fig 4, the top 10 worst products present in the market as per review data was achieved in Fig 8. , then price vs rating distribution was achieved in Fig 9. , features opinion mining of a particular product was achieved in Fig 11. , Best product as per price filter and sentiment analysis was achieved in Fig 10. , Average rating per Brand as per reviews was achieved in Fig 6. , Number of offerings in the market given by the brand was calculated in Fig 3. , Average rating per brand based on the reviews was calculated in Fig 7.

9. CONCLUSION:

Emotional analysis refers to the classification of texts based on the emotions they contain. This document focuses on a typical three-step emotional analysis model, namely data correction, review analysis and emotion classification, and describes the strategies of representatives involved in those measures. Product Analyst Advisor stands on all the conditions and achieved its goal.

Sensory analysis is an emerging field of archeology and computer language research and has attracted the attention of major research in the last few years. Future research will explore the complex mechanisms of the concept and output of a product feature, as well as new segment models that can handle individual label assets in a limited range. Applications using results in emotional analysis are also expected to appear soon.

10. REFERENCES:

[1] Sathis Kumar, NT Mohamed abeem, P Manoj, C K Jeyachandran, K , "Sentimental Analysis (Opinion Mining) in Social Network by Using SVM(Support Vector Machine) Algorithm", IEEE 2020.

[2] Wang, Chih-Chien Day, Min-Yuh Chen, Chien-Chang Liou, Jai-Wei . "Temporal and Sentimental Analysis of A Real Case of Fake Reviews in Taiwan'', IEEE 2017.

[3] Wang, Chih-Chien Day, Min-Yuh Chen, Chien-Chang Liou, Jai-Wei . "Temporal and Sentimental Analysis of A Real Case of Fake Reviews in Taiwan'', IEEE 2017.

[4] Singh, Nishant et. el, "Production Prediction based on News using Sentimental Analysis ", IEEE 2019

[5] Charoenkwan, Phasit, "ThaiFBDeep: A Sentimental Analysis Using Deep Learning Combined with Bag-of-Words Features on Thai Facebook Data '', IEEE 2018.

[6] Vijayaragavan, Ponnusamy, Aramudhan, "An optimal support vector machine based classification model for sentimental" Soft Computing - A Fusion of Foundations, Methodologies & Applications 2020.

[7] Wang, Zhang, Yongquan, Kangshun Zhang, Exploring mutual information-based sentimental analysis with kernel-based extreme learning machine for stock prediction" Soft Computing - A Fusion of Foundations, Methodologies & Applications 2017.

[8] Chang, Wei-Lun Chen, Li-Ming Verkholantsev, Alexey, "Revisiting Online Video Popularity: A Sentimental Analysis." Cybernetics & Systems. 2019, Vol. 50 Issue 6, p563-577. 15p

[9] G., Komarasamy, V., Vivek, "An Efficient Sentimental Analysis Mining for Unclassified Data in Big Data Analytics." Journal of Applied Information Science. Dec2019, Vol. 7 Issue 2, p10-18. 9p.

[10] Ruwa, Nelson Mao, Qirong Song, Heping Jia, Hongjie1 Dong, "Triple attention network for sentimental visual question answering" Computer Vision & Image Understanding. Dec2019, Vol. 189, pN.PAG-N.PAG .

[11] Xu, Yuanbo Yang, Yongjian Han, Jiayu1 Wang, En Zhuang, Fuzhen, Yang, Jingyuan4 Xiong, Hui, "NeuO: Exploiting the sentimental bias between ratings and reviews with neural networks." Neural Networks. Mar2019, Vol. 111, p77-88. 12p.

[12] Seo, Sanghyun Kang, Dongwann, "Study on predicting sentiment from images using categorical and sentimental keyword-based image retrieval" Journal of Applied Information Science. Dec2016, Vol. 5 Issue 2, p13-16.9p.

[13] Wang, Zhihong Guo, Yi, "Rumour events detection enhanced by encoding sentimental information into time series" Neurocomputing. Jul2020, Vol. 397, p224-243. 20p.

[14] Najma Sultana , Pintu Kumar , Monika Rani Patra , Sourabh Chandra and S.K. Safikul Alam "Sentimental Analysis of Product Review " IJSC 2019

[15]https://monkeylearn.com/blog/sentiment-analysis-of-product-reviews/