SENTIMENT ANALYSIS ON PRODUCT FEATURES BASED ON LEXICON APPROACH USING NATURAL LANGUAGE PROCESSING

Ameya Yerpude, Akshay Phirke, Ayush Agrawal and Atharva Deshmukh

Department of Computer Science and Engineering, RCOEM, Nagpur, India

ABSTRACT

Sentiment analysis has played an important role in identifying what other people think and what their behavior is. Text can be used to analyze the sentiment and classified as positive, negative or neutral. Applying the sentiment analysis on the product reviews on e-market helps not only the customer but also the industry people for taking decision. The method which provides sentiment analysis about the individual product's features is discussed here. This paper presents the use of Natural Language Processing and SentiWordNet in this interesting application in Python: 1. Sentiment Analysis on Product review [Domain: Electronic]2. sentiment analysis regarding the product's feature present in the product review [Sub Domain: Mobile Phones]. It usesa lexicon based approach in which text is tokenized for calculating the sentiment analysis of the product reviews on a e-market. The first part of paper includessentiment analyzer whichclassifiesthe sentiment present in product reviews into positive, negative or neutral depending on the polarity. The second part of the paper is an extension to the first part in which the customer review's containing product's features will be segregated and then these separated reviews are classified into positive, negative and neutral using sentiment analysis. Here, mobile phones are used as the product with features as screen, processors, etc. This gives a business solution for users and industries for effective product decisions.

KEYWORDS

Sentiment Analysis, Natural Language Processing, SentiWordNet, lexicon based approach

1. INTRODUCTION

Sentiment depicts a person's perspective and expression about any particular thing.Nowadays ,e-markets have a growing business and have became revolutionary in the terms of purchasing goods online.Thus, people show their reactions and attitude towards the product in their product reviews. Now, other customers on the basis of previous customers experience's and reviews can think of buying the product on e-market or not. These product reviews depicts the persons perception about the product in the product reviews.Now for instance, the product seems nice, depicts a person's positive attitude towards the product. In the same way, a customer may have different attitude toward the product and its features and customer can post the reviews online. Sentiment analysis is a process of describing a piece of text into positive,negative or neutral context. Thus, sentiment analysis[24] on product reviews helps the customer to check how many positive reaction and negative reaction have been done on the product . It is not only consumers who gets help by this summary but using this analysis the company also getsinformationfor improvisation on the product.

DOI: 10.5121/ijnlc.2019.8301

Sentiment analysis has many number of applications, this paper include the use of sentiment analysis on the product reviews. Sentiment analysis uses a part of Natural Languageprocessing which thus helps in preprocessing of text. Sentiment analysis can be done with the help of twoapproaches : Supervised learning and Unsupervised learning . In supervised learning approach contains many different methods and one of them is dictionary basedmethod which uses an extractive based algorithm, it follows a certain pattern in which word and phrases are identified and worked upon and based on the algorithm and with the help of dictionary, the sentiment analysis is done. Unsupervised approach, is approach where learning techniques such as naive bayes, support vector machine ,etc are used. The paper include the application of Sentiment analysis in :

- Product reviews
  Domain: Electronics
- Product's feature based sentiment analysis on product reviews
  Domain : Mobile phones.

Here, first of sentiment analysis will be done on the complete reviews present in the datasets. Using this analysis , the statistics will be shown in the form of visual representation such as pie and bar graphs, which will show the affection, liking and disliking of a product in electronics domain. Nowadays, people buy products on the basis of the popularity of the product's feature, like camera of mobile phone, thus it would make it easy for the customers for selection of any product if the pros and cons of the product are known.Sometimes customer wants to buy the product on the basis of the popularity of that product's feature example, if a person wants a mobile phone with best operating system, then customers need to have research analysis manually to get the best option for the product which is a tiresome process as customers need to check both positive and negative reviews of customers to get an idea about the product's feature which customer want to buy.Thus, on the basis of the customers reviews , sentiment analysis can be done on the features of the product and segregate the reviews on the basis of the features of the product into positive , negative or neutral reviews regarding the features of the product i.e the camera of the mobile phone is good but processor is not unto the mark ,this shows that this reviews is positive on the basis of camera and negative onto basis of processor. Thus, providing the pros and cons of the product. Thus, applying the sentiment analysis on those feature will help in getting the pros and cons and help in further suggesting customers on the basis of the features.

Hence, by using the sentiment analysis the customers can avoid the tiresome procedure to manually get the reviews of product's feature and check what other customer reviews suggest about the product and its features. Hence, this paper provides the sentiment of customers on an individual product and its features too. Product's features based sentiment analysis on product reviews will focus on the product reviews of the mobile phones domains in which depending on the certain features fixed in mobile phone domains such as ram, speed, camera, etc. customer's liking and disliking will be checked.On the basis of the analysis, a bar graph will be shown which will depict the features liking and disliking in a product on the basis of product reviews. This will help customers to select the product on the basis of feature and using this statistics it will visible to them that how many customers have liked this feature in this product . Also using this feature based sentiment analysis, further pros and cons of a product will be shown which will be helpful for customers, as customer won't have to search through every product and compare products manually. Thus, making easy for customers to choose the products on the basis of features in the product on the basis of product reviews. Companies will also be benefited as it would help the companies to improvise the products and its features.

The paper provides a way for sentiment analysis using python. Python[4] being a scripting language provides a better and efficient platform for sentiment analysis because of the tools and dictionary present . The tools include NLTK[10] (natural language toolkit), TextBlob[17] used for data pre processing. The python becomes a easy platform for using SentiWordNet[23] dictionary which contains max number of words. Sentiwordnet provides a score of every word present in that dicionary. Hence, using python to extract words from the dictionary becomes very easy and using the dictionary are retrieved.Before extraction of words, text processing, text categorization[9],part of speech tagging [15],text classification[13] is required for pre-processing. It includes various categories like syntactic analysis[16]where sequence of words are converted to structures that shows relation between the words, morphological analysis[5] where punctuations are grouped and removed etc. Using an algorithm, the sentiment score of the reviews is calculated and using the scores, customers reviews are segregated into positive, negative or neutral reviews .Also , python[20] contains Graphical user interdate tools such as tkinter which makes its easy for providing statistical data visually easily.

This paper provides a way to ease the effort of customer. Most of companies just extract the features present in the customer reviews and just provides a searching technique to retrieve the customer reviews containing that features. But if we applied the sentiment analysis based on that feature then it is possible to easily associate the reviews containing that features with negative, positive or neutral reviews. Hence , making it easy to provide the pros and cons of the product features. It would help in easing the efforts of customers to get sentiment about the feature of the product and overall sentiment of the product.

2. RELATED WORK

NLP[1] is an important area of research in many different application problems of machine learning , text classification , text tagging, part of speech tagging, text summarization, etc.NLP provides a major platform for sentiment classification andSentiwordNet, which is an open reliable resources, provides a method for sentiment classification or analysis. SentiWordNet provides an extension to the word net such that all synsets are associated with some value concerning positive, negative and objective connotations. This value each category from 0.0 to 1.0.

The project has two parts Overall Sentiment analysis of the products reviews and other is product’s feature based sentiment analysis in the products reviews where both requires extraction of text. The First part include overall product reviews sentiment analysis for classifying the reviews into positive, negative and neutral and the second part focuses on specific product type of reviews in which features of that product might be present and taking out the gist of the review what the customer is trying to convey i.esentiments customers about the product’s feature example, take mobile phone reviews given by customer may contain some features of mobile phone such as camera , processor, etc. , in this work we can check the sentiment about the review on this feature basis of any product. There can be a case where the customer wants a good camera phone so, the customer can search for reviews regarding the camera and not the general reviews of the mobile.

Bo pang and LiallanLee[25] provided with different approach for opinion mining and sentiment classification, these different approach included the lexicon approach with used WordNet. On the similar lines, this paper used the lexicon approach for sentiment analysis with a change in the corpse, uses SentiwordNet which has more accurate scores for the words in the corpse present in the SentiWordNet.

Akshaya Garje and Karbhari Kale[22] described different sentiment analysis level of classification which are divided as Document, Sentence and Aspect. Document based level of sentiment analysis defines the overall opinion of the document as a whole classifying it into positive, negative or neutral sentiment.On the similar lines, Sentence based level defines the sentence opinion and Aspect level defines the classification of features in the corpse into the sentiments and provides the opinion of document as a whole.Thiswork, improved this method by selecting the features present in the sentence for calculation of sentiment of the reviews as whole and targeted the components of the product present in the reviews and calculated the sentiment of the reviews regarding the component of that product in the reviews.

This mainly focuses on the feature extraction for sentiment analysis and also modifies the way in which the product sentiment analysis is done that is it provides a product component based sentiment analysis in which the products component opinion is calculated from the reviews and on the basis of the product's component the customer opinion about it is also classified into positive , negative and neutral sentiment.

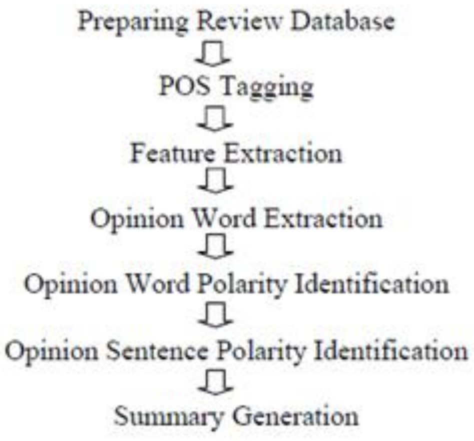
Preparing Review Database
POS Tagging
Feature Extraction
Opinion Word Extraction
Opinion Word Polarity Identification
Opinion Sentence Polarity Identification
Summary Generation

Fig 1: Basic step of Feature Extraction[22]

Fig1 shows the generic flow of the research for calculating the sentiment analysis. For product components, mobile phones customers reviews have been used as dataset. This feature based sentiment analysis is added onto this model for improvisation, which was needed in this work, as the products components need to be extracted from the reviews using regular expression and applying the above the model for sentiment classification to get the opinion of the customer on the product's component, which gives a detailed sentiment analysis of the overall reviews regarding the overall product review and its components too.Python[14] language was found out to be the best language to perform different NLP operations as the NLTK toolkit was provided easily in libraries of Python.Python was really helpful in data pre-processing and data handling and made it easy to access the database for storing the components reviews present in the product reviews.

3. COMPONENTS USED

3.1 Components for Data Processing

1. Collections: Collections provides with different types of data structures to store data in different form out of which defaultdict(x) is used to declare and define a variable of any data type 'x'.This data structure uses of keys and their corresponding values as a pair and stores them accordingly in the (key,value) pair. Associative arrays and hash tables also make use of python dictionaries where functions are mapped with their pointer values as addresses. The general syntax of a dictionary is given below:

```
dict = {p(key):x(value)}
Example: dict2 = {'hello': 'NN', 'Lord': ‘NN'}
```

2. CSVReader: CSVReader is an python module helpful in reading the csv files. The customer reviews were retrieved from different platform in csv format with cells format with rows and columns. So, python provides this module which provides different function such as reader(file,delimiter) where file is the file object which is to be read and delimiter is the parameter which provides the way the cells are separated from one another. The syntax in pythonis :

```
>>with open('E:\data.csv', encoding="utf8") as csvfile
>>readCSV = csv.reader(csvfile, delimiter=‘,’)
```

3. Nltk.tokenize: Tokens are the individual terms present in the whole text. A sentence is a sentence because of the words presents in it. Hence, words are nothing but terms of the text. Hence, tokenize function helps in splitting the sentence into substrings. A python program doesn’t know the concept of words unless it is specified how a word is formed. Hence, ntlk.tokenize module helps splits the whole input text into sentences and then sentences into words using its algorithm.It has sent_tokenize and word_tokenize functions. Each function has different algorithms for splitting, sent_tokenize splits the whole text given to sentences and word_tokenize splits the sentences to words.If a sentence is ‘This product is quite expensive.’

```
>>word_sents = [word_tokenize(se.lower()) for se in sents]
>> print word_sents[['This', ‘product', 'is', ‘quite', 'expensive', ‘.']
```

4. Nltk.corpus: Corpora contain a large set of structured data and provide an already defined english words dataset. In Python nth module, a collection of corpus contains various classes which can be used to access these large set of data. Stopwords are most common words such as the, is, on, at, etc. In Sentiment Analysis, this stop words mean nothing as they don’t provide any support in calculation of sentiment score and thus, is not helpful in differentiating the customer reviews . The method call stopwords.words(‘english’) was used to remove these unimportant words.

For Example:

```
>> from nltk.corpus import stopwords
>> a= set(stopwords.words('english') + list(punctuation))
>> print a set([u'all', u'just', u'being', '-', u'over', u'both', u'through', u'yourselves', u'its', u'before',
'$', u'herself', u'had', ',', u'should', u'to', u'only', u'under', u'ours', u'has', '<', u'do', u'them', u'his',
u'very', u'they', u'not', u'during', u'now', u'him', u'nor', '`', u'did', '^']) and many more.
```

5. Nltk.pos_tag: In python, nltk module provides part of speech tagging which tags the word with its part of speech on the basis of its usage in sentences using the pos_tag function. Providing a (word,tuple) pair results with each word having a part of speech tag in tuple.It takes tokens as input and results in (word,tags) pair.In sentiment analysis, part of speech tag of word is important to get the score of the word which will be used to calculate the overall sentiment score.

For Example:

```
>>import nltk
>>nltk.pos_tag([‘hello’,'Lord'])
[('hello', 'NN'), ('Lord', ‘NN’)]
```

6. Nltk.WordNetLemmatizer : In python, nltk provides a module for lemmatization of words get the root word of any word, like root word for excitement is excite, irrespective of the inflected forms of the word, using lemmatization it can analyzed as an single word. In a large a same word can be used in many inflected form and proposing the same meaning moreover in the whole text. Hence to group all these word together and analyze all the word as a single word only, lemmatization module is used. Nltk module has WordNetLemmatizer as function which provides object containing a set of function for lemmatization, one of them is lemmatize which is to be used in the project.For Example:

```
>>from nltk.stem import WordNetLemmatizer
>>lemmatizer = WordNetLemmatizer()
>>print("rocks :", lemmatizer.lemmatize("rocks"))
>>print("corpora :", lemmatizer.lemmatize("corpora"))
rocks : rock
corpora : corpus
```

3.2 Component for Sentiment Analysis

SentiWordNet: SentiWordNet is an open source dictionary providing the sentiment scores of the word present in this lexical source. SentiWordNet provides with the sentiment scores of the word on the basis of the part of speech of that word. The dictionary can contain different score for same word for different part of speech tag of that word. A word can be used differently in each sentences on the basis of the part of speech tags. So, scores are stored in SentiWordNet on the basis of the part of speech of the words.SentiWordNet provides withe the score and these score are to be retrieved from this lexical source.Python provides this sources which makes it easy to retrieve the data from the dictionary based source. SentiWordNet has certain function for retrieving the scores.

1. syn_sets(word,tag): this method provides with the scores of the synonyms of the word with the help of the part of speech tag provided as the parameter. this methods returns positive , negative and neutral score of all the synonyms of the word if stored in the dictionary.

Example:

```
>>breakdown = swn.senti_synset(‘breakdown','n')
>>> print(breakdown)
<breakdown.n: PosScore=0.0 NegScore=0.25>
```

2.pos_score() : this methods returns the positive score from the each word of the list returned by above senti_synset() function.

3.neg_score():this methods returns the negative score from the each word of the list returned by above senti_synset() function.

3.3 Component for Database Connectivity

1.MySQLdb Module :Python provide MySQLdb[8] module which is a compatible interface to MySQL database server that connects database present in MySQL by creating a connection link. The next step to using MySQL in a Python script is to make a connection to the database present in the MySQL Server [12]. All Python Database-API 2.0 modules provide a function 'database_name.connect'. This is the function that is used to connect to the database, in our case MySQL.

```
>>>db(anyname)=MySQLdb.connect(host=MY_HOST_NAME,user=MY_USER_NAME,passwd
=PA SSWORD, db=DB_NAME)
```

In order to use connection to use the database ,there was a need to create a cursor object to iterate the data present in the database. The cursor object is used to work with the tables of database specified in the Python Database-API 2.0. It gives us the ability to have multiple separate working environments and makes it easy to iterate different databases through the same connection to the database. To create a cursor by executing the 'cursor' function of a database object.

```
>>>cur(name) = db.cursor()
```

Execution of queries is done by using execute() method.

3.4 Components for Graphical User Interface

Tkintermodule :Tkinter[] is the Python module for implementing GUI programming where it provides functions like windows, frames, button for navigation as well as for action to be performed, message and dialogue boxes for text entering, scrollbars, text widgets and design templates for GUI. Tkinter mainly provides framework for all the GUI component required for user interaction. Text widget is where multiple lines can be written in a text box and Tkinter provides flexibility for working with widgets. Distributions of TK module are available for Unix as well as Windows.

Example: To create a new widget,

```
>>import Tkinter
>>new = Tkinter.Tk()
>>new.mainloop()
```

3.5 Components for Graph Plotting

Matplotlib: Python provides a 2D plotting library for data visualization like line graphs, bar graphs, pie charts, histograms, scatterplots etc. Matplotlib can be used in Python shell as well as script, html servers and GUI toolkits[26]. Simple plotting can be combined with iPython provides a Matlab type interface. This module lets you deal with the object oriented concepts thereby letting user a full control over its working. PlotPy is imported for Matplotlib which provides collection of various styles for plotting. Functions like constructing a graph, changing variables, plotting area and lines in that area, labels can be easily implemented in PlotPy. A bar and line graph is used to store statistical information about the number positive and negative reviews about the features of an product. A pie chart is used to show the division of customer reviews into positive , negative and neutral in a visualization.

Example: To plot a line graph

```
>>>plt.plot([1,2,3,4,5,6],[0.5,1.5,2.5,3.5])
>>>plt.ylabel(‘numbers')
>>>plt.xlabel('decimal numbers')
>>>plt.show()
```

4. EXPERIMENTATION DETAILS

The customer reviews play a vital role for any e-commerce website as it tells about the popularity of any product and further helps in decision making for other customers as well as for the manufacturing company or the distributer and seller of the product. These reviews can be either positive, negative or neutral in nature and depends on the customer whether customer was satisfied or not with the product.But this takes a lot of manual effort if there is no automation present to detect the nature of the review.So, sentiment analysis is to be done on those reviews . Sentiment analysis helps in classifying the nature of the reviews into positive ,negative or neutral reviews. The machine learning based methods used different algorithm to predict the sentiment of the text on the basis of some previously known sentiments of the text. It is like predicting the new text on the basis of some previously known sentiments of the text. It uses a labelled dataset of text with known sentiments to predict the sentiments of the new texts.The lexicon based approaches uses the dictionary of sentiment scores of words already present to calculate the new sentiment score of the whole text.In this paper, there are two parts for sentiment analysis , first part tells about the overall sentiment analysis of the products reviews and classifying the reviews of the products into positive ,negative and neutral reviews and plotting the numbers to a pie chart, and second part is specific to a single domain, mobile phones, which shows sentiment analysis of these product reviews on the basis of the products feature that is the reviews about the product’s feature will be classified into positive , negative and neutral reviews and the count will be stored into the database which will be used to reflect the pros and cons of the product and the suggestion for the product on the basis of these reviews.

Part 1

Flow of working Algorithm

input: customer reviews output: BarGraph

1. Provide the Amazon customer reviews dataset to the program
2. The dataset csv file containing different attribute with reviews being one of them will be read.
3. The reviews text will be split to sentences.These sentence then split into words.
4. These words are processed by removing stopwords and punctuations.
5. Each word will be lemmatized to get the root word.
6. The sentiwordnet will be used to retrieve positive and negative score for the word.
7. Calculate the average of all these positive as well as negative score of these words in a sentence.
8. Then again, calculate the average of all the values calculated for each sentence in the text.
9. if avg_positive>avg_negative , then it will be classified as positive review
10. if avg_positive<avg_negative , then it will be classified as negative review
11. if avg_positive - avg_negative=0 , then it will be classified as neutral review

Detailed Explanation

Take the dataset as input which is an csv file. Using CSVReader module it will reader row by row. Take input text from the reviews attribute of the dataset and tokenism it using nltk.tokenise module into sentences and then sentences into words.Now applying data preprocessing to the tokens, filtering the words by removing stopwords and removing the punctuations. The input text will split to sentences using nltk.sent_tokenize() function and these sentences are further operated by nltk.word_tokenize() to form words from the sentences. Stopwords are the words likes articles, to be verbs (am, is are, was, were, etc) and also the punctuations of which if the frequency is calculated will just increase the complexity of the code. So these words are to be neglected as soon as the text is entered.

```
sentences = sent_tokenize(input_text)        // ’sentences’ contains sentences
words_sentences= [word_tokenize(sentences)]  // ’word_sentences’ contains words
a= set(stopwords.words('english') + list(punctuation))  // a contains stop words only
```

Now, the next step is part of speech tagging of the words formed in tokenizing step.

```
word_tag=nltk.pos_tag(stoken) // part of speech tag will be generated for each word
```

Now, there might be same root words but with different form, example form and formation has the same form as root word. Hence lemmatization will be done to get the root word. Nltk.WordNetLemmetizer module with lemmetize(word) function will be used for lemmetization .

```
wnl = nltk.WordNetLemmatizer()
lemmatized = wnl.lemmatize(word)
```

Now, the next part is to get the positive and negative score for the root word with tag from the SentiWordNet library. This will be done with the syn_set(words,tag) function from the SentiWordNet Library in Python. This functions return list containing the positive and negative scores of synonyms of the word with the tag as the parameters.

synsets = list(swn.senti_synsets(lemmatized, newtag))

Now, the score is calculated for each word by iterating the above list as follows :
for syn in synsets:

score += syn.pos_score() - syn.neg_score()

The further part is to store this score to a list and calculate average of all the score in this list to get the score of a sentence and store this into another list. Now , again calculate the average of all the score in this list to get the score of complete text and update the counters of positive and negative reviews .

```
if score > 0:
positive1[count] = positive1[count] + 1
if score < 0:
negative1[count] = negative1[count] - 1
```

Now, plot the pie chart using these counter as parameter. This will be done by pie function in matplotlib.pyplot library in python.

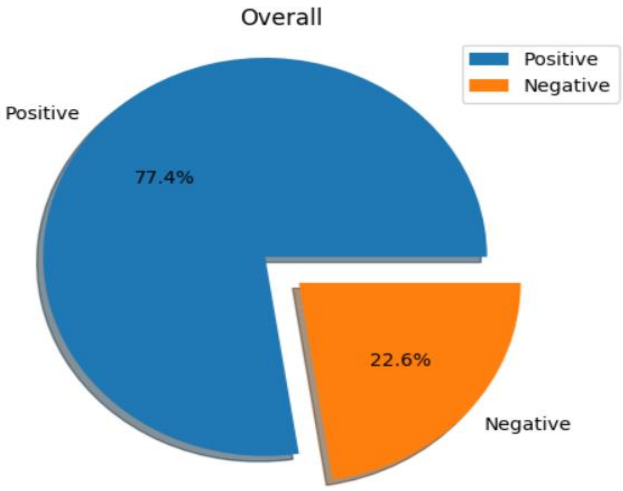

Fig 2: Classifying of the reviews

Part 2

In previous part the only focus is on the overall classification of reviews. In this part, focuses ison the features of the product also. Nowadays, only product's features are extracted from the customer reviews but there is no sentiment analysis regarding the product’s features. To explain it further, take example that mobile phones have so many different features such as processors, camera, etc. Now, by using sentiment analysis on the features of the mobile phones it can classify thereviews into positive , negative or neutral review on the basis of the mobile phone features i.eit can judge what customer says about the feature of the product regardless of what the sentiment of the overall review was. It means if a customer writes a positive review about the camera of the phone but writes the negative reviews as a whole about the phone, then the sentiment analysis on the product’s feature will make these reviews positive in terms of camera and negative in terms of overall.Thus, sentiment analysis on the products feature will help in finding out the positive reviews of the customers for the features of the product irrespective of what the overall reviews of the product was.It helps in getting the positive as well as negative reviews of the customer regarding the features of a product. It will be helpful for customers who are willing to buy any product on the basis of the product’s feature and not on the basis of overall reviews of the product. There are different features of many different products. Hence it was limited to mobile phones for demonstration. Here, customer reviews dataset with mobile phones was used as domain which contains customer reviews for different mobile phones sold on the e-commerce website.

Flow of the Algorithm

Input : Mobile phones customer review
Output: bar graph showing positive and negative reviews regarding the feature of the phone, pros and cons of the phones

1. Input the product’s features list
2. Extract the customer reviews containing the features in the reviews.
3. For each feature, apply above part 1 procedure for sentiment analysis for each customer review
4. for each review, classify that review into positive, negative or neutral review and update the respective counter for each feature.
5. Store the positive , negative and neutral reviews counter for each feature in database.
6. retrieve the values for each counter and using that plot bar graph showing the number of positive or negative reviews
7. Retrieve the values to show pros and cons of the product.

Detailed Explanation

1. Here, the main focus is on the sentiment of the reviews regarding the phone’s feature. As, domain is mobile phones, for different phones the feature list will be same. Using the list of features, extract the reviews containing the feature in it.That means extracting all those reviews in which any feature in the list is mentioned in the reviews.

*features = ["CAMERA", "CHARGER", "DISPLAY", "BATTERY", "PROCESSOR", "RAM",
"SPEAKER", "TOUCH", "SCREEN", "BUTTONS",
"SIM CARD”]*

2. Generate a list with different mobile phones present in the dataset.

```
with open('E:\data.csv', encoding="utf8") as csvfile:
readCSV = csv.reader(csvfile, delimiter=',')
rows = [r[0] for r in readCSV]
temp = rows[0]
if rows[0]:
for row in rows:
if row not in temp:
list1.append(row)
temp = row
```

3. For each feature , applying the same data pre-processing on the reviews retrieved that was applied in the previous part that is tokenizing the text to sentences and sentences to word removing the stop words, punctuation, part of speech tagging of the words, lemmatization of the words. This process will return those reviews from customer reviews of each phones present in the mobile phone dataset of customer reviews with features of feature list present in it.

4. Now, for each feature, the sentiment analysis is done and counter for each feature is maintained, those counters are positive reviews, negative reviews and neutral reviews.

5. Then, it is maintained for each phone’s customer reviews with the features list present and stored in database.

6. For each feature, database has 3 attributes which are phones, positive reviews and negative reviews. Phones contain different mobile by different company sold by e-commerce website by amazon as tool uses the dataset from the amazon website. Positive reviews contains the number of positive reviews and similarly Negative review hold the number of negative reviews.

Table 1: Databases

Phones	Positive review	Negative review
Samsung a50	30	50
Iphone X	60	20
Honor Play	10	50

7. Now, GUI was used in this tool for selection any mobile phone present in the dataset. Hence a drop down menu was created using Tkinter module’s combobox function as follows :

```
>>box = ttk.Combobox(self.parent, textvariable=self.box_value, state='readonly',
width=75)
>>self.box.bind("<<ComboboxSelected>>", self.choosePhone)
```

8. Now after selecting the phone each feature database is accessed to the count and for individual phone, a bar graph is generated showing the positive and negative reviews with x axis showing the features and y axis showing just number of positive and negative reviews.

```
>>plt.figure(figsize=(20, 10))
>>ax = plt.subplot()
```

```
>>ax.set_title('Features')
>>ax.set_xlabel("List of Features")
>>ax.set_ylabel("Number of Reviews")
>>ax.bar(x, negative_data, width=1, color='r', edgecolor='0', label='Negative Review')
>>ax.bar(x, positive_data, width=1, color='g', edgecolor='0', label='Positive Review')
```

9. Now using the same selected mobile phone value is used to show the pros and cons in which the count is checked , if number of positive reviews > number of negative reviews , then those features are pros for that phone, else cons for that mobile phones. For that tkinter module was used to create an window which shows the pros and cons.

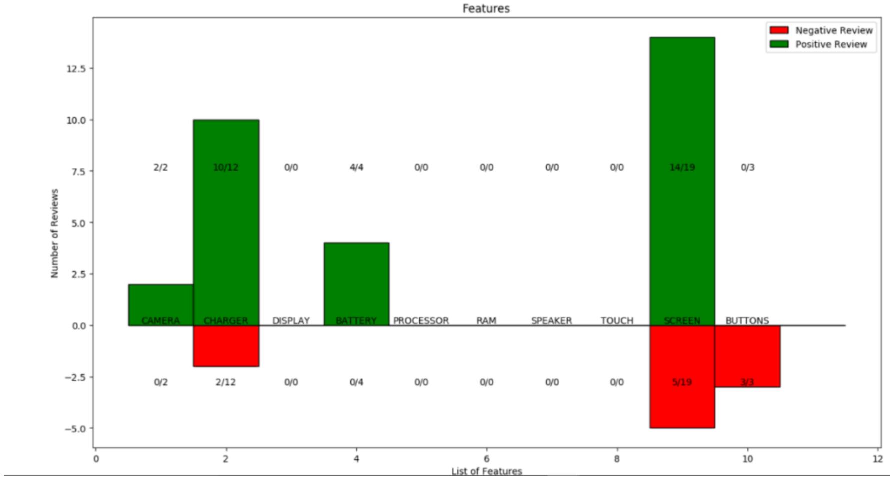

Fig 3: Classifying of the reviews according to the features of products

5. CONCLUSION

In this paper, sentiment analysis on product reviews is performed and results are provided in the form of bar graph and pie chart. Here, pie chart was displayed representing the overall positive and negatives reviews of a product. The overall analysis was done for any generalized by just extracting those words which forms the sentiment of the review, for every customer.The above Fig 2 shows the first part results with classification of all the customer reviews present in the dataset into positive, negative, and neutral review in the form of pie chart. With many different productfor same domain with different company labels, it becomes tough for customer to guess which product to buy. Hence by classifying the reviews into positive and negative reviews of that product on the basis of the features will make it easy for customer to buy the product with features of that product being the best. So, it was decided upon a small set of features commonly found in all devices. It checked the polarity of each feature in every review of a specific product and displayed a bar graph which showed the total count of positive and negative mentions of a feature. The above fig 3 shows the result of the classification of segregated customer reviews with respect to the product’s features. The bar graphs with product’s features on x axis and y axis contains classified customer reviews containing the product’s features on x-axis into positive reviews and negative reviews present in the dataset containing the product’s feature. The fig 3 shows the details with number of positive and negative reviews containing the product’s features and the other customer reviews which does not contains the product’s features are not considered in the second part but considered in the first part. The fig 3 gives the summary with x-axis showing the features and y-axis providing with the classification of reviews containing the feature present on the x-axis. Thus, providing with the count of reviews with positive sentiment of customers towards the features of the product too.

Sentiment analysis methods till now have been used to detect the polarity in the thoughts and opinions of all the users that access social media and many other platforms that are expressed by them in form of texts. Researchers and Businesses are very interested to understand the thoughts of people and how they respond to everything happening around them with the data present on the platforms. Companies can use this to evaluate their advertisement campaigns and to improve their products.

REFERENCES

[1] Allen, James, "Natural Language Understanding", Second edition (Redwood City: Benjamin/Cummings, 1995).

[2] Baxendale, P. (1958). Machine-made index for technical literature - an experiment. IBM Journal of Research Development, 2(4):354–361. [2, 3, 5]

[3] Bird Steven, Klein Ewan, Loper Edward June 2009, "Natural Language Processing with Python", Pages 16,27,79

[4] Cortez Eli, Altigran S da da Silva 2013, " Unsupervised Information Extraction by Text Segmentation", Ch 3

[5] Kumar Ela, "Artificial Intelligence", Pages 313-315

[6] Goddard Cliff Second edition 2011,"Semantic Analysis: A practical introduction ", Section 1.1- 1.5

[7] Lukaszewski Albert 2010, "MySQL for Python", Ch 1,2,3

[8] Manning Christopher D., SchützeHinrich Sixth Edition 2003,"Foundations of Statistical Natural Language Processing", Ch 4 Page no. 575

[9] Martelli Alex Second edition July 2006, "Python in a Nutshell", Pages 44,201.

[10] Natural Language Toolkit, Retrieved from http://www.nltk.org

[11] Pattern 2.6, Retrieved from http://www.clips.ua.ac.be/pattern

[12] Prasad Reshma, Mary Priya Sebastian, International Journal on Natural Language Computing (IJNLC) Vol. 3, No.2, April 2014, " A survey on phrase structure learning methods for text classification"

[13] Python Language, Retrieved from https://www.python.org/

[14] Rodrigues Mário , Teixeira António , "Advanced Applications of Natural Language Processing for Performing ", Ch 1,2,4

[15] Sobin Nicholas 2011, "Syntactic Analysis: The Basics", Ch 1,2

[16] Swaroop C H, “A Byte of Python: Basics and Syntax of Python”, Ch 5,8,9,10

[17] TextBlob: Simplified Text Processing, Retrieved from http://textblob.readthedocs.org/en/dev

[18] ThanosCostantino ,"Research and Advanced Technology for Digital Libraries", Page 338-362

[19] Tosi Sandro November 2009, "Matplotlib for Python Developers", Ch 2,3

[20] Aashutosh Bhatt et al, / (IJCSIT) International Journal of Computer Science and Information Technologies, Vol. 6 (6) , 2015, 5107-5110

[21] Akshaya R. Garjeet al, International Journal of Advanced Research in Computer Science, 8(9), Nov–Dec, 2017,554-557

[22] Andrea Esuli and Fabrizio Sebastiani. 2006. SENTIWORDNET:A publicly available lexical resource for opinion mining. In Proceedings of the 5th Conference on Language Resources and Evaluation (LREC’06), pages 417–422, Genova, IT.

[23] Bo Pang and Lillian Lee. 2008. Opinion mining and sentiment analysis. Foundations and Trends in Information Retrieval, 2(1/2):1–135.

AUTHORS

Ameya R. Yerpude, is pursuing his Bachelor of Engineering 2015-19,Computer Science and Engineering from Shri Ramdeobaba College Of Engineering and Management, Nagpur. He has been working on a domain of Natural Language Processing and Machine Learning from two years. His interests include Natural Language Processing, Databases Artificial Intelligence, Machine Learning, Deep Learning, Internet of things.Email: yerpudear@rknec.edu

Akshay S. Phirke, is pursuing his Bachelor of Engineering 2015-19,Computer Science and Engineering from Shri Ramdeobaba College Of Engineering and Management, Nagpur. He has been working on a domain of Machine Learning and Natural Language Processing from two years. His interests include Natural Language Processing, Regular Expressions Artificial Intelligence, Blockchain, Machine Learning . Email: phirkeas@rknec.edu

Ayush S. Agrawal, is pursuing his Bachelor of Engineering 2015-19,Computer Science and Engineering from Shri Ramdeobaba College Of Engineering and Management, Nagpur. He has been working on a domain of Machine Learning and Natural Language Processing from two years. His interests include Natural Language Processing, Regular Expressions Artificial Intelligence, Blockchain, Machine Learning, Internet of things. Email: agrawalas13@rknec.edu

Atharva H. Deshmukh, is pursuing his Bachelor of Engineering 2015-19,Computer Science and Engineering from Shri Ramdeobaba College Of Engineering and Management, Nagpur. He has been working on a domain of Machine Learning and Natural Language Processing from two years. His interests include Natural Language Processing, Regular Expressions ,Database. Email: deshmukhah@rknec.edu