

ABSA ANALYZER FOR EFFICIENT ASPECT EXTRACTION AND SUMMARIZATION

Kotagiri.Srividya¹, A. Mary Sowjanya²

¹Research Scholar Department of CS&SE, Andhra University College of Engineering (A), Visakhapatnam, AP, India, Department of CSE, GMR Institute of Technology, Rajam, India.

²Department of CS&SE, Andhra University College of Engineering (A), Visakhapatnam, AP, India.

Email ¹id:srividya.kotagiri@gmail.com, ²sowmaa@gmail.com

Received: 11.03.2020 Revised: 12.04.2020 Accepted: 28.05.2020

ABSTRACT:With the rapid growth in e-commerce websites, consumers are also interested in reading consumer reviews that share customer feelings about a product's different features before making buying decisions. Within this paper we present ABSA (Aspect-Based Sentiment Analyzer), based on customer feedback for the performance of aspect-level sentiment analysis. In the proposed approach, users will train models on the target domain for the output of aspect level sentiment analysis tasks in the first step. In the second step, a web application is made available from which an end-user may post comments to evaluate feelings at aspect level. In the third step tweets and reviews are produced which are important to the user defined keyword and which correspond to interesting aspects. The framework is developed using state-of-the-art approaches to aspect term extraction, aspect category detection and the detection of polarity aspects.

KEYWORDS: Aspect level opinion mining, sentiment analysis, Aspect term extraction, Aspect Category detection, Aspect Term polarity detection, Twitter

I. INTRODUCTION

Sentiment Analysis (SA), known as Opinion Mining (OM), is the study wherein people's perspectives, feelings, perceptions, sentiments, practices and feelings towards people of different types expressed in the document are analyzed. An organization can work in various ways, systems, organizations, individuals, problems, events, topics and aspects. Opinion mining or sentiment analysis has spread from healthcare, financials, consumer goods, infrastructure, telecommunications and e-commerce to social activities, political campaigns and elections to every possible domain. Aspect Based Sentiment Analysis is very important in extracting and summarizing opinions from any kind of datasets. Many Aspect Based Sentiment Analysis systems have been built in the last decade for a wide range of organizations, such as Movie Reviews, Travels, Digital Cameras, Utilities, Computers and Restaurants. With the rapid growth of e-commerce websites, we are seeing a rising increase in online sales and product and service purchases. This has resulted in a rise in review sites for various products and services, such as Amazon and many others. A user often has an interest in knowing what other people are saying about a product of interest before making a purchase decision. Customer reviews are equally relevant for businesses to know about the pitfalls of their goods and services. Nonetheless, having insights into different aspects of the product / service the consumer is dissatisfied with, or where the change is needed, is a difficult challenge from a large number of unstructured comments. Analysis of feelings at the aspect-level can reveal this useful insight as it will extract food and service as aspects and then evaluate feelings expressed for those aspects. Opinion Mining collects and analyzes the views of people regarding an individual while Sentiment Analysis then analyzes the feelings conveyed in a document. Sentiment Analysis or Opinion Mining can be viewed as a multi-step classification problem from the Data Mining standpoint.

II. BACKGROUND KNOWLEDGE

Nazir et al. [1], Reviewed and presented a detailed and in-depth review of the issues and challenges of the extraction dimension, the analysis of sentimental aspects and the analysis of aspects and their assembled solutions.

The writers have proposed a new approach that would trigger some serious thoughts and considerations to boost the solutions present. Akshat Bakliwal et al[2], developed a method that can be used in an aggregated set of tweets to find an opinion. Authors used two separate datasets in this approach which are developed using emoticons and list of suggestive terms as noisy labels, respectively. This method of scoring "Popularity Score" allows to assess the popularity score at the individual words level. Chandra Pandey et al.[3] propose a novel metaheuristic method (CSK) based on K-means and the search for cuckoo. The proposed approach was utilized to locate the optimum cluster heads from Twitter dataset's. The viability of the proposed technique was assessed on different Twitter datasets and stood out from molecule swarm improvement, differential development, cuckoo search.

Ana Salwa Shafie et al[4] conducted a preliminary study on the output of different type of dependency relationship with different patterns of POS tags in pre-extracting candidate aspect from customer analysis. Farza Nurifan et al[5] proposes a restaurant-based aspect study that involves physical setting, quality of food, quality of service and fairness of price. Aspect Term Extraction (ATE), Aspect Keyword Extraction (AKE), Aspect Categorization (AC), and Sentiment Analysis (SA) are part of the analysis steps. ATE uses the modification of the Double Propagation process and other forms of topic Modeling. Feng Xu and others[6]. Presents an ongoing naïve Bayes learning system for the classification of sentiments on broad and multi-domain e-commerce product review site. They expand the parameter estimation process in naïve Bayes to a continuous learning style while retaining the high computational efficiency of the conventional naïve Bayes model. K Schouten et al.[7] reviewed a range of supervised and unsupervised machine learning techniques for aspect extraction and classification of sentiments. The writers emphasized the need for consistency and standardization in terms of methods and data sets for assessment in order to draw firm conclusions. K Schouten et al.[8] proposed an unsupervised and supervised method that would be able to identify aspect categories based on co-occurrence frequencies. Kashfia Sailunaza et al[9] detect and evaluate the feelings and emotions conveyed in their twitter posts by people through text, and use them to produce recommendations. They collected tweets and answers on a few different subjects, and generated a dataset with text, user, emotion, details on feelings, etc. L. Shu, H. Xu, B. Liu et al[10] provides a concentrated contribution to the extraction of supervised aspects. Mohd Ridzwan Yaakub et al[11] introduces the techniques used in sentiment analysis by previous researchers which are Machine Learning and Natural Language Processing (NLP) to solve the classification problem. The comparison between these two main approaches reveals the techniques of machine learning. Svetlana Kiritchenko et al[12] explains what stood first in the detection of categories of aspect, first in the detection of feelings towards categories of aspect, third in the detection of aspects, and secondly in the detection of feelings towards aspects in the laptop and restaurant domains. S. Poria et al.[13], suggested a seven-layer deep convolutionary neural network that would mark each word as either aspect or non-aspect word in opinionated sentences. In a similar purpose, they have built up a lot of semantic instances, and combined them with the neural network. Saumya Chaturvedi[14] suggests the use of classification of sentiment analysis as an effective method for examining textual data from various online resources. Tri Doan[15] online random forest variant for conducting sentiment analysis on reviews of customers. Their model can achieve a similar accuracy to offline methods and equivalent to other online models.

III. PROBLEM STATEMENT

The Analyzer is divided into three stages. The end user will train custom models on the target domain in the first step for aspect-based sentiment analysis. During this phase user acquires the labeled dataset. Models are then conditioned for aspect term extraction, aspect category detection, and aspect polarity classification after preprocessing and characteristic extraction. A web application is created in the second step using Flask to provide the end-user with an interactive user interface to perform aspect-based sentiment analysis. Flask is a micro framework for building web applications with Python. We need to pull out tweets and develop user-specified keyword reviews in the third phase and display relevant aspects.

IV. PROPOSED SYSTEM

The proposed system workflow

1. Train the model with the dataset
2. Enter some text then perform preprocessing and feature extraction
3. Input the text to the trained model it will extract the aspect terms
4. For Aspect category detection; whatever be the aspect terms we get from the trained model
5. given as input we will get the category
6. In aspect polarity detection above the above steps will be repeated.
7. For twitter the input data will be fetched from the twitter using tweepy API based on the search keyword entered from the user. The data fetched will be dynamic and real time.

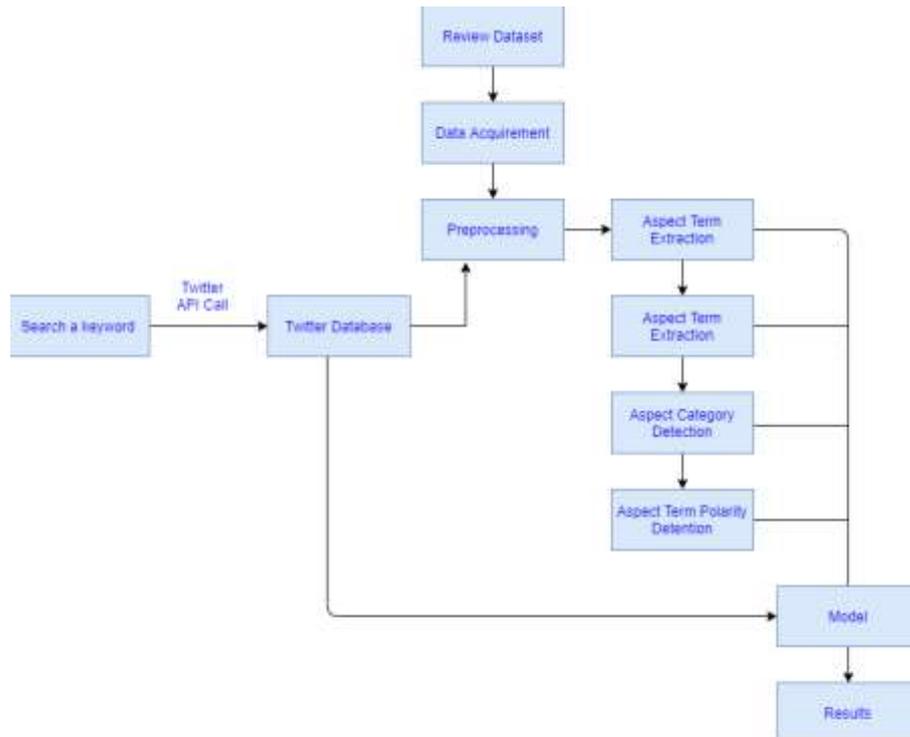


Figure 1: Architecture

A. Data acquirement

Interpreted datasets have been used for training purposes. The datasets used for training and testing must be labeled dataset. Here we used CSV file format as input file.

- (i) Aspect Term: Include expressions of aspects in the sentence analysis. Each term of aspect thought to be separated by a semicolon. The point of utilizing a semi-colon as a separator is to permit expressions of multi-word aspect.
- (ii) Aspect Category: Contains classifications of aspects found in the review sentence. They should separate every class by space.
- (iii) Aspect Polarity: Contains polarity for each aspect
- (iv) Text: Contains the review sentence
- (v) Twitter reviews: Contains text with user reviews

B. Preprocessing

For any text mining application the preprocessing is a fundamental step. Comparing various papers and discussions, text is regularly unstructured and noisier. People want to prepare a proper analysis. Data cleaning is in this manner required before the modeling. Preprocessing is applied to the column text after the information is read by the machine.

- (i) Acquire the dataset: To make and improve Machine Learning models, first we have to use the appropriate dataset. The dataset will comprise of information gathered from different and various sources which are then joined to make data set in an appropriate format. You can likewise make a dataset by gathering information through various Python APIs. Once the dataset is prepared, you should place it in a CSV, or HTML, or XLSX file formats
- (ii) Pre-processing text: This module eliminates the noise and unwanted data from the text and organizes the text to detect aspects in order to retrieve positive and negative aspects. It performs POS, Stemming and tokenization.
- (iii) Tokenization: In a given document each sentence is divided into tokens and through way unwanted data from the text. The text is divided into various tokens.

- (iv) Parts of speech: It is task of categorizing each word in a sentence with proper part of speech. The parts of speech will comprise of nouns, verb, adverbs, adjectives, pronouns, conjunction and their sub-categories.

(v) Lemmatization: Lemmatization is utilized to understand the abbreviations of word which are utilized in reviews.

C. Aspect term extraction:

In ABS Analyzer, Aspect Term Extraction aims to understand explicit aspects recorded in an document. This process is essentially a function of sequence labeling. A review sentence is given as an input sequence during the aspect term extraction stage, and each token is considered to be an element. The paper utilizes a Random Conditional Field (CRF) to prepare the term dimension.

D. Feature Extraction methods:

(i) **Morphological Features** By using this feature we can specify the nature of a word. These are represented as vectors for each word in a sentence .It will specify the starting letter is capital letter or not. Capital letters are presented in the word. How much percentage of capital letters and small letters are present. Numeric letters are present.

(ii)POS Tagging

In tagging the words in a sentence are tagged with proper set of labels. The labels are called Tag-set. The word that gives same meaning will be partitioned into same class.

(iii)Word Embeddings

In order to represent document vocabulary word-embedding were used. It catches the perspective of word in a sentence, meaning and syntactic similarities and the relation between the words.Word2Vec is one of the most famous method to learn word embedding's utilizing shallow neural networks

E. Aspect category detection

The Aspect Category Detection module attempts to classify categories of latent aspects from a review document. Aspect categories of target area are typically predefined. Aspect categories may be for the Restaurant domain {food, service, atmosphere, general, quality, location}. Aspect categories are more general than aspect terms, and are not generally defined in a text as aspects.

F. Twitter Aspect-Based Feature Extraction

In order to find the explicit aspects can be this task can be used. We can identify the explicit aspects that appears in the noun and noun phrases in the tweets by using tagging(POS).The important features or aspects can be retrieved by the comments posted by the users in the twitter and we can assign ranks to those tweets by the frequency of the words used.

G. Classification algorithms

The analyzer offers the option to choose from three different classification algorithms to train binary classifiers including Vector Machine Support (SVM), Decision Tree, Navie Bayes. The selection of algorithms was focused on their effectiveness in various text classification tasks including sentiment analysis, classification of documents.

(i)Support Vector Machine (SVM)

A Support vector machine (SVM) is a directed model of AI, which utilizes classification algorithms to classify into two classes.We are able to categorize new text after giving an SVM model sets of named training data for each category. It is a procedure of classification dependent on Bayes ' theorem .

(ii)Naive Bayes:

Navie Bayes is a straightforward procedure for developing classifiers: models that assign class labels to problem instances

(iii) A decision tree: A decision tree is a map of a series of related choices that may result. This helps an person or organization, based on their costs, risks and advantages, to balance potential actions against each other. As the name suggests, this uses a tree-like model of decisions.

H. Aspect polarity classification

ABSAnalyzer includes training data set annotated with discrete {positive, negative, neutral} sentiments. Using the classification algorithm a binary classifier is trained for each sentiment . The polarity is allocated, given a summary sentence and aspect terms, for which the respective classifier yields the highest score Term of aspect.

V. EXPERIMENTAL SETUP

A few Application Programming Interface (API) based services are available for aspect based sentiment analysis. API based services users to develop their own interface. As far as we could possibly know, there is no open-source structure accessible for aspect based sentiment analysis. The introduced analyzer creates the reviews by

utilizing the different datasets and by utilizing twitter utilizing tweepy API dependent on the search keyword entered from the user. The data fetched will be dynamic and real time.

VI. RESULTS AND DISCUSSIONS

The experimental results are obtained by using different domains of Datasets including restaurant, laptops, Twitter using tweepy API to evaluate the performance of ABSAnalyzer.

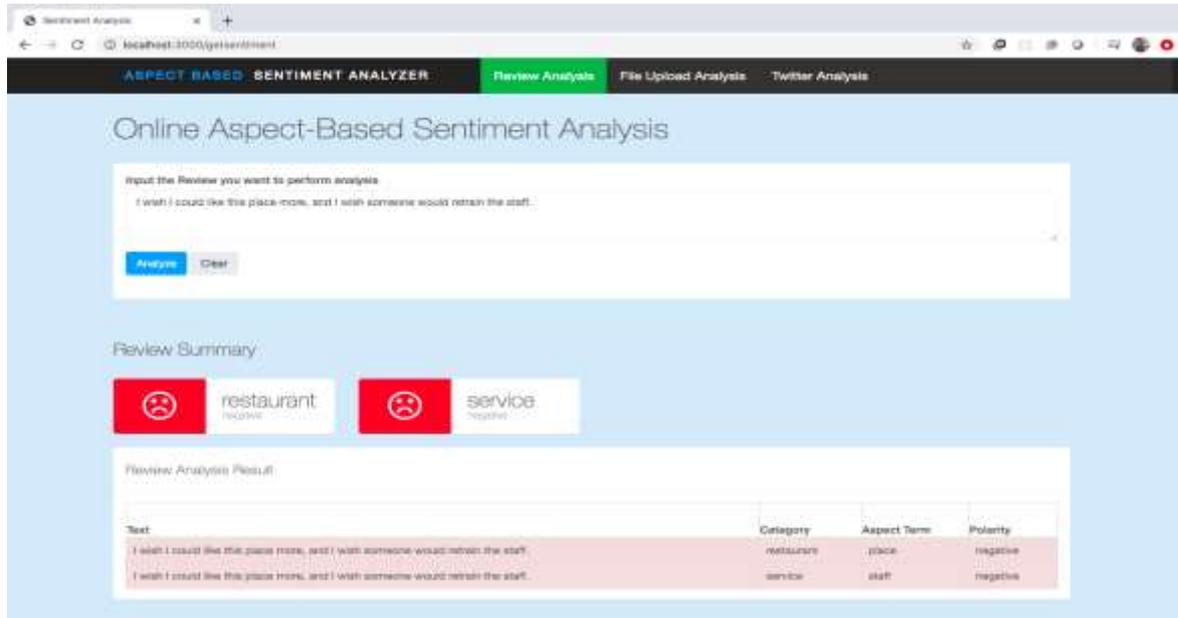


Figure 2: Review Analysis

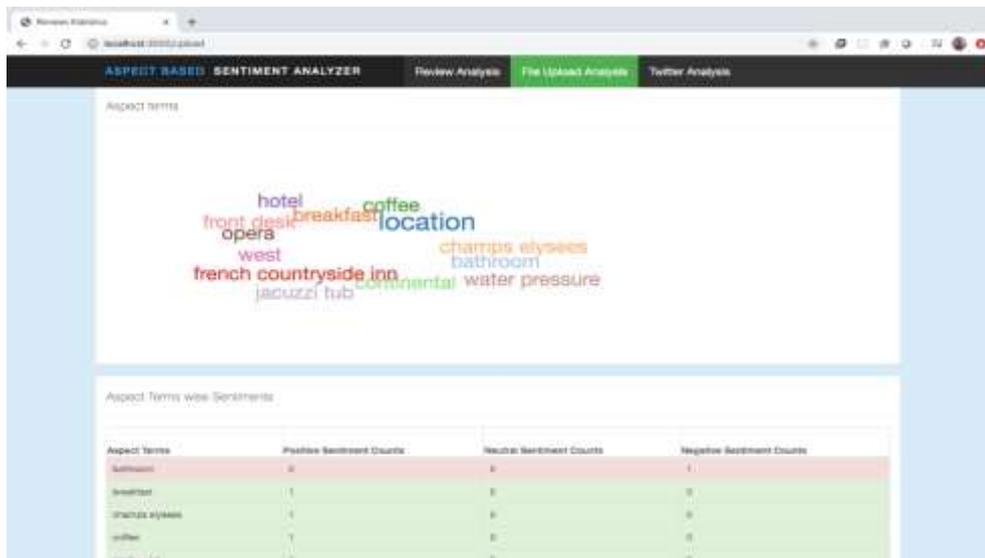


Figure 3: Aspect Terms

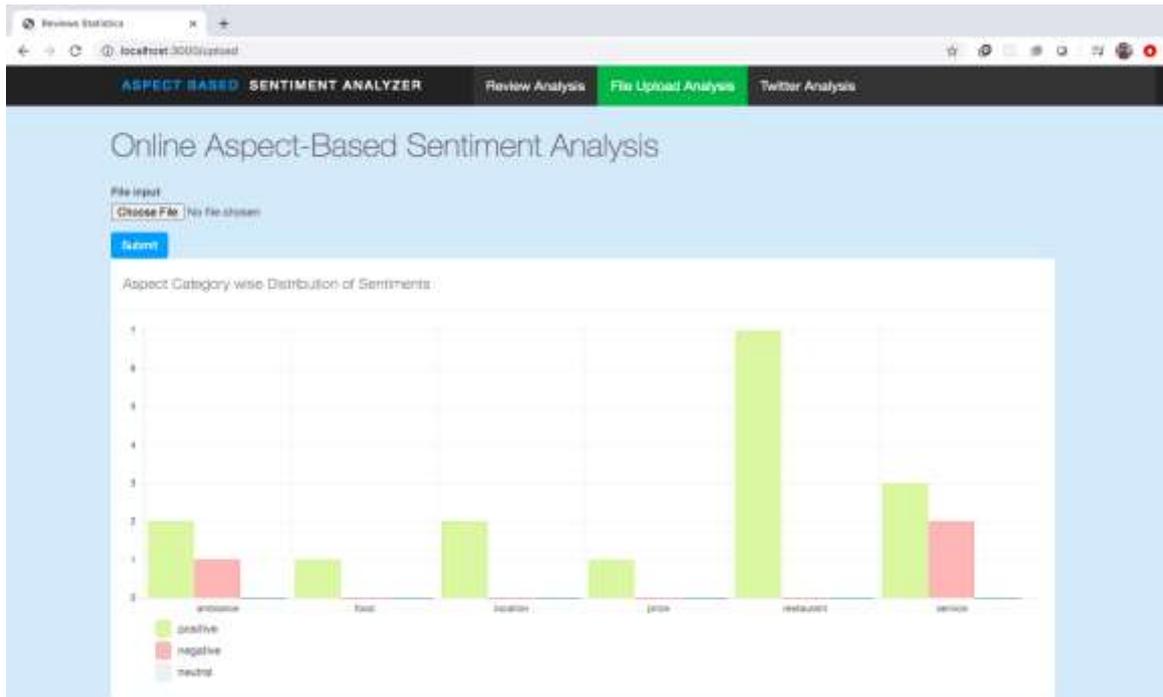


Figure 4: Aspect category

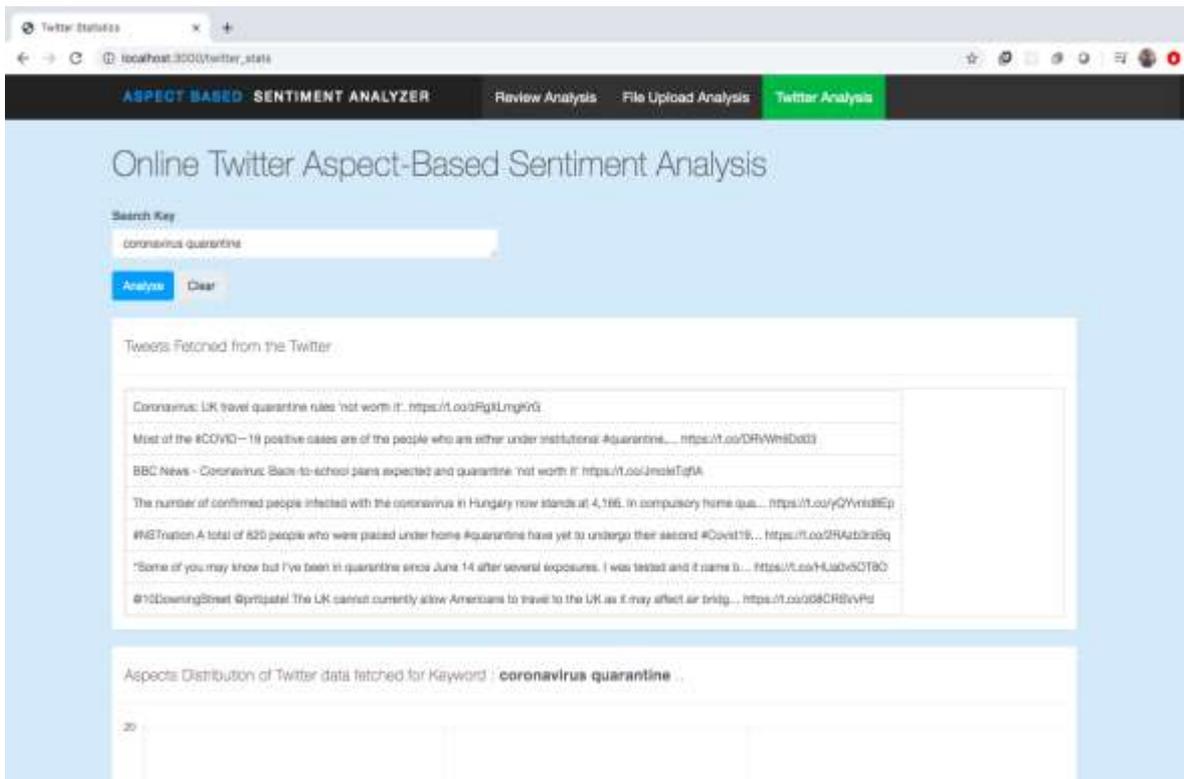


Figure 5: Keyword search from Twitter

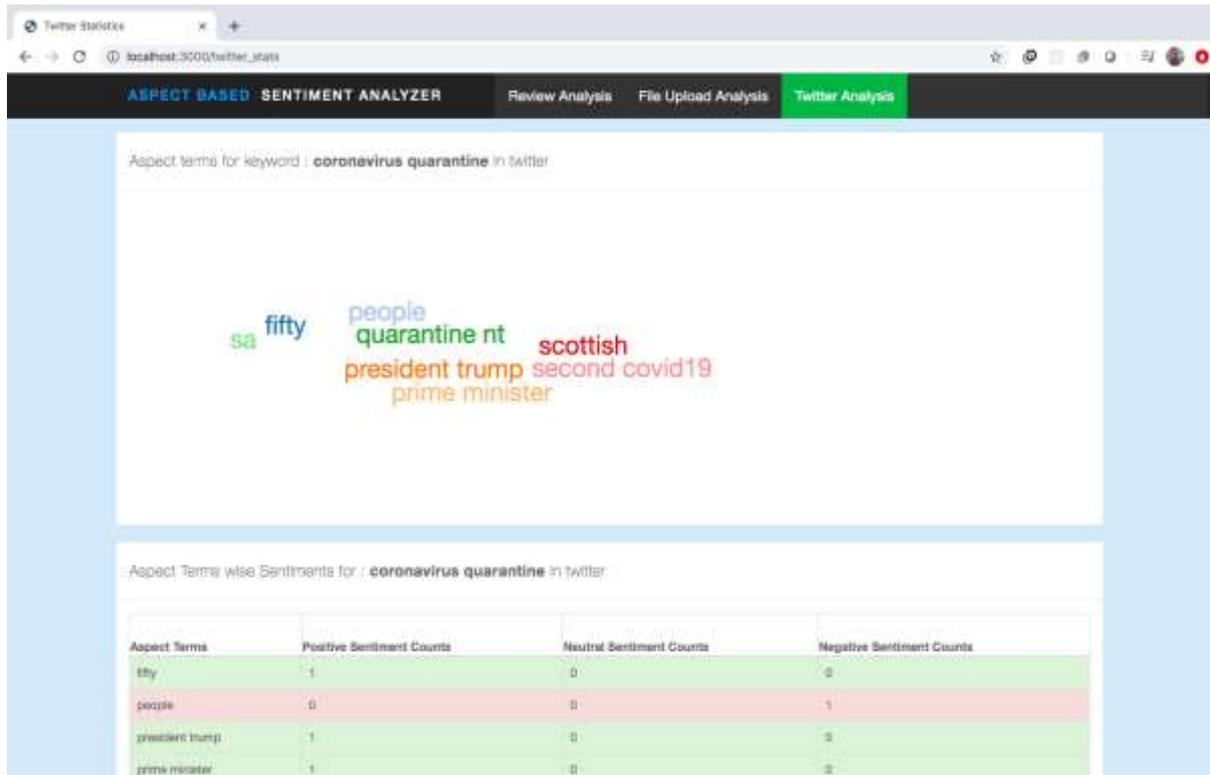


Figure 6: Aspect Terms for search Key

VII. CONCLUSION AND FUTURE WORK

In order to analyze and summarize customer reviews on social media this framework can be used. To predict the labels in each aspect word embedding’s used. The results show that ABSA analyzer gives better performance to get the aspects from the target domain. The future work will work on upgrading the analyzer by utilizing deep learning procedures to conquer and confinement of gathering commented on datasets for preparing.

VIII. ACKNOWLEDGEMENTS

This work is supported by the ministry of electronics and information Technology, Government of India under the Visvesvaraya PhD scheme for young faculty research fellows implemented by digital India Corporation (formerly Media Lab Asia).

IX. REFERENCES

- [1] A. Nazir, Y. Rao, L. Wu and L. Sun, "Issues and Challenges of Aspect-based Sentiment Analysis: A Comprehensive Survey," IEEE Transactions on Affective Computing, doi: 10.1109/TAFFC.2020.2970399.
- [2] Akshat Bakliwal, Piyush Arora, Senthil Madhappan, Nikhil Kapre, Mukesh Singh, and Vasudeva Varma, "Mining Sentiments from Tweets", Proceedings of the 3rd Workshop in Computational Approaches to Subjectivity and Sentiment Analysis, WASSA '12, pages 11–18, Stroudsburg, PA, USA, 2012
- [3] Avinash Chandra Pandey , Dharmveer Singh Rajpoot, Mukesh Saraswat "Twitter sentiment analysis using hybrid cuckoo search method", Information Processing and Management 53 (2017) 764–779
- [4] Ana Salwa Shafie , Nurfadhlina Mohd SharefMasrah Azrifah Azmi MuradAzreen Azman "Aspect Extraction Performance With POS Tag Patternof Dependency Relation in Aspect-based SentimentAnalysis", 2018 Fourth International Conference on Information Retrieval and Knowledge Management

- [5] Farza Nurifan1 Riyanarto Sarno Kelly Rossa Sungkono ,“Aspect Based Sentiment Analysis for Restaurant Reviews Using Hybrid ELMo-Wikipedia and Hybrid Expanded Opinion Lexicon-SentiCircle”, International Journal of Intelligent Engineering and Systems, Vol.12, No.6, 2019
- [6] Feng Xu, Zhenchun Pan, Rui Xia ,Feng Xu, Zhenchun Pan, Rui XiaFeng Xu, Zhenchun Pan, Rui Xia,E-commerce product review sentiment classification based on anaïve Bayes continuous learning framework,Information Processing and Management
- [7] K. Schouten, O. van der Weijde, F. Frasincar and R. Dekker, "Supervised and Unsupervised Aspect Category Detection for Sentiment Analysis with Co-occurrence Data," in IEEE Transactions on Cybernetics, vol. 48, no. 4, pp. 1263-1275, April 2018, doi,10.1109/TCYB.2017.2688801.
- [8] K. Schouten and F. Frasincar, "Survey on Aspect-Level Sentiment Analysis," in IEEE Transactions on Knowledge and Data Engineering, vol. 28, no. 3, pp. 813-830, 1 March 2016, doi: 10.1109/TKDE.2015.2485209.
- [9] Kashfia Sailunaz, Reda Alhaji ,” Emotion and sentiment analysis from Twitter text”, Journal of Computational Science, 1877-7503/© 2019 Elsevier
- [10] L. Shu, H. Xu, B. Liu, “Lifelong Learning CRF for Supervised Aspect Extraction” Proceedings of Meeting of the Association for Computational Linguistics,2, 2017, pp. 148–154.
- [11] Mohd Ridzwan Yaakub, Muhammad Iqbal Abu Latiffi , Liyana Safra Zaabar,”A Review on Sentiment Analysis Techniques and Applications”, Joint Conference on Green Engineering Technology & Applied Computing 2019, IOP Conf. Series: Materials Science and Engineering 551 (2019) 012070
- [12] Svetlana Kiritchenko, Xiaodan Zhu, Colin Cherry, and Saif M. Mohammad, “NRC-Canada-2014: Detecting Aspects and Sentiment in Customer Reviews” Proceedings of the 8th International Workshop on Semantic Evaluation (SemEval 2014), pages 437–442, Dublin, Ireland, August 23-24, 2014.
- [13] S. Poria, E. Cambria and A. Gelbukh, Aspect Extraction for Opinion Mining with a Deep Convolutional Neural Network, Knowledge-Based Systems 108 (2016) 42–49.
- [14] Saumya Chaturvedi , Vimal Mishra , Nitin Mishra ,“Sentiment Analysis using Machine Learning forBusiness Intelligence”, IEEE International Conference on Power, Control, Signals and Instrumentation Engineering (ICPCSI-2017)
- [15] Tri Doan and Jugal Kalita,”Sentiment Analysis of Restaurant Reviews on Yelp with Incremental Learning”, 2016 15th IEEE International Conference on Machine Learning and Applications