**SENTIMENTAL ANALYSIS**

**A Project Report**

## *Submitted by*

Deep Patel (190303105018)

Jeel Patel (190303105064)

Pranav Pal(190303105493)

***In fulfillment for the award of***

***the degree of***

**Bachelor of Engineering**

in

## COMPUTER SCIENCE AND ENGINEERING

***Under the guidance of***

*Prof. Mridul Mishra.*



**Department of Computer Science and Engineering**

**Parul University, Vadodara**

2021-22



**Parul University**

***Certificate***

This is to Certify that Project-II -Subject code 203105350 of 6th Semester entitled “Sentimental Analysis” of Group No. 39 has been successfully completed by

Deep Patel (190303105018)

Jeel Patel (190303105064)

Pranav Pal (190303105493)

under my guidance in partial fulfillment of the Bachelor of Technology (B.TECH) in

Computer Science and Engineering of Parul University in the Academic Year 2021-2022.

**Project Guide Project Coordinator Prof. Mridul Mishra Asst.Prof. Khushali Mistry**

**Head of Department External Examiner**

**CSE**

**Dr. Amit Barve**

# ACKNOWLEDGEMENT

Behind any major work undertaken by an individual there lies the contribution of the people who helped him to cross all the hurdles to achieve his goal. It gives me immense pleasure to express my sense of sincere gratitude towards my respected guide Mridul Mishra Sir, Senior Assistant Professor for his persistent, outstanding, invaluable co- operation and guidance. It is my achievement to be guided under him. He is a constant source of encouragement and momentum that any intricacy becomes simple. I gained a lot of invaluable guidance and prompt suggestions from him during the entire project work. I will be indebted to him forever and I take pride in working under him.

I also express my deep sense of regard and thanks to Dr. Amit Barve, Professor and Head of CSE Engineering Department. I feel very privileged to have had their precious advice, guidance and leadership. Last but not the least, my humble thanks to the Almighty God.

|  |  |  |
| --- | --- | --- |
| Place : Vadodara | Student Name | Enrollment No. |
| Date : 16-10-22 | Deep Patel | 190303105018 |
|  | Jeel Patel | 190303105064 |
|  | Pranav Pal | 190303105493 |
|  |  |  |

# ABSTRACT

Human Computer interaction is a very powerful and most current area of research because the human world is getting more digitised. This needs the digital systems to imitate the human behaviour correctly. Emotion is one aspect of human behaviour which plays an important role in human computer interaction, the computer interfaces need to recognize the emotion of the users in order to exhibit a truly intelligent behaviour.There is still a question on how to detect emotion from a text input. To solve this problem, this project generates an Emotion Recognition Model to extract emotion from text at the sentence level. Our method detects emotion from a text-input by using different deep learning algorithms like CNN, LSTM, SVM and HAN. The experiments show that the method could generate a good result for emotion detection from text input. To recognize emotion from text we have considered six emotions class such as joy, sadness, anger, love, fear, surprise).

**TABLE OF CONTENTS**

[Acknowledgement i](#_gjdgxs)

[Abstract ii](#_30j0zll)

List of Symbol, Abbreviation and Nomenclature Table of Content v

Chapter: 1 Introduction

1.1 Overview

1.2 Problem Statement

1.3 Aim and Objective

1.4 Scope

Chapter: 2 Literature Survey

Chapter: 3 Process flow and Methodology

3.1 Implementation Project Flow

3.1.1 Step 1: Setting up developing environment

3.1.2 Step 2: Choosing your Dataset

3.1.3 Step 3: Processing your data

3.1.4 Step 4: Feature Extraction

3.1.3 Step 5: Training Our Models

3.1.6 Step 6: Testing

Chapter: 4 Partial Implementation

4.1 Context on which the project is implemented

4.2 Tool and technologies which we used till now

4.3 Actual Implementation

Reference

**CHAPTER 1**

**INTRODUCTION**

### **Overview**

Emotion is one type of affect, other type of being mood, temperament and sensation. Emotions have been widely studied in psychology and in behavior sciences, as they are an important element of human nature. Our emotions influence every aspect of our lives, from how we connect with each other, to how we make decisions, to our health and well-being. Your emotional state can affect a very simple decision, like what you’re having for breakfast, to very big decisions, like what house you’re going to buy or who you’re going to marry. People who have high EQ, who are able to translate their emotion reading and sensing skills into their day-to-day behavior, tend to be more likable as human beings. They tend to be more successful in their professional lives and they actually tend to be healthier. They live longer and happier lives in general. Nowadays it is important in the field of human computer interactions. Emotions are also an important aspect in the interaction and communication between people. The exchange of emotions through text messages and posts of personal blogs poses the informal style of writing challenge for researches. Extraction of emotions from text can be applied for deciding the human computer interaction which governs communication and many more. Emotions may be expressed by a person’s speech, facial expressions and text based emotions respectively. Emotions may be expressed by one word or a bunch of words. Sentence level emotion detection method plays a crucial role to trace emotions or to search out the cues for generating such emotions. Sentences are the 1 essential information units of any document. For that reason, the document level emotion detection method depends on the emotion expressed by the individual sentences of that document that successively relies on the emotions expressed by the individual words. In computational linguistics, the recognition of emotions from texts is becoming crucial from an application point of view. The examples are affective computing, the tasks of opinion mining and market analysis, or natural language interfaces such as e-learning environments or educational and edutainment games. This needs computational approaches to successfully analyze this online content, recognize, and draw useful conclusions and detection of emotions.

### **1.2 Problem Statement**

### In today’s internet world, human expresses their emotions,sentiments, feelings via text or comments, emojis, likes and dislikes.Understanding the true meanings behind these electronic text is very crucial. The available approaches are work in the direction of recognizing the polarity of sentiment. The sentiment maybe positive or negative. Among the less explored sentiment areas is the recognition of types of emotions from text documents. Recognizing emotions conveyed by a text can provide an insight into the author’s intent and may lead to better understanding of the text’s content.

### **1.3 Aim and Objective**

### **Aim : -** To recognize the emotion behind a text phrase

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

### **1.4 Scope**

Sufficient amount of work has been done related to speech and facial emotion detection but text based emotion recognition system still requires attraction of researchers. The short messaging language have the ability to interrupt and falsify Natural language processing tasks done on text data. To illustrate that ability, consider an example, ”At de moment he cnt just put me in da better zone though. happy bday mic, ur a legend”. At this moment when going through this sentence, it will recognize some terms which doesn’t belong to decent plain text. But while going through these sentences, then and there human brain will resolve the short messaging language word to a meaningful word or phrase. When human see ”cnt” and its neighboring words ”he” and ”just”, human know that it is ”can’t”. That’s because human brain is trained with previous experiences. But when it comes to Natural Language Processing tools, they are trained and adopted to work properly with plain text. Mapping short messaging language words to plain text words can be very sensitive at some cases. A wrong mapping can result in alternations of the meaning or it may destroy semantics under the applied context. When considering the sub phrase ”ur a legend” in above example, ”ur” can be considered as ”your” or ”you are”. Humans can understand that its ”you’re a legend” and not ”your a legend”. But a direct mapping from a language tool would not. Hence it depends on the context which the word is used.

**Chapter: 2**

**Literature Survey**

***AIM: - Semantic-Emotion Neural Network for Emotion Recognition From Text***

**AUTHOR:-** ERDENEBILEG BATBAATAR , KEUN HO RYU , MEIJING LI

**PUBLISHER:-**IEEE , 2019

**INTRODUCTION:-** Emotion recognition will play a promising role in the field of artificial intelligence and human-computer interaction. Various types of techniques are used to detect emotions from a human being like facial expressions, body movements , blood pressure, heartbeat and textual information.

Within the internet , there is enormous amount of textual data and can be used for recognizing the emotion of a particular text.

Improving the previous result and emotion recognition using real world data still remains a huge challenge for several reasons.

Some of the reasons are most machine learning methods rely on features which requires a lot of manual design and adjustment which is time consuming and costly. This paper gives some architecture models which makes it easier to recognize emotions from text.

**CONCLUSION:-** This paper explored an emotion recognition method from text based on the combined network which consists of CNN based emotion encoder and BiLSTM based semantic encoder called SENN, a novel model is proposed and applied on ten real-world datasets.For the SENN model, BiLSTM is designed to capture contextual information and CNN is designed to extract emotional information effectively. Experimental results show that the SENN model outperforms most of the baseline methods and state-of-the-art approaches. Compared with traditional machine learning models, authors proved that deep learning based models outperformed the machine learning models as reported in previous studies. Logistic regression and support vector machine shows the comparative result using bag-of-word and tf-idf vectors. Compared with the state-ofthe-art models in emotion classification, SENN gives the best performance on nine out of ten datasets except Tales-Emotion dataset. It performs F1-scores of 84.8%, 51.1%, 61.3%, 74.6%, 91.0%, 56.3%,59.3%, 98.8% and 70.8% on real-world datasets. And CNN gives the best result on Tales-Emotion dataset using FastText word embedding.

***AIM:-*** ***Classification Model To Determine The Polarity Of Movie Review Using Logistic Regression.***

**AUTHOR:-** Priyanka HS , Ramya BV , Dr. Ashok Kumar

**PUBLISHER:-** International Research Journal Of Computer Science (IRJCS)

**INTRODUCTION:-** Decision making place an important role in human life, a good and correct decision will make our life better. Decisions can made based on others opinions. In case of deciding on which movie to buy or watch, people look at websites, which provides reviews and ratings of the movies. This makes people to decide on which movie to buy or watch. People may have good or bad opinion about the movie, they express there opinion through reviews in the websites like amazon, BookMyShow etc. Where people look at the reviews and ratings to decide which movie to watch or buy. Good opinions about the movies are classified as positive and bad opinions are classified as negative. Review containing “good”, “wonderful”, “enjoyed” like keywords are called as positive. Review containing “not nice”, “bad” like keywords are called as negative. Movie ratings are calculated based on the positive opinion about the movie. Sentiment analysis is a technique used to classify positive and negative opinion of the movie review. Where sentiment analysis is text classification tool, which analyses the text and identifies the polarity of the text.

**CONCLUSION:-** In this paper, feature extraction has been done using bag of words specifically bi-grams, which has powerful impact on determining the polarity of the movie review. The model is then trained using logistic regression machine learning classification algorithm , which is showing the accuracy of 88% .When a new set of reviews are feed into the model the model will predict the polarity of the movie reviews. This can enhance the feature extraction process so that model can have more precise features and also by applying different machine learning classification algorithm, accuracy of the model can be improved.

***AIM:-*** ***Emotion Analysis: A Survey***

**AUTHOR:-** Nida Manzoor Hakak , Mohsin Mohd , Mahira Kirmani , Mudasir mohd

**PUBLISHER:-** Manipal University Jaipur , IEEE

**INTRODUCTION:-**. The emotional analysis forms a fundamental part of the affective computing. “Affect” means emotion and “computing” means to calculate or measure. Affective computing is all that takes to design the devices or systems that process, recognize, interpret and simulate the human affect, thus making it possible to analyze the human and machine interactions. This data can be the text, voice, facial expressions etc. Analyzing the emotions and sentiments of various textual data over the Internet has its own significance, for example, we can measure the well being of a community, it can prevent suicides , and also it can be very helpful for organizations to measure the degree of satisfaction of their customers by analyzing the comments or the feedback they provide The emotion and the sentiment analysis also provide a way for opinion mining for the business organizations: in other words, it can explore the text extracted from e-learning environment and can use that for emotion analysis .

**CONCLUSION:-** Upon reviewing the previous works in the domain of the emotion analysis, it was conclude that much of the work has been done in the field especially in the domain of textual datasets. The experimentation result of some of the works for various computational models with their overall system accuracy was high compared to other models. It was noted that there has been a significant improvement in the system accuracies over the time with the improvement or modification of traditional computational approaches, the lexical resources and the features generated.

***AIM:-*** . ***EmoTxt: A toolkit for emotion recognition from text***

**AUTHOR:-** Fabio Calefato , Filippo Lanubile , Nicole Novielli

**PUBLISHER:-** University of Bari , IEEE

**INTRODUCTION:-** Sentiment analysis is regarded as a crucial task for several application domains, including business, social well-being, politics, security, and software engineering .To date, several off-the-shelf tools are freely available for classifying the sentiment polarity of an input text, that is its positive, negative, or neutral semantic orientation. However, none of them supports the recognition of specific emotions, such as joy, love, and anger. In this paper, they presented EmoTxt, the first open-source toolkit for emotion recognition from text. The toolkit can be used by researchers for detecting emotions from input text as well as for training a custom emotion classifier from scratch, based on manually annotated data. The system is completely developed in Java and distributed under the MIT open-source license 1 . With the toolkit, they released the classification models trained on our gold standard datasets , which can be used for emotion detection from text.

**CONCLUSION:-**

Its training approach leverages a suite of features that are independent of the theoretical model adopted for labeling the data. They provided empirical evidence that EmoTxt achieves comparable performance with different datasets. As future work, they planned to validate it on gold standards from different sources, to further assess the generality and robustness of the approach implemented.

***AIM:- A Corpus for Sentiment Analysis and Emotion Recognition for a Learning Environment***

**AUTHOR:-** Raúl Oramas-Bustillos, Maria Lucia Barron-Estrada, Ramon Zatarain-Cabada, Sandra Lucia Ramírez-Ávila.

**PUBLISHER:-** Technological Institute of Culiacan , Spain

**INTRODUCTION:-** In this paper, authors describe the development of an Educational Resources Assessment System, which is used to capture textual opinions about learning objects in the field of computer programming. This system produces a dataset (corpus), which contains the opinions provided by the students during the learning process when he/she access the educational resources. This corpus is used to train a machine learning classifier to recognize learning-centered emotions in student texts. The recognizer is then incorporated into an Intelligent Learning Environment.

**CONCLUSION:-** The corpus generated with the ERAS system contains 851 textual opinions. The system remains available for more participants to express their opinions on educational resources. This model will helped Intelligent Tutoring Systems to detect emotions through text and make the teaching process more efficient for students, adjusting the content to the particular needs of each of them. [6]

***AIM:-*** ***Emotion Recognition from Text Based on Automatically Generated Rules***

***Level Sensing.***

**AUTHOR:-** Shadi Shaheen , Wassim El-Hajj , Hazem Hajj , Shady Elbassuoni

**PUBLISHER:-** American University of Beirut , Lebanon

**INTRODUCTION:-** The problem of emotion recognition or emotion detection from text to the problem of finding relations between the input sentence and the emotional content within it. Intuitively, finding these relations relies on discovering specific terms (emotional keywords, verbs, nouns, etc.) in the input sentence and other deeper inferences that are related to the meaning of the sentence. Once these terms and their relation to the meaning of the sentence are found, they can be generalized and considered as emotion recognition rules (ERRs). For example, consider the sentence ”I received many gifts on Christmas Eve”; Assuming that this sentence reflects a happy emotion, by analyzing the sentence we can reach to the conclusion that the verb “received” and the noun ”gifts” are the most important parts of the sentence, and consequently we can come up with a rule that says ”receiving gifts” reflects the emotion happy.

**CONCLUSION:-** In this work, authors introduced a new approach for classifying emotions from textual data based on a fine grained level. Their contribution lies in performing complex syntactic and semantic analysis of the sentence and using various ontologies such as Wordnet and ConceptNet in the process of emotion recognition. Syntactic and semantic analysis of the sentence makes their classifier context sensitive, while using Wordent and ConceptNet helps the classifier generalize the training set, which leads to better coverage of emotion rules.

***AIM:- Emotion Detection From Text Documents***

**AUTHOR:-** [SN Shivhare](https://scholar.google.com/citations?user=sQEvrD8AAAAJ&hl=en&oi=sra) , [SK Saritha](https://scholar.google.com/citations?user=gyu7KAgAAAAJ&hl=en&oi=sra)

**PUBLISHER:-** International Journal of Data Mining

**INTRODUCTION:-** Detecting emotional state of a person by analyzing a text document written by him/her appear challenging but also essential many times due to the fact that most of the times textual expressions are not only direct using emotion words but also result from the interpretation of the meaning of concepts and interaction of concepts which are described in the text document. Recognizing the emotion of the text plays a key role in the human-computer interaction [1]. Emotions may be expressed by a person’s speech, face expression and written text known as speech, facial and text based emotion respectively. Sufficient amount of work has been done regarding to speech and facial emotion recognition but text based emotion recognition system still needs attraction of researchers [14]. In computational linguistics, the detection of human emotions in text is becoming increasingly important from an applicative point of view.

**CONCLUSION:-** In this paper, existing research of emotion detection based on textual data is surveyed and limitations of existing methods are reviewed. System architecture is proposed to improve detection capabilities and perform the task efficiently. Proposed system by the authors is based on keyword spotting technique as well as having rich features of ontology. Not all the limitations of existing methods are overcome by this architecture but use of ontology improves the detection capability by applying semantic approach.

This model is based on keyword spotting technique, apart from that it also uses the concept of ontology. Use of ontology makes this model more efficient than other methods in recognizing emotions from text input. This has been created to overcome following limitations:

• Ambiguity in Keyword Definitions: The meanings of keywords could be 4 multiple and vague, as most of the words could change their meanings according to different usages and contexts.

• Incapability of Recognizing Sentences without Keywords: ”I passed my qualify exam today” and ”Hooray! I passed my qualify exam today” should imply the same emotion (joy), but the former sentence without “hooray” could remain undetected if ”hooray” is the only keyword to detect this emotion.

• Lack of Linguistic Information: Syntax structures and semantics also have influences on expressed emotions. For example, ”I laughed at him” and ”He laughed at me” would suggest different emotions from the first person’s perspective. As a result, ignoring linguistic information also poses a problem to keyword-based methods.

• Difficulties in Determining Emotion Indicators Learning-based methods can automatically determine the probabilities between features and emotions but the methods still need keywords in the form of features. The most intuitive features may be emoticons which can be seen as author’s emotion annotations in the texts. [8]

***AIM:-*** ***Text Based Emotion Recognition: A Survey***

**AUTHOR:-** ECC Kao, [CC Liu](https://scholar.google.com/citations?user=2EWeXbAAAAAJ&hl=en&oi=sra), TH Yang, CT Hsieh

**PUBLISHER:-** IEEE

**INTRODUCTION:-** This paper is mainly focused on an overview of emotion detection from text and describes the emotion detection methods. These methods are divided into the following four main categories: keyword-based, Lexical Affinity method, learning based, and hybrid based approach. Limitations of these emotion recognition methods are presented in this paper and also, addresses the text normalization using different handling techniques for both plain text and short messaging language. This approach is easy to implement and intuitive since it involves identifying words to search for in text. These words are classified into categories such as disgust, sadness, happy, anger, fear, surprise etc. Existing system make use of plain text only. This paper describes the different text based emotion recognition methods and their limitations. The problems are faced by the emotion recognition system while processing raw text which contain both plain text and short messaging language. This paper addresses the existing different approaches for resolving processing of 5 raw textual data which contain combination of both plain text and short messaging language.

***AIM:- Sentiment Analysis of Twitter Data***

**AUTHOR:-** Apoorv Agarwal , Boyi Xie , Ilia Vovsha , Owen Rambow , Rebecca Passonneau

**PUBLISHER:-**Columbia University , New York

**INTRODUCTION:-** Microblogging websites have evolved to become a source of varied kind of information. This is due to nature of microblogs on which people post real time messages about their opinions on a variety of topics, discuss current issues, complain, and express positive sentiment for products they use in daily life. In fact, companies manufacturing such products have started to poll these microblogs to get a sense of general sentiment for their product. Many times these companies study user reactions and reply to users on microblogs. One challenge is to build technology to detect and summarize an overall sentiment.

In this paper, authors look at one such popular microblog called Twitter and build models for classifying “tweets” into positive, negative and neutral sentiment. They built models for two classification tasks: a binary task of classifying sentiment into positive and negative classes and a 3-way task of classifying sentiment into positive, negative and neutral classes.

They experimented with three types of models: unigram model, a feature based model and a tree kernel based model. For the feature based model they used some of the features proposed in past literature and propose new features. For the tree kernel based model they designed a new tree representation for tweets. At last they used a unigram model, previously shown to work well for sentiment analysis for Twitter data, as their baseline.

**CONCLUSION:-** They presented results for sentiment analysis on Twitter. Use previously proposed state-of-the-art unigram model as baseline and report an overall gain of over 4% for two classification tasks: a binary, positive versus negative and a 3-way positive versus negative versus neutral. They presented a comprehensive set of experiments for both these tasks on manually annotated data that is a random sample of stream of tweets and investigated two kinds of models: tree kernel and feature based models and demonstrate that both these models outperform the unigram baseline. For feature-based approach, they did feature analysis which reveals that the most important features are those that combine the prior polarity of words and their parts-of-speech tags. They tentatively concluded that sentiment analysis for Twitter data is not that different from sentiment analysis for other genres.

***AIM:-Sentiment Analysis Algorithms And Applications : A Survey***

**AUTHOR:-** [W Medhat](https://scholar.google.com/citations?user=59Wiuq8AAAAJ&hl=en&oi=sra), [A Hassan](https://scholar.google.com/citations?user=uyrs20AAAAAJ&hl=en&oi=sra), H Korashy

**PUBLISHER:-** Ain Shams engineering journal

**INTRODUCTION:-** This survey can be useful for new comer researchers in this field as it covers the most famous SA techniques and applications in one research paper. This survey uniquely gives are fined categorization to the various SA techniques which is not found in other surveys. It discusses also new related fields in SA which have attracted the researchers lately and their corresponding articles. These fields include Emotion Detection (ED), Building Resources (BR) and Transfer Learning (TL). Emotion detection aims to extract and analyze emotions, while the emotions could be explicit or implicit in the sentences. Transfer learning or Cross-Domain classification is concerned with analyzing data from one domain and then using the results in a target domain. Building Resources aims at creating lexica, corpora in which opinion expressions are annotated according to their polarity, and sometimes dictionaries. In this paper, the authors give a closer look on these fields.

**CONCLUSION:-** This survey paper presented an overview on the recent updates in SA algorithms and applications. Fifty-four of the recently published and cited articles were categorized and summarized. These articles give contributions to many SA related fields that use SA techniques for various real-world applications. After analyzing these articles, it is clear that the enhancements of SC and FS algorithms are still an open field for research. Naïve Bayes and Support Vector Machines are the most frequently used ML algorithms for solving SC problem. They are considered a reference model where many proposed algorithms are compared to. The interest in languages other than English in this field is growing as there is still a lack of resources and researches concerning these languages. The most common lexicon source used is WordNet which exists in languages other than English. Building resources, used in SA tasks, is still needed for many natural languages.

***AIM: - Survey on Aspect-Level Sentiment Analysis***

**AUTHOR:-** Kim Schouten and Flavius Frasincar

**PUBLISHER:-** IEEE

**INTRODUCTION:-** The digital age, also referred to as the information society, is characterized by ever growing volumes of information. Driven by the current generation of web applications, the nearly limitless connectivity, and an insatiable desire for sharing information, in particular among younger generations, the volume of user-generated social media content is growing rapidly and likely to increase even more in the near future. People using the Web are constantly invited to share their opinions and preferences with the rest of the world, which has led to an explosion of opinionated blogs, reviews of products and services, and comments on virtually anything. This type of web-based content is more and more recognized as a source of data that has added value for multiple application domains.

**CONCLUSION:-** From the overview of the state-of-the-art in aspect-level sentiment analysis presented in this survey, it was clear that the field is transcending its early stages. While in some cases, a holistic approach was presented that was able to jointly perform aspect detection and sentiment analysis, in others dedicated algorithms for each of those two tasks are provided. Most approaches that are described in this survey are using machine learning to model language, which is not surprising given the fact that language is a non-random, very complex phenomenon for which a lot of data is available. The latter is especially true for unsupervised models, which are very well represented in this survey.

***AIM:- A survey on sentiment analysis challenges***

**AUTHOR:-** Doaa Mohey El-Din Mohamed Hussein

**PUBLISHER:-** Faculty of Computers and Information, Cairo University, Cairo, Egypt

**INTRODUCTION:-** Sentiment analysis uses the natural language processing text analysis and computational techniques to automate the extraction or classification of sentiment from sentiment reviews. Hundreds of thousands of users depend on online sentiment reviews. 90% of customer’s decisions depended on online Reviews in April 2013.The main goal of analyzing sentiment is to analyze the reviews and examine the scores of sentiments. This analysis is divided into many documents level, sentence level, word/term level or aspect level. The sequence processes are of sentiment analysis evaluation and detection of the sentiment polarity.

The evaluation sentiment drawbacks that Reflected in language coverage. This paper summarizes keys of sentiment challenges with respect to the type of review structure. It also divides the challenges into two types to ease to deal with them and focus on the degree of accurate meaning. This research discusses these sentiment challenges, the factors affecting them, and their importance.

As a result a large number of studies and research have helped monitor the trending new research increasing year by year. The focus in this research, has been to achieve the most suitable challenges facing sentiment evaluation to be useful for researchers and facilitate their relationships.

**CONCLUSION:-** This survey discusses the importance and effects of sentiment analysis challenges in sentiment evaluation based on two com parisons among forty-seven papers. The first comparison is based on the relationship between the sentiment review structure and sentiment analysis challenges. The result of this comparison reveals another essential factor to recognize the sentiment challenges which is domain dependence. Moreover, the negation challenge became popular in all types of reviews structured just differs in implicit or explicit meaning

***AIM:- Music Sentiment and Stock Returns Around the World***

**AUTHOR:-** Alex Edmans, Adrian Fernandez-Perez, Alexandre Garel, Ivan Indriawan

**PUBLISHER:-** Journal of Financial Economics (JFE), Forthcoming

**INTRODUCTION:-** This paper introduces a real-time, continuous measure of national sentiment that is language-free and thus comparable globally: the positivity of songs that individuals choose to listen to. This is a direct measure of mood that does not pre-specify certain mood-affecting events nor assume the extent of their impact on investors. We validate our music-based sentiment measure by correlating it with mood swings induced by seasonal factors, weather conditions, and COVID-related restrictions. We find that music sentiment is positively correlated with same-week equity market returns and negatively correlated with next-week returns, consistent with sentiment-induced temporary mispricing. Results also hold under a daily analysis and are stronger when trading restrictions limit arbitrage. Music sentiment also predicts increases in net mutual fund flows, and absolute sentiment precedes a rise in stock market volatility. It is negatively associated with government bond returns, consistent with a flight to safety.

**CONCLUSION:-**This study introduces a novel measure of investor sentiment, which captures actual sentiment rather than shocks to sentiment. Our main result is a positive and significant relation between music sentiment and contemporaneous market returns, controlling for world market returns, seasonality, and macroeconomic variables. We also find a significant price reversal the following week. Taken together, our findings are consistent with sentiment-induced temporary mispricing that subsequently reverses. We show that the relationship between music sentiment and market returns is stronger when countries implemented trading restrictions such as short-sale bans during the COVID-19 pandemic, consistent with greater limits to arbitrage. Music sentiment also predicts increases in net mutual fund flows and decreases in government bond returns, and absolute sentiment precedes a rise in stock market volatility. Overall, our study provides evidence that a proxy for the actual sentiment of a country’s citizens is significantly correlated with asset prices.

***AIM:-*** ***Sentiment analysis: A combined approach***

**AUTHOR:-** Rudy Prabowo , Mike Thelwall

**PUBLISHER:-** School of Computing and Information Technology, University of Wolverhampton Wulfruna Street

**INTRODUCTION:-** The sentiment found within comments, feedback or critiques provide useful indicators for many different purposes. These sentiments can be categorised either into two categories: positive and negative or into an n-point scale, e.g. very good, good, satisfactory, bad, very bad. In this respect, a sentiment analysis task can be interpreted as a classification task where each category represents a sentiment. Sentiment analysis provides companies with a means to estimate the extent of product acceptance and to determine strategies to improve product quality. It also facilitates policy makers or politicians to analyse public sentiments with respect to policies, public services or political issues. This paper presents the empirical results of a comparative study that evaluates the effectiveness of different classifiers, and shows that the use of multiple classifiers in a hybrid manner can improve the effectiveness of sentiment analysis. The procedure is that if one classifier fails to classify a document, the classifier will pass the document onto the next classifier, until the document is classified or no other classifier exists.

**CONCLUSION:-** The use of multiple classifiers in a hybrid manner can result in a better effectiveness in terms of micro and macro-averaged F1 than any individual classifier. By using a Sentiment Analysis Tool, we can apply a semi-automatic, complementary approach. The induction algorithm can generate a set of induced antecedents that are too sparse for a deeper analysis. Therefore, in a real-world scenario, it is desirable to have two rule sets, one is the original set, and another one is the induced rule set.

***AIM:- Sentiment analysis: A review and comparative analysis of web services***

**AUTHOR:-** Jesus Serrano-Guerrero , Jose A. Olivas a , Francisco P. Romero a , Enrique Herrera-Viedma

**PUBLISHER:-** Department of Computer Science and Artificial Intelligence, University of Granada, 18071 Granada, Spain

**INTRODUCTION**:- Sentiment Analysis, also called Opinion Mining, is one of the most recent research topics within the field of Information Processing. Textual information retrieval techniques are mainly focused on processing, searching or mining factual information. Facts have an objective component; however, there are other textual elements which express subjective characteristics. These elements are mainly opinions, sentiments, appraisals, attitudes, and emotions, which are the focus of Sentiment Analysis. All of them are closely related, however, they present slight differences. This fact involves the birth of many related tasks in this new research field, such as opinion mining, subjectivity analysis, emotion detection or opinion spam detection, among others. Sentiment Analysis offers many opportunities to develop new applications, especially due to the huge growth of available information in sources such as blogs and social networks. For example, recommendations of items proposed by any recommender system can be computed taking into account aspects such as positive or negative opinions about those items. Review- and opinion-aggregation websites could collect information from different sources in order to summary or compose an opinion about a candidate, product, etc., thus replacing systems which require explicitly opinions or summaries. Question answering systems represent another field where opinions play an important role. Detection of opinion-oriented questions and possible answers, and its treatment are essential to compute good answers. Detection of subjective information is really important in fields related to argumentation where objective sentences are usually more valuable. But certainly, one of the most important fields where Sentiment Analysis has a greater impact is in the industrial field.

**CONCLUSION**:- This work presents a detailed review of 15 web services which include functionalities related to Sentiment Analysis. Some of these services belong to private companies, but even so, they allow restricted free access to their functionalities, and the others are totally free services. This fact is interesting to those users who desire to include Sentiment Analysis capabilities within their own platforms without having to develop their own algorithms; hence, these tools are especially interesting for researching purposes and rapid prototyping. Besides, due to the fact that the selected services can work as web services, the inclusion of them into any platform is really easy.

**Chapter: 3**

**Methodology And Process Flow**

Emotion classification can be divided into two different categories:

1.coarse-grained level (positive or negative) which can be accurately perceived from text.

2.fine grained level(the six Ekman emotions) it requires semantic and syntactic analysis of the sentence and can be done using three methods viz.,Keyword-based detection,Learning-based detection, and Hybrid detection.

1. Keyword-based detection : classifying emotions is done by searching for the emotional keywords in the input sentence .Such methods suffer from the ambiguity in the keyword definitions in the sense that a word can have different meanings according to usage and context, the incapability of recognizing emotions within sentences that do not contain emotional keywords.
2. Learning-based detection: In these methods, the emotion is detected by using classification approaches based on a training dataset.
3. Hybrid detection: In hybrid methods, emotions are detected by using a combination of emotional keywords and learning patterns collected from training datasets, in addition to information from different sciences, like human psychology, system starts by looking for emotional abbreviations and emoticons.

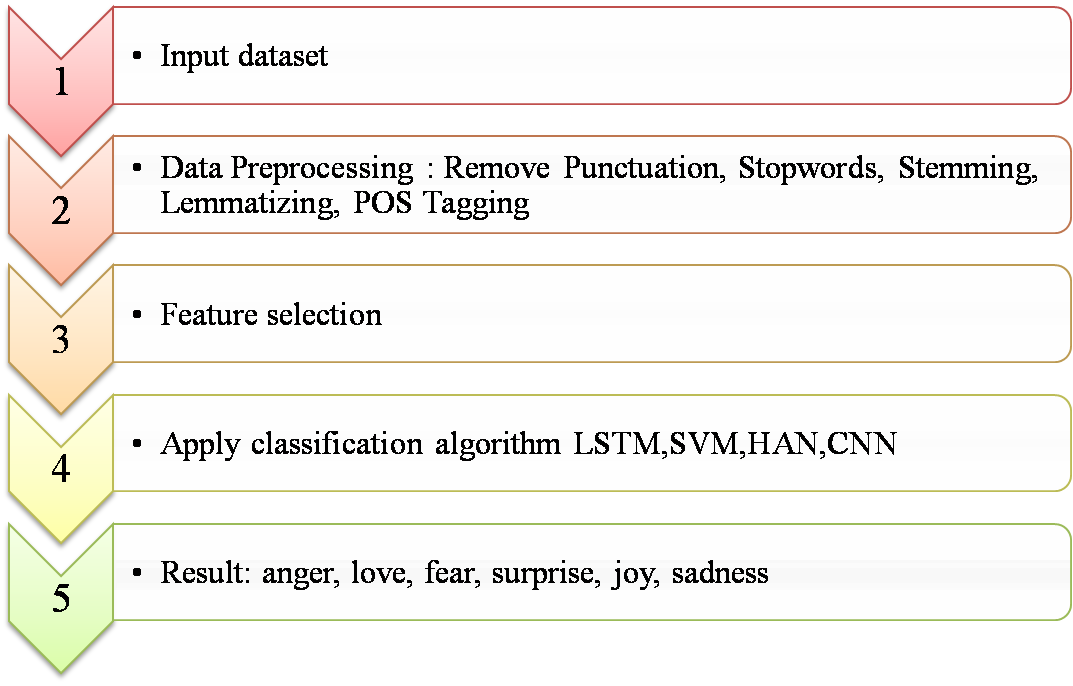


Figure 3.1: process flow

**3.1 Implementation Process flow:**

## 3.1.1 Step 1: Setting up the Development Environment

We will be using Spyder IDE that comes along with the anaconda installation to do all our programming.

## 3.1.2 Step 2: Choosing Your Dataset

We first need data. we can get textual data from any website like a movie review website, or Amazon product reviews, and so on. Here, we’ll be using a labelled textual dataset. The dataset we are using has 416809 tweets in total, labelled into different human sentiments.We import the basic libraries and then read the dataset.

If we want to perform emotion analysis on real-time tweets we have connect our twitter API to python program.In order to use Twitter’s API, we have to create a developer account on the Twitter apps site as per following steps.

1. Log in or make a Twitter account at https://developer.twitter.com
2. Create a new app.
3. Fill in the app creation page with a unique name, a website name, and a project description.
4. Once the application is created , open the ’Keys and Access Token’ tab.
5. Copy ‘Consumer Key’, ‘Consumer Secret’, ‘Access token’ and ‘Access Token

Secret’.

1. Establish the connection with Twitter API.

## 3.1.3 Step 3: Preprocessing the Data

First, we have to bring some uniformity to the text by making everything lowercase, removing punctuation, and stop words (like prepositions), using tokenization, stemming , finding synonyms and antonyms lemmatizing and by speech tagging.

* Making all letters lowercase:
* Removing Punctuation, Symbols:
* Lemmatisation:

To gain any proper insight, we need to get all the words to their root form, i.e the variants of a word within the text (for example plural forms, past tense, etc) must all be converted to the base word it represents. This is called lemmatisation. Along with that, we have added code to revert repetition of letters in a word with the assumption that hardly any word has letters repeated more than twice, consecutively. Though not very accurate, it can help in some corrections.

* Stemming stemming involves removing affixes from words and returning the root. Search engines like Google use this to efficiently index pages. The most common algorithm for stemming is the PorterStemmer. Let’s take an example.

*stemmer.stem(’writes’)*

‘write’ *stemmer.stem(’writing’)*

‘write’ *stemmer.stem(’write’)*

‘write’

* Finding antonyms and Synonyms:

WordNet is an NLP database with synonyms, antonyms, and brief definitions. We have downloaded this with the NLTK downloader.

* Finding Stop words:

We can filter NLTK stop words from text before processing it.

example: from nltk.corpus import stopwords text=”Today is a great day. It is even better than yesterday. And yesterday was the best day ever!”

output: [‘Today’, ‘great’, ‘day’, ‘.’, ‘It’, ‘even’, ‘better’, ‘yesterday’, ‘.’, ‘And’, ‘yesterday’, ‘best’, ‘day’, ‘ever’, ‘!’]

* Tokenizing Text:

Before processing the text in NLTK Python Tutorial, you should tokenize it. What we mean is you should split it into smaller parts- paragraphs to sentences, sentences to words. We have two kinds of tokenizers- for sentences and for words.for example,

\*Sentence Tokenizer text=”Today is a great day. It is even better than yesterday. And yesterday was the best day ever.” *from nltk.tokenize import sent tokenize sent tokenize(text)*

[‘Today is a great day.’, ‘It is even better than yesterday.’, ‘And yesterday was the best day ever.’] \*Word Tokenizer *nltk.word tokenize(text)*

[‘Today’, ‘is’, ‘a’, ‘great’, ‘day’, ‘.’, ‘It’, ‘is’, ‘even’, ‘better’, ‘than’, ‘yesterday’,

‘.’,‘And’, ‘yesterday’, ‘was’, ‘the’, ‘best’, ‘day’, ‘ever’,‘.’]

• Speech Tagging

NLTK can classify words as verbs, nouns, adjectives, and more into one of the following classes:

1. CC coordinating conjunction
2. CD cardinal digit
3. DT determiner
4. EX existential there
5. FW foreign word
6. IN preposition/subordinating conjunction
7. JJ adjective ‘big’
8. JJR adjective, comparative ‘bigger’
9. JJS adjective, superlative ‘biggest’
10. LS list marker 1)
11. MD modal could, will
12. NN noun, singular ‘desk’
13. NNS noun plural ‘desks’
14. NNP proper noun, singular ‘Harrison’
15. NNPS proper noun, plural ‘Americans’
16. PDT predeterminer ‘all the kids’ 17. POS possessive ending parent‘s
17. PRP personal pronoun I, he, she
18. PRP$ possessive pronoun my, his, hers
19. RB adverb very, silently,
20. RBR adverb, comparative better
21. RBS adverb, superlative best
22. RP particle give up
23. TO to go ‘to‘ the store.
24. UH interjection
25. VB verb, base form take
26. VBD verb, past tense took
27. VBG verb, gerund/present participle taking
28. VBN verb, past participle taken
29. VBP verb, sing. present, non-3d take
30. VBZ verb, 3rd person sing. present takes
31. WDT wh-determiner which
32. WP wh-pronoun who, what
33. WP$ possessive wh-pronoun whose
34. WRB wh-abverb where, when

For example, text=’I am a human being, capable of doing terrible things’ sentences=nltk.sent tokenize(text) for sent in sentences:

print(nltk.pos tag(nltk.word tokenize(sent)))

[(‘I’, ‘PRP’), (‘am’, ‘VBP’),(‘a’, ‘DT’), (‘human’, ‘JJ’), (‘being’, ‘VBG’), (‘,’,

‘,’),(‘capable’, ‘JJ’), (‘of’, ‘IN’), (‘doing’, ‘VBG’), (‘terrible’, ‘JJ’), (‘things’, ‘NNS’)]

## 3.1.4 Step 4: Feature Extraction

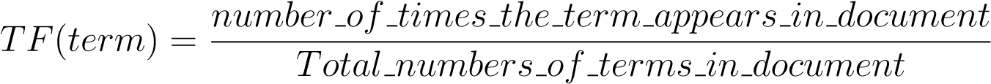
Once you make the text data clean, precise, and error-free, each sentence is represented by a group of keywords. Now, we need to perform ‘Feature Extraction’, i.e extracting some parameters from the data that can be presented numerically. We consider two different features, TF-IDF and Count Vectors.

Split the data into training and testing parts before performing feature extraction. **Term Frequency-Inverse Document Frequency (TF-IDF):**

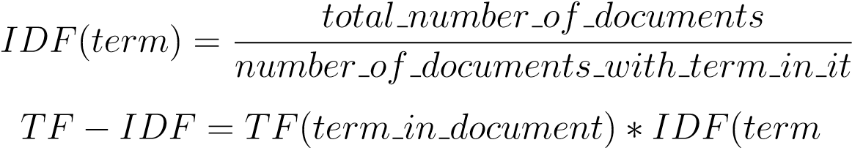
TF-IDF an acronym than stands for “Term Frequency – Inverse Document Frequency”, is used for Word vectorization i.e., general process of turning a collection of text documents into numerical feature vectors. TF-IDF are word frequency scores that try to highlight words that are more interesting, e.g. frequent in a document but not across documents. This parameter gives the relative importance of a term in the data and is a measure of how frequently and rarely it appears in the text.

Term Frequency: This summarizes how often a given word appears within a document.

Inverse Document Frequency: This down scales words that appear a lot across documents.

 (3.1)

(3.2)

 ) (3.3)

**Count Vectors:**

This is another feature we consider and as the name suggests we transform our tweet into an array having the count of appearances of each word in it. The intuition here is that the text that conveys similar emotions may have the same words repeated over and over again.

## 3.1.5 Step 5: Training Our Models

With the numerical representations of sentences ready, we can directly use them as inputs for some classic machine learning models. Here, we trained our learning model using 4 different algorithms such as Long Short-Term Memory (LSTM), Convolutional Neural Network(CNN), Support vector machine(SVM) and Hierarchical Attention Network (HAN) . These methods can, in fact, be used for tackling any kind of classification problem. In our case, we want to classify if a given tweet into one of the 6 emotions. For this we are using a dataset which contains 416809 tweets to train our model.

**3.1.6 Step 6 :Testing**

Now test how it performs in reality by giving this model some random text input.

**Chapter: 4**

**Implementation till now**

**4.1 Context on which the model is in creation**

Data and information is one of the key to gaining more knowledge about any field , and in this generation we can access data from anywhere in the world via internet.

For example , political data can be accessed from many different sources such as social media, news channels and newspapers and then we can figure out for ourselves what people must be feeling about a particular political event or any individual.

But what if we don’t need to scan through all the data to come to a conclusion regarding sentiment of people over any event , what if we can implement a model which does this for us.

This is what our team is modeling , we are creating a deep learning model which will show what is the emotion of public with respect to any event in context of politics .

This model can be used during elections to see what people feel about any politician , party or any event occurring in real time.

**4.1 Tool and Technologies used**

**Programming Language** – We are using Python as programming language because it offers many libraries which we can use for fetching , data cleaning , tabulating , storing , plotting and many different purposes which we will be needing to build the model.

**Data Source** – We need huge amount of data for training the model to recognize emotions so we will be using Twitter API which will give us access to millions of tweets ,then these tweets will be used for training the model to recognize emotions.

**4.1 Steps performed for implementation**

**Step 1 –** Creating twitter developers account and getting the api key , api secret key and access to tweets .

Twitter Developer account can be created with steps as follows :

1. Go to <https://developer.twitter.com/en>
2. Create a twitter account and if you already have one , just click on sign up.
3. After that give a brief of how are going to use the tweets , select the package which is suitable to your model and you must have access to at least 500k tweets.

**Step 2-** Fetching the tweets in our code and storing it in tabular form in .csv file.

Used tweepy library for authentication of twitter developer account and fetching the data.

The important columns which we need were filtered out and stored in .csv file , some of the columns were , the tweet , emojis , hashtags etc.

**Step 3** – Data Cleaning to collect useful information which will be used in training the model.

Tweets from the .csv file is used to clean the data and add subjectivity and polarity.

**References**

1. Chetan R. Chopade. Text based emotion recognition: A survey. *International Journal of Science and Research (IJSR)*, 2013.
2. M. Li E. Batbaatar and K. H. Ryu. Semantic-emotion neural network for emotion recognition from text. *IEEE Access*, 2019.
3. F. Lanubile F. Calefato and N. Novielli. In *EmoTxt: A toolkit for emotion recognition from text*, 2017.
4. Mahira Kirmani Mudasir mohd Nida Manzoor Hakak, Mohsin Mohd. *Emotion Analysis: A Survey*. International Conference on Computer, Communications and Electronics

(Comptelix)Manipal University Jaipur, Malaviya National Institute of Technology Jaipur and IRISWORLD, 2017.

1. Dr. Ashok Kumar R Priyanka H S, Ramya B V. Classification model to determine the polarity of movie review using logistic regression. *International Research Journal of Computer Science (IRJCS)*, 2019.
2. R. Zatarain-Cabada R. Oramas-Bustillos, M. L. Barron-Estrada and S. L. Ram´ırez-Avila. A corpus for sentiment analysis and´ emotion recognition for a learning environment. 2018.
3. H. Hajj S. Shaheen, W. El-Hajj and S. Elbassuoni. Emotion recognition from text based on automatically generated rules. *IEEE International Conference on Data Mining Workshop, 2014*, 2014.

S. N. Shivhare and S. K. Saritha. Emotion detection from text documents. In *International Journal of Data Mining and Knowledge Management Process*, 2014.

[16] P. Rodriguez, A. Ortigosa, and R. M. Carro, ‘‘Extracting emotions from texts in e-learning environments,’’ in Proc. 6th Int. Conf. Complex, Intell., Softw. Intensive Syst., Jul. 2012, pp. 887–892

[17] A. Jamshidnejad and A. Jamshidined, ‘‘Facial emotion recognition for human computer interaction using a fuzzy model in the e-business,’’ in Proc. Innov. Technol. Intell. Syst. Ind. Appl. (CITISIA), Jul. 2009, pp. 202–204.

[18] Rosalind W. Picard-Affective computing: MIT Press Cambridge, MA, USAc, (1997).

[19] A Calvo and Senior Member-Affect Detection: An Interdisciplinary Review of Models, Methods, and Their Applications. IEEE Transactions on Affective Computing, 1(1):18–37, (2010).

[20] Bart Desmet and V´eronique Hoste- Emotion detection in suicide notes. Expert Systems with Applications, 40(16):6351–6358, November, (2013).

[21] P. Basile, V. Basile, M. Nissim, N. Novielli, V. Patti. “Sentiment Analysis of Microblogging Data”. To appear in Encyclopedia of Social Network Analysis and Mining, Springer. In press.

[22] M. Thelwall, K. Buckley, G. Paltoglou. “Sentiment strength detection for the social web”. J Am Soc Inf Sci Technol 63(1):163–173, 2012

[23] Zahra Khalili, and Mohammad Hasan Moradi. "Emotion recognition system using brain and peripheral signals: using correlation dimension to improve the results of EEG." Neural Networks, 2009. IJCNN 2009. International Joint Conference on. IEEE, 2009.

[24] Jeffrey F. Cohn, and Gary S. Katz. "Bimodal expression of emotion by face and voice." Proceedings of the sixth ACM international conference on Multimedia: Face/gesture recognition and their applications. ACM, 1998.

[25] Liyanage C. De Silva, and Pei Chi Ng. "Bimodal emotion recognition." Automatic Face and Gesture Recognition, 2000. Proceedings. Fourth IEEE International Conference on. IEEE, 2000.

[26] Torao Yanaru, "An emotion processing system based on fuzzy inference and subjective observations." Artificial Neural Networks and Expert Systems, 1995. Proceedings., Second New Zealand International Two-Stream Conference on. IEEE, 1995

[27] R. Cowie, E. Douglas-Cowie, N. Tsapatsoulis, G. Votsis, S. Kollias, “Emotion recognition in humancomputer interaction,” in IEEE Signal Processing Magazine, vol. 18(1), Jan. 2001, pp. 32-80, doi: 10.1109/79.911197

[28] Nicu Sebea, Ira Cohenb, Theo Geversa, and Thomas S. Huangc “Multimodal Approaches for Emotion Recognition: A Survey”, USA

[29] S. D’Mello, T. Jackson, S. Craig, B. Morgan, P. Chipman, H. White,... & A. Graesser, “AutoTutor detects and responds to learners affective and cognitive states”, in Workshop on emotional and cognitive issues at the international conference on intelligent tutoring systems, pp. 306-308, 2008. [30] P. Ekman, “An argument for basic emotions”, Cognition and Emotion, vol. 6, pp. 169-200, 1992. [31] R. S. Baker, S. K. D'Mello, M. M. T. Rodrigo, & A. C. Graesser, “Better to be frustrated than bored: The incidence, persistence, and impact of learners' cognitive affective states during interactions with three different computer-based learning environments”, Int. J. Hum.-Comput. Stud., 68(4), pp. 223-241, 2010. [4] S. D’Mello, & A. Graesser, “Dynamics of affective states during complex learning”, Learning and Instruction, 22(2), pp. 145-157, 2012

[32] Mikalai Tsytsarau, Themis Palpanas Survey on mining subjective data on the web ,Data Min Knowl Discov, 24 (2012), pp. 478-514

[33] B. Liu, Sentiment analysis and subjectivity, Handbook Nat. Lang. Process. 5 (1) (2010) 1–38

[34] M. McGlohon, N. Glance, Z. Reiter, Star quality: aggregating reviews to rank products and merchants, in: Proceedings of Fourth International Conference on Weblogs and Social Media (ICWSM), 2010, pp. 114–121.

[35] A. Tumasjan, T.O. Sprenger, P.G. Sandner, M. Isabell Welpe, Predicting elections with twitter: What 140 characters reveal about political sentiment, in: Proceedings of International Conference on Weblogs and Social Media (ICWSM-2010), 2010, pp. 178–185.

[36] B. Chen, L. Zhu, D. Kifer, D. Lee, What is an opinion about? Exploring political standpoints using opinion scoring model, in: Proceedings of AAAI Conference on Artificial Intelligence (AAAI-2010), 2010, pp. 1007–1012

[37]Baker, S. R., Bloom, N., Davis, S. J., 2016. Measuring economic policy uncertainty. Q. J. Econ.131, 1593–1636.

[38]Ben-Rephael, A., Kandel, S., Wohl, A., 2011. The price pressure of aggregate mutual fund flows. J. Financial Quant. Anal. 46, 585–603.

[39] B. Bickart and R. M. Schindler, “Internet forums as influential sources of consumer information,” J. Interactive Marketing, vol. 15, no. 3, pp. 31–40, 2001.

[40] E. van Kleef, H. C. M. van Trijp, and P. Luning, “Consumer research in the early Stages of new product development: A critical review of methods and techniques,” Food Quality Preference, vol. 16, no. 3, pp. 181–201, 2005. SCHOUTEN AND FRASINCAR: SURVEY ON ASPECT-LEVEL SENTIMENT ANALYSIS 827

[41] B. Pang and L. Lee, “Opinion mining and sentiment analysis,” Found. Trends Inf. Retrieval, vol. 2, no. 1-2, pp. 1–135, 2008.