Real-Time Sentiment Analysis

Anand Sinha[1], Anubhav Bansal[1], Himani Gulati[1], Ms. Upasna Joshi[2]

[1]*Student, Department of Computer Science and Engineering, Delhi Technical Campus, Greater Noida, U.P., India*

[2]*Professor, Department of Computer Science and Engineering, Delhi Technical Campus, Greater Noida, U.P., India*

A. Abstract

Natural Language Processing (NLP) is a computerized method to text interpretation that is founded on a set of ideas as well as a set of technology. This is a very active field of study and development, there is no universally accepted definition that satisfies everyone, but there are several elements that any intelligent person would consider a person's interpretation.

Sentiment analysis or opinion mining is one of the important aspects of NLP (Natural Language Processing). In recent years, sentiment analysis has received a lot of attention. Sentiment analysis has a wide range of applications, including social media analytics, which simply means creating opinions for individuals on social media by analyzing their sentiments or ideas, which they express through text. It is also an example of how you can review customer feedback and responses. Thus identify the negative comments and reasons why the customers have issues with your product or service. Sentiment analysis enables you to respond to matters promptly before the customer leaves you altogether! In this paper, we aim to tackle the problem of Real-Time sentiment analysis, which is one of the dominant applications of sentiment analysis. Real-time sentiment analysis is **an AI-powered solution** to analyze the input provided by the user with the help of a pre-trained Machine Learning model. We have used a dataset of 1600000 tweets with 6 different attributes to build our model with the help of machine learning algorithms such as Naïve Bayesian classifier, Neural networks, and VADER. This paper is about the techniques and machine learning algorithms used to analyze the effectiveness of such models.

Keywords: *Sentiment Analysis; Machine Learning Algorithms; Naïve Bayesian classifier; Neutral Networks; VADER.*

1. INTRODUCTION

In today's world, smartphones have become a foremost part of our lives. Whether we are talking about our health, finance, and even education in these difficult times, smartphones help us in everything. In short, everything we want is in our palms. Yet, there is still one thing our smartphones cannot do. Tell us, how are we feeling right now? “Hey Siri, how am I doing today?”, “Ok, Google. Tell me, what are my emotions today?” Your phone won't be able to answer that. Consider this sentence, "I am quite happy today!" If we ask someone to rate this sentence, that person will rate this a "10" on the happiness scale. If we remove the word "quite" from the above statement, that person will not rate it a complete "10". Instead, he will give it something near "7"."I hated my day." Whoever said this feels negative about their day and will give a "0" on the happiness scale. Imagine if machine learning could perceive your emotions for you. Sentiment Analysis is a process of collecting and analysing data based upon the person's feelings, reviews, and thoughts. It is often known as opinion mining as it mines the dominant features from people's opinions. Various ML

Statistical models are used to carry out sentiment analysis, along with Natural language processing (NLP) extracting features from big datasets to study the subjective information in each expression. Sentiment Analysis has various applications, including social media analytics which in simpler terms is the ability to gather and find meaning in data gathered from social channels to support business decisions and measure the performance of actions based on those decisions through social media. Another application of Sentiment Analysis is to monitor customer reviews, by identifying negative comments and figuring out the reasons as to why these customers had issues with your products or services. Sentiment analysis enables you to resolve these issues promptly before the customer leaves you altogether! Real-time sentiment analysis is an AI-powered solution to analyse sentiments attached to the data provided by a user with the help of a Machine Learning Model. We aim to visualize the opinion mined with the help of visualizing libraries and make analysis easier on large sequential data.

2. LITERATURE SURVEY

A lot of researchers from the past few decades are using different machine learning algorithms and datasets so we can perform Real-time Sentiment Analysis. It is an ongoing process from many years and different accuracy has been attained by different researchers with the help of their models. Following are the researches.

Pang, L. Lee, S. Vaidyanathan ***et al*** [8] were the first to begin work on sentimental analysis. Their prior aim was to classify a text as an overall sentiment. e.g., Classifying a movie review to either a positive or negative sentiment. They applied machine learning algorithms on review databases which resulted in the algorithms that they use are Naïve-Bayes, maximum entropy, and support vector machines. They also concluded by examining various factors that classification of sentiment is very challenging. They show supervised machine learning algorithms are the base for sentimental analysis.
Wang, D. Can, F. Bar, S. Narayana *et al* [11]
They were the researchers who proposed a system for real-time analysis of public responses for the 2012 presidential elections in the U.S. They collected the responses from Twitter. People's responses on Twitter for election candidates in the U.S. created a large amount of data, which helps to create a sentiment for each candidate and also creates a prediction of who is winning.
A relation is created between sentiments that arise from people's responses on Twitter with the complete election events. They also explore how sentiment analysis affects these public events. They also show

that live sentiment analysis is very fast as compared to traditional content analysis which takes many days or up to some weeks to complete. The system they demonstrated analyses the sentiment of entire Twitter data about the election, candidates, promotions, etc., and delivers results at a continuous rate. It offers media, politicians, and researchers a new way that is time effective which is completely based on public opinion.

O. Almatrafi, S. Parack, B. Chavan *et al* [5] They are the specialists who proposed a framework in view of thearea. As per them, Sentiment Analysis is done by Natural

Language Processing (NLP) and a machine-learning algorithm to separate a slant from a content unit that is from a specific area. They ponder different uses of area-based sentiment analysis by utilizing an information source in which information can be separated from various areas effectively.

Sun, V. Ng, *et al* [9] Numerous endeavors have been done to accumulate data from informal organizations to perform sentiment analysis on web clients. Their point is to demonstrate how wistful investigation impacts informal community posts and they likewise analyze the outcome on different themes on various online networking stages. An expansive measure of information is created each day; individuals are likewise exceptionally inquisitive in finding other comparable individuals among them. Many scientists measure the impact of any post through the quantity preferences and answers it got, however, they don't know whether the impact is certain or negative on other posts. In their exploration, a few inquiries are raised and new systems are set up for a wistful impact of a post.

A Kumar, D Gupta, *et al* [29] They worked on Sentiment Analysis of Twitter User Data on Punjab Legislative Assembly Election. The number of people who share their opinions on social media has expanded considerably in the last decade, the opportunity to use these data for good has increased as well. It's important to know what others think. Initially, Websites and social networking have grown exponentially during the last decade. Twitter, Facebook, Tumbler, and other social media platforms have exploded in popularity. Which is the most popular website in the world? Twitter is used all around the world. According to Twitter's statistics, it has Around 6000 tweets per second have been recorded.350000 tweets are sent every minute, and roughly [29].

There are around 200 billion tweets per day and 500 million tweets per day. Individuals may use Twitter to express themselves.

It's a difficult task to deal with a large dataset, but with NLTK and Text Blob, they easily categorized their data and offered more accurate results using severalclassifiers.

3. REQUIREMENT ANALYSIS

A. Python

This is an interpreted high-level general-purpose programming language that emphasizes code readability with its use of significant indentation. Its language elements and object-oriented approach are designed to help programmers in writing clear, logical code for both small and large-scale projects. Python is garbage-collected and conditionally typed. It supports a myriad of programming paradigms, including structured (especially procedural) programming, object-oriented programming, and functional programming. And for its enormous standard library, it’s usually said as a "batteries included" language. Guido van Rossum collaborated on Python within the late 1980s as a replacement for the ABC Programming Language, and Python 0.9.0 was launched in 1991. [33] Python 2.0 was launched in 2000, and it included new capabilities including list comprehensions and a garbage pickup mechanism that detects loops (in addition to reference counting). Python
3.0 debuted in 2008, and it absolutely was a fundamental shift of the language that wasn’t backwards compatible. Python 2 was deprecated in 2020, with version 2.7.18.

B. Deep Learning

Deep learning is a subset of machine learning except that it’s a layered structure designed to emulate a personality’s brain’s neural network. These neural networks aim to enable the model to "learn" from enormous quantities of knowledge, although they fall well wanting its capabilities. A single-layer neural network produces approximate predictions, but more hidden layers assist the model to optimize and improve for accuracy. Many computer science (AI) apps and services depend upon deep learning to extend automation by executing analytical and physical activities without the necessity for human participation. Everyday goods and services (such as digital assistants, voice-enabled TV remotes, and master card fraud detection) moreover as upcoming innovations use deep learning technology (such as self-driving cars).

C. Neural Networks (RNN)

In a neural network as the name suggests, we have an artificial neural network wherein connections between nodes do not form a cycle. These neural networks are thecore of our deep learning algorithms. These deep learning algorithms are often employed for ordinal or temporal issues like language translation, natural language processing (NLP), speech recognition, and picture captioning, and they're utilized in popular apps like Siri, voice search, and Google Translate. Recurrent neural networks, like feedforward and convolutional neural networks (CNNs), learn from training input.
They are characterized by their "memory," which allowsthem to impact current input and output by using knowledge from previous inputs.
While typical deep neural networks presume that inputsand outputs are independent of one another, recurrent neural networks' output is reliant on the sequence's priorcomponents. While future occurrences may be useful indefining a sequence's output, unidirectional recurrent neural networks cannot account for them in their predictions.
Recurrent networks are further distinguished by the fact that their parameters are shared across all layers of the network. While each node in a feedforward network has avariable weight, each layer of a recurrent

neural network has the same weight parameter. To allow reinforcement learning, these weights are still modified through the processes of backpropagation and gradient descent.

D. Machine Learning

Machine learning automates analytical model building by automating data analysis. It is a branch of artificial intelligence based on the idea that systems can learn from data, identify patterns, and make decisions without human intervention. The same dynamics that have made data mining and Bayesian analysis more popular than ever are driving renewed interest in machine learning. Things include increasing data quantities and variety, cheaper and more powerful computing processing, and economical data storage. All of this means that models that can evaluate more, more complicated data and offer faster, more accurate answers – even on a massive scale
– can be created rapidly and automatically. An organization's chances of recognizing profitable possibilities – or avoiding unforeseen hazards – are improved by developing detailed models.

4. PROPOSED METHODOLOGY

To achieve our objective mentioned above, we haveused the following methodology:

- A thorough study of existing approaches and techniquesin the field of sentiment analysis.
- Collection of Sequential Data from Twitter, Facebook,and Amazon.
- Performing feature engineering of collected data.
- Building a classifier that is based on differentsupervised ML techniques.
- Training and testing on built models usingpre-processed data.
- Evaluating and Hyperparameter Tuning on Classifiers.
- Computing the results of each classifier.
- Saving and shifting model to web-based UI.
- Further:
- Analyze the input data from the user.
- Web Scrape thread data from Twitter to mine real-timeopinion and analyze the sentiment trend.

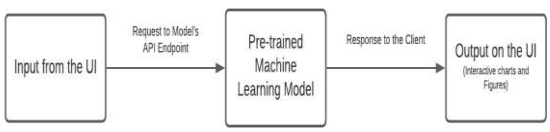

Fig.1 0-Level Data Flow Diagram

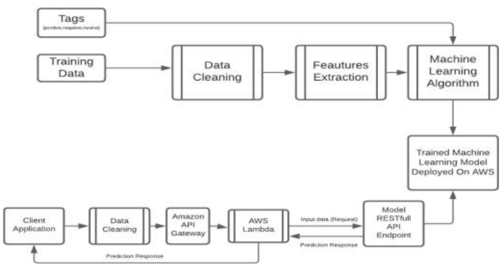

Fig.2 1-Level Data Flow Diagram

5. ALGORITHMS AND TECHNIQUES USED

A. Decision Tree

Decision Trees are a famous data mining techniquewhich uses a tree-shape structure to produce results based on inputs. It works as a flowchart where the inner blocks are

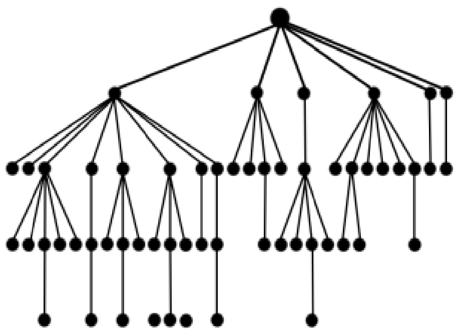

Fig. Decision Tree

the attributes from the dataset and the outer nodes represent the outcome. This type of method has the capability of managing non-uniform as well as missing data. Also, the classifications are processed without much computation. It maps out the possible outcomes from different binary choices. Typically, it starts with a single block, which leads into possible outputs. Each of the blocks leads to additional blocks, which jumps to other possibilities.

B. Text Blob.

Text Blob may be a python library for Natural Language Processing (NLP). Text Blob uses NLTK to accomplish its tasks.
NLTK gives easy accessibility to lots of lexical resources and allows users to figure with categorization, classification, and plenty of other tasks. Text Blob may be a simple library that supports complex analysis and operations on textual data.
For lexicon-based approaches, a sentiment is defined by its semantic orientation and therefore the intensity of every word in the sentence. This needs a pre-defined dictionary classifying negative and positive words. Generally, a text message is represented by a bag of words or a sparse vector. After assigning individual scores to all the words, the final sentiment is calculated by some pooling operation liketaking an average of all the sentiments.
The polarity and subjectivity of a statement are returned by Text Blob. The range of polarity is [-1,1], with -1 indicating a negative feeling and 1 indicating positive sentiment. Negative words are used to change the polarity of a sentence. Semantic labels in Text Blob aid in fine-grained analysis. Emoticons, exclamation marks, and emojis, for example. Between [0,1] is subjectivity. The degree of personal opinion and factual information in a text is measured by subjectivity. Because of the text's heightened subjectivity, it provides personal opinions rather than factual facts.

C. VADER (Valence Aware Dictionary and Sentiment Reasoner)

VADER is a rule-based/lexicon model used for text sentiment analysis which provides the polarity (positive/negative) and the intensity (strong) of emotion. It's immanent in the NLTK package and can be used on unprocessed text data right away.
The sentiment analysis under VADER is based on a lexicon's rules that map lexical elements to emotion intensities, which are referred to as sentiment scores. A text's sentiment score may be calculated by adding the intensity of each word in the text. Words like 'love,' 'enjoy,' 'glad,' and 'like,' for example, all express a pleasant attitude. VADER is also smart enough to recognize the underlying meaning of certain phrases, such as "did not love" as a negative remark. It also recognizes the importance of capitalization and punctuation, as in the phrase "Wonderful".

D. *NAIVE BAYES*

The Naive Bayes Algo, is a multinomial classification technique. It works on the principle of conditional probability as encapsulated by Bayes theorem. Classifying text is a high use case for Naive Bayes Algorithm.

E. CONVOLUTION NEURAL NETWORK

provide various functions and methods to process data in the form of a table. For sentiment analysis, we are working with text data hence we have Neural networks are a subset of Machine learning and lie at the core of Deep Learning. They process data in a layered structure resembling the working of neurons in a brain.
Deep Learning emulates the Human brain which makes the application of deep learning into reading human text much more fruitful.

Data Collection

TWEEPY
Python is an inter active, object-oriented programming language with modules and dynamic data types and classes. This makes collection of twitter data from python docile and hassle free. Tweepy is an open source python library to access twitter API. It helps to access the API from a python application. There are other packages to collect twitter data such as Twython and SNScrape. In the contemporary scenario, tweepy is the official and most appropriate library that is permitted by twitter to access their content without any security issues. Though, tweepy has some drawbacks such as providing data only a week old and a data limit of 32K tweets at a time.

Twitter Data
We'll go through how to construct a Twitter Applica- tion, how to authenticate a user, and how to make basic API requests in this portion of the paper. First, we'll head to https://dev.twitter.com/apps and create an account. We receive applications when the registration procedure is completed. From the application details tab, we can access the following keys:

- API KEY
- API KEY SECRET
- ACCESS LEVEL
- OWNER
- OWNER ID (Optional)

Data Pre-Processing

We have used python's numpy and pandas modules toclean and manipulate data. Numpy and pandas used the following steps:

- After loading the datasets, we have started offwith analyzing and removing stop words.
- We further convert our text into lowercase so as to avoid repetition.
- Removing URLs to avoid unnecessary traffic in data. This can be done using the regular expression module provided by python.
- Further, removing slang and Abbreviations that don't provide any valuable insights to our results.
- Removing Smileys and Emoticons; Both mentioned steps can be done using modules which can be found already created or can be created.
- Removing any extra noise i.e twitter handles, punctuation, extra spaces, numbers and special characters.
- Next, we normalize the text using two techniques i.e Stemming and Lemmatizing.
- The next step is to convert text into numbers. i.e. Converting our training data to Sparse Matrix. The idea behind this is to convert text into matrix format so a computer can analyze it and produce some valuable insights.
- Finally, we build our model and depending upon the model we convert the data into tensors or sparse matrices.

6. RESULT AND ANALYSIS

Input	Prediction	Actual	Accuracy
This film...terrible	0.002	0	100%
This film...great!	0.993	1	99.4%
They are...nicely.	0.945	1	94.6%
I had a great day!	0.948	1	94.8%
We...not...away.	0.339	0	66.1%

Table 1: Comparison between original and predicted values

The accuracy of our model in test data is 88.08%. For simplicity, we tested our model with 5 more sentences for real-time accuracy which turned out to be 90.98%, slightly more than our test accuracy.

7. ACKNOWLEDGMENT

We are thankful to Delhi Technical Campus and we owe special debt of gratitude to our project coordinator Ms. Medhavi Pandey (Professor), The Department of Computer Science Engineering, Delhi Technical Campus. It is only because of her consistent efforts that we have performed our project successfully for their support and providing us with the necessary facilities to perform our project work successfully.

D. CONCLUSION

Sentiment analysis has been one of the most active study fields in the last ten years for a variety of reasons. For starters, sentiment analysis offers a wide range of uses in practically every field. Second, it presents a variety of difficult research questions that have never been investigated previously. Third, because of big data technology, we now have a massive amount of opinionated data that is freely available in digital formats on the internet. With Real-time Sentiment Analysis, we can actually solve a problem that every business encounters almost every day. Even if it's an enormous firm like Amazon or a small online startup, every business wants to know the opinions of their customers and what their customer wants. Our project helps small businesses with just that. On our platform, anyone can come and input the reviews and opinions of their customers, and our Machine Learning Model will give them rich insight into their customer's feedback. With our detailed feedback and analysis, every business can focus on what their customers want.

9.References

[1] Apurb Rajdhan, Milan Sai, Avi Agarwal, Dundigalla Ravi, Dr. Poonam Ghuli, "Heart Disease Prediction using Machine Learning", *International Journal of Engineering Research & Technology (IJERT)*, Vol. 9, pp.659-662, 2020.

[2] Shan Xu ,Tiangang Zhu, Zhen Zang, Daoxian Wang, Junfeng Hu and Xiaohui Duan et al. "Cardiovascular Risk Prediction Method Based on CFS Subset Evaluation and Random Forest Classification Framework", *2017 IEEE 2nd International Conference on Big Data Analysis*.

[3] Jaymin Patel, Prof. Tejal Upadhyay, Dr. Samir Patel, "Heart Disease Prediction Using Machine Learning and Data Mining Technique", *International Journal of Computer Science & Communication (IJCSC)*, Vol. 7, pp.129-137, 2016.

[4] Seyedamin Pouriyeh, Sara Vahid, Giovanna Sannino, Giuseppe De Pietro, Hamid Arabnia, Juan Gutierrez et al. "A Comprehensive Investigation and Comparison of Machine Learning Techniques in the Domain of Heart Disease", *22nd IEEE Symposium on Computers and Communication*.

[5] V.V. Ramalingam, Ayantan Dandapath, M Karthik Raja, "Heart disease prediction using machine learning techniques", *International Journal of Engineering & Technology (IJET)*, Vol. 7, pp.684-687, 2018.

[6] Seyedamin Pouriyeh, Sara Vahid, Giovanna Sannino, Giuseppe De Pietro, Hamid Arabnia, Juan Gutierrez et al. "A Comprehensive Investigation and Comparison of Machine Learning Techniques in the Domain of Heart Disease", *22nd IEEE Symposium on Computers and Communication (ISCC 2017): Workshops -ICTS4eHealth 2017*.

[7] Youness Khourdifi, Mohamed Bahaj, "Heart Disease Prediction and Classification Using Machine Learning Algorithms Optimized by Particle Swarm Optimization and Ant Colony Optimization", *International Journal of Intelligent Engineering & Systems (INASS)*, Vol. 12, pp.242-252, 2019.

[8] Quazi Abidur Rahman, Larisa G. Tereshchenko, Matthew Kongkatong, Theodore Abraham, M. Roselle Abraham, and Hagit Shatkay et al. "Utilizing ECG- based Heartbeat Classification forHypertrophic Cardiomyopathy Identification",

DOI 10.1109/TNB.2015.2426213, IEEE Transactions on Nano Bioscience TNB-00035-2015.

[9] Avinash Golande, Pavan Kumar T, "Heart Disease Prediction Using Effective Machine Learning Techniques", *International Journal of Recent Technology and Engineering*, Vol 8, pp.944-950, 2019.

[10] Fahd Saled Alotaibi, "Implementation of Machine Learning Model to Predict Heart Failure Disease", *International Journal of Advanced Computer Science and Applications,* Vol. 10, pp.261-268, 2019.

[11] T. Padmapriya and V.Saminadan, "Handoff Decision for Multi-user Multiclass Traffic in MIMO- LTE-A Networks", *2nd International Conference on Intelligent Computing,Communication & Convergence (ICCC-2016) – Elsevier - PROCEDIA OF COMPUTER SCIENCE, vol. 92, pp: 410-417, August 2016.*

[12] S.V.Manikanthan and D.Sugandhi "Interference Alignment Techniques For Mimo Multicell Based On Relay Interference Broadcast Channel" *International Journal of Emerging Technology in Computer Science & Electronics(IJETCSE)* ISSN: 0976-1353 Volume- 7, Issue 1 –MARCH 2014.

[13] Y. E. Shao, C.-D. Hou, and C.-C. Chiu, "Hybrid intelligent modelling, schemes for heart disease classification," *Applied Soft Computing,* vol. 14, pp. 47–52, 2014.

[14] C. S. Dangare and S. S. Apte, "Improved study of heart disease prediction system using data mining classification techniques," *International Journal of Computer Applications,* vol. 47, no. 10, pp. 44–48, 2012.

[15] J. Nahar, T. Imam, K. S. Tickle, and Y.-P. P. Chen, "Computational intelligence for heart disease diagnosis: A medical Knowledge driven approach," *Expert Systems with Applications*, vol. 40, no. 1, pp. 96–104, 2013.

[16] Theresa Princy R,J. Thomas,'Human heart Disease Prediction System using Data Mining Techniques', *International Conference on Circuit Power and ComputingTechnologies*,Bangalore,2016.

[17] Combination data mining methods with new medical data to predicting outcome of coronary heart disease," *in Convergence Information Technology, 2007. International Conference on. IEEE*, 2007, pp. 868–872.

[18] Y. Xing, J. Wang, Z. Zhao, and Y. Gao, "Combination data mining methods with new medical data to predicting outcome of coronary heart disease," pp. 868–872, 2007.

[19] P. V. Ankur Makwana, "Identify the patients at high risk of re-admission in hospital in the next year," *International Journal of Science and Research,* vol. 4, pp. 2431–2434, 2015.

[20] M. Shouman, T. Turner, and R. Stocker, "Using data mining techniques in heart disease diagnosis and treatment," pp. 173–177, 2012

[21] S. Palaniappan and R. Awang, "Intelligent heart disease prediction system using data mining techniques," pp.108–115, 2008.

[22] Kim S-M, Hovy E (2004) Determining the sentiment of opinions. In: Proceedings of the 20th international conference on Computational Linguistics, page 1367. Association for Computational Linguistics, Stroudsburg, PA, USA

[23] Liu B (2010) Sentiment analysis and subjectivity. In: Handbook of Natural Language Processing, Second Edition. Taylor and Francis Group, Boca

[24] Liu B, Hu M, Cheng J (2005) Opinion observer: Analyzing and comparing opinions on the web. In: Proceedings of the 14th International Conference on World Wide Web, WWW '05. ACM, New York, NY, USA. pp 342–351

[25] Pak A, Paroubek P (2010) Twitter as a corpus for sentiment analysis and opinion mining. In: Proceedings of the Seventh conference on International Language Resources and Evaluation. European Languages Resources Association, Valletta, Malta

[26] Pang B, Lee L (2004) A sentimental education: Sentiment analysis using subjectivity summarization based on minimum cuts. In: Proceedings of the 42Nd Annual Meeting on Association for Computational Linguistics, ACL '04. Association for Computational Linguistics, Stroudsburg, PA, USA

[27] Pang B, Lee L (2008) Opinion mining and sentiment analysis. Found Trends Inf Retr 2(1-2):1–135 7. Turney PD (2002) Thumbs up or thumbs down? Semantic orientation applied to unsupervised classification of reviews. In: Proceedings of the 40th Annual Meeting on Association for Computational Linguistics, ACL '02. Association for Computational Linguistics, Stroudsburg, PA, USA. pp 417–424

[28] Whitelaw C, Garg N, Argamon S (2005) Using appraisal groups for sentiment analysis. In: Proceedings of the 14th ACM International Conference on Information and Knowledge Management, CIKM '05. ACM, New York, NY, USA. pp 625–631

[29] Twitter (2014) Twitter apis.https://dev.twitter.com/start 10. Liu B (2014) Thescience of detecting fake reviews.http://content26.com/blog/bing-liu-the-science-of- detectingfake-

reviews/

[30] Jindal N, Liu B (2008) Opinion spam and analysis. In: Proceedings of the 2008 International Conference on, Web Search and Data Mining, WSDM '08. ACM, New York, NY, USA. pp 219–230

[31] Mukherjee A, Liu B, Glance N (2012) Spotting fake reviewer groups in consumer reviews. In: Proceedings of the 21st, International Conference on World Wide Web, WWW '12. ACM, New York, NY, USA. pp 191–200

[32] Stanford (2014) Sentiment 140http://www.sentiment140.com/

[33] www.twitter.com

[34] Go A, Bhayani R, Huang L (2009) Twittersentiment classification using distant supervision. CS224N Project Report, Stanford. pp 1–12

[35] Lin Y, Zhang J, Wang X, Zhou A (2012) An information theoretic approach to sentiment polarity classification. In: Proceedings of the 2Nd Joint WICOW/AIRWeb Workshop on Web Quality, WebQuality '12. ACM, New York, NY, USA. pp 35–40

[36] Sarvabhotla K, Pingali P, Varma V (2011) Sentiment classification: a lexical similarity-based approach for extracting subjectivity in documents. Inf Retrieval 14(3):337–353

[37] Wilson T, Wiebe J, Hoffmann P (2005) Recognizing contextual polarity in phrase-level sentiment analysis. In: Proceedings of the conference on human language technology and empirical methods in natural language processing. Association for Computational Linguistics, Stroudsburg, PA, USA. pp 347–354

[38] Yu H, Hatzivassiloglou V (2003) Towards answering opinion questions: Separating facts from opinions and identifying the polarity of opinion sentences. In: Proceedings of the 2003 conference on, Empirical methods in natural language processing. Association for Computational Linguistics, Stroudsburg, PA, USA.pp 129–136

[39] Zhang Y, Xiang X, Yin C, Shang L (2013) Parallel sentiment polarity classification method with substring feature reduction. In: Trends and Applications in Knowledge Discovery and Data Mining, volume 7867 of Lecture Notes in Computer Science. Springer Berlin Heidelberg, Heidelberg, Germany. pp 121–132

[40] Zhou S, Chen Q, Wang X (2013) Active deep learning method for semi-supervised sentiment classification. Neurocomputing 120(0):536–546.
Image Feature Detection and Description

[41] Chesley P, Vincent B, Xu L, Srihari RK (2006) Using verbs and adjectives to automatically classify blog sentiment. Training 580(263):233

[42] Choi Y, Cardie C (2009) Adapting a polarity lexicon using integer linear programming for domain-specific sentiment classification. In: Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing: Volume 2 - Volume 2, EMNLP '09. Association for Computational Linguistics, Stroudsburg, PA, USA.pp 590–598

[43] Jiang L, Yu M, Zhou M, Liu X, Zhao T (2011) Target-dependent twitter sentiment classification. In: Proceedings of the 49th, Annual Meeting of the Association for Computational Linguistics: Human Language Technologies-Volume 1. Association for Computational Linguistics, Stroudsburg, PA, USA.pp 151–160

[44] Tan LK-W, Na J-C, Theng Y-L, Chang K (2011) Sentence-level sentiment polarity classification using a linguistic approach. In:Digital Libraries: For Cultural Heritage, Knowledge Dissemination, and Future Creation. Springer, Heidelberg, Germany. pp 77–87

[45] Liu B (2012) Sentiment Analysis and Opinion Mining. Synthesis Lectures on Human Language Technologies. Morgan & Claypool Publishers

[46] Hu M, Liu B (2004) Mining and summarizing customer reviews. In: Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining. ACM, New York, NY, USA. pp 168–177

[47] Gann W-JK, Day J, Zhou S (2014) Twitter analytics for insider trading fraud detection system. In: Proceedings of the sencond ASE international conference on Big Data. ASE

[48] Roth D, Zelenko D (1998) Part of speech tagging using a network of linear separators. In: Coling-Acl, The 17th International Conference on Computational Linguistics. pp 1136–1142

[49] Kristina T (2003) Stanford log-linearpart-of-speech taggehttp://nlp.stanford.edu/software/tagger.shtml

[50] Marcus M (1996) Upenn part of speechtagger. http://www.cis.upenn.edu/~treebank/home.html

[51] Han J, Kamber M, Pei J (2006) Data Mining: Concepts and Techniques, Second Edition (The Morgan Kaufmann Series in Data Management Systems), 2nd ed. Morgan Kaufmann, San Francisco, CA, USA

[52] (2014) Scikit-learn.http://scikit-learn.org/stable/