

## Machine Learning and Artificial Intelligence based Analysis for Top Organization

**Anil Kumar Singh**

Department of Computer Science & Eng.  
BN College of Engineering  
& Technology  
*anilsinghrkg@gmail.com*

**Dr. APJ Abdul Kalam**

Technical University Lucknow, India  
*er.jaswantsingh786@gmail.com*

**Dr. APJ Abdul Kalam**

Technical University Lucknow, India  
*vinay.bncet@gmail.com*

**Vinay Kumar**

Department of Computer Science & Eng.  
BN College of Engineering  
& Technology

**Dr. Jasvant Kumar**

Department of Computer Science & Eng .  
BN College of Engineering  
& Technology  
*vinay.bncet@gmail.com*

**Dr. APJ Abdul Kalam**

Technical University Lucknow, India  
*anilsinghrkg@gmail.com*

### Abstract—

Top organizations doing surveys so they can get to know their rating in between other similar organizations. The Internet provides a platform for organizations to display the details of an organization on their website which is accessible throughout the world, people can see all details about the organization on the internet. They can view the ranking of top organizations and able to compare and analyze the ranking of different top organizations. Clients first see the rating of the organization while joining them and from the customer's point of view organization rating shows the growth of the organization. This proposed paper concept introduces a data mining technique to provide a rating of organization based on customer feedbacks and plot the comparison graph of different organization, for this data mining analysis machine learning (ML) is used for data pre-processing and for self-learning maximum entropy artificial intelligence (AI) algorithm is used, which is helpful to the machine to self-decide word meaning sense like positive, negative or neutral. To get the desired result first step which applied on input data is preprocessing which is mainly used to filter the data based on Machine Learning (ML) concept means it is removing unnecessary part of data

from input data. Here it will remove words like is, am, are, was, were, will, will be, etc. Apart from this, it will remove all special characters from a sentence. If any URL is mention in a sentence that is also removed from the sentence. The proposed system of the paper using a maximum entropy algorithm for self-learning of people feedback sentiment to provide ratings of top organizations. In this algorithm, first, we have to pass keyword (meaning full words) it generates numerical values for all keywords and based on threshold value in creating categories range for positive, negative, and neutral sentiment. Based on these categories machine-self recognize the sentiment of the word.

***Keywords— Sentiment analysis, Machine learning, Artificial intelligence, Supervise learning.***

## I. INTRODUCTION

Nowadays more people using the internet, in all countries use of the internet is spreading exponentially, people using internet services for their daily basic needs like buying food, medicine and to check the ranking of organizations, etc [3][7]. Here people can compare top organizations' facilities, education quality, previous year employment offer records, and facility details [5]. Subsequently, individuals, input data is of great relevance as it very well may be used as a piece of various circumstances where associations or brands can utilize a quick relationship with all of their client or customer and along these lines, upgrade their thing.

The Internet provides a platform for organizations to display the details of the organization on their website which is accessible throughout the world, people can see all details about the organization on the internet. They can view the ranking of top organizations and able to compare and analyze the ranking of different top organizations [5].

Here people can give feedback on any organization, which visible to other people who are checking details about the organization. This feedback plays an important role for others to decide which organization is better because if organization services are good in that case people feedback also is positive for the organization, but if organization services are not good as per expectation in that case people feedback is negative.

Thusly, whether or not the accomplice's Opinion is sure/negative about their contribution transforms into a fundamental and crushing inquiry for the relationship to pose and screen [1]. Moreover associations in the business place. In like way, this assessment looks for after a comparative target by performing purchaser things according to proposed paper algorithms and once-over in the territory of mobile phones and data set.

Therefore, we truly remarked on Twitter data for our assessments [2]. End assessment is the place the dataset contains sentiments, demeanors or assessment which considers the way a human thinks [1]. In a sentence, endeavoring to appreciate the positive and the negative edge is an incredibly inconvenient task. The features used to portray the sentences should have a very strong modifier to consolidate the study. These substances are even written in different strategies which are not really inferred by the customers or the associations making it difficult to gather them. Notion assessment impacts customers to aggregate the details of organization and these things are important for people who are searching top listed organizations.

Organizations using this detail examination to fathom the things or organizations with the goal that it might be offered by the customer's needs. There are two sorts of AI frameworks that are normally used for thought assessment, one is solo and the other is regulated [2]. Solo learning doesn't involve a class and they don't outfit with the privilege centers at all and henceforth lead gathering.

Regulated learning relies upon stamped dataset and along these lines, the names are given to the model in the midst of the system. These stamped datasets are set up to convey reasonable yields when experienced in the midst of essential authority.

To help us with understanding the idea examination harshly, this investigation paper relies upon the oversaw AI. In once-over, this paper shows a careful and all-around fundamental evaluation of 15 Sentiment Analysis web gadgets that has never been done.

To truly play out this examination, a set-up of appraisal models and most likely comprehended data gatherings from the field of Sentiment Analysis has been decided to empower the per client to explore the upsides and drawbacks of the usage of these instruments seeing points of view, for instance,

disclosure of feelings inside short and long messages, the revelation of disjointedness or count of limit assessments, among others.

Besides these standard data collections, these devices have moreover been studied by mirroring an even more real circumstance, in which the sufficiency for recommending movies from authentic customers' comments has been had a go at using the information assembled from the striking site IMDb1.

This paper concept based on the rating of top organizations, Nowadays, all top organizations doing surveys so they can get to know their rating in between other similar organizations. Clients first see the rating of the organization while joining them and from the customer's point of view organization rating shows the growth of the organization. This proposed paper concept introduces a data mining technique to provide a rating of organization based on customer feedbacks and plot the comparison graph of different organization, for this data mining analysis machine learning (ML) is used for data pre-processing and for self-learning maximum entropy artificial intelligence (AI) algorithm is used, which is helpful to the machine to self-decide word meaning sense like positive, negative or neutral.

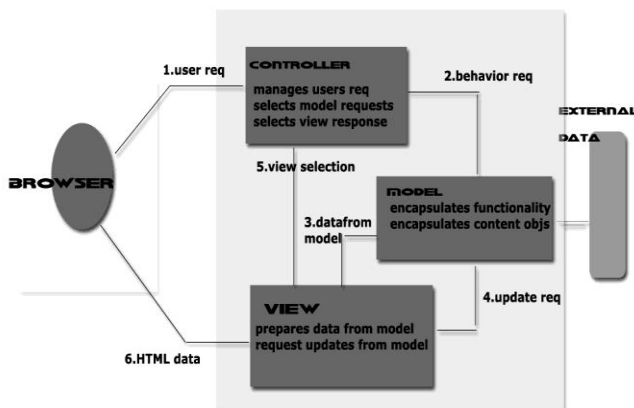


Fig 1 J2EE uses MVC Architecture

This web application developed to provide a rating of the top organizations, this proposed paper concept introduces a data mining technique to provide a rating of organization based on customer feedbacks and plot the comparison graph of different organization, for this data mining analysis machine learning (ML) is used for data pre-processing and for self-learning maximum entropy artificial intelligence (AI) algorithm is used, which is helpful to the machine to self-decide word meaning sense like positive, negative or neutral.

This template, modified in MS Word 2007 and saved as a “Word 97-2003 Document” for the PC, provides authors with most of the formatting specifications needed for preparing electronic versions of their papers. All standard paper components have been specified for three reasons: (1) ease of use when formatting individual papers, (2) automatic compliance to electronic requirements that facilitate the concurrent or later production of electronic products, and (3) conformity of style throughout the conference proceedings. Margins, column widths, line spacing, and type styles are built-in; examples of the type styles are provided throughout this document and are identified in italic type, within parentheses, following the example. Some components, such as multi-leveled equations, graphics, and tables are not prescribed, although the various table text styles are provided. The formatter will need to create these components, incorporating the applicable criteria that follow.

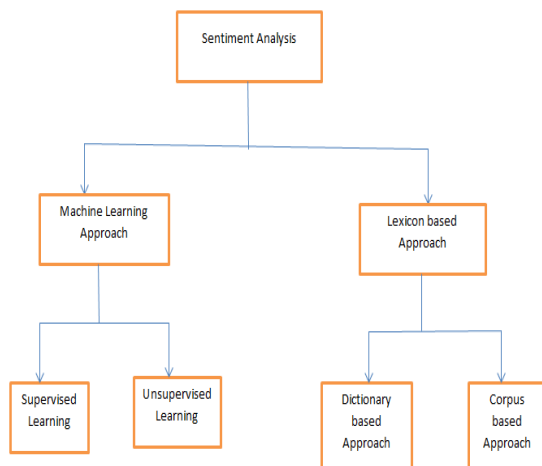
## II. RELATED WORK

There are 2.4 billion dynamic online clients, who compose and read online around the globe [1]. In spite of the fact that the logical space is colossal as a major universe of diaries and meetings, there are more than 4000 appraised gatherings and 5000 positioned diaries [2]. Notably, an enormous section of WWW scientists makes their substance open, permitting analysts, social orders, colleges, and

enterprises to utilize and investigate information. As indicated by another study led by dimensional examination, April 2013: 90% of client's choices rely upon online surveys [3].

As indicated by a 2013 Study [4]: 79% of client's certainty depends on online individual proposal surveys. As the outcome, countless investigations and examinations have observed the slanting increment of online exploration assets step by step. In this proposal, we attempt to accomplish believed logical audits assessment to be valuable for scientists and encourage the determination of papers that coordinate their examination course. We make another proposed method for assessing logical papers dependent on supposition investigation and space parameters. We present another proposed smaller than normal vocabulary. As of late, a few sites urge specialists to communicate and trade their perspectives, recommendations, and feelings identified with logical papers. Assessment examination targets deciding the demeanor of an author as for certain themes or the general slant extremity of a book, for example, positive or negative. Supposition investigation [5] relies upon two issues notion extremity and conclusion score. Slant extremity [5] is a parallel worth either positive or negative. Then again, the feeling score depends on one of three models.

Those models are the Bag-of-words model (BOW) [6], grammatical form (POS) [7], and semantic connections [8]. BOW [6] model is the most well-known for analysts and dependent on the portrayal of terms. Here it is using the word collection model in this it is not considering the language structure or word proper sequencing, it is going to arrange based on algorithm steps. In that, it is going to remove all verbs and meaningless words from the collection of all words. Here it is collecting only meaning full word, which is used to categorize the sentence based on these meaning full words. Finally, it collects all meaning full word in separate word collection, after that based on these input data it will generate the details output, which is help client to understand the performance of their business. Checking the recorded data in the audit process to release the data analysis report based on input data. These data generated based on client service to the service user. The goal of Sentiment Analysis is to assess the suppositions and assessments of an author individually, in one subject area or multi-theme space. It figures the total opinion extremity of online genuine audits for one subject dependent on notion grouping levels, for example, positive or negative. Existing investigation ways to deal with assessment surveys can be assembled into four primary classes: word level, sentence level, report level, and viewpoint/element level.



**Fig 2.** Techniques of Sentiment classification

The machine learning approach work based on machine training, in that the first thing, is the training of the machine, based on the machine training, the machine will work. In machine learning also has two sections, the first section is supervised learning, the second section is unsupervised learning. Supervised learning consists of four classifiers, like decision tree classifier, linear classifier, etc. All these classifiers are used to detect people's sentiments, which is used to categorize them based on emotion and sentiments.

Bush V. D, Racherla P., and King R. A., 2014, [3] Here it introduces the technique to get to know major user search details based on a lot of different user search behavior. It learns the behavior pattern of users and provides a result of the most frequently searched details. Many commercial organizations need this type of result to get the present market trends. So these results help them to make products based on market trends.

Neville S.W, Li K.F, and Arora D., 2015, [6] The author focus to product review, which is having a major impact on product selling, nowadays customer see the product review before buying the product. Present time customer mindset is to verify the product before buying, so which product is genuine and good in quality they have a lot of good review feedback and rating and new customer see this review and buy these products.

Kanakaraj M., Guddeti R M.R., 2015, [8] In this paper explain how to extract information data available on the internet. Daily people activities data storing on internet, from these data using natural language processing (NLP) mines the data which is beneficial for us. Using NLP extract the data which is related to our requirement and analyze these data to get the desired result. These analysis results used in many sectors to improve the growth and product manufacturing based on demands.

Rajasree R., Neethu M. S., 2013, [12] author provides details on how to get user feeling assumptions. Based on the sentiment analysis algorithm it is analyzing meaning full word sense like work is a positive sense, negative sense, or neutral sense. Based on this word sense result user feeling is marked positive, negative or neutral. So by applying this technique, able to assume the user's feelings.

Shokoufandeh A, Bai B., Qi Y., and Bepalov D., 2011, [13] here author proposes a thought to get a product review analysis which helps the customer to choose the better product based on reviews. Here the author explains how to extract information data available on the internet. Daily people activities data storing on internet, from these data using natural language processing (NLP) mines the data which is beneficial for us. Using NLP extract the data which is related to our requirement and analyze these data to get the desired result. These analysis results used in many sectors to improve the growth and product manufacturing based on demands.

### III. PROPOSED METHODOLOGY

The proposed system is to give top organization ratings based on people sentiment classification. The proposed paper concept is more accurate and more efficient compare to the existing system. It is using people feedback for top organization rating based on people sentiment analysis algorithm which is providing analysis of all collected people feedback data.

#### A. *Preprocessing*

The preprocessing mainly used to filter the data based on Machine Learning (ML) concept, means it is removing unnecessary part of data from input data. Here it will remove words like is, am, are, was, were, will, will be, etc. Apart from this, it will remove all special characters from a sentence. If any URL is mention in a sentence that is also removed from the sentence.

Define abbreviations and acronyms the first time they are used in the text, even after they have been defined in the abstract. Abbreviations such as IEEE, SI, MKS, CGS, sc, dc, and rms do not have to be defined. Do not use abbreviations in the title or heads unless they are unavoidable.

#### B. *Self-Learning and word standardization System*

The proposed system of the paper using a maximum entropy algorithm for self-learning of people feedback sentiment to provide ratings of top organizations. In this algorithm, first, we have to pass keyword (meaning full words) it generates numerical values for all keywords and based on threshold value in creating categories range for positive, negative, and neutral sentiment. Based on these categories machines self-recognize the sentiment of the word.

### C. Algorithm

#### Step 1: Preprocessing:

Input image

$$I(A) = \sum_{i=0}^i \sum_{j=0}^j xy$$

P, the set windows declared positive by cascade

Set P = {[i, i+e-1]} \* [j, j+e-1] ∈ N

For L = 1 to L do, 0 < y <= 1

Every window in P do

End

Normalizetheweight

$$W_i = \frac{W_i}{\sum_i W_i}$$

If c/W<sub>i</sub> > 0.4 then

Return true

Else return false

End if

Return P

#### Step 2: Keyword Identification:

The cosine similarity between the two text t and s is given below:

$$Cos(t, s) = \frac{\sum_{a \in b} Q_t(a) Q_s(a)}{\sqrt{(\sum_{a \in b} Q_t^2(a)) (\sum_{a \in b} Q_s^2(a))}}$$

#### Step 3: Self Learning:

Initialize the value of Q

While Q is not converged do

r ← R(S, a)

s' ← T(S, a)

For

Choose A and S using policy divided from Q

Take action A and observe R, S

$$R(A) = \sum_{i=0}^i \sum_{j=0}^j yr, 0 < y <= 1$$

Where r = Q(s', a') - Q(s, a)

Receive reward R<sub>t</sub> and new state S<sub>t+1</sub>

t ← t+1

Q, (S<sub>0</sub>, a<sub>0</sub>) ← Q<sub>0</sub>, (S, a) + yR<sub>t</sub>

end

#### Step 4: Sentiment classification:

Classification is based on distance metric:

$$D(f_x, i_y) = \frac{d(f_x, NN_2(f_x, i_y))}{d(f_x, NN_1(f_x, i_y))}$$

$$m(f_x, i_y) = \begin{cases} 1, & \text{if } D(f_x, i_y) < z \\ 0, & \text{otherwise} \end{cases}$$

This proposed paper concept provides the data analysis which important data for users whoever selecting organization based on rating on the internet.

A proposed methodology using the following step to get the desired result:

**Step 1:** First thing required is survey data based on this survey data further step applied to get an analysis. Here in this paper 500 people feedback is used for data analysis. These survey feedbacks have written in an excel sheet.

**Step 2:** Upload the excel file in the proposed system application. It will read data from an excel sheet and store it in the database.

**Step 3:** now preprocessing is used on survey data, preprocessing mainly used to filter the data based on Machine Learning (ML) concept, means it is removing unnecessary part of data from input data. Here it will remove words like is, am, are, was, were, will, will be, etc. Apart from this, it will remove all special characters from a sentence. If any URL is mention in a sentence that is also removed from the sentence.

**Step 4:** Self-learning: here system automatically decides the sense of feedback, for this proposed system of the paper using a maximum entropy algorithm for self-learning of people's feedback sentiment to provide a rating of top organizations. In this algorithm, first, we have to pass keyword (meaning full words) it generates numerical values for all keywords and based on threshold value in creating categories range for positive, negative, and neutral sentiment. Based on these categories machines self-recognize the sentiment of the word.

**Step 5:** Finally it will classify the feedback based on negative, positive, and neutral word counts.

After the data analysis which organization having more positive feedback that organization rating better than others. So here organization rating depends on positive feedback given by people in the survey.

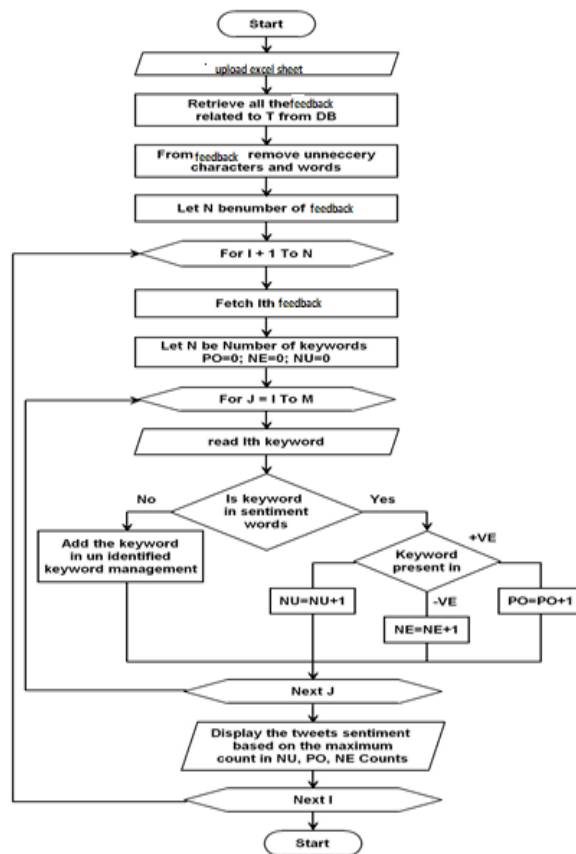


Fig 3. Flow chart of sentiment analysis

#### IV. RESULTS

Nowadays, all top organizations doing surveys so they can get to know their rating in between other similar organizations. Clients first see the rating of the organization while joining them and from customer's point of view organization rating shows the growth of the organization. This proposed paper concept introduces a data mining technique to provide a rating of organization based on customer feedbacks and plot the comparison graph of different organization, for this data mining analysis machine learning (ML) is used for data pre-processing and for self-learning maximum entropy artificial intelligence (AI) algorithm is used, which is helpful to the machine to self-decide word meaning sense like positive, negative or neutral.

In this proposed paper 2000 personal survey data is used to get data analysis results, in the survey people answered a question about the NIT, IIT, and AIIMS. People gave their feedback on this organization which is stored in an excel sheet.

TABLE I. Details of people review in Existing System

College	Ratio of positive to negative feedback	Average positive feedback
NIT	1.48	2.91
IIT	1.52	3.04
AIIMS	1.77	5.12

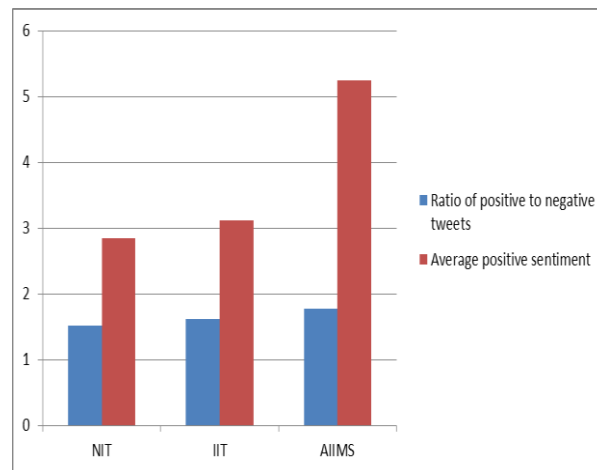


Fig 4. Details of people feedback in Existing System

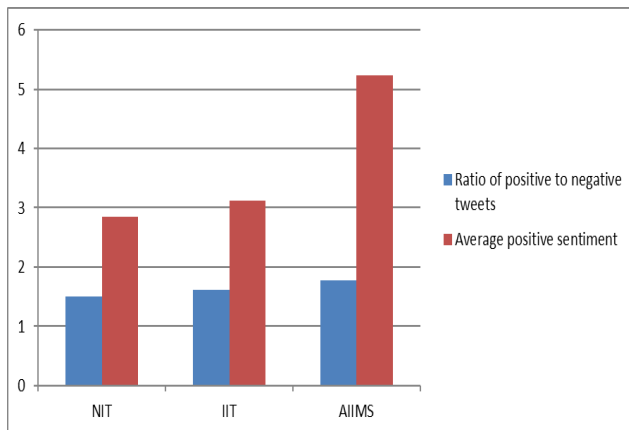
The general sentiment derived from the excel sheet dataset regarding the three organizations AIIMS, IIT, and NIT were, as follows: a total of 400 feedbacks were regarded as positive, 229 as negative, and 670 as neutral for AIIMS. 559 feedbacks were classified as positive, 367 as negative, and 868 as neutral for IIT. 503 feedbacks were considered as positive, 351 as negative, and 985 as neutral for NIT. Table I displays a few of the statistics obtained. Table 1 provides the details of comparisons.

TABLE II details of people feedback in Proposed System

College	Ratio of positive to negative feedback	Average positive sentiment
NIT	1.53	2.82
IIT	1.65	3.14
AIIMS	1.77	5.27



AIIMS had the highest positive average sentiment and the ratio for positive to negative survey feedback. This translates to the observation that the positive survey feedback about AIIMS are more positive in the magnitude of their sentiment and also indicates that AIIMS is talked about positively more than it is talked about negatively the most among the three institutions.



**Fig 5.** Details of feedback in Proposed System

## V. CONCLUSION AND FUTURE WORK

The Internet provides a platform for organizations to display the details of the organization on their website which is accessible throughout the world, people can see all details about the organization on the internet. They can view the ranking of top organizations and able to compare and analyze the ranking of different top organizations. Here people can give feedback on any organization, which visible to other people who are checking details about the organization. This feedbacks play an important role for others to decide which organization is better because if organization services are good in that case people feedback also is positive for the organization, but if organization services are not good as per expectation that case people feedback is negative.

In the future proposed system can be used in the mobile app, or it can be used on social media applications. In a proposed system using excel data for data analysis, but in the future input data can be getting from a web application that is running on the internet.

## REFERENCES

- [1] Anuj, S. & Shubhamoy, D., "Performance Investigation of Feature Selection Methods and Sentiment Lexicons for Sentiment Analysis", Special Issue of International Journal of Computer Applications (0975 – 8887) on Advanced Computing and Communication Technologies for HPC Applications - ACCTHPCA, June 2012
- [2] Posting a tweet, <https://support.twitter.com/articles/15367-posting-atweet>
- [3] King R. A., Racherla P. and Bush V. D., What We Know and Don't Know about Online Word-of-Mouth: A Review and Syndissertation of the Literature, Journal of Interactive Marketing, vol. 28, issue 3, pp. 167- 183, August 2014
- [4] Ministry of Human Resource Development, <http://mhrd.gov.in/statist>
- [5] India's Best Colleges, 2015, <http://indiatoday.intoday.in/bestcolleges/2015/>
- [6] Arora D., Li K.F. and Neville S.W., Consumers' sentiment analysis of popular phone brands and operating system preference using Twitter data: A feasibility study, 29th IEEE International

Anil Kumar Singh, Dr. APJ Abdul Kalam, Dr. APJ Abdul Kalam, Vinay Kumar, Dr. Jasvant Kumar,  
Dr. APJ Abdul Kalam

Conference on Advanced Information Networking and Applications, pp. 680-686, Gwangju, South Korea, March 2015

[7]Choi C., Lee J., Park G., Na J. and Cho W., Voice of customer analysis for internet shopping malls, International Journal of Smart Home: IJSH, vol. 7, no. 5, pp. 291-304, September 2013

[8]Kanakaraj M., Guddeti R.M.R., Performance Analysis of Ensemble Methods on Twitter Sentiment Analysis using NLP Techniques, 9<sup>th</sup> IEEE International Conference on Semantic Computing, pp. 169-170, Anaheim, California, 2015

[9]Bahrainian S.-A., Dengel A., Sentiment Analysis and Summarization of Twitter Data”, 16th IEEE International Conference on Computational Science and Engineering, pp. 227-234, Sydney, Australia, December 2013

[10]Pak A. and Paroubek P., Twitter as a Corpus for Sentiment Analysis and Opinion Mining, 7th International Conference on Language Resources and Evaluation, pp. 1320-1326, Valletta, Malta, May 2010

[11]Shahheidari S., Dong H., Bin Daud M.N.R., Twitter sentiment mining: A multidomain analysis, 7th IEEE International Conference on Complex, Intelligent and Software Intensive Systems, pp.144-149, Taichung, Taiwan, July 2013

[12]Neethu M. S. and Rajasree R., Sentiment Analysis in Twitter using Machine Learning Techniques, 4th IEEE International Conference on Computing, Communications and Networking Technologies, pp. 1-5, Tiruchengode, India, 2013

[13]Bespalov D., Bai B., Qi Y., and Shokoufandeh A., Sentiment classification based on supervised latent n-gram analysis, 20th ACM international conference on Information and knowledge management, pp. 375-382, New York, USA, 2011

[14]Jotheeswaran J. and Koteeswaran S., Decision Tree Based Feature Selection and Multilayer Perceptron for Sentiment Analysis, Journal of Engineering and Applied Sciences, vol. 10, issue 14, pp. 5883-5894, January 2015

[15]Socher R., et al, Recursive Deep Models for Semantic Compositionality Over a Sentiment Treebank, Conference on Empirical Methods in Natural Language Processing, Seattle, Washington, October 2013.

[16]dos Santos C. N. and Gatti M., Deep Convolutional Neural Networks for Sentiment Analysis of Short Texts, 25th International Conference on Computational Linguistics, pp. 69-78, Dublin, Ireland, August 2014.

[17]Segaran T. and Hammerbacher J., Beautiful Data: The Stories behind Elegant Data Solutions, Beijing: O'Reilly, 2009

[18]Nielsen F.A., Making sense of microposts, Finn Årup Nielsen blog, <https://finnaarupnielsen.wordpress.com/tag/sentimentanalysis/>

- [19]Koto F. and Adriani M., A Comparative Study on Twitter Sentiment Analysis: Which Features are Good?, *Natural Language Processing and Information Systems, Lecture Notes in Computer Science* vol. 9103, pp. 453-457, June 2015
- [20]Ng A.Y., Jordan M. I., On Discriminative vs. Generative classifiers: A comparison of logistic regression and naive Bayes, *Advances in Neural Information Processing Systems* vol. 14, pp. 841-848, 2002.
- [21]Karatzoglou A., Meyer D., Hornik K., Support vector machines in R., *Journal of Statistical Software*, vol. 15, issue: 9, April 2006
- [22]Rajendran S., Kalpana B., A Comparative Study and Choice of an Appropriate Kernel for Support Vector Machines., *International Journal of Soft Computing and Engineering (IJSCE)*, vol. 1, issue: 5, November 2011
- [23]Abakar K. A. A., Yu C., Performance of SVM based on PUK kernel in comparison to SVM based on RBF kernel in prediction of yarn tenacity, *Indian Journal of Fibre and Textile Research*, vol. 39, pp. 55-59, March 2014
- [24]Salazar D. A., Velez J. I., Salazar J. C., Comparison between SVM and Logistic Regression: Which One is Better to Discriminate?, *Colombian Journal of Statistics, Special Issue Biostatistics*, vol. 35, no. 2, pp. 223- 237, June 2012
- [25]Hsu C.-W., Chang C.-C., Lin C.-J., A practical guide to support vector classification, National Taiwan University, Taipei, April 2010
- [26]Girma H., A Tutorial on Support Vector Machine, Center of Experimental Mechanics, University of Ljubljana, 2009
- [27]Multilayer perceptron, [https://en.wikipedia.org/wiki/Multilayer\\_perceptron](https://en.wikipedia.org/wiki/Multilayer_perceptron)
- [28]Manning C. and Raghavan P., *Introduction to information retrieval*, New York: Cambridge University Press, 2008.