

Sentiment Analysis Of Product Reviews – A Survey

Dishi Jain, Bitra Harsha Vardhan, Saravanakumar Kandasamy

Abstract: With expanding development of the web network, informal communities, online gateways, audits, reviews, suggestions, evaluations, and input are produced by clients and it can be about anything like books, individuals, items, explore, occasions, and so on. These opinions become advantageous for organizations, political body, and people. While this is intended to be valuable, a greater part of this client produced content require utilizing the opinion mining methods or sentiment analysis. Sentiment Analysis is a field in which we study about feelings, conclusions and subjectivity of opinions. This review paper is a similar investigation of numerous as of late proposed calculations' upgrades and different slant examination applications. Mainly for product reviews that how sentiment analysis can be used to recommend a product based on reviews by the client. The related fields to estimation investigation that pulled in specialists as of late are examined. The fundamental objective of this review is to give nearly full picture of sentiment analysis, its sorts and characterization. The primary commitments of this paper include the complex orders of late articles and the outline of the ongoing pattern of research in the sentiment analysis and its related territories.

Index Terms: computational linguistic, feature extraction, feature selection techniques, lexicon-based approach, natural language processing, machine learning algorithm, sentiment analysis.

1. INTRODUCTION

The Sentiment Analysis is the computational investigation of individuals' feelings, frames of mind and feelings toward a substance. The element can speak about any people, occasions or themes. These subjects are destined to be secured by surveys. This recognizes the feeling communicated in a book at that point investigates it. Consequently, the objective of feeling examination is to discover sentiments, distinguish the notions they express, and afterward arrange their extremity as appeared in Fig. 1 [5]. Sentiment Analysis can be viewed as an arrangement procedure. There are three primary characterization levels in Sentiment Analysis: document level, sentence-level, and aspect level. Document level plans to characterize a conclusion archive as communicating a positive or negative supposition or estimation while Sentence-level intends to arrange feeling clarified in each sentence. The initial step is to distinguish whether the sentence is abstract or goal. On the off chance that the sentence is emotional, Sentence-level supposition examination will decide if the sentence communicates positive or negative sentiments. Be that as it may, there is no essential contrast among document and sentence level orders since sentences are simply short archives [5]. Perspective level sentiment analysis means to group the opinion as for the particular parts of substances. The initial step is to distinguish the substances and their viewpoints. The assessment holders can give various conclusions for various parts of a similar substance like this sentence "The voice nature of this telephone isn't great, yet the battery life is long". This study handles the initial two sorts of opinion examination. Sentiment analysis utilizes the NLP content investigation and computational procedures to computerize the extraction or arrangement of slant from assessment surveys. Investigation of these assumptions and sentiments has spread crosswise over numerous fields, for example, Consumer data, Marketing, books, applications, sites, and Social. Notion examination turns into a hot region in basic leadership [37]. A huge number of clients rely upon online feeling surveys. The primary objective of investigating supposition is to break down the surveys and look at the scores of notions. This investigation is separated into numerous levels [38]: document level [39], sentence-level [40], word/term level [41] or aspect level [42]. The succession

procedure of sentiment analysis is assessment and recognition of the sentiment polarity [43]. The authors in the papers [44],[45],[46], have highlighted the conclusion challenges. The datasets utilized in estimation examination are a significant issue in this field. The primary wellsprings of information are from the item surveys. These surveys are imperative to the business holders as they can take business choices as indicated by the investigation aftereffects of clients' sentiments about their items. The audits sources are fundamentally survey destinations like amazon. [5] Sentiment examination isn't just applied on item audits however can likewise be applied on financial exchanges, news stories, or political discussions. In political discussions, we could make sense of individuals' sentiments on a specific political race up-and-comers or ideological groups. The interpersonal organization destinations and miniaturized scale blogging locales are viewed as a generally excellent wellspring of data since individuals share and talk about their assessments about a specific theme uninhibitedly [4]. They are likewise utilized as information sources in the conclusion investigation process. The Feature choice procedures are given in subtleties in fig.1 alongside their related articles alluding to some beginning references. The Sentiment Classification methods are talked about with more subtleties representing related articles and starting references too. This paper talks about new related fields in sentiment analysis which have pulled in the analysts recently and their comparing articles. These fields incorporate Emotion Detection, Building Resources and Transfer learning. Feeling location intends to separate and investigate feelings, while the feelings could be unequivocal or certain in the sentences [3]. Move learning or Cross-Domain grouping manages investigating information from one space and after that utilizing the outcomes in an objective area [5]. Building Resources centers around making lexica, corpora in which supposition articulations are commented on as indicated by their extremity, and now and then lexicons [1]. The commitment of this overview is critical for some reasons. To start with, this overview gives modern order of ongoing articles as indicated by the strategies utilized. This point could help the specialists who know about specific systems to utilize them in the slant examination field and pick the proper procedure for a specific application. Second, the different strategies of sentiment analysis are ordered with brief subtleties of the

calculations and their starting references. This can push new comers to the sentiment analysis field to have an all-encompassing perspective on the whole field.

2 ARCHITECTURE

The main application that is discussed in various papers is the reviews given by customers to the products like amazon, or commenting about a movie on social media platforms. To understand these product reviews, sentiment analysis comes into picture. How the text is pre-processed, the feature selection methods used and what are the classifier modals needed is discussed in further sections. These reviews are converted into simple English language which is easily understandable and classify whether the review is positive, negative or neutral. The feature extraction methods and classifiers are the main steps in sentiment analysis.

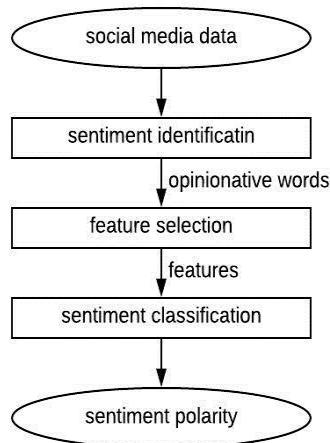


Figure 1. Process flow for sentiment analysis [5]

3 METHODOLOGY

The study approach is as per the following: brief clarification to the renowned feature selection methods and sentiment classification calculations speaking to some related fields to sentiment analysis are examined. At that point the commitment of these articles to these calculations is displayed representing how they utilize these calculations to tackle uncommon issues in sentiment analysis. The principle focus of this study is to see how sentiment analysis is used in product reviews.

4 FEATURE SELECTION METHOD

The initial phase in the opinion examination is to separate and choose content highlights. A portion of the present highlights are:

- Terms nearness and recurrence: This gives various events of a word in an info content.
- Part of Speech (POS): labeling the word in given content and discovering descriptors, as they are significant pointers of assessments.
- Feeling words and expressions: these are words generally used to express sentiments including

fortunate or unfortunate, as or loathe. While a few expressions express opinions lacking utilizing conclusion words.

- Refutations: the presence of negative words may change the conclusion significance like not great is equal to awful.

Feature Selection strategies can be isolated into vocabulary-based techniques that need human explanation, and measurable strategies which are programmed techniques which are utilized as often as possible. Dictionary based methodologies typically start with a little arrangement of 'seed' words. At that point they bootstrap this set through equivalent word location or on-line assets to get a bigger dictionary. While factual methodologies, are completely programmed. The feature selection techniques treat the reports either as gathering of words (Bag of Words (BOWs)), or as a string which holds the succession of words in the archive. Bag of Words is regularly utilized in view of it's a basic methodology for the order procedure. The most widely recognized component choice advance is the expulsion of stop-words and stemming. Three of the most as often as possible utilized factual strategies in highlight determination are examined in the following area and their related articles.

4.1 Point-wise Mutual Information (PMI)

To process the pointwise basic information for satisfied portrayal and to request tweets in classes, PMI is utilized. A dataset of tweets and a dictionary for each class of words which have a spot with that characterization is available. Given this information, how is it possible to register the PMI for each order per tweet, to orchestrate a tweet in one of these classes. PMI is an extent of connection between a segment and an arrangement, not between a record like a tweet and a characterization. The condition is open on Wikipedia. In that condition, X is the discretionary variable that models the occasion of a word, and Y models the occasion of a class. For a given word x and a given class y, one can use PMI to pick if a component is illuminating or not. Having less highpoints as often as possible improves the show of your request count and speeds it up altogether. The portrayal step, regardless, is discrete PMI admirable motivation you select better features to reinforce into your learning count.

$$pmi(x;y) \equiv \log \frac{p(x,y)}{p(x)p(y)} = \log \frac{p(x|y)}{p(x)} = \log \frac{p(y|x)}{p(y)} \quad (1)$$

Whenever x and y are impeccably connected, $P(x|y) = P(y|x) = 1$, so $pmi(x,y) = 1/P(x)$. Less successive x-es (words) will have a higher PMI score than incessant x-es, irrespective of whether both are excellently corresponded with y.

4.2 Chi-Square (χ^2)

The Chi-Square estimation is most routinely used to survey Tests of Independence when using a crosstabulation (generally called a bivariate table). Crosstabulation presents the spreads of two obvious factors at the same time, with the unions of the orders of the elements appearing in the phones of the table. The Test of Independence assesses whether a connection exists between the two factors by taking a gander at the watched case of responses in the cells to the model that would be ordinary if the components were self-ruling of each other. Processing the Chi-Square estimation and taking a gander at it against an essential motivation from the Chi-Square movement empowers the investigator to overview

whether the watched cell counts are generally one of a kind in connection to the typical cell checks. The estimation of the Chi-Square estimation is straight-forward and characteristic:

(2)

where f_o = the observed frequency

f_e = the expected frequency if no relationship existed between the variables

4.3 Latent Semantic Indexing

Latent semantic indexing (LSI) is an expansion of the vector space model that attempts to conquer these insufficiencies by fusing semantic data. It begins with the development of a term-archive lattice in which every section demonstrates the quantity of events of a particular term in a particular report. This grid is deteriorated into three frameworks by a procedure called singular value decomposition (SVD). These new lattices connote records and terms as vectors of factor esteems that catch the example of term utilization among reports. By lessening the quantity of variables esteems the hidden semantic structure is uncovered while unimportant commotion is sifted through. Note that Latent semantic ordering doesn't endeavor to comprehend the importance of the elements yet just uses them to show reports and vectors. From the all-out gathering of reports a term-file cross section X is confined with t lines (one for each term that appears in the game plan of records) and d portions (one for each record in the set). An SVD of cross section A results into the aftereffect of three frameworks:

$$X = USV^T \quad (3)$$

$\chi^2 = \sum \frac{(f_0 - f_e)^2}{f_e}$ These three matrices involve a $t \times m$ grid U that contains the term vectors, a $m \times m$ corner to corner arrange S of positive private characteristics, and a $d \times m$ system V that includes the record vectors. The singular characteristics in S are then orchestrated by size and simply the k greatest explicit characteristics are kept in S while the others are set to zero. Worth k is settled by methods for examination. The cross sections U and V are in addition altered with the ultimate objective that solitary the relating k areas in U and V remain. With these structures extra cross section $X_k = U_k S_k V_k^T$ can be registered that be similar to the first framework X. The aftereffect of the SVD can be translated geometrically as a k dimensional space where the lines of U and V are taken as directions of focuses that speak to terms and documentations individually.

The matrix $X_k X_k^T$ contains the dot products of every term:

$$X_k X_k^T = U_k S_k^2 V_k^T \quad (4)$$

The dot products of all terms are given by the following matrix:

$$X^T X_k = D_k S_k^2 V_k^T \quad (5)$$

The significant favored situation of LSI is the way that terms in reports and profiles can be by and large extraordinary yet simultaneously produce a match dependent on the semantic development. An impediment is the computational arbitrariness of the structure estimations which could decrease run-time execution to an unacceptable level.

5 SENTIMENT CLASSIFICATION TECHNIQUES

Sentiment Classification methods can be separated into machine learning approach and lexicon-based methodology.

The Machine Learning Approach uses AI calculations and etymological highlights while the Lexicon-put together Approach depends with respect to feeling dictionary, a gathering of known and precompiled estimation terms. It is isolated into lexicon-based methodology and corpus-based methodology which utilize factual or semantic techniques to discover estimation extremity. The content arrangement techniques utilizing AI approach are of two sorts managed and unaided learning strategies. The administered techniques utilize enormous number of named preparing archives and the classifiers utilized are Probabilistic, Decision tree, Rule-based and so forth while the unaided strategies are utilized when it is hard to locate these labelled preparing records. The vocabulary put together approach centers with respect to finding the feeling dictionary which aides in investigating the content. There are two techniques in this methodology. The lexicon put together approach which depends with respect to discovering sentiment seed words, and after that searches the word reference of their equivalent words and antonyms. The corpus-based methodology starts with a rundown of conclusion words, and afterward finds other opinion words in a huge corpus to help in discovering sentiment words with setting explicit directions. This should be possible by the assistance of measurable or semantic strategies.

5.1 Machine Learning Approach

Machine learning approach is one of the most noticeable methods picking up enthusiasm of scientists because of its flexibility and precision. In assumption examination, generally the managed learning variations of this strategy are utilized. It includes three phases: Data gathering, Pre-handling, Training information, Classification and plotting results. In the preparation information, an accumulation of labelled corpora is given. The Classifier is introduced a progression of highlight vectors from the past information. A model is made dependent on the preparation informational collection which is utilized over the new/concealed content for characterization reason. In machine learning strategy, the way to exactness of a classifier is the determination of suitable highlights. For the most part, unigrams (single word phrases), bi-grams (two continuous expressions), tri-grams (three successive expressions) are chosen as highlight vectors. There is an assortment of proposed includes number of positive words, number of negative words, length of the report, Support Vector Machines (SVM) [24], and Naïve Bayes (NB) calculation [5]. Exactness is accounted for to change from 63% to 80% contingent on the mix of different features chosen.

Machine learning approach follows the given steps below:

1. Data Collecting – in this stage information to be broke down to analyse from different sources like Blogs, Social systems (Twitter, Myspace, and so on.) contingent on the zone of use.
2. Pre-handling – In this stage, the gained information is cleaned and prepared for bolstering it into the classifier. Cleaning incorporates extraction of catchphrases what's more, images. For example – Emoticons are the smiley utilized in printed structure to speak to feelings for example ":-)", ":", "=)", ":", "D", ".:", ";:", and so on.. Amending the all capitalized and all lowercase to a typical case, expelling the non-English (or proffered language writings), evacuating un-vital blank areas and tabs, and so forth.

3. Training Data – A hand-labelled accumulation of information is set up by most normally utilized publicly supporting technique. This information is the fuel for the classifier; it will be nourished to the calculation for learning reason.
4. Classification – This is the core of the entire procedure. Depending upon the prerequisite of the application SVM or Naïve Bayes is sent for investigation. The classifier (in the wake of finishing the preparation) is fit to be sent to the constant tweets/content for supposition extraction reason.
5. Results – Results are plotted dependent on the sort of portrayal chosen for example diagrams, charts, and so forth. Execution tuning is done preceding the arrival of the calculation.

5.2 Lexicon Based Approach

In unsupervised technique, order is finished by looking at the highlights of a given book against sentiment lexicon whose assumption esteems are decided preceding their utilization. Assessment vocabulary contains arrangements of words and articulations used to express individuals' abstract emotions and sentiments [1]. For instance, start with positive and negative word dictionaries, investigate the record for which assumption need to discover. At that point if the archive has increasingly positive word dictionaries, it is certain, else it is negative. The dictionary-based systems to Sentiment examination is unaided learning since it doesn't require earlier preparing to characterize the information.

The essential steps of the lexicon-based methods are listed below:

1. Pre-process every content (for example remove HTML labels, noisy characters)
2. Initialize the absolute sentiment score: $s = 0$.
3. Tokenize content. For every token, check if it is available in an opinion word reference. (an) If token is available in word reference,
 - i. On the off chance that token is certain, at that point $s = s + w$.
 - ii. On the off chance that token is negative, at that point $s = s - w$.
4. Consider the total sentiment score s , (an) If $s > \text{limit}$, at that point group the content as positive.
5. There are three strategies to develop an assessment dictionary: manual development, corpus-based techniques also, dictionary-based strategies. The manual development of assessment vocabulary is a troublesome undertaking. In lexicon-based systems the thought is to first gather a little arrangement of assessment words physically with known directions, and afterward to develop this set by looking in the WordNet lexicon for their equivalent words and antonyms. The recently discovered words are added to the seed list. The following emphasis begins. The iterative procedure stops when not any newer words are found. Supposition words share a similar direction as their equivalent words and inverse directions as their antonyms.
6. Corpus put together procedures depend with respect to syntactic examples in enormous corpora. Corpus-based techniques can create feeling words with moderately high precision. A large portion of these corpus-based techniques need extremely huge marked preparing information. This methodology has a significant bit of that

the word reference-based methodology doesn't have. It can help discover space explicit assessment words and their directions.

6 ACTUAL SURVEY

6.1 Problem Discussed

Sentiment analysis in social media is hard than in alternative kinds of text thanks to limits like abbreviations, short text for words, and references to existing content or ideas. Social media provides noticeably info apart from text, for instance hooked up media, user reactions, and relations between users. this is often referred to as social context [6]. Recent works have with success gain advantage from the fusion of text with social context for sentiment analysis tasks. However, these works are sometimes restricted to specific aspects of social context, and there haven't been any tries to investigate and apply social context consistently.

6.2 Traditional Methods and their disadvantages

Ancient approaches use the Bag Of Words model, wherever a document is mapped to a feature vector, and so classified by machine learning techniques. Though the Bag Of Words approach is easy and economical, a good amount of information from original language is lost. So, numerous kinds of options are exploited, like higher order n-grams. Another quite feature that may be used is an element Of Speech tagging, that is often used throughout a syntactical analysis method. This option is understood as surface forms. Lexical resources are very useful in the field of Sentiment analysis, as they present a resource that gives sentimental information directly [12]. Normally, sentiment lexica are used to find polarity by word matching and their associated lexicon sentiment polarities within the text. Such resources have disadvantages in vocabulary coverage and domain adaptation.

6.3 Modern Approaches

Besides this, several recent techniques exploit the idea of distributed linguistics, ordinarily through word embeddings. A linguistics similarity metric is computed between text words and lexica vocabulary in [1]. Victimization this metric, this paper proposes a sentiment classification model that uses the linguistics similarity live together with embedding representations [13]. Since there's some downside concerning the distinction between opinion, sentiment and feeling, they outlined opinion as a shift idea that shows the perspective towards that entity. Feeling words square measure utilized in various opinion preparation undertakings. Positive conclusion words square measure utilized to specific some ideal states, whereas negative feeling words square measure utilized to specific some unsought states. There square measure likewise assessment expressions and figures of speech that along square measure known as opinion lexicon [3].

6.3.1 Deep learning in Sentiment Analysis

Deep learning techniques for Sentiment Analysis is turning into extremely popular [2]. It provides automatic feature extraction, higher depiction capabilities and sensible presentation as compared to ancient feature-based techniques. These long-established approaches will yield robust baselines, and their prophetic capabilities are often employed in conjunction with the arising deep learning strategies [2]. during this paper, the authors look for to boost

the performance of deep learning techniques group action them with ancient surface approaches supported manually extracted options. As of late, mining web-based life

Table 1. Comparison of first six papers destinations like

meth ods	Algorithm used	Work scope	Dataset
[1]	Semantic similarity evaluation Embedding text representation SIMilarity-based sentiment projectiON (SIMON)	Web pages, ecommerce sites	SemEval20 13 SemEval20 14 Vader STS-Gold Sentiment-140
[2]	deep neural networks, automatic feature extraction, traditional machine learning approach	Movie reviews, Product review	IMDBs
[3]	Clustering algorithm, Feature selection method, Machine learning algorithm	Movie, Product Reviews	IMDB, Amazon.co m
[4]	Dictionary and corpus-based algorithms, Porter's stemming quality.	Document reviews, product reviews	N/A
[5]	Latent-semantic algorithm, Machine learning algorithm	Web surveys	Amazon reviews
[6]	Social context in sentiment analysis, POS tagging, Lexicon approaches	Comparative study	RT Mind OMD SemEval datasets Health Care Reform (HCR)

Twitter, Facebook have bowed into a hot center. Twitter is one among the foremost fashionable microblogging administrations that enables shoppers to precise their views on web-based business sites, sports, legislative problems, gift day advancements, motion photos, etc. Feeling are often thought-about as a characteristic articulation of watcher's observation. it's arduous to acknowledge the sensation regarding associate degree item or occasion by gathering and transcription microblog assignments physically [4]. Also, these days, fashionable abbreviated words square measure notably utilized in content informing as a way for simplicity of composing a protracted message. The present techniques do eliminate these cutting-edge pattern words. To conquest the

impediment, a marking model is planned [4], that fuses the utilization of abbreviated words and emojis. With the top goal of this work, we've gathered summing up of abbreviated words and emojis.

6.3.2 Machine learning in Sentiment Analysis

Utilizing both machine learning and Lexicon-based methodology was introduced by Rathor [5]. This paper centers around taking a gander at the viability of three AI procedures (Naive Bayes (NB), Support Vector Machines (SVM), and Maximum Entropy (ME)) for classification of online reviews using a web model using managed learning methods. The fundamental issue of this slant examination of client surveys is that characterization of assessments of clients as negative, positive or nonpartisan is a repetitive and extreme employment to be practiced. The outcomes have demonstrated that machine learning computations work honorably on weighted unigrams and SVM has come about most prominent precision. This isn't helpful for purchasers those need to glance through the things going before purchase yet also for associations those need to watch the open's reaction to their things. In web time, advance being used of online life goals, for instance, twitter, Facebook, and review website page conveys a tremendous proportion of printed information. The printed information fills a basic source to perceive customer's presumption towards ideological gathering, things or an event. The inclinations imparted by customers are in the structure of positive, negative or fair limit. The printed information in online life goals accept a noteworthy activity in decision genuinely steady systems and individual decision producers. The technique of mechanizing familiar evidence of appraisal in a substance is known as Sentiment Analysis. The online web is significantly changing and the assessments of each client is on the web. There are different continuing examinations and proposed modals for such sentiment analysis. This survey paper talks about the current takes and shot at opinion mining and conclusion characterization of client criticism and audits on the web, and assesses the various methods utilized for the procedure. This look on the zones encased by the assessed papers, brings up the territories that are pleasantly secured by different specialists and regions that are ignored in conclusion mining and slant order which are open for future research works.

7 FINDINGS

8 DISCUSSION AND ANALYSIS

Datasets are very important thing in sentiment analysis. IMDB and Amazon.com are acclaimed information wellsprings of survey information. IMDB is a wellspring of motion picture surveys while amazon.com is a wellspring of numerous item audits. These information sources are utilized in sentiment analysis and sentiment classification errands. It is seen that twitter was utilized every now and again in the most recent year. Twitter is an extremely acclaimed interpersonal organization site where its tweets express individuals' conclusions and its length is greatest 140 characters. Machine learning calculations are normally used to tackle the sentiment classification issue for its straightforwardness and the capacity to utilize the preparation information which gives it the benefit of space flexibility. Lexicon based calculations are as often as possible used to tackle general sentiment analysis issues in

view of their adaptability. They are additionally basic and computationally efficient. The number and level of articles that utilization machine learning and the Lexicon-based calculations are changing among years. The general work for the ongoing couple of years demonstrates that the analysts are utilizing vocabulary-based methodology all the more every now and again. This is on the grounds that it explains numerous sentiment analysis errands in spite of its high multifaceted nature. Machine learning methodologies are as yet an open field of pursuit.

Table 2. Datasets

Methods	Data Set/Source
[2]	Twitter
[35]	Sentiment dictionaries
[10]	Live Journals Blogs, text, Annotated Blogs
[23]	Online product reviews
[22]	Article reviews
[16]	TSCCA, PCA, C&W, CBOW
[21]	Blogs, sentiment dictionaries
[5]	amazon.com
[18]	Sentiwordnet
[4]	Twitter
[28]	amazon.com, opinions, blogs
[8]	Twitter and review data From social platforms
[3]	amazon.com
[17]	Twitter
[1]	IMDB
[13]	WS353-Sim, M&C datasets, R&G
[27]	MLL, CNS, TR11, TR23
[19]	Twitter
[20]	Twitter and review data From social sites

9 CONCLUSION

This survey paper gives an impact on the ongoing educates in sentiment analysis calculations and applications. These articles offer commitments to numerous feeling examination related fields that utilizations Sentiment analysis systems for genuine application. In the wake of contrasting these articles,

plainly the enhancements of notion investigation calculations are as yet an open field for examine. The data given by these assumptions can be incorporated into existing informal organization investigation. To discover potentially radicalized clients and to highpoint expressively charged substance, vocabulary-based strategies can be utilized for assessment investigation. A comprehensive vocabulary-based way to deal with assessment mining is additionally another way which can be actualized. Appropriated portrayals of sentences and reports likewise should be possible. Utilizing computational investigation in sentiment analysis is another method for approach. Naïve Bayes and Support Vector Machines are the most used machine learning computations for dealing with sentiment classification problems. Curtailed dictionary and emoticons list area unit accessorial in the orchestrated evaluation model to expand the precision of picking incline furthest point. To discover the tweets are designated positive, negative or fair-minded tweets.

ACKNOWLEDGMENT

We would like to thank Mr. Saravanakumar K for introducing us to this topic and giving us an opportunity to explore such an interesting topic.

REFERENCES

- [1] Araque, Oscar, et al. "A Semantic Similarity-Based Perspective of Affect Lexicons for Sentiment Analysis." *Knowledge-Based Systems*, vol. 165, 2019, pp. 346–359., doi:10.1016/j.knosys.2018.12.005.
- [2] Araque, Oscar, et al. "Enhancing Deep Learning Sentiment Analysis with Ensemble Techniques in Social Applications." *Expert Systems with Applications*, vol. 77, 2017, pp. 236–246., doi:10.1016/j.eswa.2017.02.002.
- [3] Kumar, H. M., & Harish, B. S. (2018). Sarcasm classification: A novel approach by using Content Based Feature Selection Method. *Procedia Computer Science*,143, 378-386. doi:10.1016/j.procs.2018.10.409
- [4] J, A., & G, J. (2018). Sentiment Classification of Tweets with Non-Language Features. *Procedia Computer Science*,143, 426-433. doi:10.1016/j.procs.2018.10.414
- [5] Rathor, A. S., Agarwal, A., & Dimri, P. (2018). Comparative Study of Machine Learning Approaches for Amazon Reviews. *Procedia Computer Science*,132, 1552-1561. doi:10.1016/j.procs.2018.05.119
- [6] Sánchez-Rada, J. F., & Iglesias, C. A. (2019). Social context in sentiment analysis: Formal definition, overview of current trends and framework for comparison. *Information Fusion*, 52, 344-356. doi:10.1016/j.inffus.2019.05.003
- [7] Tommasel, Antonela, and Daniela Godoy. "A Social-Aware Online Short-Text Feature Selection Technique for Social Media." *Information Fusion*, vol. 40, 2018, pp. 1–17., doi:10.1016/j.inffus.2017.05.003.
- [8] Yang, Yi, and Jacob Eisenstein. "Overcoming Language Variation in Sentiment Analysis with Social Attention." *Transactions of the Association for Computational Linguistics*, vol. 5, 2017, pp. 295–307., doi:10.1162/tacl_a_00062.
- [9] Dong, Yucheng, et al. "A Survey on the Fusion Process in Opinion Dynamics." *Information Fusion*, vol. 43, 2018, pp. 57–65., doi:10.1016/j.inffus.2017.11.009.
- [10] Cambria, Erik, et al. "Affective Computing and Sentiment Analysis." *A Practical Guide to Sentiment Analysis Socio-*

- Affective Computing, 2017, pp. 1–10., doi:10.1007/978-3-319-55394-8_1.
- [11] Xiaomei, Zou, et al. "Microblog Sentiment Analysis with Weak Dependency Connections." *Knowledge-Based Systems*, vol. 142, 2018, pp. 170–180., doi:10.1016/j.knosys.2017.11.035.
- [12] Hailong, Zhang, et al. "Machine Learning and Lexicon Based Methods for Sentiment Classification: A Survey." 2014 11th Web Information System and Application Conference, 2014, doi:10.1109/wisa.2014.55.
- [13] Wanjari, Harshal. "A Hybrid Approach for Computing Semantic Similarity of Concepts in Knowledge Graphs." *International Journal for Research in Applied Science and Engineering Technology*, vol. 6, no. 3, 2018, pp. 2186–2189., doi:10.22214/ijraset.2018.3345.
- [14] Sánchez, David, et al. "Ontology-Based Semantic Similarity: A New Feature-Based Approach." *Expert Systems with Applications*, vol. 39, no. 9, 2012, pp. 7718–7728., doi:10.1016/j.eswa.2012.01.082.
- [15] Liu, Man. "EmoNLP at IEST 2018: An Ensemble of Deep Learning Models and Gradient Boosting Regression Tree for Implicit Emotion Prediction in Tweets." Proceedings of the 9th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis, 2018, doi:10.18653/v1/w18-6228.
- [16] Schnabel, Tobias, et al. "Evaluation Methods for Unsupervised Word Embeddings." Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing, 2015, doi:10.18653/v1/d15-1036.
- [17] Nakov, P. (2016). Sentiment Analysis in Twitter: A SemEval Perspective. Proceedings of the 7th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis. doi: 10.18653/v1/w16-0427
- [18] Ohana, Bruno, and Brendan Tierney. "Opinion Mining with Sent WordNet." *Knowledge Discovery Practices and Emerging Applications of Data Mining*, doi:10.4018/9781609600679.ch013.
- [19] Gimpel, Kevin, et al. "Part-of-Speech Tagging for Twitter: Annotation, Features, and Experiments." Jan. 2010, doi:10.21236/ada547371.
- [20] Severyn, Aliaksei, and Alessandro Moschitti. "UNITN: Training Deep Convolutional Neural Network for Twitter Sentiment Classification." Proceedings of the 9th International Workshop on Semantic Evaluation (SemEval 2015), 2015, doi:10.18653/v1/s15-2079.
- [21] Melville, Prem, et al. "Sentiment Analysis of Blogs by Combining Lexical Knowledge with Text Classification." Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining - KDD '09, 2009, doi:10.1145/1557019.1557156.
- [22] Kaur, B., & Kumari, N. (2016). A Hybrid Approach to Sentiment Analysis of Technical Article Reviews. *International Journal of Education and Management Engineering*,6(6), 1-11.doi:10.5815/ijeme.2016.06.01
- [23] Chen, C. C., & Tseng, Y. (2011). Quality evaluation of product reviews using an information quality framework. *Decision Support Systems*, 50(4), 755–768.doi:10.1016/j.dss.2010.08.023
- [24] Pan, H., Li, X., Jin, L., & Xia, L. (2011). A Binary Particle Swarm Optimization and Support Vector Machine-based Algorithm for Object Detection. *Journal of Electronics & Information Technology*,33(1), 117-121. doi:10.3724/sp.j.1146.2010.00251
- [25] Lo, Y., & Potdar, V. (2009). A review of opinion mining and sentiment classification framework in social networks. 2009 3rd IEEE International Conference on Digital Ecosystems and Technologies. doi:10.1109/dest.2009.5276705
- [26] Figueira, Á., & Oliveira, L. (2016). Analyzing Social Media Discourse - An Approach using Semi-supervised Learning. *Proceedings of the 12th International Conference on Web Information Systems and Technologies*. doi:10.5220/0005786601880195
- [27] Chormunge, S., & Jena, S. (2018). Correlation based feature selection with clustering for high dimensional data. *Journal of Electrical Systems and Information Technology*,5(3), 542-549. doi:10.1016/j.jesit.2017.06.004
- [28] Nagwanshi, P., & Madhavan, C. E. (2014). Sarcasm Detection using Sentiment and Semantic Features. *Proceedings of the International Conference on Knowledge Discovery and Information Retrieval*. doi:10.5220/0005153504180424
- [29] Mukherjee, S., & Bala, P. K. (2017). Sarcasm detection in microblogs using Naïve Bayes and fuzzy clustering. *Technology in Society*,48, 19-27. doi:10.1016/j.techsoc.2016.10.003
- [30] Tripathy, A., Agrawal, A., & Rath, S. K. (2016). Classification of sentiment reviews using n-gram machine learning approach. *Expert Systems with Applications*,57, 117-126. doi:10.1016/j.eswa.2016.03.028
- [31] Wang, G., Sun, J., Ma, J., Xu, K., & Gu, J. (2014). Sentiment classification: The contribution of ensemble learning. *Decision Support Systems*,57, 77-93. doi:10.1016/j.dss.2013.08.002
- [32] Nakov, P. (2017). Semantic Sentiment Analysis of Twitter Data. *Encyclopedia of Social Network Analysis and Mining*,1-12. doi:10.1007/978-1-4614-7163-9_110167-1
- [33] Vilares, D., Alonso, M. A., & Gómez-Rodríguez, C. (2015). Sentiment Analysis on Monolingual, Multilingual and Code-Switching Twitter Corpora. *Proceedings of the 6th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis*. doi:10.18653/v1/w15-2902
- [34] Williams, P., Sennrich, R., Post, M., & Koehn, P. (2016). Syntax-based Statistical Machine Translation. *Synthesis Lectures on Human Language Technologies*,9(4), 1-208. doi:10.2200/s00716ed1v04y201604hlt033
- [35] Cernian, A., Sgariu, V., & Martin, B. (2015). Sentiment analysis from product reviews using SentiWordNet as lexical resource. 2015 7th International Conference on Electronics, Computers and Artificial Intelligence (ECAI). doi:10.1109/ecai.2015.7301224
- [36] Fersini, E., Messina, E., & Pozzi, F. (2016). Expressive signals in social media languages to Improve polarity detection. *Information Processing & Management*,52(1), 20-35. doi:10.1016/j.ipm.2015.04.004
- [37] Chalothom, Tawunrat, and Jeremy Ellman. "Simple Approaches of Sentiment Analysis via Ensemble Learning." *Lecture Notes in Electrical Engineering Information Science and Applications*, 2015, pp. 631–639., doi:10.1007/978-3-662-46578-3_74.
- [38] "A Study on Sentiment Analysis: Methods and Tools." *International Journal of Science and Research*

- (IJSR), vol. 4, no. 12, May 2015, pp. 287–292., doi:10.21275/v4i12.nov151832.
- [39] Yin, Yichun, et al. "Document-Level Multi-Aspect Sentiment Classification as Machine Comprehension." Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing, 2017, doi:10.18653/v1/d17-1217.
- [40] Farra, Noura, et al. "Sentence-Level and Document-Level Sentiment Mining for Arabic Texts." 2010 IEEE International Conference on Data Mining Workshops, 2010, doi:10.1109/icdmw.2010.95.
- [41] E. Nikos, L. Angeliki , P . Georgios, C. Konstantinos ELS: a word-level method for entity-level sentiment analysis WIMS'11 Proceedings of the International Conference on Web Intelligence, Mining and Semantics (2011) doi:10.1145/1988688.1988703.
- [42] Li, Chunping, et al. "Modeling Sentiment and Aspect Using Syntax: A Topic Model Approach." Proceedings of The 5th IIAE International Conference on Industrial Application Engineering 2017, 2017, doi:10.12792/icisip2017.036.
- [43] Khan, Khairullah, et al. "Mining Opinion Components from Unstructured Reviews: A Review." Journal of King Saud University - Computer and Information Sciences, vol. 26, no. 3, 2014, pp. 258–275., doi:10.1016/j.jksuci.2014.03.009.
- [44] Pozzi, F.a., et al. "Challenges of Sentiment Analysis in Social Networks." Sentiment Analysis in Social Networks, 2017, pp. 1–11., doi:10.1016/b978-0-12-804412-4.00001-2.)
- [45] Liu, Bing, and Lei Zhang. "A Survey of Opinion Mining and Sentiment Analysis." Mining Text Data, 2012, pp. 415–463., doi:10.1007/978-1-4614-3223-4_13.
- [46] Mukherjee, Arjun, et al. "Spotting Fake Reviewer Groups in Consumer Reviews." Proceedings of the 21st International Conference on World Wide Web - WWW 12, 2012, doi:10.1145/2187836.2187863.