Evaluation of product rating using data mining

Priti Kumari[#1], Upasana Dugal [*2]

[#]Dept. of Computer Science & Engineering, BBD University, Uttar Pradesh, India

[1]kprit806@gmail.com

[2]upasana_gupta31@@bbdu.ac.in

***Abstract*—** **Sentiment Analysis (SA) is a progressing field of research in content mining field. SA is the computational treatment of feelings, conclusions and subjectivity of content. This paper handles an exhaustive diagram of the last refresh in this field. Numerous as of late proposed calculations' upgrades and different SA applications are researched and exhibited quickly in this study. These articles are arranged by their commitments in the different SA strategies. The related fields to SA (exchange learning, feeling location, and building assets) that pulled in analysts as of late are talked about. The principle focus of this paper is to give almost full picture of SA strategies and the related fields with brief subtleties. This study covers procedures and methodologies that guarantee to straightforwardly empower supposition arranged data looking for frameworks. Our attention is on techniques that look to address the new difficulties raised by supposition mindful applications, when contrasted with those that are as of now present in increasingly conventional truth based investigation. We incorporate material on synopsis of evaluative content and on more extensive issues with respect to protection, control, and financial effect that the improvement of feeling focused twitter data get to administrations offers ascend to. To encourage future work, a dialog of accessible assets, benchmark datasets, and assessment crusades is additionally given.**

Keyword: ***Sentiment Analysis*, methodologies, *twitter data*, framework.**

I. Introduction

With the expanding prominence of long range interpersonal communication, blogging and miniaturized scale blogging sites, each day an enormous measure of casual abstract content articulations are made accessible on the web. The data caught from these writings, could be utilized for logical overviews from a social or political point of view [9]. Organizations and item proprietors who plan to enhance their items/administrations may emphatically profit by the rich input [6], [15]. Then again, clients could likewise find out about inspiration or cynicism of various highlights of items/administrations as indicated by clients' sentiments, to make an informed buy. Besides, applications like rating motion pictures dependent on online motion picture surveys [12] couldn't develop without making utilization of these information. A Sentiment Summarization framework takes as information a lot of records that contain suppositions about some substance of intrigue. Thusly, it forms all the given archives and produces an outline of all the information reports. This rundown ought to speak to the normal supposition of the considerable number of reports and critical parts of the objective of SA tended to in those records. This enables the two clients and organizations to have simple access to general's conclusion with respect to specific things/items. There are two primary ways to deal with creating printed rundowns. The primary technique, known as extractive-based outline, is removing some as far as anyone knows vital parts of the correct messages in a corpus of records and introducing them as a rundown of that corpus. The second strategy, known as abstractive outline, is producing a printed rundown in which the utilized words are not really the ones utilized in the corpus. There has been a lot of past research on different strategies to use the web innovation to expand the advantages of clients and in addition organizations in the commercial center [2]. In like manner, this examination seeks after a similar objective by performing buyer items based SA and rundown in the area of cell phones on Twitter information. For this reason, we physically commented on Twitter information for our examinations. Conclusion examination is where the dataset comprises of feelings, dispositions or evaluation which considers the way a human thinks [1]. In a sentence, attempting to comprehend the positive and the negative angle is an extremely troublesome errand. The highlights used to characterize the sentences ought to have an extremely solid modifier so as to condense the survey. These substances are even written in various methodologies which are not actually derived by the clients or the organizations making it hard to group them. Supposition examination impacts clients to group whether the data about the item is tasteful or not before they gain it. Advertisers and firms utilize this investigation to comprehend about their items or administrations so that it very well may be offered according to the client's needs. There are two sorts of machine learning systems which are commonly utilized for notion examination, one is unsupervised and the other is administered [2]. Unsupervised learning does not comprise of a class and they don't furnish with the right focuses at all and hence lead grouping. Administered learning depends on marked dataset and therefore the names are given to the model amid the procedure. These marked dataset are prepared to deliver sensible yields when experienced amid basic leadership. To assist us with understanding the notion investigation bitterly, this exploration paper depends on the managed machine learning. In rundown, this paper shows a thorough and inside and out basic appraisal of 15 Sentiment Analysis web devices that has never been finished. To legitimately play out this appraisal, a suite of assessment criteria and surely understood information accumulations from the field of Sentiment Analysis has been chosen to enable the per user to investigate the upsides and downsides of the utilization of these instruments viewing perspectives, for example, revelation of opinions inside short and long messages, discovery of incongruity or calculation of extremity evaluations, among others. Aside from these standard information accumulations, these apparatuses have likewise been surveyed by imitating an all the more genuine situation, in which the adequacy for prescribing motion pictures from genuine clients' remarks has been tried utilizing data gathered from the notable site IMDb1. The rest of the work is sorted out as pursues: Section 2 exhibits the primary ideas identified with Sentiment Analysis examined in a few late works. Segment 3 demonstrates the data description. Area 4 exhibits a few techniques that have been performed so as to think about the Web administrations remarked on the past segment, and also the outcomes got. At last, Section 5 points out a few ends.

II. Literature Survey

Conclusion investigation has been taken care of as a Natural Language Processing errand at numerous dimensions of granularity. Beginning from being a report level order assignment (Turney, 2002; Pang and Lee, 2004), it has been dealt with at the sentence level (Hu and Liu, 2004; Kim and Hovy, 2004) and all the more as of late at the expression level (Wilson et al., 2005; Agarwal et al., 2009). Microblog information like Twitter, on which clients present ongoing responses on and sentiments about "everything", presents more current and diverse difficulties. A portion of the early and ongoing outcomes on assessment investigation of Twitter information are by Go et al. (2009), (Bermingham and Smeaton, 2010) and Pak and Paroubek (2010). Go et al. (2009) utilize far off figuring out how to gain conclusion information. They use tweets finishing off with positive emojis like ":)" ":-)" as positive and negative emojis like ":(" ":- (" as negative. They assemble models utilizing Naive Bayes, MaxEnt and Support Vector Machines (SVM), and they report SVM beats different classifiers. Regarding highlight space, they attempt a Unigram, Bigram demonstrate related to parts-of-discourse (POS) highlights. They take note of that the unigram show outflanks every other model. In particular, bigrams and POS highlights don't help. Pak and Paroubek (2010) gather information following a comparative far off learning worldview. They play out an alternate grouping assignment however: emotional versus objective. For emotional information they gather the tweets finishing with emojis in indistinguishable way from Go et al. (2009). For target information they creep twitter records of famous papers like "New York Times", "Washington Posts" and so on. They report that POS and bigrams both help (in spite of results introduced by Go et al. (2009)). Both these methodologies, be that as it may, are fundamentally founded on ngram models. Besides, the information they use for preparing and testing is gathered via seek inquiries and is consequently one-sided. Conversely, we present highlights that accomplish a huge increase over a unigram benchmark. Moreover we investigate an alternate strategy for information portrayal and report critical enhancement over the unigram models. Another commitment of this paper is that we report results on physically commented on information that does not experience the ill effects of any known predispositions. Our information is an irregular example of gushing tweets not at all like information gathered by utilizing explicit inquiries. The span of our hand-named information enables us to perform cross approval investigations and check for the change in execution of the classifier crosswise over folds. Another huge exertion for slant characterization on Twitter information is by Barbosa and Feng (2010). They use extremity forecasts from three sites as uproarious marks to prepare a model and utilize 1000 physically named tweets for tuning and another 1000 physically named tweets for testing. They anyway don't specify how they gather their test information. They propose the utilization of linguistic structure highlights of tweets like retweet, hashtags, connection, accentuation and outcry checks related to highlights like earlier extremity of words and POS of words. We expand their methodology by utilizing genuine esteemed earlier extremity, and by consolidating earlier extremity with POS. Our outcomes demonstrate that the highlights that upgrade the execution of our classifiers the most are highlights that consolidate earlier extremity of words with their parts of discourse. The tweet language structure highlights help yet just imperceptibly. Gamon (2004) perform notion examination on feeadback information from Global Support Services overview. One point of their paper is to break down the job of phonetic highlights like POS labels. They perform broad component investigation and highlight determination and exhibit that dynamic etymological examination highlights adds to the classifier precision.

III. Data Description

Twitter is a person to person communication and small scale blogging administration that enables clients to post constant messages, called tweets. Tweets are short messages, confined to 140 characters long. Because of the idea of this micro blogging administration (fast and short messages), individuals use abbreviations, commit spelling errors, use emojis and different characters that express extraordinary implications. Following is a concise wording related with tweets. Emojis: These are outward appearances pictorially spoken to utilizing accentuation and letters; they express the client's temperament. Target: Users of Twitter utilize the "@" image to allude to different clients on the microblog. Alluding to different clients as such consequently cautions them. Hashtags: Users more often than not utilize hashtags to stamp subjects. This is essentially done to expand the perceivability of their tweets.

IV. TECHNIQUES

A few works endeavor to demonstrate the diverse systems connected to Sentiment Analysis. A large portion of them bunch the works from the perspective of the distinctive applications/challenges that can be found in SA as in (Pang and Lee, 2008) and (Liu and Zhang, 2012). Different works like (Tsytsarau and Palpanas, 2011) or (Feldman, 2013) are centered around the primary points of SA. In this manner, Feldman bunches all works under five primary gatherings: report level supposition investigation, sentence-level feeling examination, perspective based slant investigation, similar assumption examination and, conclusion vocabulary obtaining (Feldman, 2013). What's more, then again, Tsytsarau and Palpanas for the most part center around conclusion total, feeling spam and logical inconsistencies investigation, particularly connected to Web administrations, for instance, microblogs or gushing information, among others (Tsytsarau and Palpanas, 2011). They present four unique focuses concerning past attempts to order Sentiment Analysis methods: machine learning, word reference based, factual and semantic. Potentially, the most fascinating work from the perspective of the SA systems is (Medhat et al., 2014), which displays a refined arrangement of surely understood SA procedures (see Fig. 1) including new patterns, for example, Emotion Detection (Rao et al., 2014), Building Resources and Transfer Learning.

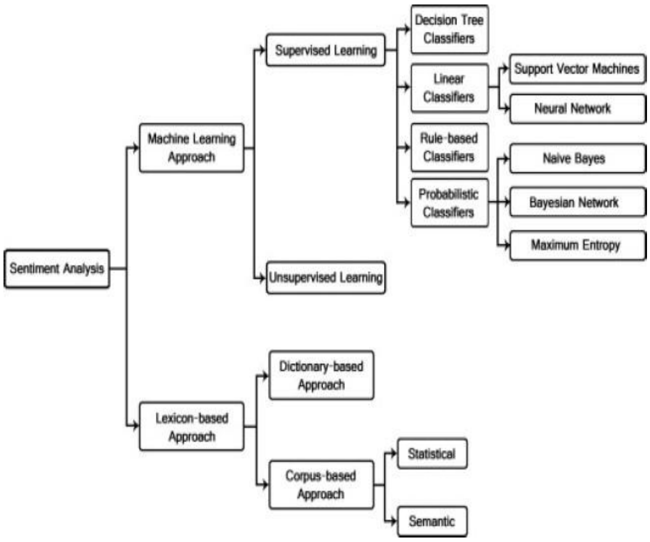

Figure 1: Sentiment classification techniques (Medhat et al., 2014)

A. Machine learning approaches

They can be assembled in two fundamental classifications: managed and unsupervised methods. The accomplishment of both is for the most part dependent on the determination and extraction of the suitable arrangement of highlights used to recognize opinions. In this errand Natural Language Processing strategies assume an imperative job since probably the most critical highlights utilized are for instance: (1) terms (words or n-grams) and their recurrence; (2) grammatical form data, descriptive words assume a vital job yet things can be noteworthy; (3) refutations can change the importance of any sentence; (4) syntactic conditions (tree parsing) can decide the significance of sentence; among others (Liu and Zhang, 2012; Chenlo and Losada, 2014). As for regulated procedures, bolster vector machines (SVM), Naive Bayes, Maximum Entropy are the absolute most normal methods utilized (Ye et al., 2009; Rushdi-Saleh et al., 2011; Montejo-Raez et al., 2014). While semi-directed and unsupervised methods are proposed when it is beyond the realm of imagination to expect to have an underlying arrangement of marked reports/suppositions to group whatever remains of things (He and Zhou, 2011; Xianghua et al., 2013). In addition, crossover approaches, consolidating directed and unsupervised procedures, or even semi-regulated strategies, can be utilized to order feelings (Kim and Lee, 2014; Konig and Brill, 2006).

B. Lexicon-based approaches

Dictionary put together methodologies for the most part depend with respect to a feeling vocabulary, i.e., a gathering of known and precompiled supposition terms, states and even figures of speech, produced for customary types of correspondence, for example, the Opinion Finder dictionary (Wilson et al., 2005a); be that as it may, considerably progressively complex structures like ontologies (Kontopoulos et al., 2013), or lexicons estimating the semantic introduction of words or expressions (Turney, 2002; Taboada et al., 2011) can be utilized for this reason. Two sub characterizations can be found here: Dictionary-based and Corpus based methodologies. The previous is normally founded on the utilization of an underlying arrangement of terms (seeds) that are typically gathered and explained physically. This set develops via looking through the equivalent words and antonyms of a lexicon. A case of that lexicon may be WordNet (Miller, 1995), which was utilized to build up a thesaurus called SentiWordNet (Baccianella et al., 2010). The principle downside of this sort of methodologies is the lack of ability to manage space and setting explicit introductions; all things being equal, it may be an intriguing arrangement relying upon the issue (Martin-Valdivia et al., 2012; Montejo-Raez et al., 2014). The corpus-based strategies emerge with the target of giving word references identified with an explicit area. These lexicons are created from a lot of seed sentiment terms that becomes through the pursuit of related words by methods for the utilization of either measurable or semantic systems. Strategies dependent on measurements, for example, Latent Semantic Analysis (LSA) (Deerwester et al., 1990), or basically the recurrence of event of the words inside a gathering of records can be utilized (Cao et al., 2011). What's more, on other hand, semantic strategies, for example, the utilization of equivalent words and antonyms or connections from thesaurus like WordNet may likewise speak to an intriguing arrangement (Zhang et al., 2012). 2.4. Regular Language Processing and Information Retrieval in Sentiment Analysis According to Cambria, Sentiment Analysis can be considered as an extremely limited NLP issue, where it is just important to comprehend the positive or negative estimations concerning each sentence as well as the objective elements or themes (Cambria et al., 2013). In any case, regardless of being a limited issue, all works in this field, and all works in Information Retrieval, dependably battle with NLPs uncertain issues (invalidation taking care of, named-element acknowledgment, word-sense disambiguation,) which are fundamental to recognize scholarly gadgets, for example, incongruity or mockery (Reyes and Rosso, 2012; Reyes et al., 2012), and thus, to discover and rate conclusions. One of the primary angles that NLP needs to manage is the distinctive dimensions of investigation. Contingent upon whether the objective of study is an entire content or archive, one or a few connected sentences, or one or a few substances or parts of those elements, diverse NLP and Sentiment Analysis undertakings can be performed. Thus, it is important to recognize three dimensions of investigation that will obviously decide the distinctive undertakings of Sentiment Analysis: (I) report level, (ii) sentence level and (iii) element/angle level. Report level thinks about that a record is an assessment on a substance or part of it. This dimension is related with the undertaking called report level opinion characterization (Zhang et al., 2009; Moraes et al., 2013; Duric and Song, 2012; He and Zhou, 2011; Zhou et al., 2010; Yessenalina et al., 2010; Paltoglou and Thelwall, 2010; Li and Liu, 2012). Notwithstanding, in the event that a report gives a few sentences managing distinctive viewpoints or elements, the sentence level is progressively appropriate. Sentence level is firmly identified with the assignment called subjectivity order, which recognizes sentences that express verifiable data from sentences that express emotional perspectives and sentiments (Wilson et al., 2005b, 2009; Agarwal et al., 2009; Remus and Hianig, 2011). Lastly, when increasingly exact data is vital, at that point the element/perspective dimension emerges. It is the _nest-grained level, it considers an objective on which the conclusion holder communicates a positive or negative feeling. This last dimension is conceivably the most perplexing in light of the fact that it is important to separate with high exactness numerous highlights, for example, dates or time ranges, the distinctive highlights/perspectives and elements to be obstinate, alongside the relations between them, the assessment holders and their qualities, and so on. It is firmly identified with undertakings like Feature-based Opinion Mining and Opinion Summarization (Thet et al., 2010; Ojokoh and Kayode, 2012; Vechtomova, 2010). A significant number of these papers pursue indistinguishable general procedures from other Information Retrieval works did previously, however supplanting a few factual or semantic factors for angles identified with assumptions. For instance, Vechtomova and Karamuftuoglu (2008) propose the utilization of lexical attachment, i.e., the \physical" remove between collocations to rank archives while in feeling positioning Vechtomova proposes a comparable strategy, yet estimating the separation between abstract words (Vechtomova, 2010). In this way, the principle distinction between these works is the element determination process. Another model may be the work introduced in (Moraes et al., 2013), that applies surely understood regulated techniques for the Artificial Neural Networks and Support Vector Machines to Sentiment Classification, which have been utilized a huge number of times in Information Retrieval. For this situation once more, the distinction with different deals with Information Retrieval (Zhang et al., 2008) is the element determination. The trials were done after the Abassi's thoughts, who suggests that highlight choice strategies ought to be custom fitted to conclusion examination by consolidating syntactic properties of content highlights with supposition related semantic data extricated, for instance, from sources like SentiWordNet (Abbasi et al., 2011; Abbasi, 2010). As can be seen, the utilization of dictionaries is ordinarily important to help a few of these NLP exercises (Gerani et al., 2012; Thet et al., 2010; Loia and Senatore, 2014). What's more, besides, the issue of investigating the diverse syntactic dimensions can be much progressively complex working with writings written in various dialects (Abbasi et al., 2008; Kim et al., 2010; Banea et al., 2010).

V. RESULTS

The general assessment got from the dataset with respect to the three universities AIIMS, IIT and NIT were, as per the following: a sum of 371 tweets were viewed as positive, 231 as negative and 542 as impartial for AIIMS. 612 tweets were named positive, 332 as negative and 798 as unbiased for IIT. 542 tweets were considered as positive, 327 as negative and 924 as unbiased for NIT. As recently referenced, a greatness of 0 was considered as impartial valence, more prominent than 0 was considered as positive valence while under zero was considered as negative valence. Table I shows a couple of the insights acquired.

TABLE I. STATISTICS ON THE SENTIMENTS EXTRACTED FROM TWEETS

College	**Ratio of positive to negative tweets**	**Average positive sentiment**
NIT	1.51	2.84
IIT	1.62	3.12
AIIMS	1.78	5.24

AIIMS had the most noteworthy positive normal assumption and the proportion for positive to negative tweets. This means the perception that the positive tweets about AIIMS are progressively positive in the greatness of their feeling and furthermore shows that AIIMS is discussed decidedly more than it is discussed adversely the most among the three organizations. The expectations made by the AI calculations demonstrated high exactness. For estimating exactness, ROC bends were built which plot the genuine positive rate as an element of the bogus positive rate at different edge settings. Basically, genuine positive rate portrays the quantity of tests anticipated to be sure which were likewise positive in fact. It is registered as the proportion of genuine positives to add up to positives. While, false positive rate means the quantity of tests which were really negative, yet were anticipated to be sure and is characterized as the proportion of false positives to add up to negatives.

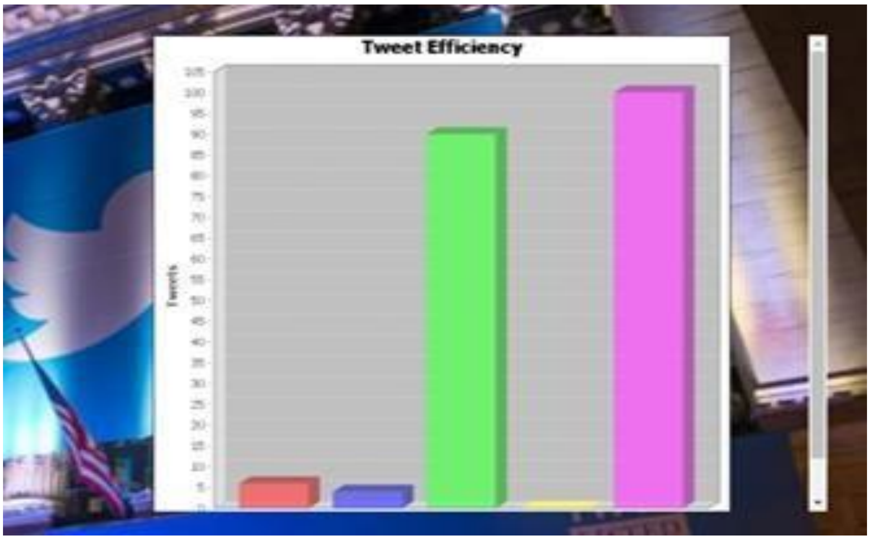

Figure 2. Shows the tweets efficiency graph.

VI. CONCLUSIONS

This work introduces a point by point audit of some web administrations which incorporate functionalities identified with Sentiment Analysis. A portion of these administrations have a place with privately owned businesses, yet all things considered, they enable confined free access to their functionalities, and the others are thoroughly free administrations. This reality is fascinating to those clients/specialists who want to incorporate Sentiment Analysis abilities inside their very own stages without building up their very own calculations; thus, these apparatuses are particularly intriguing for inquiring about purposes and fast prototyping. Additionally, because of the way that the chose administrations can fill in as web benefits, the incorporation of them into any stage is extremely simple. In addition, so as to encourage the undertaking of choosing the most proper administration relying upon the client needs, the capacities of these administrations identified with Sentiment Analysis have been evaluated under various conditions. For this reason three unique accumulations have been picked containing data of an alternate sort: huge and short messages; positive, negative, nonpartisan data; and even amusing substance. Such accumulations have been utilized to evaluate the characterization and extremity rating abilities of the proposed devices. From the outcomes acquired, administrations, for example, Alchemy and Semantria could be considered for any sort of content. Conclusion Analysis might be extremely intriguing to the client if the broke down writings are very vast and you need to group them as positive or negative. Musicmetric and Uclassify are different instruments that could be considered. Every one of these apparatuses could likewise be viewed as a decent alternative if the writings contain amusing sentences. Then again, the discoveries of this work may dispose of devices like Wingify or Viralheat on account of the got outcomes. It is likewise important to remark that the primary drawback of every one of these apparatuses is that the0y can't get attractive outcomes working with impartial writings. What's more, besides, these devices still need to manage a few difficulties, for example, the express location of subjectivity inside writings. To support these announcements, a genuine situation has been proposed to check whether the outcomes and ends extricated from the trials utilizing standard accumulations. This new examination has shown that those apparatuses that should be the best regarding the earlier tests, were extremely the most fascinating instruments for the proposed situation. Along these lines, from this work any client/scientist has enough data about the administrations offered and the conceivable outcomes anticipated from them, to choose the most proper one for his/her advantages. As a last finishing up comment, it is important to remark that these apparatuses have a great deal of difficulties ahead. They experience the ill effects of issues like the over the top straightforwardness while characterizing, by and large, just positive, negative or impartial classifications are utilized; or the lack of ability to total evaluations from various sentences or sections, so as to get a general rate about a total conclusion.

REFERENCES

[1] Apoorv Agarwal, Fadi Biadsy, and Kathleen Mckeown. 2009. Contextual phrase-level polarity analysis using lexical affect scoring and syntactic n-grams. Proceedings of the 12th Conference of the European Chapter of the ACL (EACL 2009), pages 24–32, March.

[2] Luciano Barbosa and Junlan Feng. 2010. Robust sentiment detection on twitter from biased and noisy data. Proceedings of the 23rd International Conference on Computational Linguistics: Posters, pages 36–44.

[3] Adam Bermingham and Alan Smeaton. 2010. Classifying sentiment in microblogs: is brevity an advantage is brevity an advantage? ACM, pages 1833–1836.

[4] C. Fellbaum. 1998. Wordnet, an electronic lexical database. MIT Press. Michael Gamon. 2004. Sentiment classification on customer feedback data: noisy data, large feature vectors, and the role of linguistic analysis. Proceedings of the 20th international conference on Computational Linguistics.

[5] Alec Go, Richa Bhayani, and Lei Huang. 2009. Twitter sentiment classification using distant supervision. Technical report, Stanford.

[6] David Haussler. 1999. Convolution kernels on discrete structures. Technical report, University of California at Santa Cruz.

[7] M Hu and B Liu. 2004. Mining and summarizing customer reviews. KDD.

[8] S M Kim and E Hovy. 2004. Determining the sentiment of opinions. Coling.

[9] Dan Klein and Christopher D. Manning. 2003. Accurateunlexicalized parsing. Proceedings of the 41st Meeting of the Association for Computational Linguistics, pages 423–430.

[10] Alessandro Moschitti. 2006. Efficient convolution kernels for dependency and constituent syntactic trees. In Proceedings of the 17th European Conference on Machine Learning.

[11] Alexander Pak and Patrick Paroubek. 2010. Twitter as a corpus for sentiment analysis and opinion mining. Proceedings of LREC.

[12] B. Pang and L. Lee. 2004. A sentimental education: Sentiment
analysis using subjectivity analysis using subjectivity summarization based on minimum cuts. ACL.

[13] P. Turney. 2002. Thumbs up or thumbs down? Semantic orientation applied to unsupervised classification of reviews. ACL.

[14] Abbasi, A., Chen, H., Salem, A., 2008. Sentiment analysis in multiple languages. ACM Transactions on Information Systems 26, 1{34.

[15] Baccianella, S., Esuli, A., Sebastiani, F., 2010. SentiWordNet 3.0: An Enhanced Lexical Resource for Sentiment Analysis and Opinion Mining, in: Calzolari, N., Choukri, K., Maegaard, B., Mariani, J., Odijk, J., Piperidis, S., Rosner, M., Tapias, D. (Eds.), Proceedings of the Seventh International Conference on Language Resources and Evaluation (LREC'10), European Language Resources Association (ELRA), Valletta, Malta. pp. 2200{2204.

[16] Banea, C., Mihalcea, R., Wiebe, J., 2010. Multilingual subjectivity: are more languages better?, in: Proceedings of the 23rd International Conference on Computational Linguistics (COLING '10), pp. 28{36.

[17] Bar-Haim, R., Dinur, E., Feldman, R., Fresko, M., Goldstein, G., 2011. Identifying and following expert investors in stock microblogs, in: Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP '11), pp. 1310{1319.

[18] Barbosa, L., Feng, J., 2010. Robust sentiment detection on Twitter from biased and noisy data, in: Proceedings of the 23rd International Conference on Computational Linguistics (COLING '10), pp. 36{44.

[19] Beineke, P., Hastie, T., Manning, C., Vaithyanathan, S., 2004. Exploring Sentiment Summarization, in: Shanahan, J.G., Wiebe, J., Qu, Y. (Eds.), Proceedings of the AAAI Spring Symposium on Exploring Attitude and Affect in Text Theories and Applications, AAAI Press. pp. 1{4.

[20] Boiy, E., Moens, M.F., 2008. A machine learning approach to sentiment analysis in multilingual Web texts. Information Retrieval 12, 526{558.

[21]Bollen, J., Mao, H., 2011. Twitter Mood as a Stock Market Predictor. Journal of Computational Science 44, 91{94.

[22] Cambria, E., Schuller, B., Xia, Y., Havasi, C., 2013. New Avenues in Opinion Mining and Sentiment Analysis. IEEE Intelligent Systems 28, 15{21.

[23] Cao, Q., Duan, W., Gan, Q., 2011. Exploring determinants of voting for the helpfulness of online user reviews: A text mining approach. Decision Support Systems 50, 511{521.

[24] Castellanos, M., Dayal, U., Hsu, M., Ghosh, R., Dekhil, M., Lu, Y., Zhang, L., Schreiman, M., 2011. LCI: a social channel analysis platform for live customer intelligence, in: Proceedings of the 2011 international conference on Management of data - SIGMOD '11, ACM Press, New York, New York, USA. pp. 1049{1058.

[25] Chen, B., Zhu, L., Kifer, D., Lee, D., 2010. What is an Opinion About? Exploring Political Standpoints using Opinion Scoring Model, in: Proceeedings of AAAI Conference on Artificial Intelligence (AAAI-2010), pp. 1007{1012.

[26] Chenlo, J.M., Losada, D.E., 2014. An empirical study of sentence features for subjectivity and polarity classification. Information Sciences 280, 275{288.

[27] Deerwester, S.C., Dumais, S.T., Landauer, T.K., Furnas, G.W., Harshman, R.A., 1990. Indexing by Latent Semantic Analysis. Journal of the American Society of Information Science 41, 391{407.

[28] Esuli, A., Sebastiani, F., 2006. Determining Term Subjectivity and Term Orientation for Opinion Mining, in: Proceedings of the 11th Conference of the European Chapter of the Association for Computational Linguistics (EACL06), pp. 193{200.

[29] Feldman, R., 2013. Techniques and applications for sentiment analysis. Communications of the ACM 56, 82{89.

[30] Filatova, E., 2012. Irony and Sarcasm: Corpus Generation and Analysis Using Crowdsourcing, in: Proceedings of the Eighth International Conference on Language Resources and Evaluation (LREC-2012), Istanbul, Turkey. pp. 392{398.

[31] Finn, A., Kushmerick, N., 2006. Learning to classify documents according to genre: Special Topic Section on Computational Analysis of Style. Journal of the American Society for Information Science and Technology 57, 1506{ 1518.

[32] Ganesan, K., Zhai, C., Han, J., 2010. Opinosis: a graph-based approach to abstractive summarization of highly redundant opinions, in: Proceedings of the 23rd International Conference on Computational Linguistics (COLING '10), pp. 340{348.

[33] Ganesan, K., Zhai, C., Viegas, E., 2012. Micropinion generation: An Unsupervised Approach to Generating Ultra-Concise Summaries of Opinions, in: Proceedings of the 21st international conference on World Wide Web - WWW '12, ACM Press. pp. 869 { 878.
Garcia Esparza, S., OMahony, M., Smyth, B., 2012. Mining the real-time web: A novel approach to product recommendation. Knowledge-Based Systems 29, 3{11.

[34] Gerani, S., Carman, M., Crestani, F., 2012. Aggregation Methods for Proximity-Based Opinion Retrieval. ACM Transactions on Information Systems 30, 1{36.

[35] Go, A., Bhayani, R., Huang, L., 2009. Twitter sentiment classification using distant supervision. Technical Report. Standford University.

[36] Groh, G., Hau_a, J., 2011. Characterizing Social Relations Via NLP-Based Sentiment Analysis, in: Adamic, L.A., Baeza-Yates, R.A., Counts, S. (Eds.), ICWSM, The AAAI Press. pp. 502{505.

[37] Guo, L., Wan, X., 2012. Exploiting syntactic and semantic relationships between terms for opinion retrieval. Journal of the American Society for Information Science and Technology 63, 2269{2282.

[38] He, Y., Lin, C., Alani, H., 2011. Automatically extracting polarity-bearing topics for cross-domain sentiment classification, in: Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies (HLT '11), pp. 123{131.