# Sentiment Analysis in E-commerce Using SVM on Roman Urdu Text

Faiza Noor, Maheen Bakhtyar(✉), and Junaid Baber

Department of Computer Science and Information Technology,
University of Balochistan, Quetta, Pakistan
noorfaiza84@gmail.com,maheenbakhtyar@um.uob.edu.pk,junaidbaber@ieee.org

**Abstract.** The usefulness and importance of sentiment analysis task is a widely discussed and effective technique in e-commerce. E-commerce is a very convenient way to buy things online. It saves a lot of time that is usually spent traveling and buying by visiting the shops. E-commerce provides an efficient and effective way to shop sitting right in front of one's computer/mobile at home. For a given product, sentiment analysis captures the users views; their feelings and opinion related to that product. The reviews are categorized into three basic classes i.e. negative, positive, and neutral. This paper focuses on *Urdu Roman reviews* that are obtained by one of the most famous and accessed e-commerce website of Pakistan–Daraz.pk. There are total 20.286 K reviews which are annotated into three classes by three different experts. Vector space model, a.k.a bag of word model is applied for feature extraction which are later passed to Support Vector Machines (SVM) for sentiment classification. Experiments are conducted on MATLAB Linux server. The dataset is kept public for future use and experiments.

**Keywords:** Roman Urdu · Sentiment analysis · Opinion mining · SVM · E-commerce · Reviews

## 1 Introduction

Internet has widely become user centric. People are busy in sharing their views by using different platforms. Similarly, online shopping has also become very common and one of the most convenient ways to shop. Products are being purchased online avoiding the real hectic process to visit the shops in person. Customers get the products at their door step and secure payment is made online conveniently. Whenever a people plan to buy a product, they tend to read product respective product reviews provided by other people who have already used the same product. Online reviews provide a way to check product popularity and usefulness prior to buying.

People show their positive or negative attitude towards the products through their comments below the product description. Products are getting hundreds of


M. H. Miraz et al. (Eds.): iCETiC 2019, LNICST 285, pp. 213–222, 2019.
https://doi.org/10.1007/978-3-030-23943-5_16

reviews online affecting the overall perception of the company and their products. We examine the reviews of DarazPK and observe that in Pakistan, customers conventionally post the comments either in English language or Roman Urdu language. Roman Urdu, Urdu being native language of Pakistan, is widely used, making it easy for people to correctly express their feelings. Daraz.pk is an e-commerce website which brings all the cultural aspects of Pakistani nation along with its products. The products being sold belong to the interests of Pakistani customers.

Various other e-commerce sites are available targeting Pakistani users such as symbios.pk, homeshopping.pk, shophive.com, ishopping.pk, and 24hours.pk/. These sites either do not have an option to review the products, hence, unavailability of review data, or the reviews are not filled by the customers most of the times. Daraz.pk is the most commonly used site therefore, the reviews are being filled and thus data is available.

This research deals mainly with sentiments analysis on Pakistani products based on comments/reviews in English and Roman Urdu languages. Reviews are mainly either positive, negative, or neutral and they determines and examines the user perception assisting the sellers to increase and enhance their products availability and quality hence, affecting online shopping positively.

In Pakistan business ideas has completely transformed and people prefer to buy items such as electronics, ladies clothing, gents wear, kids wear, accessories, home appliances, etc. online.

Roman Urdu language is being used now a days. Mostly in Pakistan and India, people express their words by typing in roman Urdu. Even now a days social sites are based on reviews that comprises of Roman Urdu texts. If a person buy any product online, he/she shares the sentiments by writing in their own native language. In English Language the sentiment analysis has been well explored. In every area research has been done any analyzing the sentiments. Machine learning and lexicon based learning has been done at great extent. These can be used in Roman Urdu reviews. Arabic language is vast language and being used in Arab countries. This language has achieved its goals by analyzing the sentiments in Arabic texts. Many research works has been done in Arabic language [1,9,12,20]. Persian language is the old language and the first language in Muslim World which was in competition with Arabic Language. Persian has a very good history in Persian literature. Many great poets has also written in Persian script. Sentiment analysis has been discussed and experimented in Persian language [6,16,25]. Local languages have also been making use of sentiment analysis to analyze the user opinions. For example, Pashto language sentiment analysis [19] and Sindhi language sentiment analysis [3,4] are making use of sentiment analysis approaches.

Roman text is usually very challenging to process. Roman Urdu has no standards and no rules therefore, understanding such language is not so easy. For example word *Mein* میں means I/Me, and can be written in various ways such as, *Mn, Me, Main, Ma, Men, M,* etc. Non-standard word forms make it difficult to process and understand. Further discussions are given in the sections to follow.

Section 2 discusses the related work, Sect. 3 provides the methodology of the framework, and Sect. 3 discusses the experiments and results of the system.

## 2 Related Work

The aim of sentiment analysis and opinion mining is to differentiate a user like/dislike review on a particular product. Big data was collected from amazon and reviews. Recommendation system was used to check the users reviews priority and qualitative analysis was done to check sentiment analysis on large data reported by [30] Bootstrapping method was used to extract adopter information reviews from site. Maximum likelihood was used to check reviews and matrix factorization for recommendation of product [32]. Data collected from amazon and drawn a distribution curve of the products. Products were divided into categories like product category, number of product, review of product and mean of product. Compared the models and at last proved their work by lemma, proof of proposition and proof of corollary [15]. Gathered data from different platforms which were available easily. And four types of platforms were identified and experimented. A survey was done using Qualtrics, a questionnaire. Data was displayed in the form of table and compared the result [18]. Sentiment analysis was done and the reviews were extracted by preprocessing method, then part of speech tagging was applied and feature score was calculated using opinion mining [24]. Amazon data was collected and positive and negative reviews were extracted. Reviews were tested by test classifier. Logistic regression and L2 regularization was used as baseline of classifier. Drawn ROC curve, F1 measure was also calculated, also measured precision and recall, unigram, n-gram, and histogram [7]. Web crawler (web spider) was used to extract data and collect them from amazon. Locospider was used as a web crawler. Data pre-processing (including segmentation, POS tagger), text analysis (labeling noun, phrases, feature detection), multiple linear regression was used [10]. Data was collected from regarding the reviews of products from cnet.com, ciao.co.uk and shopping.yahoo.com. Three classifiers were used and discussed, 1: Passive Aggressive algorithm 2: Language Modeling, 3: Winnow Classifier. N-gram was also used and high order n-gram improves the performance of classifiers, especially negative instances [11]. Online product reviews data was gathered from amazon. Analysis was done for sentences and reviews with hopeful conclusion. Reviews were about electronics, book, beauty and home and were categorized into tables. Negatives phrases were separated. Sentences were divided into tokens and POS tagger was applied. Histogram was drawn, F1 measure was calculated and ROC curve was shown. Application used is scikit-learn, python open source. Models used were Nave Bayesian, SVM, and Random Forest [13] Polarity of texts was found. They trained the model depending on bootstrap aggregating algorithm and state of art model. This improved F measure result. SentiME system was developed. Stanford sentiment system was used to classify the sarcasm of sentences. Cross validation test was done. Sentiment analysis, precision and recall was measured [28]. Data was gathered about product reviews from Twitter. Opinion lexicon dictionary was used to find polarity of words either positive, negative or neutral. Twitter has its own features like emoticons, abbreviation, hashtags etc, this creates less recall for lexicon based method. Classifier was built to find positive, negative or neutral words. Sentences were detected like declarative sentences, interrogative sentences and imperative sentences. Hashtags words like # fail was included in their lexicon. Score equation formula was used to collect the scores. POS tagger was used for comparative sentences like iPhone is better than Samsung. And also used pearson chi square test for identifying the indicators. $F$ measure, recall and precision was used for evaluation measure [31] Aim was to perform replication using webis source code to see if it produces comparable result. Evaluated sentiMe on amazon reviews. Semantic evaluation was done. Positive, negative, or neutral words were separated from twitter reviews. Top scoring system between 2013–2015 on semEval was discussed in paper. F score was used to compare the pre trained model and retrained model. Selected linear averaging function because it predicts in a very simple way [27]. Twitter is popular communication platform. Blogs were create to communicate between users for variety of topics. Data was collected from twitter API v1.0 and corpus of tweet was related to Justin Bieber brand. Sentiment analysis was done. POS tagger, negative sentences and positive sentences were separated, n-gram was applied. Also focused on sad and happy emoticons because users also comment through their emotions. DAN2 architecture was also discussed and SVM was also implement and at last by comparing the results of DAN2 and SVM, it was concluded that DAN2 produces better recall [14]. Sentiment analysis on twitter to check the reviews on different topics. Sentences polarity was measured including subjective or objective. Emoticons are used to differentiate between positive or negative tweets. Opinion mining was also done. Experiments were performed (NaveBayes algorithm with supervised classification and it was then displayed with map SOM method). Vector space model was applied (stemmer or stopper process). TF/IDF was also calculated at the end [23]. There is a gap between description and review of a product because mostly customers buy a product by reading the reviews written by early users. Data was collected from 63 participants (male/female) both. HOM (high order mean) was calculated and shown by graph (description and experience rating). Online consumer rating and WOM was also discussed [29]. Emotions help a lot in purchasing any product from online stores. Emotions express positive or negative sentiments. Cognitive appraisal theory is discussed which shows that some sentiments are linked with reliability, and some are allied with unreliability. Latent semantic analysis (LSA) was used. Also discussed Tobit regression for analyzing the model [2]. Objects with high rating provides much information about the product. Data was about 4000 books from amazon.com and data was extracted from it. Sentiment polarity was done to check the positive or negative language in text. Feature based sentiment analysis was applied (to check subjectivity or objectivity of sentences) [17]. Customers share their views in their native language. In this paper Roman Urdu and English views were extracted using easy extractor software. Extracted data was then filtered in WEKA2. Also used three different algorithms Nave Bayes algorithm, decision trees, and KNN. And in the end the results of each algorithm was compared. Which has showed that nave Bayes provides good result as compared to decision tree because of its simplicity and easyuse [8]. In this paper the authors have first created a dictionary for Persian words. They have added persian words in it which was the difficult task of their work. Sentiment analysis was done to check the positive or negative reviews by the viewers. Data was collected from the most visited website digikala by using web crawler.dataset contains reviews about different products with different categories. Preprossesising was done by spell checker, lemmitizer, pos, stop words removal etc. they have also used various classification methods like F measure, Bayesnet, KNN, LibLinear, SMO. Results were matched with each classification method. Persian reviews are analyzed just like other languages. FarsNet is the first Persian WordNet containing Persian words, its latest version is also available for researchers. Words are first matched in the dictionary and then the extraction is done. The authors have used n-gram, bi-gram, uni-gram methods, TF/IDF, precision and recall. And in the end they have just compared the results [5] Just like other languages Chinese language reviews are also analyzed by the researchers. They have collected the mobile phone reviews from a known website, www.360buy.com. Java crawler was created to find negative and positive reviews. Then preprocessing is done and Boolean weighting is done to calculate feature weight of the products. Authors have used statistical machine learning methods to find online Chinese reviews. They choose ICTCLASS system for word segmentation and POS tagging. Used DF/IDF algorithm for their experiment on Chinese reviews. At last compared their results and showed it through graph and charts [33] Based on multilingual sentiment analysis. Viewed and worked on different languages like Chinese, English, Spanish, German, Swedish, Romanian, French, Japanese. Investigated both lexicon and corpus based approaches. lemmatization, tokenization, POS tagging, F-measure, unigram, bigram, n-gram, is done and compared. Translation from languages (formal to informal) is also done and results are compared. References from various papers have been pinpointed in their works. Which has helped them a lot. Authors have looked at wide range of tools and methods used in sentiment analysis [21] Sentiment analysis was investigated and reviewed. Total number of published papers on sentiment analysis and number of citations for that papers. Tools used in that papers are analyzed and compared with others. Publications on Sentiment analysis started from 2004. And today it is one of the fast growing field in research area of computer science. Top most cited papers are also discussed. Positive, negative reviews, their polarity, languages, citation, everything has been deeply studied and compared. Qualitative and quantitative analysis has been done. The top most citation from scopus and google scholar and publication venues area also discussed and shown in table. Different areas of research has been discussed like languages, Humans(emotions, interaction, language, behavior etc), tools (nlp, machine learning etc). 20 most cited papers per year in scopus and google scholar has been highlighted and their work has been discussed. How sentiments are analyzed, which tool to use, polarity checking, preprocessing etc. and it has been observed that with the passage of time citations have been increased per year. And it might increase in next few years too [22].

**Table 1.** Review Dataset description. Total 20285 reviews.

| Review sentiments | Training set instances | Test set instances |
|---|---|---|
| Neutral ($\mathcal{N}$) | 7119 | 1780 |
| Positive ($\mathcal{P}$) | 4880 | 1220 |
| Negative ($\mathcal{E}$) | 4228 | 1058 |
| Total | 16227 | 4058 |

## 3 Methodology

This section explains the methodology for experiments and data generation. Different query terms are used on DarazPK portal and then the reviews of the users are stored in raw text file. The products used for the reviews are listed in the Table 1.

Reviews are represented by vector space model (VSM), a.k.a bag of word model. In VSM, each review is represented by the normalized frequency of the words, known as term frequency (TF), from the set of a vocabulary ($\mathcal{V} = \{v_1, v_2, \ldots, v_d\}$) by its weight ($\mathcal{W} = \{w_1, w_2, \ldots, w_d\}$); inverse document frequency (IDF) is widely used to find the weights $\mathcal{W}$.

The basic flow of the system is shown in Fig. 1.

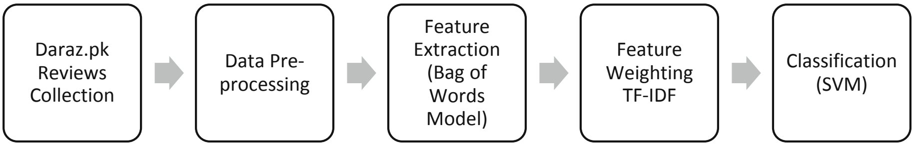



**Fig. 1.** Basic flow of the framework

There are total 20285 reviews which are denoted by $\mathcal{R} = \{r_1, r_2, \ldots, r_n\}$, where $n = 20285$. Three post graduate students, who have sufficient knowledge of sentiment analysis, are requested to label each review with either of following sentiments $\mathcal{S} = \{\mathcal{N}, \mathcal{P}, \mathcal{E}\}$; Neutral ($\mathcal{N}$), Positive ($\mathcal{P}$), or Negative ($\mathcal{E}$). Finally, each review $r_i$ is labeled as $s \in \mathcal{S}$ if two of the students voted as $s$. So, the representation of reviews $\mathcal{R}$ is extended as $\mathcal{R} = \{(r_1, s_1), (r_2, s_2), \ldots, (r_n, s_n)\}$, where $s_i \in \mathcal{S}$. Each review $r_i \in \mathcal{R}$ is feature vector, a.k.a bag of word, which is $\mathbb{R}^d$ dimensions, where $d$ indicates the number of words.

There are total 13662 words taken into account during the experiments. To generate the vocabulary, 25 thousand reviews are randomly crawled from Darazpk, frequency of each word is computed and all words present in at least 80% of the reviews are removed (marked as stop words), such as 'a', 'aaaaa', and 'yar'. Once the vocabulary is chosen, then TF-IDF is applied.

Since, the reviews are mostly short sentences, comprises of few words; that makes $r_i$ very sparse. For example, some of the reviews contains only two words such as 'Bahtareen hai yar' which is labeled as $\mathcal{P}$, in English its mean 'its wonderful buddy', where the word 'yar' (in English buddy) is treated as stop word and removed.

Support Vector Machine (SVM) classifier is used with two kernels: linear kernel and cubic kernel. SVM is inherently a binary classifier. To enable SVM to classify multiple classes, one of the following two approaches are used, namely; (i) one-vs-one (OVO) (ii) one-vs-all (OVA). OVA trains, for $q$ different classes, ($q > 2$,) $q$ different classifiers. For each class $i$, it assumes $i$ as positive and the remaining as negative. In case of sentiment analysis, for the sentiment $\mathcal{E}$ recognition, there will be 65% which give more weight to negative instances.

Commonly, OVO approach is better in accuracy than OVA [26]. $q(q-1)/2$ binary classifiers are trained in case of OVO approach. In case of sentiment analysis of Roman Urdu would have three distinct binary classifiers. The decision to label the given instance $r_j$ is taken using the ensemble/voting approach. $r_j$ is fed to 21 various binary classifiers, and the label showing the highest frequency selected. Approach of OVO is used in this paper.

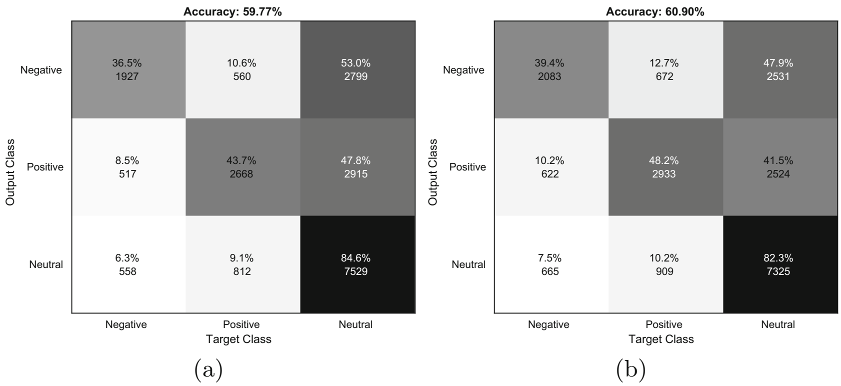
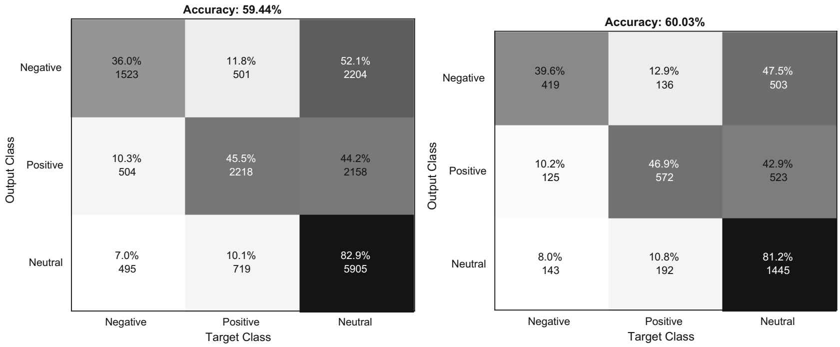



**Fig. 2.** SVM accuracy on whole dataset. (a) shows the Linear kernel, and (b) shows the cubic kernel.

## 4 Experiments and Results

Figure 2 shows the accuracy of SVM on whole dataset; dataset is not divided into training set and test set. In literature, cubic kernel out performs linear kernel in multi-class classification, but in current situation, it is just marginally better. In both kernels, the accuracy of sentiment $\mathcal{N}$ is very high, whereas the sentiment $\mathcal{E}$ remains challenging. It is because there are number of ways to show negative feelings for any product, where natural intelligent person can identify




(a) (b)

**Fig. 3.** SVM accuracy using cubic kernel. (a) shows the training set accuracy, and (b) shows the test set accuracy.

but it remains very difficult for artificial intelligent algorithm, unless complex and advance algorithms are not applied such as semantic analysis using natural language processing.

Figure 3 shows the accuracy of cubic kernel, when the dataset is split into training set and test set; 20% of each sentiment are randomly selected from the main dataset and labeled as test set, and the remaining 80% is labeled as training set. It is quite interesting to observe that neither of the kernel over fits the data, in case of facial expression recognition cubic kernel over fits the training set [26]. It is also interesting to observe that $\mathcal{N}$ has highest accuracy and also it has highest false positive scores against rest of the sentiments. Mostly, in live implementation of sentiment analysis on any application, if the score of any unknown reviews is low, then that review should be treated as neutral unless manually reported as negative by any user. In our experiment, this constraint is already intact. The dataset is available publicly online[1] for future use and research. The source code can be provided if requested to the corresponding author.

## 5 Conclusion

This paper classifies the reviews on DarazPK, an e-commerce portal, into three different classes which are widely known as sentiment analysis. More than 20 K reviews are obtained and annotated by experts into three classes. The dataset is later divided into two sets; training set (80%) and test set (20%). Vector space model is used for feature extraction. Different kernels of SVM are used for classification. Cubic Kernel achieves highest accuracy on given dataset.

[1] http://www.maheenbakhtyar.com/links.

## References


1. Abbasi, A., Chen, H., Salem, A.: Sentiment analysis in multiple languages: feature selection for opinion classification in web forums. ACM Trans. Inf. Syst. **26**(3), 12:1–12:34 (2008). https://doi.org/10.1145/1361684.1361685. http://doi.acm.org/10.1145/1361684.1361685
2. Ahmad, S.N., Laroche, M.: How do expressed emotions affect the helpfulness of a product review? Evidence from reviews using latent semantic analysis. Int. J. Electron. Commer. **20**(1), 76–111 (2015)
3. Dootio, M., Wagan, A.: Sentiment summarization and analysis of Sindhi text. Int. J. Adv. Comput. Sci. Appl. **8**, 296–300 (2017). https://doi.org/10.14569/IJACSA.2017.081038
4. Ali, M., Wagan, A.I.: An analysis of sindhi annotated corpus using supervised machine learning methods. Mehran Univ. Res. J. Eng. Technol. **38**(1), 185–196 (2019). https://doi.org/10.22581/muet1982.1901.15. http://publications.muet.edu.pk/index.php/muetrj/article/view/754
5. Asgarian, E., Kahani, M., Sharifi, S.: The impact of sentiment features on the sentiment polarity classification in Persian reviews. Cogn. Comput. **10**(1), 117–135 (2018)
6. Basiri, M.E., Naghsh-Nilchi, A.R., Ghassem-Aghaee, N.: A framework for sentiment analysis in Persian. Open Trans. Inf. Process. **1**(3), 1–14 (2014)
7. Bhat, S.K., Culotta, A.: Identifying leading indicators of product recalls from online reviews using positive unlabeled learning and domain adaptation. arXiv preprint arXiv:1703.00518 (2017)
8. Bilal, M., Israr, H., Shahid, M., Khan, A.: Sentiment classification of Roman-Urdu opinions using naïve Bayesian, decision tree and KNN classification techniques. J. King Saud Univ.-Comput. Inf. Sci. **28**(3), 330–344 (2016)
9. Boudad, N., Faizi, R., Thami, R.O.H., Chiheb, R.: Sentiment analysis in Arabic: a review of the literature. Ain Shams Eng. J. **9**(4), 2479–2490 (2018). https://doi.org/10.1016/j.asej.2017.04.007. http://www.sciencedirect.com/science/article/pii/S2090447917300862
10. Chen, X., Sheng, J., Wang, X., Deng, J.: Exploring determinants of attraction and helpfulness of online product review: a consumer behaviour perspective. Discret. Dyn. Nat. Soc. **2016**, 19 Pages (2016)
11. Cui, H., Mittal, V., Datar, M.: Comparative experiments on sentiment classification for online product reviews. AAAI **6**, 1265–1270 (2006)
12. El-Masri, M., Altrabsheh, N., Mansour, H., Ramsay, A.: A web-based tool for Arabic sentiment analysis. Procedia Comput. Sci. **117**, 38–45 (2017). https://doi.org/10.1016/j.procs.2017.10.092. http://www.sciencedirect.com/science/article/pii/S187705091732149X. Arabic Computational Linguistics
13. Fang, X., Zhan, J.: Sentiment analysis using product review data. J. Big Data **2**(1), 5 (2015)
14. Ghiassi, M., Skinner, J., Zimbra, D.: Twitter brand sentiment analysis: a hybrid system using n-gram analysis and dynamic artificial neural network. Expert. Syst. Appl. **40**(16), 6266–6282 (2013)
15. Holleschovsky, N.I., Constantinides, E.: Impact of online product reviews on purchasing decisions. In: WEBIST (2016)
16. Hosseini, P., Ramaki, A.A., Maleki, H., Anvari, M., Mirroshandel, S.A.: Sentipers: a sentiment analysis corpus for Persian. CoRR abs/1801.07737 (2018)

17. Hu, N., Koh, N.S., Reddy, S.K.: Ratings lead you to the product, reviews help you clinch it? The mediating role of online review sentiments on product sales. Decis. Support. Syst. **57**, 42–53 (2014)
18. Hu, N., Pavlou, P.A., Zhang, J.: On self-selection biases in online product reviews. MIS Q. **41**(2), 449–471 (2017)
19. Kamal, U., Siddiqi, I., Afzal, H., Rahman, A.U.: Pashto sentiment analysis using lexical features, November 2016. https://doi.org/10.1145/3038884.3038904
20. Kiritchenko, S., Mohammad, S.M., Salameh, M.: Semeval-2016 task 7: determining sentiment intensity of English and Arabic phrases. In: Proceedings of the International Workshop on Semantic Evaluation, SemEval 2016, San Diego, California, June 2016
21. Lo, S.L., Cambria, E., Chiong, R., Cornforth, D.: Multilingual sentiment analysis: from formal to informal and scarce resource languages. Artif. Intell. Rev. **48**(4), 499–527 (2017)
22. Mäntylä, M.V., Graziotin, D., Kuutila, M.: The evolution of sentiment analysis—a review of research topics, venues, and top cited papers. Comput. Sci. Rev. **27**, 16–32 (2018)
23. Martínez-Cámara, E., Martín-Valdivia, M.T., Urena-López, L.A., Montejo-Ráez, A.R.: Sentiment analysis in Twitter. Nat. Lang. Eng. **20**(1), 1–28 (2014)
24. Pednekar, S., Patil, K., Sawant, R., Shah, T.: Sentiment analysis on online product reviews. Int. Educ. Res. J. **3**(3), 130–134 (2017)
25. Roshanfekr, B., Khadivi, S., Rahmati, M.: Sentiment analysis using deep learning on Persian texts. In: 2017 Iranian Conference on Electrical Engineering (ICEE), Tehran, pp. 1503–1508, May 2017. https://doi.org/10.1109/IranianCEE.2017.7985281
26. Saeeda, S., Baber, J., Bakhtyar, M.: Empirical evaluation of SVM for facial expression recognition. Int. J. Adv. Comput. Sci. Appl. **9**(11), 670–673 (2018)
27. Sygkounas, E., Rizzo, G., Troncy, R.: A replication study of the top performing systems in semeval Twitter sentiment analysis. In: Groth, P., et al. (eds.) ISWC 2016. LNCS, vol. 9982, pp. 204–219. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-46547-0_22
28. Sygkounas, E., Rizzo, G., Troncy, R.: Sentiment polarity detection from amazon reviews: an experimental study. In: Sack, H., Dietze, S., Tordai, A., Lange, C. (eds.) SemWebEval 2016. CCIS, vol. 641, pp. 108–120. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-46565-4_8
29. Wulff, D.U., Hills, T.T., Hertwig, R.: Online product reviews and the description-experience gap. J. Behav. Decis. Mak. **28**(3), 214–223 (2015)
30. Yengi, Y., Omurca, S.İ.: Distributed recommender systems with sentiment analysis. EJOSAT: Eur. J. Sci. Technol. Avrupa Bilim ve Teknoloji Dergisi **4**(7) (2016)
31. Zhang, L., Ghosh, R., Dekhil, M., Hsu, M., Liu, B.: Combining lexicon based and learning-based methods for Twitter sentiment analysis. HP Laboratories, Technical Report HPL-2011, vol. 89 (2011)
32. Zhao, W.X., Wang, J., He, Y., Wen, J.R., Chang, E.Y., Li, X.: Mining product adopter information from online reviews for improving product recommendation. ACM Trans. Knowl. Discov. Data (TKDD) **10**(3), 29 (2016)
33. Zheng, L., Wang, H., Gao, S.: Sentimental feature selection for sentiment analysis of Chinese online reviews. Int. J. Mach. Learn. Cybern. **9**(1), 75–84 (2018)