I.J. Information Engineering and Electronic Business, **2022, 4, 28-38**
Published Online August 2022 in MECS (http://www.mecs-press.org/)
DOI: 10.5815/ijieeb.2022.04.04

A Comparison of Opinion Mining Algorithms by Using Product Review Data

Sumaiya Sultana, Sumaiya Rahman Eva, Nayeem Hasan Moon
Department of Computer Science, American International University Bangladesh, Dhaka, 1229, Bangladesh
E-mail: {sumaiyasultanaa99, sumaiyarahmaneva.1999, moonnayeem88}@gmail.com

Akinul Islam Jony and Dip Nandi
Department of Computer Science, American International University Bangladesh, Dhaka, 1229, Bangladesh
E-mail: {akinul, dip.nandi}@aiub.edu

Received: 20 April 2022; Revised: 03 May 2022; Accepted: 23 May 2022; Published: 08 August 2022

Abstract: After release of Web 2.0 in 2004 user spawned contents on the internet eminently in abundant review sites, online forums, online blogs, and many other sites. Entire user generated contents are considerable bunches of unorganized text written in different languages that encompass user emotions about one or more entities. Mainly predictive analysis exerts the existing data to forecast future outcomes. Currently, a massive amount of researches are being engrossed in the area of opinion mining, also called sentiment analysis, opinion extraction, review analysis, subjective analysis, emotion analysis, and mood extraction. It can be an utmost choice whilst perceiving the meaning and patterns in prevailing data. Most of the time, there are various algorithms available to work with polling. There are contradictory opinions among researchers regarding the effectiveness of algorithms. We have compared different opinion mining algorithms and presented the findings in this paper.

Index Terms: Sentiment analysis, Opinion Mining, Supervised Learning Algorithms, Unsupervised Learning Algorithms, Boosting Algorithms

1. Introduction

Opinion mining or sentiment analysis is a text analysis method that utilizes computational linguistics and natural language processing to spontaneously identify and extricate emotions or opinions from texts, i.e., positive, negative, neutral, etc. [1]. In this multidisciplinary and multifaceted Artificial intelligence problem, people's perspectives and reactions toward an entity are studied in computational form [2, 3]. The term sentiment analysis is the method of obtaining useful information from an opinion. Sentiment analysis is a promising area defined as the crossroads of information retrieval and computational linguistic techniques for processing documented opinions [4].

Sentiment Analysis has three levels of granularities and it can be applied in any level [5].

Everyday a vast number of texts are generated online through people containing their emotions over various kinds of entities. These emotions can be positive or negative, depending on their written expressions. Sometimes they also use emoji to express emotions. These people generated text can be beneficial for companies and organizations to get insights about their products or services through such reviews, recommendations, and texts. So, lots of researchers are working on this field and finding new techniques and ways to extract emotions from enormous text data. For example, [6] worked on movie review data, [7] analyzed new algorithms for twitter review data, [8] used medical data to work on sentiment analysis. Due to COVID-19 pandemic in 2019, most of the country decided to go on shutdown to secure people from this deadly Coronavirus. Because this virus is contagious and people started to maintain social distance to stay safe. So, they moved toward online shopping for their daily needs to avoid going out. Online purchasing of products and daily goods increased. That is the reason an online product review dataset was being picked to find out the opinion of people toward online products through different reviews. By this it can be learnt which algorithm or method works best for online product review dataset and the online purchasing experiences of customers will be known as well.

1. What is the need for comparison in distinct algorithms?
2. What will be the outcome if these algorithms are applied in a dataset?

In the talk of sentiment analysis or opinion mining a general question always arises, which algorithm works best in which domain or dataset. There are different opinion mining algorithms available, some of them are supervised, some are

unsupervised and some are boosting algorithms which are discussed on the latter portion of the paper. Works of all the algorithms are similar but some work faster, some give better accuracy and some algorithms help to generate rules for sentiment analysis. So, comparison is required to find out the better working algorithms regarding dataset size in sentiment analysis. As it is not possible to implement all the algorithms on so many datasets at once, that is why it had been decided to pick a dataset which is a product review dataset. In this research, three datasets will be adopted from Amazon as it is one of the biggest websites doing an e-commerce business. After that, several research papers relevant to this research on sentiment analysis and opinion mining will be reviewed to have a clear idea and picked up mostly used sentiment analysis algorithms and techniques for comparison. Then selected and available opinion analysis algorithms and techniques will be used on the datasets to do a comparison review. Google Colaboratory cloud system will be used to implement the algorithms on the datasets. At the end, a comparison will be made based on implementation results. It will be found after this research those which algorithms work well on large datasets and which algorithms work better on small datasets.

The paper orientation is being demonstrated thusly, where the introduction is provided in Section 1. Then, Section 2 describes literature review. Next, details of methodology are given in Section 3. Later, Section 4, the algorithms used for opinion mining are presented. Section 5 highlights the comparison and discussion subsequently. Lastly, Section 6 is finished with a conclusion.

Table 1. Levels of sentiment analysis

Level Name	Description
Document Level	The solicitude is to analyze the whole document to discover whether the opinion is affirmative or non-affirmative. As a whole document works like an individual entity, inappropriate results can be generated sometimes from this level as conflicting sentiment occurs.
Sentence Level	This level is more fine-grained than the previous one as each sentence acts as an entity here. Each sentence can be categorized by emotions or polarities as positive emotion, negative emotion, and neutral emotion. The summarized result of sentences provides an overall result of the document.
Feature Level or Aspect Level	Analyzing product features to determine document sentiment is known as aspect- based sentiment analysis. From here affirmative, non-affirmative, or neutral opinions can be easily identified from extracted features. Among all other models, it is good and a fine-grained analysis model.

2. Literature Review

The resoluteness of this paper is to make a comparison of some well-known opinion mining algorithms. Usually, opinion mining algorithms are used for text analysis to bring out emotions written in text format. Apart from algorithms several other techniques, tools, and ecosystems can be used to do sentiment analysis.

In [9] the author talked about sentiment analysis and the Hadoop ecosystem. Hadoop is an open-source framework where a large dataset can be stored and processed. This paper described the hierarchy of big data, sentiment analysis classification and described the Hadoop systems in detail. [10] used the Hadoop framework to store and analyze large data collected from amazon reviews, did predictive analysis using those data, and made a matchbox recommendation system for the customers of amazon which will provide a recommendation based on previous ratings. Here, they used machine learning to do predictive analysis and did various calculations based on ratings of product-based, category-based, review-based, etc. Leximancer is used in [11] to analyze the reviews of 4 top-grossing games and made ten text files for analysis. By using Leximancer they created a theme map for each file to examine and analyze them. [12] took multi-dataset on different languages to evaluate two supervised learning approaches which are Decision Tree and Naive Bayes for cognizing best results. The evaluation was based on accuracy and runtime parameters. They used RapidMiner tools to perform these experiments.

Authors of the paper [13,14] worked on different opinion mining techniques and classifiers to analyze student performance. Paper [13] did a literature survey on performance prediction of students and found that researchers mostly use Naïve Bayes, Decision Tree, and Rule-Based algorithms for the prediction of students academic performances. Paper [14] took university student data to implement five classifiers: Neural Network, J48, ID3, Bayesian Network and Naïve Bayes. They made a comparison between these five classifiers based on error measures and found out that Bayesian Network classifier accuracy was higher among other classifiers. An experiment was done on the spam dataset to classify spam emails and found that the Random Tree classifier works best with the accuracy of 99.72% for spam mail classification [15]. Classification and Regression Tree (CART), Association Rule Mining, Regression, Clustering and Classification are the most extensive techniques of data mining used in the health care domain for decision making and identification [16].

A study and comparison of several methods of assessing an article's reputation using sentiment analysis is performed in [17]. They classified sentiment analysis based on techniques, approaches, and rating methods. [4, 18] discussed opinion mining areas, issues, technologies, and challenges. They describe opinion mining approaches such as supervised, machine learning, unsupervised, and CRB [Case Based Reasoning]. Key issues and challenges of opinion mining have also been highlighted. The first challenge for researchers is data collection. Accuracy of data if required to find out the useful insides. Reviews are written in different languages like English, Arabic, Chinese, etc. All the data

extraction techniques did not work well in all languages. Moreover, as the data is user-generated, this data contains spam data, unfinished data, noisy data, and unstructured data. So, filtering this data is a challenge for researchers. After filtering, the classification of sentiment is another challenge. Some words can have different meanings based on the domain [i.e., word "High" can be positive for battery life but can be negative for pricing]. Object identification, feature extraction, identifying and grouping synonyms, integration, identifying comparative words, people's writing styles, and misleading options are some of the issues and challenges for researchers in opinion mining. This two-paper discussed these issues and tried to give solutions based on these problems.

A brief review of competitive opinion mining is given in paper [1]. They had studied previous research on general opinion mining along with comparative opinion mining to show differences between them. From two distinct perspectives, they presented opinion mining. A practical viewpoint [i.e., a machine learning approach, a rule mining approach, and an approach to natural language processing], and an element-based viewpoint of opinion [i.e., comparative opinion recognition, entity recognition, relationship recognition, feature recognition]. They had shown all the past research on this and pointed out the past problems that the researcher faced and talked about the future possibilities on this comparative opinion mining domain. [19] used hotel review data for opinion analysis and made a prototype to visualize the output of sentiment analysis. They used knowledge-based approaches, SentiWordNet 3.0 [based on WordNet 3.0] for sentiment classification and feature extraction to find detailed aspects. They also used temporal opinion mining to find out the sudden changes in opinions which is helpful to find out valuable information about the entities.

In the field of temporal opinion mining, the most common methods for predicting and estimating changes in opinion are opinion lexicons and statistical modeling. These changes include time and recent events. [19]. They employed burst detection as an identifier to find out the changes over time. The main problem they describe and the challenge of temporal separation is to find out when the intensity of a particular property has increased. An innovative approach named AMOD is proposed in [20]. AMOD approach can automatically extract opinions from the internet for specific domains. The author of this paper claimed that AMOD approaches can extract opinions better than lexicon approaches and can give higher accuracy than lexicon. They also discussed the drawback of the AMOD approach. In AMOD at first, a learning dataset is extracted and after that from the learning dataset, new positive and negative adjectives are extracted. This process is repeated till no newer adjective is learned. They also showed the comparison with traditional classification methods CopyVote. They used the FScore measure to compare the results and the FScore result of AMOD was higher than CopyVote. If right approaches can be selected depending on the domain and if the issues of selected approaches can be overcome, then some valuable findings can come out from opinion mining which will be beneficial.

3. The Algorithms for Opinion Mining

Nowadays, with the proliferation of contemporary digital-based economies, substantial amounts of information are available in the form of textual data, and classification or grouping into predefined classes is often easier to use.

Several online activities like blogging, micro-blogging, e-commerce, social media communications, click streams etc. create an exceedingly large amount of data which is denoted as Big Data [21]. These data can be of any type – structured or unstructured, and these need to be extracted, transformed, loaded, and analyzed. To do such things, Opinion Mining is needed. There are two fundamental approaches of Opinion Mining [5]. The first one is the quite common one which is machine learning (ML). It is based on learning techniques of three types - supervised learning, unsupervised learning, and semi-supervised learning [22]. Supervised learning captures labeled datasets, while unsupervised learning uses unlabeled data. On the other hand, if the dataset is a combination of labeled and unlabeled examples, the semi-supervised approach is used [23]. Usage of present lexicon with words, expressions, or phrases is the second approach of Opinion Mining. Although based on the dataset these approaches are applied, ML is the most used approach. There are several types of machine learning algorithms such as Decision Tree, Maximum Entropy, Passive Aggressive, Adaptive Boosting, Logistic Regression, Ridge Regression, Support Vector Machine, Naïve Bayes, Viterbi algorithm, Dynamic Artificial Neural Network (DAN2), K-Nearest Neighbor etc. Naive Bayes is great for features that rely heavily on it, but if the conditional independence assumptions are not met, the maximum entropy will be more appropriate [24]. [24] proves that Support Vector Machine works best. [5] worked on four different datasets where Passive Aggressive with a unigram performs best. Paper [3] made a survey which is about sentiment analysis algorithms and showed better accuracy with supervised learning algorithms. In paper [25], they classified tweets on twitter using three supervised learning algorithms which are - Decision Tree, Naive Bayes, and Support Vector Machine. Which are also supervised learning algorithms. Neural Network was used in [26,27], which is an unsupervised learning algorithm also known as learn supervised algorithm. Paper [28] used boosting algorithms Xgboost and got a satisfying accuracy rate. Clustering is the most effective unsupervised technique. There are various algorithms to do clustering such as - K-Means Clustering, Hierarchical Clustering, DBSCAN (Density Based Spatial Clustering of Application with Noise) Clustering, Optics, Sting, SOM (self-organized map). All these clustering algorithms cannot be used in the selected datasets as clustered text data cannot provide proper accuracy information. So, a total of eight algorithms were implemented with classification reports and the confusion metrics were also evaluated. These eight algorithms are:

1. SVM [Support Vector Machine]
2. Decision Tree Classifier
3. Naive Bayes
4. K-Nearest Neighbor
5. XG BOOST
6. Neural Network
7. AdaBoost Classifier
8. Random Forest

The paper [29] conducted a survey on sentiment analysis and summarizing user feedback on the Web. Various algorithms are used here, and this approach is suitable for opinion analysis in specific domains such as movies, products, hotels, etc. In this paper eight algorithms are selected because of their versatile usage. Several advantages in definite papers based on opinion in different reviews were found.

For both classification and regression challenges, a supervised machine learning algorithm, Support Vector Machine (SVM) can be used. [30]. In paper [31], SVM and NLP algorithms are used for Opinion Mining on newspaper headlines because of their satisfactory performance. Its performance is incredibly good in experimental results and is independent of the dimensions of the dataset. All of these, SVM is good for biological reading and interpretation [32]. It offers more advantages at the textual content type whilst excessive-dimensional areas are being used. It is used extensively in diverse real-time packages with an excessive scope in comparing appropriate outcomes. It is used in Text categorization, Image classification, Medicine, Bioinformatics, Signature/Handwriting recognition, Pattern recognition, Email spam categorization. Because of these types of benefits, we select SVM algorithms.

Decision trees are a decision support tool that uses a tree-like selection model and outcomes such as accident outcomes, resource costs, and benefits. In paper [33], fake product review was monitoring using opinion mining algorithm. The Decision Tree algorithm was used to identify its outcomes, as it performed well. It is one way to display an algorithm that contains only conditional control statements. It is easy to interpret and explain. That is why we selected the Decision Tree algorithm.

Naïve Bayes is a super simple algorithm, which just needs to do a bunch of counts. [34] used three different classification methods to classify the emotion of Roman-Urdu opinions. Naive Bayes excels in higher accuracy, higher retraction, and higher F-measure value. We selected this because of its model which is quite easy to interpret and has an efficient computation. Naïve Bayes is the most suitable for textual classification [32]. In paper [35], The study not only concentrates on the sentiment of reviews but also predicts the rating of the movie using opinion mining algorithms. Different algorithms are used to compare their accuracy. Out of all these algorithms, Naive Bayes gave a superior performance. It is used in Ongoing Prediction, Multi-Class Prediction, Content Grouping/ Spam Filtering/ Sentiment Analysis, Suggestion Systems [36].

Supervised learning technology supports a few machine learning algorithms. K-Nearest Neighbor is one of them. This presumes similarities between new and available cases. After that, places the new cases in the category that most closely resembles the available categories. In the paper [37], it shows that K-Nearest Neighbor Algorithm gave high accuracy in most of the cases in medical datasets. This is a surprisingly good classifier, but when applied to text (nominal) data, all performance parameters change based on the size of the dataset which is a good side of an opinion mining algorithm in textual dataset. So, we use it for this paper. Being a non-parametric method, it is widely being practiced for classification and regression. This algorithm is used in distinct types of fields like Finance, Medicine, and Agriculture.

XGBoost, a decision tree-based ensemble machine learning algorithm is used for supervised learning problems. This uses a gradient boosting framework. In the article, [38] worked with product recommendations while implementing XGBoost classifier in content-based filtering. Compared with other algorithms, XGBoost algorithms provided higher output in the recommendation system. XGBoost is an ameliorated boosting library that is extremely efficient, flexible, and portable. It provides parallel tree boosts to solve many data science problems quickly and accurately. XG- Boost is an efficient and easy-to-use algorithm that offers high performance and accuracy compared to other algorithms. So, we also chose this algorithm for its high performance.

A classifier that takes specific dataset to make a set of decision trees and averages them to improve the predictive accuracy of that dataset, is called Random Forest Classifier. These algorithms can not only be used for classification, but also for regression tasks. Cross validation provides higher accuracy. In paper [39], different algorithms are used in their research work for performing comparative Opinion Mining. Random Forest provides higher accuracy in different tasks and performed very well. Random Forest Algorithms is used in Movie reviews [40]. Here also it performed good as well. In another article [41], where pertinency of different algorithms for presenting comparative opinion mining is differentiated, it is also good to classify comparative opinions into nine polar classes and random forests. Also, in paper [42], used these algorithms on a Twitter data stream and as well as Random Forest Algorithm displayed acceptable performance. In a substantial proportion of data, random forest classifier will maintain accuracy and handle the missing values. That is the reason we selected this algorithm.

Neural networks are designed to work just like the human brain does. It contains a set of algorithms which aim to detect relationships between datasets. The paper [3] conducted a survey on sentiment analysis algorithms. In the research field, sentiment analysis has become an immensely popular field. Supervised techniques provide better accuracy. Neural Network is also a supervised algorithm. In paper [43], a neural network-based model for discovering overall aspect weights in sentiment analysis, demonstrates the excellent performance of neural network classification algorithms. Neural networks are great for discovering and estimating existing patterns in your data. While working with colossal neural networks, Gradient descent algorithm is recommended. Their performance in predicting future pattern changes is not very impressive. It has also been used in our research for its benefits.

The AdaBoost classifier starts by adjusting the classifier to the original dataset and then an additional copy of the classifier to the same data set. By adjusting the weights of the misclassified instances, such that - a meta estimator that allows subsequent classifiers to focus on more difficult cases. AdaBoost classifier algorithm is used in Novel Opinion Mining System for movie reviews in Turkish and performed very well with good accuracy in paper [44]. In another paper [45], the emotional classification of customer write-ups about cars in Roman Urdu used the AdaBoost classifier.

Here also it provided a good outcome and its accuracy was also good. Another use of the AdaBoost Algorithm is on customer reviews depending on their score by using Machine Learning Techniques [46]. To compare other algorithms, AdaBoost provides higher accuracy than other ones. Using AdaBoost in any machine learning algorithm, the performance can be improved. Also it is great for weak learners. These are models that achieve greater accuracy than chance when faced with classification problems. The most appropriate and therefore most frequently used algorithm in AdaBoost is a decision tree with one level. After seeing its usages and benefits we picked it for our research paper.

4. Sample Dataset

For the basic study of this research, relevant papers were reviewed to have a clear understanding of sentiment analysis and the techniques used by other researchers to analyze emotions. It was decided to work in eight selected algorithms. Some algorithms are supervised learning algorithms, others are unsupervised learning algorithms, and some are boosting algorithms that are widely used in other studies.

After selecting algorithms for the research amazon product review datasets were selected for the implementation process. The Google Colaboratory cloud system, also known as Google Colab was used for the working environment. Google Colab is more portable and easy to use as it is easier to set up than others. It has several useful features like sharing, versioning, code snippets, etc. It takes a small amount of time in its preprocessing and classification stages and also provides high accuracy for a short time as other platform that requires any expensive hardware [47]. Any person with a Gmail account can access this source, as it is an open platform of Google [48]. Google Colab works when all data sets are on Google Drive [49]. Codes written in python language can be executed using a browser on Google Colab. So, python programming languages were used for this research.

In order to work with any data sets, prepossessing of data set is needed to clean the data. Because the data set contains missing values, null rows, URLs, emails, special characters, accented characters, HTML tags and the wrong spelling of words. If the data set is used without prepossessing, then the accuracy rate of algorithms will be not satisfying. So, in this research Kgptalkie package was used for data prepossessing and cleaning the data before implementation of algorithms. Pandas, NumPy, Matplotlib, Xgboost, Seaborn and Scikit Learn python libraries have been used for implementation, finding out the classification report and confusion metrics of the algorithms. A total of eight algorithms were implemented with those classification reports and confusion metrics were evaluated also. The implementation was done on all three different data sets of amazon to get better outcomes.

A lot of datasets are available online. Researchers choose those based on their preferred topic to conduct the work. Considering the increased number of online shopping due to Covid-19 pandemic product review data set had been selected for implementation. Amazon is one of the biggest e-commerce websites, and an uncountable number of reviews are there. For this reason, the datasets were collected from there containing three individual sets of data with different numbers of reviews. The first two datasets have a substantial amount of data which are 34,661 and 28,333 in number respectively, whereas the last dataset has only 5001 data. However, all the datasets contain 25 individual columns where the information of products is given along with review and rating information. Such as- product manufacturer details, product category, product id, product image URL (Uniform Resource Locators), uploading and updating date of the product, brand name, source URL etc. In terms of review, there are given review id, review date and time, recommendation of the review, review rating, review source URL, review giver username, title of the review and the texts written as review. Despite having immense details in each product, only review texts and review ratings were used for analysis. Review texts were given by the customers of amazon and ratings are based on the scale of one to five. As Google Colab is totally based on Google Drive so we have uploaded out datasets in Google Drive from where we implemented it.

5. Analysis and Discussion

In the above section, several opinion mining algorithms have been introduced. Distinct algorithms use definite methods and techniques to implement, analysis, train, and test. A few of them work faster, where some give better accuracy, and many help to generate rules, etc. For this reason, there are various features to compare among algorithms. With the help of comparison and contrast, the features can be understood easily.

In the following table certain features are given

Features	Support Vector Machine	Decision Tree Classifier	Naïve Bayes	K-Nearest Neighbor	Neural Network	Random Forest
Type	Supervised	Supervised	Supervised	Supervised	Learn Supervised	Supervised
Learning based on	Instance	Case	Probability	Instance	Memory	Tree
Simplicity	Very Simple	Simple	Simple	Simple	Moderate	Simple
Memory Efficiency	High	High	Moderate	Very Low	Low	Very Low
Good for	Classification and regression problem	Categorical and numerical data	Multiclass Prediction Problem	Multiclass Classification	Discovering existing patterns in data and extrapolating them	Classification and regression problem
Time Required for Training	Arbitrary long	Least	Less	Least	Less	Recurrent Learning with every dataset
Generative or Discriminative Model	Discriminative	Discriminative	Generative	Discriminative	Both	Discriminative
Advantage	More effective in high dimensional spaces	Less effort required for data preparation	Simple and easy to implement	Store training dataset and learn from it only at the time of making real time prediction	Ability of parallel processing, Fault tolerance	Reduce overfitting

Fig. 1. Comparison Table of Algorithms

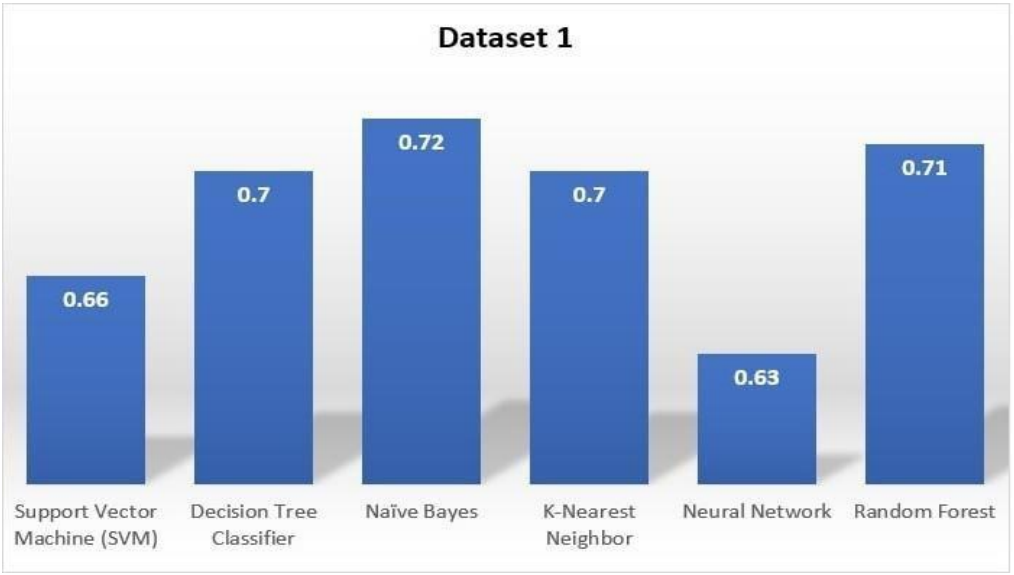

Fig. 2. Accuracy rate of algorithms for Dataset 1

The Fig.2 depicts accuracy rate of six different sentiment analysis algorithms in Amazon Product Review Dataset-1. Naïve Bayes classifier gives the highest accuracy rate which is 72 percent. However, Random Forest classifier is 1 percent less than the previous one. The third highest accuracy rate was gained by both Decision Tree Classifier and K-Nearest Neighbor which is 70 percent. For Support Vector Machine (SVM), accuracy percentage is slightly more than

65. On the other hand, Neural Network has gained the least accuracy that is 63 percent. The other two algorithms which are also included for the analysis purpose did not work well in Dataset-1. Those took plenty of time to process and train data. For these reasons, it seems inefficient to use XGBoost and ADABoost algorithm for big datasets.

Accuracy of selected opinion mining algorithms on Amazon Product Review Dataset-2 has been illustrated in Fig.3.

In this case, Random Forest Classifier is at peak with 86 percent accuracy rate.

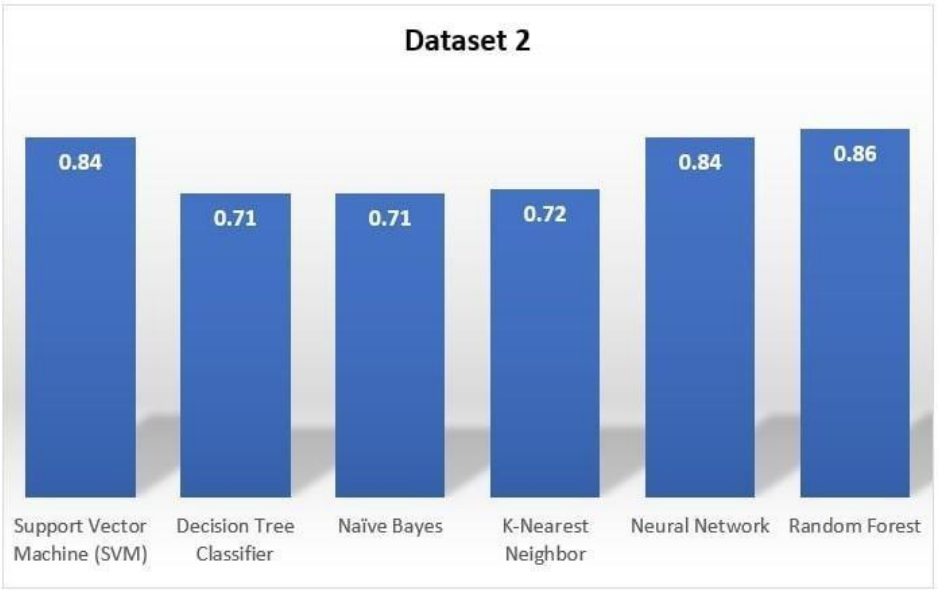

Fig. 3. Accuracy rate of algorithms for Dataset 2

Oppositely, absolute minimum is 71 percent which number represents the accuracy rate of two algorithms i.e., Decision Tree Classifier and Naïve Bayes sequentially. Nonetheless, Neural Network and Support Vector Machine (SVM) obtained the second highest accuracy rate. While on the contrary, K-Nearest Neighbor came by second least accuracy rate that is around 72 percent. XGBoost and ADABoost algorithm worked here as same as before so these two are discarded here also.

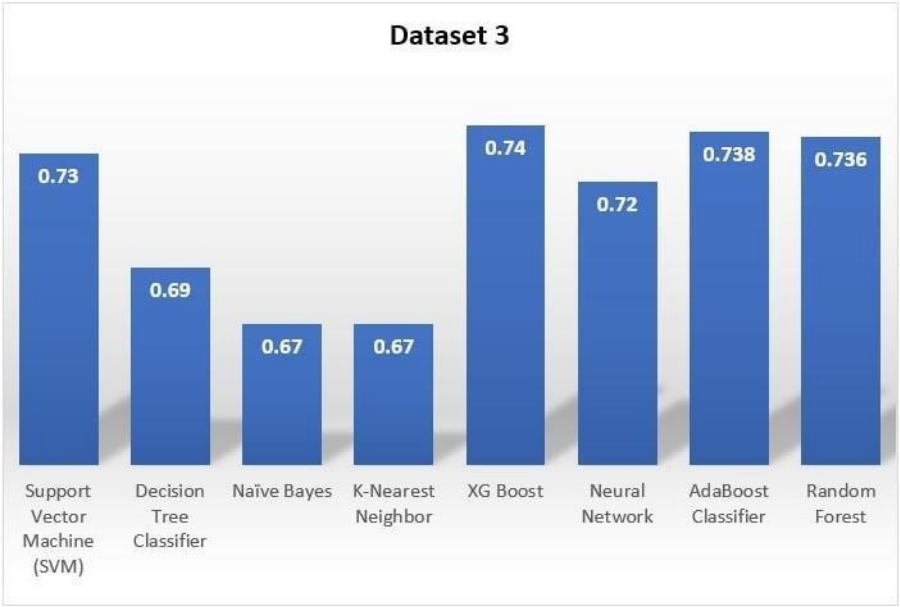

Fig. 4. Accuracy rate of algorithms for Dataset 3

Fig.4 shows accurateness of eight different sentiment mining algorithms in Amazon Product Review Dataset-3. Being a small dataset, all the selected algorithms worked very well in this. Among them, the XG Boost algorithm was introduced with an accuracy rate of 74 percent which is the highest one. After that, AdaBoost Classifier gave 73.8 percent and Random Forest gave 73.6 percent accuracy. These two are nearer to the peak. However, the percentage of

Support Vector Machine (SVM) has slightly dropped than Random Forest. Next lowest percentage is 72 which is for Neural Network. Thenceforth, accuracy rate significantly fell to 69 percent in terms of Decision Tree Classifier. Lastly, K- Nearest Neighbor and Naïve Bayes achieved the lowest accuracy which is 67 percent.

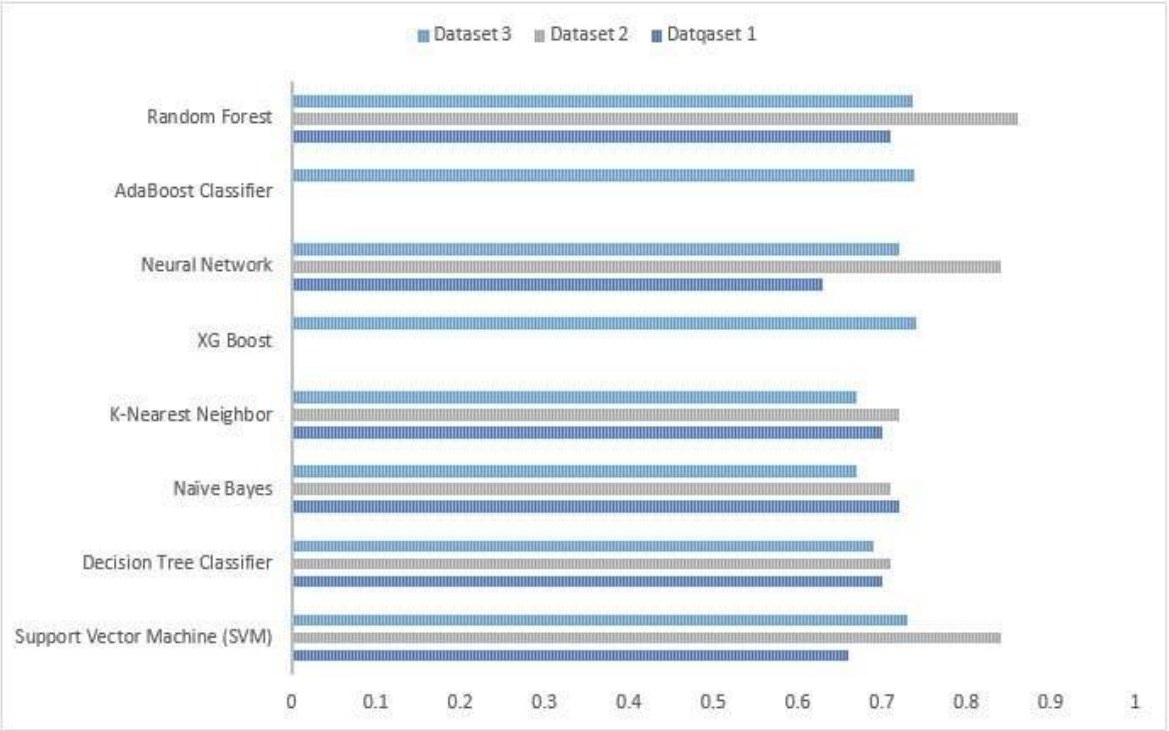

Fig. 5. Accuracy Comparison of algorithms on three distinct Amazon Product Review Dataset

The above line graph displays an accurate disparity between several machine learning algorithms in three unique datasets of Amazon. Here, the highest accuracy has been derived from Naive Bayes Classifier in Dataset-1. Subsequently in Dataset-2 Random Forest classifier gave the best result whereas XG Boost provides excellent result in terms of Dataset-3. Among the three datasets, dataset one and dataset two contained the best amount of review data. For this reason, two boosting algorithms XGBoost and ADABoost did not perform well in such a large dataset. Their training time is high, which is very inefficient to use. Naïve Bayes classifier, Random Forest classifier and XG Boost algorithms gave the highest accuracy in dataset 1, 2 and 3 sequentially. Other algorithms also worked well in dataset 3 as it contained a small amount of data.

6. Conclusion

Comparison of opinion mining algorithms in specific datasets has been done for the paper. In this research, several sentiment analysis algorithms have been used on three amazon product review datasets. They were both large and small in size, on which six supervised algorithms and two boosting algorithms were implemented that had been selected after analyzing several relevant papers. After implementing the algorithms, it has been found that on large dataset Naïve Bayes and Random Forest algorithm works better. Whereas for small datasets boosting algorithms work best, although almost all algorithms also work well in this case. This research comparison of algorithms will help other researchers in decision making so they can easily choose between algorithms, which to apply and which to avoid while working on different sized datasets.

Opinion Mining/Sentiment Analysis faces a lot of challenges. Usage of abbreviations, spam contents, use of different synonyms, usage of different languages, etc. are quite common problems [23]. Aside, it is difficult to classify sentence emotions because of different writing styles. Elimination of spam, fake and duplicate reviews is laborious. To overcome these issues, a rating system can be implied on reviews with new pre-processing and stemming techniques in the near future. Furthermore, comparison of other tools and algorithms can also be evaluated using distinct datasets on them.

References

[1] Kasturi Dewi Varathan, Anastasia Giachanou, and Fabio Crestani, “Comparative opinion mining: a review,” Journal of the Association for Information Science and Technology, vol. 68, no. 4, pp. 811–829, 2017.

[2] Amandeep Kaur and Vishal Gupta, "A survey on sentiment analysis and opinion mining techniques," Journal of Emerging Technologies in Web Intelligence, vol. 5, no. 4, pp. 367–371, 2013.
[3] Vidisha M Pradhan, Jay Vala, and Prem Balani, "A survey on sentiment analysis algorithms for opinion mining," International Journal of Computer Applications, vol. 133, no. 9, pp. 7–11, 2016.
[4] Ayesha Rashid, Naveed Anwer, Muddaser Iqbal, and Muhammad Sher, "A survey paper: areas, techniques and challenges of opinion mining," International Journal of Computer Science Issues, vol. 10, no. 6, pp. 18, 2013.
[5] Donia Gamal, Marco Alfonse, El-Sayed M El-Horbaty, and Abdel-Badeeh M Salem, "Analysis of Machine Learning Algorithms for Opinion Mining in Different Domains," Machine Learning and Knowledge Extraction, vol. 1, no. 1, pp. 224–234, 2019.
[6] Vivek Kumar Singh, Rajesh Piryani, Ashraf Uddin, and Pranav Waila, "Sentiment analysis of movie reviews: A new feature-based heuristic for aspect-level sentiment classification," in 2013 International Mutli-Conference on Automation, Computing, Communication, Control and Compressed Sensing (iMc4s), IEEE, pp. 712-717.
[7] Hajar Rehioui and Abdellah Idrissi, "New clustering algorithms for twitter sentiment analysis," IEEE Systems Journal, vol. 14, no. 1, pp. 530–537, 2019.
[8] Kerstin Denecke and Yihan Deng, "Sentiment analysis in medical settings: New opportunities and challenges," Artificial intelligence in medicine, vol. 64, no. 1, pp. 17–27, 2015.
[9] M Edison and A Aloysius, "Concepts and Methods of Sentiment Analysis on Big Data," International Journal of Innovative Research in Science, Engineering and Technology, vol. 5, no. 9, pp. 16288–16296, 2016.
[10] Monika Mishra, Jaydeep Chopde, Maitri Shah, Pankti Parikh, Rakshith Chandan Babu, and Jongwook Woo, "Big data predictive analysis of amazon product review," in KSII The 14th Asia Pacific International Conference on Information Science and Technology, APIC-IST, KSII, Beijing, China.
[11] Karen Robson, Mana Farshid, John Bredican, and Stephen Humphrey, "Making sense of online consumer reviews: A methodology," International Journal of Market Research, vol. 55, no. 4, pp. 521–537, 2013.
[12] Mohamed Elhag M Abo, Nordiana Ahmad Kharman Shah, Vimala Balakrishnan, and Ahmed Abdelaziz, "Sentiment analysis algorithms: evaluation performance of the Arabic and English language," in 2018 International Conference on Computer, Control, Electrical, and Electronics Engineering (ICCCEEE), IEEE, pp. 1–5.
[13] Mukesh Kumar, AJ Singh and Disha Handa, "Literature survey on student's performance prediction in education using data mining techniques," International Journal of Modern Education and Computer Science, vol. 7, no. 6, pp. 46-49, 2017.
[14] Almarabeh and Hilal, "Analysis of students' performance by using different data mining classifiers," International Journal of Modern Education and Computer Science, vol. 9, no. 8, pp. 9, 2017.
[15] Megha Rathi and Vikas Pareek, "Spam mail detection through data mining - A comparative performance analysis," International Journal of Modern Education and Computer Science, vol. 5, no. 12, pp. 31, 2013.
[16] Zahid Ullah, Muhammad Fayaz and Asif Iqbal, "Critical analysis of data mining techniques on medical data," International Journal of Modern Education and Computer Science, vol. 8, no. 2, pp. 42, 2016.
[17] Anaıs Collomb, Crina Costea, Damien Joyeux, Omar Hasan, and Lionel Brunie, "A study and comparison of sentiment analysis methods for reputation evaluation," Rapport de recherche RR-LIRIS-2014-002, 2014.
[18] Bakhtawar Seerat and Farouque Azam, "Opinion mining: Issues and challenges (a survey)," International Journal of Computer Applications, vol. 49, no. 9, 2012.
[19] Eivind Bjørkelund, Thomas H Burnett, and Kjetil Nørvåg, "A study of opinion mining and visualization of hotel reviews," in Proceedings of the 14th International Conference on Information Integration and Web-based Applications & Services, pp. 229–238.
[20] Ali Harb, Michel Plantié, Gerard Dray, Mathieu Roche, François Trousset, and Pascal Poncelet, "Web Opinion Mining: How to extract opinions from blogs?," in Proceedings of the 5th international conference on Soft computing as transdisciplinary science and technology, pp. 211–217.
[21] Kumar Ravi and Vadlamani Ravi, "A survey on opinion mining and sentiment analysis: tasks, approaches and applications," Knowledge-based systems, vol. 89, pp. 14–46, 2015.
[22] Bohlouli, Mahdi and Bigham, Bahram and Narimani, Zahra and Vasighi, Mahdi and Ansari, Ebrahim, "Data Science: From Research to Application," 2020. DOI: 10.1007/978-3-030-37309-2.
[23] G Sneka and CT Vidhya, "Algorithms for Opinion Mining and Sentiment Analysis: An Overview. International Journal of Advanced Research in Computer Science and Software Engineering," vol. 6, no. 2, pp. 1–5, 2016.
[24] Bo Pang, Lillian Lee, and Shivakumar Vaithyanathan, "Thumbs up? Sentiment classification using machine learning techniques," arXiv preprint cs/0205070, 2002.
[25] Sonia Anastasia and Indra Budi, "Twitter sentiment analysis of online transportation service providers," in 2016 International Conference on Advanced Computer Science and Information Systems (ICACSIS), IEEE, pp. 359–365.
[26] Anuj Sharma and Shubhamoy Dey, "A document-level sentiment analysis approach using artificial neural network and sentiment lexicons," ACM SIGAPP Applied Computing Review, vol. 12, no. 4, pp. 67–75, 2012.
[27] Yaser Maher Wazery, Hager Saleh Mohammed, and Essam Halim Houssein, "Twitter sentiment analysis using deep neural network," in 2018 14th International Computer Engineering Conference (ICENCO), IEEE, pp. 177–182.
[28] Mehmet Saygın Seyfioğlu and Mustafa Umut Demirezen, "A hierarchical approach for sentiment analysis and categorization of Turkish written customer relationship management data," in 2017 Federated Conference on Computer Science and Information Systems (FedCSIS), IEEE, pp. 361–365.
[29] Vijay B Raut and DD Londhe, "Opinion mining and summarization of hotel reviews," in 2014 International Conference on Computational Intelligence and Communication Networks, IEEE, pp. 556–559.
[30] Jayashri Khairnar and Mayura Kinikar, "Machine learning algorithms for opinion mining and sentiment classification," International Journal of Scientific and Research Publications, vol. 3, pp. 6, pp. 1–6, 2013.
[31] C Jashubhai Rameshbhai and Joy Paulose, "Opinion mining on newspaper headlines using SVM and NLP," International Journal of Electrical and Computer Engineering (IJECE), vol. 9, no. 3, pp. 2152–2163, 2019.

[32] Pravesh Kumar Singh and Mohd Shahid Husain, "Methodological study of opinion mining and sentiment analysis techniques," International Journal on Soft Computing, vol. 5, no. 1, no. 11, 2014.
[33] Anusha Sinha, Nishant Arora, Shipra Singh, Mohita Cheema, and Akthar Nazir, "Fake product review monitoring using opinion mining. International Journal of Pure and Applied Mathematics," vol. 119, no. 12, pp. 13203–13209, 2018.
[34] Muhammad Bilal, Huma Israr, Muhammad Shahid, and Amin Khan, "Sentiment classification of Roman-Urdu opinions using Naïve Bayesian, Decision Tree and KNN classification techniques," Journal of King Saud University-Computer and Information Sciences, vol. 28, no. 3, pp. 330–344, 2016.
[35] Vibhor Singh, Priyansh Saxena, Siddharth Singh, and S Rajendran, "Opinion mining and analysis of movie reviews," Indian Journal of Science and Technology, vol. 10, no. 19, pp. 1–6, 2017.
[36] Vijender Kumar Solanki, Nguyen Ha Huy Cuong, and Zonghyu Joan Lu, "Opinion mining: using machine learning techniques," in Extracting knowledge from opinion mining, IGI Global, pp. 66–82.
[37] JS Raikwal and Kanak Saxena, "Performance evaluation of SVM and k-nearest neighbor algorithm over medical data set," International Journal of Computer Applications, vol. 50, no. 14, 2012.
[38] Zeinab Shahbazi and Yung-Cheol Byun, "Product recommendation based on content-based filtering using XGBoost classifier," Int. J. Adv. Sci. Technol, vol. 29, pp. 6979–6988, 2019.
[39] Umair Younis, Muhammad Zubair Asghar, Adil Khan, Alamsher Khan, Javed Iqbal, and Nosheen Jillani, "Applying machine learning techniques for performing comparative opinion mining," Open Computer Science, vol. 10, no. 1, pp. 461–477, 2020.
[40] Palak Baid, Apoorva Gupta, and Neelam Chaplot, "Sentiment analysis of movie reviews using machine learning techniques," International Journal of Computer Applications, vol. 179, no. 7, pp. 45–49, 2017.
[41] Aurangzeb Khan, Umair Younis, Alam Sher Kundi, Muhammad Zubair Asghar, Irfan Ullah, Nida Aslam, and Imran Ahmed, "Sentiment classification of user reviews using supervised learning techniques with comparative opinion mining perspective," in Science and Information Conference, Springer, pp. 23–29.
[42] Balakrishnan Gokulakrishnan, Pavalanathan Priyanthan, Thiruchittampalam Ragavan, Nadarajah Prasath, and AShehan Perera, "Opinion mining and sentiment analysis on a twitter data stream," in International Conference on Advances in ICT for Emerging Regions (ICTer2012), IEEE, pp. 182–188.
[43] Duc-Hong Pham and Anh-Cuong Le, "A neural network based model for determining overall aspect weights in opinion mining and sentiment analysis. Indian Journal of Science and Technology," vol. 9, no. 18, pp. 1–6, 2016.
[44] Abdul Hafiz AbdulHafiz, "Novel Opinion mining System for Movie Reviews," International Journal of Intelligent Systems and Applications in Engineering, vol. 8, no. 2, pp. 94–101, 2020.
[45] Moin Khan and Kamran Malik, "Sentiment classification of customer's reviews about automobiles in roman urdu," in Future of Information and Communication Conference, Springer, pp. 630–640.
[46] Miss Lovenika Kushwaha and Mr Sunil Damodar Rathod, "Opinion mining of customer reviews based on their score using machine learning techniques," International Research Journal of Engineering and Technique, vol. 3, no. 03, pp. 2198–2203, 2016.
[47] Yilmaz, Abdurrahim and Demircali, Ali Anil and Kocaman, Sena and Uvet, Huseyin, "Comparison of Deep Learning and Traditional Machine Learning Techniques for Classification of Pap Smear Images," arXiv preprint arXiv: 2009.06366, 2020.
[48] Kanani, Pratik and Padole, Mamta, "Deep learning to detect skin cancer using google colab," International Journal of Engineering and Advanced Technology Regular Issue, vol. 8, no. 6, pp. 2167-2183, 2019.
[49] Dos Santos, Ludmila Pereira, "Using Machine Learning and Neural Networks to Optimize Car Position When It is captured by Smart Cameras," 2021.

Authors' Profiles

Sumaiya Sultana is a fresh graduate from American International University with a major in Computer Science and Engineering. She is strongly captivated to continue research and open to work in any area of Machine Learning especially Natural Language Processing and Reinforcement Learning. Besides this, Big Data and Sentiment Analysis are also the realm of her interest. She can be communicated in sumaiyasultanaa99@gmail.com.

Sumaiya Rahman Eva have recently graduated from American International University Bangladesh in Computer Science and Engineering. She has interests in Software Engineering, Management information systems, and Data Science, Data processing, Data mining algorithms, etc. She wants to do further research work in the data mining field. She can be contacted through sumaiyarahmaneva1999@gmail.com.

Nayeem Hasan Moon is from Dhaka, Bangladesh. He did his undergraduate in Computer Science & Engineering at American International University Bangladesh. He is highly interested in the development domain and learning more about various development tools and techniques. He is also extremely passionate about the data science domain. He is interested in doing further research on advanced web and software which will use data science technologies. He can be contacted at moonnayeem88@gmail.com.

Dr. Dip Nandi is currently working as an Associate Professor and Director of Faculty of Science and Technology in American International University- Bangladesh. His research interests include Software Engineering, Management Information Systems, E-learning etc. He can be contacted at dip.nandi@aiub.edu.

Akinul Islam Jony currently works as an Assistant Professor of Computer Science at American International University- Bangladesh (AIUB). All information about him can be found here (https://cs.aiub.edu/profile/akinul).

How to cite this paper: Sumaiya Sultana, Sumaiya Rahman Eva, Nayeem Hasan Moon, Akinul Islam Jony, Dip Nandi, "A Comparison of Opinion Mining Algorithms by Using Product Review Data", International Journal of Information Engineering and Electronic Business(IJIEEB), Vol.14, No.4, pp. 28-38, 2022. DOI:10.5815/ijieeb.2022.04.04