



ISSN: 0067-2904

Sentiment Analysis in Social Media using Machine Learning Techniques

Hayder A. Alatabi*, Ayad R. Abbas

Computer Science Department, University of Technology, Baghdad, Iraq

Received: 3/5/ 2019

Accepted: 18/ 8/2019

Abstract

Over the last period, social media achieved a widespread use worldwide where the statistics indicate that more than three billion people are on social media, leading to large quantities of data online. To analyze these large quantities of data, a special classification method known as sentiment analysis, is used. This paper presents a new sentiment analysis system based on machine learning techniques, which aims to create a process to extract the polarity from social media texts. By using machine learning techniques, sentiment analysis achieved a great success around the world. This paper investigates this topic and proposes a sentiment analysis system built on Bayesian Rough Decision Tree (BRDT) algorithm. The experimental results show the success of this system where the accuracy of the system is more than 95% on social media data.

Keywords: Sentiment Analysis, Opinion Mining, Sentiment Mining, Social Media.

تحليل المشاعر في وسائل التواصل الاجتماعي باستخدام تقنيات تعلم الآلة

حيدر عدنان العتابي*، إياد روضان عباس

قسم علوم الحاسوب، الجامعة التكنولوجية، بغداد، العراق.

الخلاصة

حققت وسائل التواصل الاجتماعي في الفترة الأخيرة نجاحاً كبيراً حول العالم حيث تشير الإحصائيات الخاصة بهذا الموضوع إن أكثر من ثلاث مليارات شخص حول العالم موجودون بالفعل على منصات التواصل الاجتماعي الأمر الذي أدى إلى توافر كميات هائلة من البيانات على هذه المواقع. تحليل هذه البيانات يحتاج إلى نوع خاص من عمليات التصنيف هذا النوع يسمى تحليل المشاعر. يقدم هذا البحث نظام تحليل مشاعر جديد يعتمد على تقنيات تعلم الآلة لخلق عملية استخراج قطبية نصوص وسائل التواصل الاجتماعي. بواسطة استخدام تقنيات تعلم الآلة حقق تحليل المشاعر نجاحاً منقطع النظير حول العالم. في هذا البحث تم بالبحث في هذا المجال وتم اقتراح نظام تحليل مشاعر مبني على خوارزمية BRDT. في النهاية فإن نتائج الاختبارات أثبتت نجاح هذا النظام حيث حقق هذا النظام دقة تفوق الـ 95% على بيانات وسائل التواصل.

1.Introduction

In the last decade, social media gained a massive expansion around the world and huge quantities of information have been available on the Internet as users have become more inclined to share their feelings about products, services, movies or whatever they want to share online [1].

The ability to process these large quantities of information is important because, for example, a company can introduce a new product into the market and wait for people's responses on the Internet, while later they can extract these feelings usefully and decide the future viability of this type of

*Email: h6yder@yahoo.com

products. Most of this information has not been classified and rated by any type of classification to become easy to use, making it difficult to classify these huge quantities using regular human tools. For this reason, other tools have been developed in order to confer the ability to read the text and extract features. *f e e l i n g s p o l a r i t y .*

The Natural Language Processing (NLP) is the path in Computer Science (CS), Artificial Intelligence (AI) and linguistics that seeks to give the machines the capacity to understand the natural language, e.g. English and Arabic. Inside NLP, there is the domain of Sentiment Analysis that studies the modality of using the machines to process texts and to give each one a kind of classification that we can use and understand. This domain uses the language processing algorithms for extracted features, such as word frequency, and uses supervised machine learning algorithms that learn from an initial set of data originally classified by a human [2].

Unfortunately, the machines are limited, and one must be careful with the kind of text to be processed which largely influences the size of vocabulary that the algorithm needs to learn and the size of the text to be processed. For example, micro blogging sites like Facebook or Twitter use closed and short sentences and the users can use any language in formal or informal forms. The same word can appear with lots of different representations. On the other side, sites like Internet Movie Database (IMDb) contain reviews about movies with large texts and have a more formal language [3].

In this paper, the main aim is to construct a special system and give it the ability to classify each input review as positive sentiment or negative sentiment employing a recently developed Bayesian Rough Decision Tree Algorithm, taking into account the computation time and the system accuracy.

2. Related Work

There is a very large number of research conducted in the field of opinion analysis, from the approach of rule-based and bag-of-words (BOW) approach to the techniques of machine learning. There are two major directions of sentiment analysis research, which are the document-level and the sentence-level. Usually, both these classification methods are based on identifying words that hold opinions or phrases [4].

The authors of a previous article [5] proposed a method to analyze movie and hotel review datasets where movie reviews were scanned from IMDb and hotel reviews from OpinRank. Each dataset was prepared by having 5000 positive and 5000 negative reviews. They achieved the best results when the number of reviews in the training dataset is 4500. In the movie reviews dataset, the accuracy value of Naïve Bayes was 82.43% and that of K Nearest Neighbor (KNN) was 69.81%, whereas in the hotel reviews dataset the values were 55.09% and 52.14%, respectively.

In another publication [6], the authors obtained 50000 reviews from IMDb. They labeled reviews with a score of more than 6 as positive and those with a score of less than 5 as negative. Then, they took 25000 reviews as a training dataset and 25000 reviews as a testing dataset. They applied Random forest, logistic regression and Support Vector Machine (SVM) and performed the bag-of-words model. At the end, the system achieved accuracy of up to 84.3% with the Random forest classifier on their tested data.

Authors of another article [7] used Twitter as a training data source to perform sentiment analysis, in which they collected the messages that contained the emoticons :) and :(via the Twitter API. For testing data, they used 75 negative tweets and 108 positive tweets that were manually rated. They used Naïve Bayes, Maximum entropy and SVM, with Naive Bayes classifier showing the best results using mutual information measures for feature selection. The authors obtained about 81% accuracy on the test set with Naïve Bayes.

In another publication [8], the authors used three machine learning techniques to analyze movie review datasets. They did not clarify the method that they used to collect their dataset but they used a preprocessing stage to clean the dataset to reduce noise, feature extraction stage, extract some features, and finally perform the classification algorithm. The accuracies of the used classifiers were 81.4% for the Naive Bayes, 78.65% for the Random Forest and 55.3% for the K Nearest Neighbor.

Authors of another investigation [9] proposed a system to analyze movie reviews. They obtained 1800 tweets about movie reviews, labeled as 600 tweets for each class of positive, negative and neutral. They used Naïve Bayes and SVM to train the model. Finally, They achieved accuracy values of up to 87% using Naïve Bayes and 90% using SVM.

For the sentiment analysis, author of another study [10] decided to collect the texts that contain emoticons from Usenet newsgroup, in which the training set was created by using emoticons such as smile, wink, frown and more others. The dataset was split into positive and negative by regarding the texts containing emoticons that holds happiness meaning as positive and those holding sadness or anger meanings as negative. They utilized SVM and Naïve Bayes to train the model and obtained accuracy levels of about 70.1% with SVM and 61.5% with Naïve Bayes on their tested set.

In another article [11], the authors decided to analyze movie review datasets from IMDb that consisted of 1000 positive reviews and 1000 negative reviews. They used Naïve Bayes and SVM to train the system. At the end, they achieved accuracy values of up to 87.9% using Naïve Bayes and 86.5% using SVM.

In an additional example [12], the authors decided to analyze multiple domains of reviews using a simple unsupervised learning algorithm, as recommended (thumbs up) or not recommended (thumbs down). The semantic orientation average of the phrases in the review that contained adverbs or adjectives predicted the classification of there view. The system achieved accuracy levels of up to 84% for the reviews of automobiles, 80% for the reviews of banks and 66% for the movie reviews.

3. Datasets

In this paper, two different datasets were used; Facebook data consisted of 4000 comments categorized as 2000 positive and 2000 negative [13], while movie reviews dataset consisted of 2000 reviews categorized as 1000 positive and 1000 negative reviews [14].

Table 1-Datasets Details

| Dataset | Positive | Negative | Total |
|---------------|----------|----------|-------|
| Facebook | 2000 | 2000 | 4000 |
| Movie reviews | 1000 | 1000 | 2000 |

4. Bayesian Rough Decision Tree

Bayesian Rough Decision Tree is a combination of the Bayesian Rough set and the Decision Tree. The system improves the decision tree by determining the threshold value instead of computing the entropy of the decision [15].

The threshold value (α) is in the range 0 to 1 in order to control the boundary region degree rather than using the threshold.

Where α is $0.01 \leq \alpha \leq 0.99$

Computing the posterior probability ($A_i|X_j$) and prior probability (X_j) is used instead of computing entropy for each feature of the dataset in the decision tree.

$$P(X_j|A_i) = \frac{P(A_i \cap X_j)}{P(A_i)}$$

Where (X_j) is the prior probability of the decision, A_i is the value of each discrete value i in feature A , and X_j is the value of the decision j in the dataset. Then, the result value of ($X_j|A_i$) is compared with α :

$$\begin{aligned} POS(X_j) &= \cup \{A_i: P(X_j|A_i) \geq \alpha\} \\ NEG(X_j) &= \cup \{A_i: P(X_j|A_i) \leq 1-\alpha\} \\ BND(X_j) &= \cup \{A_i: 1-\alpha < P(X_j|A_i) < \alpha\} \end{aligned}$$

where U is the finite universe of objects.

Computing the discriminant index was based on positive regions for each feature of the dataset. The discriminant index is used instead of computing information gain of the decision tree.

$$\eta = \frac{\sum card(POS(X_j))}{card(U)}$$

where η is the discriminant index.

U is a number of instances in the dataset.

$card$ is the cardinality of the dataset.

Subsequently, the highest discriminant index feature among the features is selected as the root node of the tree, and later divide the dataset according to the highest discriminant.

When the result of ($X_j|A_i$) for each branch of the highest discriminant index is (X_j) or $NEG(X_j)$ then it is a leaf node, otherwise $BND(X_j)$ is define as a non-leaf node and requires further splitting.

This process is iterated until the last feature of the dataset. When features are split and reached to the last feature in the dataset, there is a need for further splitting to distinguish positive or negative sentiment.

5. Sentiment Analysis System

The Sentiment Analysis system is an automatic system to detect the polarity of the input text. This system builds on machine learning techniques and passes through several steps; the first step of the system is to collect the data. Next, the text processing step aims to eliminate noise from text using

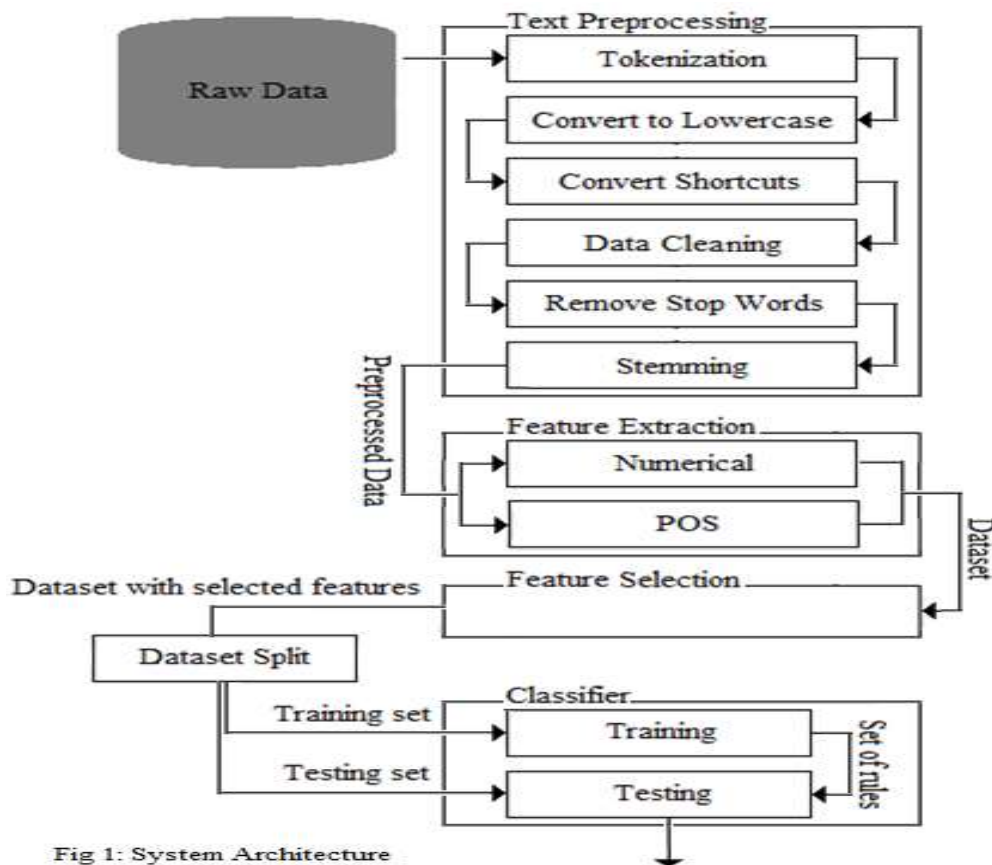


Fig 1: System Architecture

operations such as tokenization, Stop words removal, and Stemming. Feature Extraction step is to extract features from data. Feature Selection is applied to select the most useful feature subset from the feature set. Then, the classification algorithm Bayesian Rough Decision Tree is used to analyze data(Figure-1).

5.1 Data Selection

Data selection is the first stage of any classification system, targeting the selection of the most suitable data for the classification task. In this paper, two datasets were selected, one is Facebook dataset and the other is movie reviews from IMDb, as mentioned in 3.

5.2 Text preprocessing

Text preprocessing is the most important stage in any text classification system. The importance of this stage is illustrated by its benefit of filtering the text by removing unwanted words and converting words into suitable representations. Text preprocessing covers the following preprocessing operations:

a. Tokenization

Tokenization is the process of splitting a text document into small parts such as words, phrases, symbols or other elements called tokens. Tokenization gained its importance from the need for the upcoming steps to deal with each single word as an input. The process of tokenization is performed by scanning the input text and saving each separate piece of text (separated by space) as a token.

b. Data Cleaning

Data cleaning is the process of removing noise from the text by using the following steps:

1. Conversion to Lower-case: Having one text form will facilitate the process of dealing with words. This step is important because without it the same word with different characters forms will be presented as a different word, creating problems to the system when it needs to work at the possible speed and memory.
2. Conversion of Shortcuts: All shortcuts should be converted into their original words. This converting process is very important because it will reduce the number of words that the system steps need to deal with. This converting process should pass through the following:
 - 's → is 'm → am 're → are 't → not 've → have
 - 'll → will 'd → would
3. Removal of URL: The Uniform Resource Locator (URL) is the address that shows the place where a particular page can be found on the World Wide Web. URL should be removed from the text because it is not a required part of the text for the classification task. Removing the URL is performed by scanning the tokens list and removing every token that contains URL.
4. Removal of Emojis: Emojis is a digital image used in social media texts. Emojis are removed from the text because it is not an appropriate part of the text for the process of Sentiment Analysis. When the text enters the system the signs of emojis appear as a triple question marks "???", then the removal process is easy by eliminating every token that consists of triple question marks.
5. Removal of Slashes between two words: In this sub-step, a slash '/' between two words is removed from the text. This process can be performed by searching the tokens list for a slash and replacing it by a space.
6. Removal of Brackets: Removing any type of brackets because it is not appropriate for this type of analysis process. This process can be performed by scanning the token's list and removing every starting and ending brackets.
7. Removal of Numbers: Numbers are not appropriate parts of the text for the Sentiment Analysis and they should be removed from the text in this step. Removing numbers can be performed by searching and removing the digits in the tokens.
8. Removal of Mentions: In Social media, there is a tool to notify users about a post or a comment by using a mention '@' to the account. Since the name of the user account and the mention sign '@' are not appropriate parts of the text for the Sentiment Analysis, mentions are removed from the text in this step. This process can be performed by searching and removing the tokens starting with the sign '@'.
9. Removal of Hashtag sign: Hashtag is a social media tool that presents all posts used the same hashtag in one place. Since the hashtag sign is not an appropriate part of the text for the Sentiment Analysis, it is removed from the text, where as the hashtag word remains because it might contain useful information about the polarity of the text. This process can be performed by searching for a token starting with the sign '#' and, when detected, the sign will be removed while the hashtag word is left.

c. Stop word removal

Stop words are those words that appear very frequently in texts but basically do not hold any meaning as they are used in a sentence to connect words. It is a common understanding that stop words do not contribute to the context or content of text documents.

In text mining, stop words are seen as a block in the content of the understanding of the text due to their high frequency of occurrence. Stop words are those words that are used very commonly like 'is', 'are', 'and', 'this' etc. They do not give any benefit in the process of document classification and, therefore, they should be eliminated. The removing process is achieved by searching the tokens and removing each token that has stop words.

d. Word stemming

Stemming is the process of linguistic normalization, which aims to reduce the different word forms to a common form.

The process of stemming is achieved by using the Porter stemming algorithm which starts by checking the keyword and following a set of rules. The next step is to eliminate endings that turn the keyword into plurals such as '-s' and '-es', past tenses such as '-ed', or continuous tenses like '-ing'. Then, the stemmer moves on to check and convert double suffixes into a single one. Other suffixes such as '-ic', '-full', '-ness', '-ant', '-ence' are also removed. For example words such as "organize",

“organized”, “organizing”, “organization”, and “organizations” should all be represented as the common representation of “organize”.

5.3 Feature Extraction

Feature Extraction (FE) is a significant step that represents the texts in the form of a string of features to be used in the next stages. It contains the following operations:

- a. **Numerical Feature Extractor:** indicates a process for extracting essential features from the text. These features are illustrated as follows:
 1. Sentences counter: counts the number of sentences in the text.
 2. Words counter: counts the number of words in the text.
 3. Negation words counter: counts the number of negation words in the text.
 4. Positive words counter: counts the number of positive words in the text.
 5. Negative words counter: counts the number of negative words in the text.
- b. **Part of speech Feature Extractor:** indicates a process for extracting numerical grammatical features from the text. These features are illustrated as follows:
 1. Nouns counter: counts the number of nouns in the text.
 2. Verbs counter: counts the number of verbs in the text.
 3. Adjectives counter: counts the number of adjectives in the text.
 4. Adverbs counter: counts the number of adverbs in the text.

5.4 Feature Selection

Feature selection is the procedure that aims to select the most useful features to be used in the construction of the model. The feature selection methods can be used at the identification and removing of irrelevant and redundant features from data that actually do not take part in raising the accuracy of a predictive model or probably scale down the accuracy.

Model complexity can be reduced by using fewer features, since a simple model is simpler to understand and explain.

In this paper, Correlation-based Feature Selection (CFS) was used as an evaluator and Best First Search (BFS) algorithm as a searching method.

CFS is a heuristic approach used in order to evaluate the worth or merit of the features subset. It is a technique to rank the relevance of the features by measuring the correlation between features and classes and between features and the other features, in which the evaluation function bias is toward the subsets of features that have a high correlation with the class and have no correlation with each other.

Eliminating attributes with low correlation with the class, which are called Irrelevant features.

Screening out features with high correlation with one or more of the remaining features, which are called Redundant features.

The feature's acceptance depends mainly upon the extent of the feature to which it predicts classes in areas of the instance space not already predicted by other features.

BFS is an important AI search strategy that allows backtracking along the search path. Best first search moves through the search space by making local changes to the current feature subset [16].

5.5 Classifier Algorithm

In this stage, the dataset is prepared and ready to be classified. First, the value of threshold a is selected in the range of 0 to 1 in order to control the boundary degree, where a is $0.01 \leq a \leq 0.99$

In this paper, the dataset is divided using k-fold cross-validation into k groups where $k=10$, in which the dataset is split into 10 equal groups. For each fold of the 10 folds the model uses (k-1) groups as the training set and the remaining k group for testing.

Then, the evaluation measures of folds are collected to find the overall evaluation measures. For each fold, the training dataset is passed through the following operations:

1. Computing the posterior probability $(A_i|X_j)$ and prior probability $P(X_j)$ for each feature.
2. Comparing the result of $(X_j|A_i)$ with a .
 - if $(X_j|A_i) \geq 1-a \rightarrow POS(X_j)$
 - if $(X_j|A_i) \leq a \rightarrow NEG(X_j)$
 - if $1-a > P(X_j|A_i) > a \rightarrow BND(X_j)$
3. Computing the discriminant index based on positive regions for each feature of the dataset.
4. Selecting the feature that holds the highest discriminant index as the root node of the tree, then splitting the dataset according to this feature.

When the result of $(X_j|A_i)$ for each branch of the highest discriminant index is (X_j) or $NEG(X_j)$ as a leaf node. When the result of $(X_j|A_i)$ for each branch of the highest discriminant in the index is neither $POS(X_j)$ nor $NEG(X_j)$, then it is $BND(X_j)$ will be defined as a non-leaf node. Hence, $BN(X_j)$ requires further splitting.

This process is iterated until the last feature of the dataset. When features are split and reach to the last feature in the dataset, there is a need for further splitting to distinguish positive and negative.

When the training has been completed, the system will start testing by predicting the class of the test dataset without seeing the class and comparing the actual class with the predicted class to extract the evaluation measures.

6. Results and Discussion

The experimental results show the efficiency of the system. In this paper, a comparison was made between Bayesian Rough Decision Tree and Decision Tree (DT), both using K-cross validation in which the dataset is divided into K equal groups. In each fold, there are K-1 groups for training the model and the remaining group is for testing. Then, the evaluation parameters are computed to present the overall performance of the system.

Through the experiments on the Facebook dataset, the best results are gained when the threshold value is in the range of 0.01 and 0.02. Therefore this result is considered as the final result of the system on the Facebook dataset. On the other side, the best results are gained when the threshold value is in the range of 0.01 and 0.04. Therefore this result is considered as the final result of the system on the movie reviews dataset.

The performance of BRDT and DT algorithms on the Facebook dataset can be quantified using Accuracy, Precision, Recall and F1-measure.

Table 2-Evaluation of BRDT and DT on the Facebook dataset

| | Accuracy | Precision | Recall | F1-measure |
|------|----------|-----------|--------|------------|
| BRDT | 99.625% | 0.999 | 0.994 | 0.996 |
| DT | 98.95% | 0.993 | 0.988 | 0.99 |

The performance of BRDT and DT algorithms on the Movie reviews dataset can be quantified using Accuracy, Precision, Recall and F1-measure.

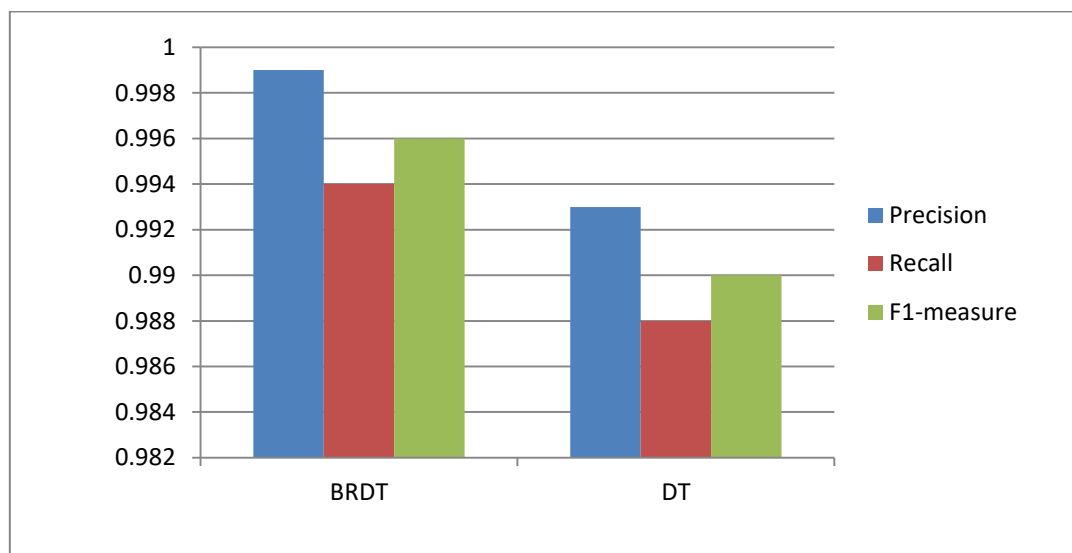


Figure 2-Graph chart of performance measures using BRDT and Decision Tree Algorithms on the Facebook dataset.

The graph chart in Figure-2 that could be created from Table-2 shows that BRDT algorithm has better results than Decision Tree algorithm, which can be easily noticed as Precision, Recall and F1-

measure values using BRDT algorithm are greater than the ones using Decision Tree algorithm on the dataset of Facebook.

Table 3-Evaluation of BRDT and DT on movie reviews dataset.

| | Accuracy | Precision | Recall | F1-measure |
|------|----------|-----------|--------|------------|
| BRDT | 96.15% | 0.968 | 0.956 | 0.962 |
| DT | 95.45% | 0.958 | 0.955 | 0.954 |

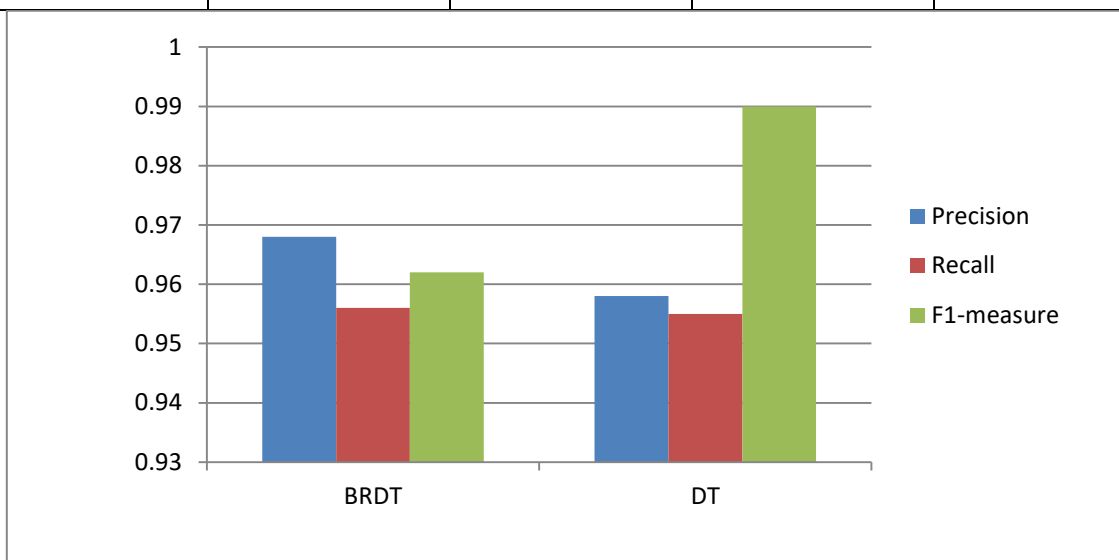


Figure 3-Graph chart of performance measures using BRDT and Decision Tree Algorithms on Movie reviews dataset.

The graph chart in Figure-3 that could be created from Table-3 shows that BRDT algorithm has better results than Decision Tree algorithm, which can be easily noticed as Precision, Recall and F1-measure values using BRDT algorithm are greater than ones using Decision Tree algorithm on the dataset of movie reviews.

7. Conclusion

Sentiment analysis is the trending technology in recent times, providing a distinctive service for individuals and institutions in a form of daily use. The need for this technology is very large when it comes to manufacturers, for example, when these companies need some kind of inquiry about their products, services, its policy, or anything else. It is also a useful service for individuals when a person needs to ask about a movie, a product to buy, a specific subject or anything else. Thus, this technique saves time and effort for anyone who needs to inquire about a particular subject.

During the execution of this system, simple notes have been noticed. For instance, operations like text preprocessing have a big impact on the accuracy of the system as this process facilitates dealing with words and the deletion of unwanted words. This would facilitate the work of Feature Extraction, which is the backbone of any algorithm used to classify, so that good work of Feature Extraction gives better system accuracy. Also, feature selection is a significant process because it reduces the number of features by selecting the most useful set of features. Decreasing the number of features means less training time.

The best possible results could be obtained using BRDT algorithm when the threshold value is in the range of 0.01 and 0.02, where the system accuracy reaches 99% on the Facebook dataset, and with threshold value within the range of 0.01 and 0.04, when the system accuracy reaches 96% on movie reviews dataset. On the other side, when the threshold value is above 0.4, the system is crashed and its accuracy reaches the lowest value of 50% on both Facebook and movie review datasets.

References

1. Digital, social and mobile worldwide in 2018. **2018**. special report. <https://wearesocial.com/blog/2018/01/global-digital-report-2018>.
2. Bo Pang and Lillian Lee. **2008**. "Opinion Mining and Sentiment Analysis", *Foundations and Trends in Information Retrieval*, **2**(1-2): 1–135.
3. Goldberg Y. **2016**. "A Primer on Neural Network Models for Natural Language Processing". *Journal of Artificial Intelligence Research*, **57**: 345–420.
4. Poornima Singh, Gayatri S Pandi (Jain), **2015**. "Opinion Mining on Social Media: Based on Unstructured Data", *International Journal of Computer Science and Mobile Computing IJCSMC*, **4**(6): 768–777.
5. Dey, L., Chakraborty, S. and Anuraag B. **2016**. "Sentiment Analysis of review dataset using Naïve Bayes and KNN classifier", *International Journal of Information Engineering and Electronic Business*, **4**: 54–62.
6. Pouransari, H. and Ghili, S. **2014**. "Deep learning for sentiment analysis of movie reviews". Technical report, Stanford University, available at: <https://cs224d.stanford.edu/reports/PouransariHadi.pdf>.
7. Alec Go, Lei Huang and RichaBhayani, "Twitter Sentiment Analysis", Final Projects from CS224N for Spring **2008/2009** at the Stanford Natural Language Processing Group.
8. PalakBaid, Apoorva Gupta, and NeelamChaplot. **2017**. "Sentiment Analysis of Movie Reviews using Machine Learning Techniques", *International Journal of Computer Applications*, **179**(7).
9. A. Amolik, N. Jivane, M. Bhandari, and M. Venkatesan, **2015**, "Twitter Sentiment Analysis of Movie Reviews using Machine Learning Techniques", *International Journal of Engineering and Technology*, **7**(6): 2038 – 2044.
10. Jonathon Read, **2005**, "Using Emoticons to Reduce Dependency in Machine Learning Techniques for Sentiment Classification", ACLstudent '05 Proceedings of the ACL Student Research Workshop, pp:43-48.
11. N. Sudha, M. Govindarajan, PhD, **2016**, "Mining Movie Reviews using Machine Learning Techniques", *International Journal of Computer Applications* (0975 – 8887), **144**(5).
12. Turney, P. **2002**. "Thumbs up or thumbs down? Semantic orientation applied to unsupervised classification of reviews", In Proc. of the ACL.
13. Facebook Data, "kaggle.com" [online]. Available: <https://www.kaggle.com/gvrlifestyle/sample-facebook-comments>.
14. Movie Review Data, "cs.cornell.com" [online]. Available: <http://www.cs.cornell.edu/people/pabo/movie-review-data>.
15. Ayat Omar Farooq and Asst. Prof. Dr. Ayad R. Abbas. **2018**. Human Skin Color Detection using Machine Learning Techniques. MSc thesis, department of computer sciences, university of technology, Iraq.
16. Pearl, J. **1984**. *Heuristics: Intelligent Search Strategies for Computer Problem Solving*, Addison-Wesley Longman Publishing Co., Inc. Boston, MA, USA.