Journal of Theoretical and Applied Information Technology
30th November 2021. Vol.99. No 22

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

SENTIMENT ANALYSIS OF AMAZON'S REVIEWS USING MACHINE LEARNING ALGORITHMS

[1]SANA NABIL, [2]JABER ELBOUHDIDI, [3]MOHAMED YASSIN CHKOURI
SIGL Laboratory, ENSA TETUAN, Department of Computer Science, Abdelmalek Essaâdi University, Morocco

E-mail : [1]sana.nabil@etu.uae.ac.ma, [2]jaber.elbouhdidi@uae.ac.ma, [3]mychkouri@uae.ac.ma

ABSTRACT

Sentiment analysis also called opinion mining, is the field of study that analyses people's opinion, sentiment, evaluations, appraisals, attitudes and emotions towards entities such us services, organizations, individuals, issues, events, topics and products. The fast evolution of Internet-based applications like websites, social networks, and blogs, leads people to generate enormous heaps of opinions and reviews about products, services, and day-to-day activities. Sentiment analysis poses as a powerful tool for businesses, governments, and researchers to extract and analyze public mood and views, gain business insight, and make better decisions.

There are many approaches to classify the sentiment, approaches based on machine learning or lexicon-based approach. In this article we will discuss the different approaches of sentiment analysis, and we will compare the performance of the different machine learning algorithms. In this comparative study we will use the naïve Bayes, Support vector Machine, the Decision Tree and the Logistic regression algorithms to analyze the sentiments in amazon's reviews data. The main objective is to analyze the large number of reviews expressed in amazons in order to deduce the different feelings expressed in it, positive, negative or neutral.

The main goal of this work is to achieve the best result of sentiment analysis. So, to analyze and classify the data we will start by preprocessing the data then the features extraction after that the sentiment classification using the machine learning algorithms and finally the evaluation of the algorithms, using Spark and Scala language for implementing the algorithms. The final results show that the SVM classifier achieved 100% accuracy, Naive Bayes classifier achieved 95% accuracy, the Logistic regression 97% and the Decision Tree classifier achieved 75% accuracy.

Keywords: *Sentiment analysis, Opinion mining, Machine learning, Big data, Lexicon-based approach, Spark, Amazon, Naïve Bayes, SVM, Decision Tree, Logistic regression.*

1. INTRODUCTION

Every day, millions of people share their reviews, opinions and evaluations about movies and products on various social media site like Facebook and Twitter, e-commerce sites like Amazon and movie reviewing sites. These reviews and opinions may hold some of uses expectation which is important to business and marketing professionals and researchers. The objective of sentiment analysis is to analyze a large amount of data in order to deduce the different feelings expressed in it, positive, negative or neutral.

Since customers express their thoughts and feelings more openly than ever before, sentiment analysis is becoming an essential tool to monitor and understand that sentiment. Automatically analyzing customer feedback, such as opinions in survey responses and social media conversations, allows brands to learn what makes customers happy or frustrated, so that they can tailor products and services to meet their customers' needs.

There are different types of approaches that can be used to perform sentiment analysis. A classical approach is a lexicon-based approach and the other

option is a machine learning-based sentiment classification.

The lexicon-based approach is subdivided into two approach dictionaries based and corpus-based approach.

Dictionary based approach find the keyword and then search for synonyms and antonyms for that keyword. The lexicon-based approach begins with tokenization of the document and extracting the sentence which holds some sentiment value by neglecting the stop word from the document.
The Machine Learning based approach can be classified into supervised and unsupervised learning approach.
Supervised learning approach use a labeled training documents for training the model. In unsupervised learning approach it is difficult to identify label in training documents to train the model.
In this article we will do a comparative study of tree machine learning algorithms, the Naïve Bayes, The Decision Tree, The Support Vector Machine and the Logistic Regression, using Amazon's reviews data to classify sentiments. The goal of this work is to identify the most performed machine learning algorithm, using the accuracy, the precision, the recall, the F-measure and the test error to evaluate the algorithms, using Spark Mlib to implement algorithms and Scala programming language.
This article is structured as follow; in the section 2 we will start with research works related with sentiment analysis. In the section 3 we will mention the different approaches of sentiment analysis. In the section 4 the general architecture of the study. In the section 5 the description of the used dataset. In the section 6 text preprocessing and feature extraction. In the section 7 the general model. In the section 8 the algorithms. In the section 9 the classification evaluation. In section 10 the results and experiments and finally the conclusion.

2. RELATED WORKS

In this section we propose many studies that proposed the different approaches to analyze sentiments.

Wang et al [1] Propose an approach that uses sentence-to-sentence attention to realize the sentence-level attention mechanism. This approach disregards sentence positional information and reduces the complexity involved in building sentence sequences.

Xiang et al [2] In this work, the authors present a novel approach for sentiment analysis by fusing external affective knowledge into neural networks. The affective knowledge is distilled from two sentiment lexicons grounded by two psychological theories, e.g., the Affect Control Theory and word affections in terms of Valence.

Baid et al [3] In this research various techniques were used to analyze sentiments. The authors analyzed the Movie reviews using various techniques like Naïve Bayes, K-Nearest Neighbour and Random Forest. The best results were given by Naïve Bayes classifier. The Naïve Bayes classifier achieved 81.45% accuracy, Random Forest classifier we achieved 78.65% accuracy, K-Nearest Neighbour classifier achieved 55.30% accuracy.

Başarslan et al [4]: In this study, the authors use the Naïve Bayes (NB), Support Vector Machines (SVM) and Artificial Neural Networks (ANN) algorithms, and for categorizing the sentiments as positive, negative and neutral two datasets were used in the experiments. The first one is the user reviews about movies from the IMDB, and the second one is the Twitter tweets, including the tweets of users about health topic in English in 2019, collected using the Twitter API. The feature extraction from the dataset was performed using Term Frequency-Inverse Document Frequency (TF-IDF) and Word2Vec (W2V) modeling techniques. According to the experimental results, Artificial Neural Network had the best accuracy of 0.90 on IMDB dataset and 0.87 on twitter dataset. The naive bayes accuracy is 0.83 on IMDB dataset and 0.72 on twitter dataset. The SVM accuracy is 0.84 on IMDB dataset and 0.84 on Twitter dataset.

Moin et al [5] The contribution of this paper includes the adoption of a hybrid approach that involves a sentiment analyzer that includes machine learning. Moreover, this paper also provides a comparison of techniques of sentiment analysis in the analysis of political views by applying supervised machine-learning algorithms such as Naïve Bayes and support vector machines (SVM). The best accuracy given by the naïve bayes in this work is 79% and the best accuracy given by the SVM in this is work is 62.76%.

Mitra [6] Throughout this research work a rule-based approaches was proposed which defines a set of rules and inputs like Classic Natural Language Processing techniques, stemming, tokenization, a region of speech tagging and parsing of machine learning for sentiment analysis which is going to be implemented by most advanced python language. The results show that the naïve bayes accuracy is 0.7044, the SKlearnBernouliieNB accuracy is 0.701,

Decision Tree accuracy is 0.52, the Random Forest accuracy is 0.80 and the KNN accuracy is 0.71.

Jagdale et al [7]: In this paper, Dataset has taken from Amazon which contains reviews of Camera, Laptops, Mobile phones, tablets, TVs, video surveillance. After preprocessing they applied machine learning algorithms to classify reviews that are positive or negative. This paper concludes that, Machine Learning Techniques gives best results to classify the Products Reviews. Naïve Bayes got accuracy 98.17% and Support Vector machine got accuracy 93.54% for Camera Reviews.

AZWA et al [8]: They provide a method known as Contextual Analysis (CA), a mechanism that constructs a relationship between words and sources that is constructed in a tree structure identified as Hierarchical Knowledge Tree (HKT). Then, Tree Similarity Index (TSI) and Tree Differences Index (TDI), a formula generate from tree structure are proposed to find similarity as well as changes between train and actual dataset. The regression analysis of datasets reveals that there is a highly significant positive relationship between TSI and SML accuracies. As a result, the prediction model created indicated estimation error within 2.75 to 3.94 and 2.30 for 3.51 for average absolute differences.

Rathi et al [9] ensemble machine learning techniques was used for increasing the efficiency and reliability of proposed approach. For the same, they merged Support Vector Machine with Decision Tree and experimental results prove that the proposed approach is providing better classification results in terms of f-measure and accuracy in contrast to individual classifiers.

Valencia et al [10] In this paper, they propose the usage of common machine learning tools and available social media data for predicting the price movement of the Bitcoin, Ethereum, Ripple and Litecoin cryptocurrency market movements. In this article the authors compare the utilization of neural networks (NN), support vector machines (SVM) and random forest (RF) while using elements from Twitter and market data as input features. The results show that it is possible to predict cryptocurrency markets using machine learning and sentiment analysis and the SVM model performed the best, using both Twitter and market data it obtained 0.66 accuracy and 0.8 precision scores.

Rahman et al [11]: Here, we have collected movie review data as well as used five kinds of machine learning classifiers to analyze these data. Hence, the considered classifiers are Bernoulli Naïve Bayes (BNB), Decision Tree (DE), Support Vector Machine (SVM), Maximum Entropy (ME), as well as Multinomial Naïve Bayes (MNB). Our analysis outlines that MNB achieves better accuracy, precision and F-score while SVM shows higher recall compared to others. The MNB accuracy is 88.50%,the Bernoulli NB accuracy 87.50%,the SVM with 87.33%,The maximum entropy with 60.67% and Decision Tree with 80.17%.

Birjali et al [12] Propose, for a better analysis, an algorithm of computing semantic analysis between tweets in training set and tweets in data set based on WordNet. Experimental results demonstrate that the method based on machine learning algorithms and semantic sentiment analysis can extract predictions of suicidal ideation using Twitter Data. In addition, this work verifies the effectiveness of performance in term of accuracy and precision on semantic sentiment analysis that could thinking of suicide. The naïve bayes precision in this article is 87.50%.

MOUDHICH et al [16] they used two machine learning approaches: First known as Long Short-Term Memory (LSTM). The second is Bidirectional Encoder Representations from Transformers (BERT). In this work, they analyzed tweets' sentiment to predict the USA elections
results. First, they collected data based on ontologies that they defined. Second, text pre-processing to clean data. Third, predicting subjectivity and sarcasm in a tweet. Fourth applying the two cited approaches to get the sentiment.

RAMANA et al [17] In this research study, a Modified Bi-directional Long Short-Term Memory (MBLSTM) is developed for predicting the sentiments. Initially, raw data is given as input to pre-processing techniques and extracted the important features using two feature extraction techniques. Text blogs are used to identify the polarity of tweets and learning rate of MBLSTM is improved by Self-Attention mechanism. The results proved that the proposed MBLSTM achieved 93.66% of accuracy and 93% of precision,

ALSUBAIE et al [18] The chosen method for this study is to use ensemble Machine Learning approach using Naïve Bayesian combined with Support Vector Machine, followed by semantic analysis to improve its accuracy. The main goal of the proposed model is to be able to determine the polarity of any given text "tweet" to generate a comprehensive statistical report regarding the public's opinion in a certain matter. In this article The highest precision was achieved by Multinomial NB classifier which is

78.14%,the CE precision was 75.53% and the SVM precision was 60.17%.

3. SENTIMENT ANALYSIS APPROACHES:

Sentiment analysis is a type of natural language processing for tracking the mood of the public about a particular product or topic. It involves in building a system to collect and examine opinions about the product made in blog posts, comments or reviews.

The different approaches of sentiment analysis are:

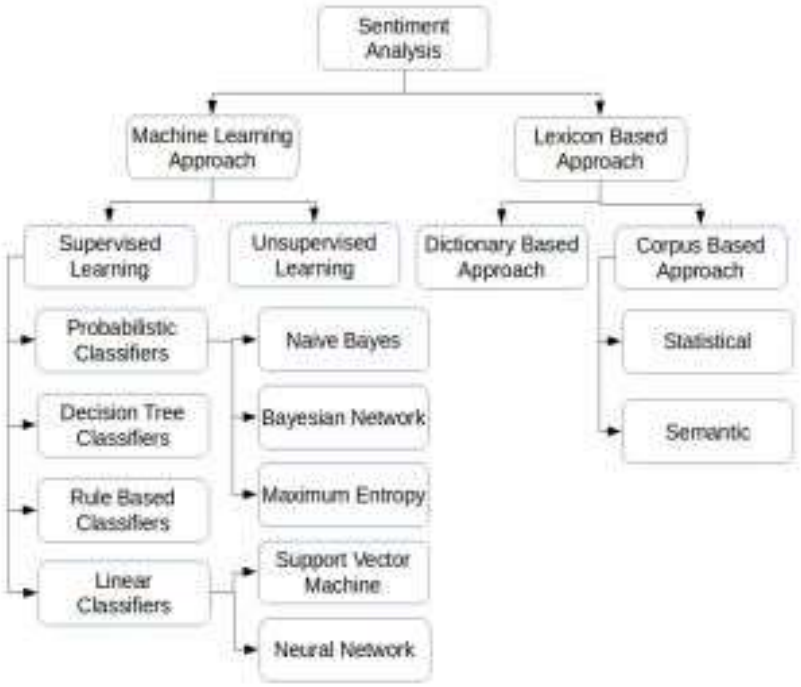

Figure 1: The different approaches of sentiment analysis

3.1 There are two main categories of analysis: lexical analysis and machine learning analysis

Sentiment Classification techniques are classified in to lexicon based, machine language approach, lexicon based and hybrid approach. The Machine learning (ML) approach uses some common ML algorithms and linguistic features. Lexicon based approach makes use sentiment lexicons. This technique is sub-classified corpus based and dictionary-based method that uses statistical approach to determine the sentiment of the sentence. Hybrid approach is a combination of both techniques the lexicon based and ML based approach.

3.1.1- The Machine Learning based approach

The Machine Learning based approach can be classified into supervised and unsupervised learning approach. Supervised learning approach use a labeled training documents for training the model. In unsupervised learning approach it is difficult to identify label in training documents to train the model.

a) Unsupervised learning

Unsupervised learning is a branch of machine learning that learns from test data that has not been labeled, classified or categorized. Instead of responding to feedback, unsupervised learning identifies commonalities in the data and reacts based on the presence or absence of such commonalities in each new piece of data.

b) The supervised learning

algorithm depends on the labeled training data. The training data contain a collection training example. Each example in supervised learning to consist of an input object and output value.

4. THE GENERAL ARCHITECTURE OF THE STUDY

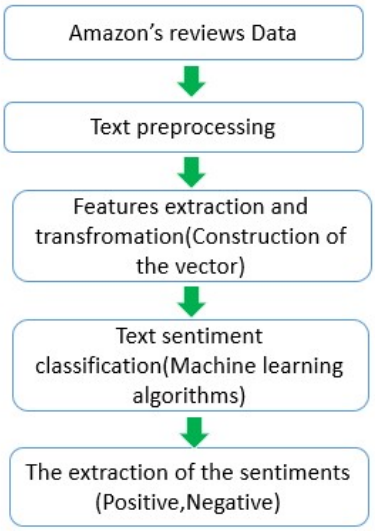

Figure 2: The general architecture of the study

5. DATASET

This dataset consists of a few million Amazon customer reviews (input text) and star ratings (output labels) for learning how to train fastText for sentiment analysis

In this case, the classes are __label__1 and __label__2, and there is only one class per row.

__label__1 corresponds to 1- and 2-star reviews, and __label__2 corresponds to 4- and 5-star reviews. (3-star reviews i.e. reviews with neutral sentiment were not included in the original),

The review titles, followed by ':' and a space, are prepended to the text.

Most of the reviews are in English, but there are a few in other languages, like Spanish

Table 1: Description of Amazon reviews dataset

Dataset Attribute	Description of the attribute
Label	Sentiment Class Label1 for negative Label2 for Positive
Text	Reviews from Amazon Dataset

6. TEXT PREPROCESSING AND FEATURE EXTRACTION

6.1 hashingTF

HashingTF is a Transformer which takes sets of terms and converts those sets into fixed-length feature vectors. In text processing, a "set of terms" might be a bag of words. HashingTF utilizes the hashing trick. A raw feature is mapped into an index (term) by applying a hash function. The hash function used here is MurmurHash . Then term frequencies are calculated based on the mapped indices. This approach avoids the need to compute a global term-to-index map, which can be expensive for a large corpus, but it suffers from potential hash collisions, where different raw features may become the same term after hashing. To reduce the chance of collision, we can increase the target feature dimension. the number of buckets of the hash table. Since a simple modulo on the hashed value is used to determine the vector index, it is advisable to use a power of two as the feature dimension, otherwise the features will not be mapped evenly to the vector indices. The default feature dimension is 218=262,144218=262,144. An optional binary toggle parameter controls term frequency count. When set to true all nonzero frequency counts are set to 1. This is especially useful for discrete probabilistic models that model binary, rather than integer, counts.

6.2 Tokenizer

Tokenization is the process of taking text (such as a sentence) and breaking it into individual terms (usually words). A simple Tokenizer class provides this functionality. The example below shows how to split sentences into sequences of words.

6.3 StopWordsRemover

Stop words are words which should be excluded from the input, typically because the words appear frequently and don't carry as much meaning.

StopWordsRemover takes as input a sequence of strings (e.g., the output of a Tokenizer) and drops all the stop words from the input sequences. The list of stopwords is specified by the stopWords parameter. Default stop words for some languages are accessible by calling StopWordsRemover.loadDefaultStopWords (language), for which available options are "danish", "dutch", "english", "finnish", "french", "german",

"hungarian", "italian", "norwegian", "portuguese", "russian", "spanish", "swedish" and "turkish". A boolean parameter caseSensitive indicates if the matches should be case sensitive (false by default).

7. GENERAL MODEL

Figure 3: Sentiment analysis model

- ✓ Loading data (data ingestion)
- ✓ Extracting features (feature extraction)
- ✓ Training model (model training)
- ✓ Evaluate (or predictionize)

8. ALGORITHMS

8.1 The Naïve Bayes Algorithm

Naïve Bayes It is a classification technique based on Bayes Theorem with an assumption of independence among predictors. In simple terms, a Naive Bayes classifier assumes that the presence of a particular feature in a class is unrelated to the presence of any other feature. Naïve Bayes mainly targets the text classification industry. It is mainly used for clustering and classification purpose depends on the conditional probability of happening [13].

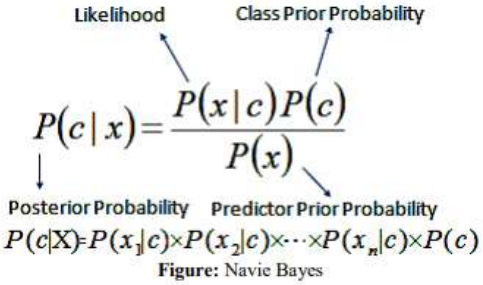

Figure 4: Naïve Bayes Model

8.2 Decision Tree Classifier

Decision tree classifier a method which is commonly used in data mining. The main idea behind the decision tree is to predict the outcome based on input variables. Decision tree classifiers [17] train itself by decomposing the data with help of condition on the attribute values. The decomposition of the data will be continued until the leaf nodes of the tree contain certain minimum number of records which can be later used for classification.

Decision tree (DT) is one of the earliest and prominent machine learning algorithms. A decision tree models the decision logics i.e., tests and corresponds outcomes for classifying data items into a tree-like structure. The nodes of a DT tree normally have multiple levels where the first or top-most node is called the root node. All internal nodes (i.e., nodes having at least one child) represent tests on input variables or attributes. Depending on the test outcome, the classification algorithm branches towards the appropriate child node where the process of test and branching repeats until it reaches the leaf node [15].

8.3 The Naïve Bayes Algorithm

The SVM can be defined as a vector space-based data mining method that finds a decision bound-ary between the two classes farthest from a random point on the training data (Song et al., 2002). An interesting feature of the SVM is its structural risk minimization in statistical learning theory. One of the main assumptions of the SVM. Support vector machine (SVM) algorithm can classify both linear and non-linear data. It first maps each data item into an n-dimensional feature space where n is the number of features. It then identifies the hyperplane that separates the data items into two classes while maximizing the marginal distance for both classes and minimizing the classification errors [14].

8.4 Logistic Regression

Logistic regression, despite its name, is a classification model rather than regression model. Logistic regression is a simple and more efficient method for binary and linear classification problems.

Logistic regression is widely used to predict a binary response. It is a linear method as described above in equation (1)(1), with the loss function in the formulation given by the logistic loss:

$$L(\mathbf{w}; \mathbf{x}, y) := \log(1 + \exp(-y\mathbf{w}^T\mathbf{x})).$$

For binary classification problems, the algorithm outputs a binary logistic regression model. Given a new data point, denoted by xx, the model makes predictions by applying the logistic function

$$\mathrm{f}(z) = \frac{1}{1 + e^{-z}}$$

9. CLASSIFICATION EVALUATION

- ✓ True Positive (TP) - label is positive and prediction is also positive
- ✓ True Negative (TN) - label is negative and prediction is also negative
- ✓ False Positive (FP) - label is negative but prediction is positive
- ✓ False Negative (FN) - label is positive but prediction is negative

9.1 Accuracy
is also used as a statistical measure of how well a binary classification test correctly identifies or

$$\text{Accuracy} = \frac{tp + tn}{tp + tn + fp + fn}$$

excludes a condition. That is, the accuracy is the proportion of correct predictions (both true positives and true negatives) among the total number of cases examined.

Precision is defined as the fraction of relevant examples (true positives) among all of the examples which were predicted to belong in a certain class.

$$\text{Precision} = \frac{tp}{tp + fp}$$

Recall is defined as the fraction of examples which were predicted to belong to a class with respect to all of the examples that truly belong in the class.

$$\text{Recall} = \frac{tp}{tp + fn}$$

F-measure
F-Measure provides a way to combine both precision and recall into a single measure that captures both properties.

$$F_1 = 2 \cdot \frac{\mathrm{precision} \cdot \mathrm{recall}}{\mathrm{precision} + \mathrm{recall}} = \frac{\mathrm{TP}}{\mathrm{TP} + \frac{1}{2}(\mathrm{FP} + \mathrm{FN})}$$

10. RESULTS AND EXPERIMENTS

The following figures demonstrate the results of analyzing 100 data of amazon's reviews using the SVM, the Naïve Bayes and Decision Tree for the classification of the sentiments. After pre-processing the data. We used the Recall, the precision, the F-measure, the test error and the accuracy to evaluate each algorithm. The figures show the results each algorithm.
The SVM gives the best results in these experiments.
The following table gives the accuracy and the precision given by the SVM, the Naïve Bayes, the Decision Tree and the Logistic regression
The experiments show that the best results are given by the SVM and the Naïve Bayes.

Table 2: Comparison Of The Accuracy And The Precision Of Each Algorithm Using 100 Data

Algorithm	Recall	F-measure	Test error
SVM	0.99	0.99	0.0
Naïve Bayes	0.99	0.98	0.01
Decision Tree	0.75	0.72	0.25
Logistic regression	0.97	0.95	0.03

Table 3: Comparison Of The Recall, The F-Measure And The Test Error Using 100 Data.

Algorithm	Accuracy	precision
SVM	0.98	0.97
Naïve Bayes	0.98	0.97
Decision Tree	0.73	0.73
Logistic regression	0.98	0.97

The following table gives the recall, the F-measure and Test error of each algorithm. The best results are also given by the SVM and the Naïve Bayes.

Table 4: Comparison of the Accuracy and the precision of each algorithm using 200 data

Algorithm	Recall	F-measure	Test error
SVM	0.98	0.97	0.015
Naïve Bayes	0.98	0.98	0.015
Decision Tree	0.73	0.73	0.27
Logistic regression	0.98	0.98	0.015

Using 200 data

The following figures demonstrate the results of analyzing 200 data of amazon's reviews using SVM,Naïve Bayes and Decision Tree for the classification of the sentiments. After pre-processing the data. We used the Recall, the precision, the F-measure, the test error and the accuracy to evaluate each algorithm. The figures show the results each algorithm.

The following tables demonstrate that the best results are given by the the SVM, The decision Tree and the Logistic Regression.

Table 5: Comparison Of The Recall, The F-Measure And The Test Error Using 200 Data.

Algorithm	Accuracy	precision
SVM	1.0	0.99
Naïve Bayes	0.99	0.98
Decision Tree	0.75	0.79
Logistic regression	0.97	0.94

Table 6: Comparison Of The Accuracy And The Precision Of Each Algorithm Using 300 Data.

Algorithm	Accuracy	precision
SVM	0.98	0.97
Naïve Bayes	0.98	0.97
Decision Tree	0.73	0.73
Logistic regression	0.98	0.97

Using 300 Data

The following figures demonstrate the results of analyzing 300 data of amazon's reviews using the SVM, the Naïve Bayes and Decision Tree for the classification of the sentiments. After pre-processing the data. We used the Recall, the precision, the F-measure, the test error and the accuracy to evaluate each algorithm. The figures show the results each algorithm.

The SVM gives the best results in these experiments.

Table 7: Comparison of the Recall, the F-measure and the Test error using 300 data

Algorithm	Recall	F-measure	Test error
SVM	0.98	0.97	0.015
Naïve Bayes	0.98	0.98	0.015
Decision Tree	0.73	0.73	0.27
Logistic regression	0.98	0.98	0.015

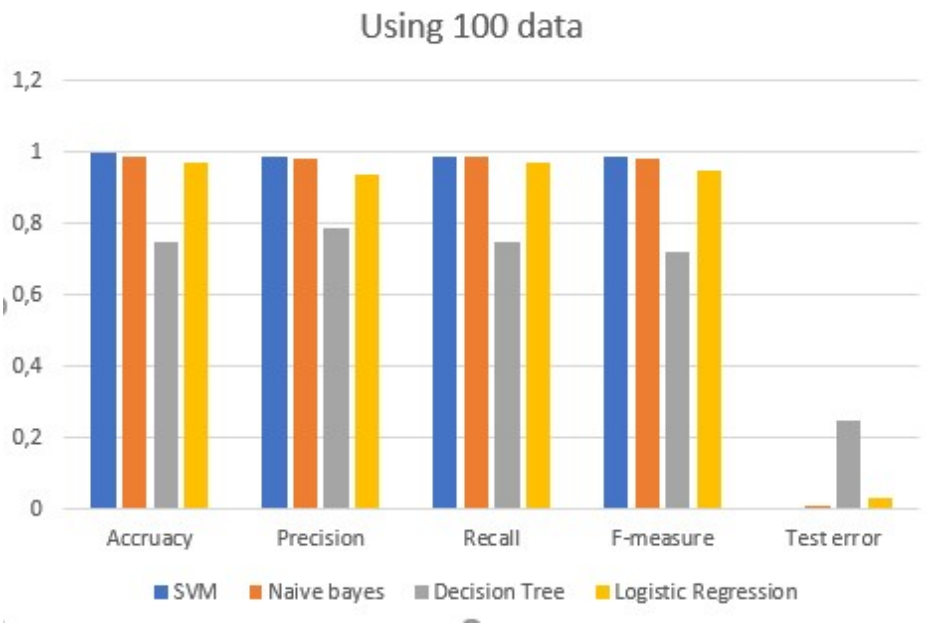

Figure 5: Graphical Representation Of Comparison Between The SVM, Naïve Bayes, Decision Tree And Logistic Regression Using 100 Data

Figure 6: Graphical Representation Of Comparison Between The SVM, Naïve Bayes, Decision Tree And Logistic Regression Using 200 Data

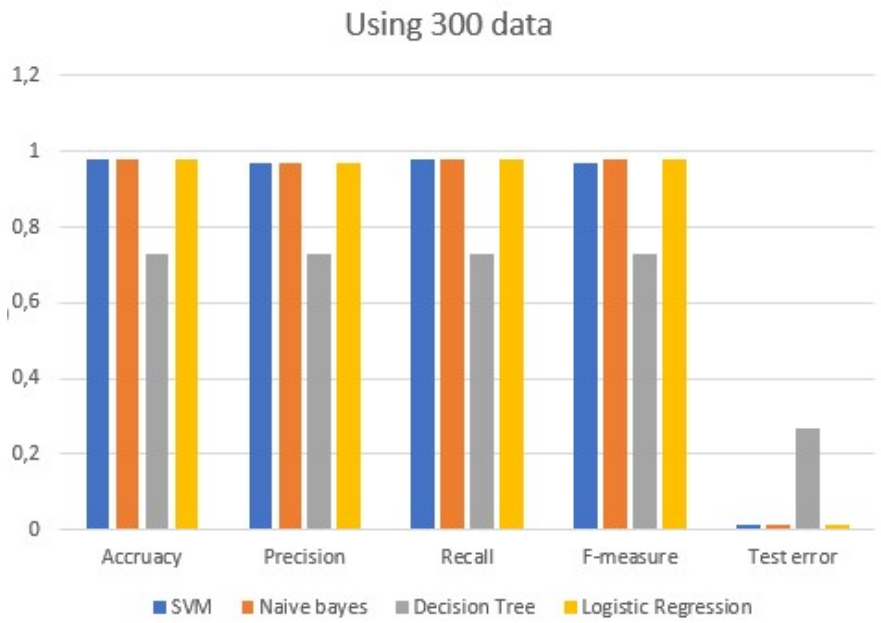

Figure 7: Graphical Representation Of Comparison Between The SVM, Naïve Bayes, Decision Tree And Logistic Regression Using 300 Data

The experiments results show that the naïve bayes accuracy using 100 data is 99%, the SVM accuracy is 100%, Logistic regression 97% and the Decision Tree 75%. Using 200 Data. The naïve bayes accuracy is 98%, the SVM accuracy is 98%, Logistic regression 98% and the Decision Tree 73%. Using 300 Data The naïve bayes ACCURACY IS 98%, the svm accuracy is 98%, Logistic regression 98% and the Decision Tree 73%.

Comparing our results (accuracy and the precision) with the results of the works cited in the section related works and that use the same machine learning algorithms, demonstrate that the best accuracy and precision is given by our results.

11. CONCLUSION

The aim of this work is to study the performance of the most performed sentiments classification algorithms and to compare the result of each algorithm. various algorithms were used to analyze the sentiments amazon's reviews. We used the Naïve Bayes, Support Vector Machine, the decision Tree and the Logistic regression. The best results were given by Support Vector Machine classifier. The SVM classifier achieved 100% accuracy, Naive Bayes classifier achieved 95% accuracy, the Logistic regression 97% and the Decision Tree classifier achieved 75% accuracy.

As a future work we plan to further improve the results of sentiments analysis by adding the treatment of the sarcasm and analyze the meanings of individual emojis or show the accuracy of sentiment analysis when emojis are included.

ACKNOWLEDGMENTS

The work is done through the support of the CNRST in part of bourse of excellent scholarship of research.

REFRENCES:

[1] Ping Wang ,Jiangnan Li , Jingrui Hou, S2SAN: A sentence-to-sentence attention network for sentiment analysis of online reviews,Decision Support System and Electronic Commerce

[2] Rong Xiang , Jing Li , Mingyu Wanc, Jinghang Gu , Qin Lu , Wenjie Li ,Chu-Ren Huang, Affective awareness in neural sentiment analysis, Knowledge-Based Systems 226 (2021) 107137.

[3] Palak Baid, Apoorva Gupta, Neelam Chaplot, Sentiment Analysis of Movie Reviews using Machine Learning Techniques, International Journal of Computer Applications (0975 – 8887) Volume 179 – No.7, December 2017

[4] Muhammet Sinan Başarslan, Fatih Kayaalp, Sentiment Analysis with Machine Learning Methods on social media, ADCAIJ: Advances in Distributed Computing and Artificial Intelligence Journal Regular Issue, Vol. 9 N. 3 (2020), 5-15 eISSN: 2255-2863

[5] Ali Hasan , Sana Moin , Ahmad Karim , Shahaboddin Shamshirband, Machine Learning-Based Sentiment Analysis for Twitter Accounts,Mathematical and Computational Applications ,Math, Comput. Appl. 2018, 23, 11

[6] Ayushi Mitra,Sentiment Analysis Using Machine Learning Approaches (Lexicon based on movie review dataset), Journal of Ubiquitous Computing and Communication Technologies (UCCT) (2020) Vol.02/ No.03 Pages: 145-152

[7]Rajkumar S. Jagdale,Vishal S. Shirsat,Sachin N. Deshmukh, Sentiment Analysis on Product Reviews Using Machine Learning Techniques, Cognitive Informatics and Soft Computing pp 639-647

[8] AZWA ABDUL AZIZ, ANDREW STARKEY, Predicting Supervise Machine Learning Performances for Sentiment Analysis Using Contextual-Based Approaches, IEEE ACCESS, Digital Object Identifier 10.1109/ACCESS.2019.2958702

[9]Megha Rathi; Aditya Malik; Daksh Varshney; Rachita Sharma; Sarthak MendirattaSentiment Analysis of Tweets Using Machine Learning Approach, 2018 Eleventh International Conference on Contemporary Computing (IC3),IEEE

[10]:Franco Valencia, Alfonso Gómez-Espinosa,Benjamín Valdés, Price Movement Prediction of Cryptocurrencies Using Sentiment Analysis and Machine Learning, Entropy 2019, 21, 589,MDPI

[11] Atiqur Rahman; Md. Sharif Hossen,Sentiment Analysis on Movie Review Data Using Machine Learning Approach, 2019 International Conference on Bangla Speech and Language Processing (ICBSLP), IEEE

[12] MarouaneBirjalia,, AbderrahimBeni-Hssane, MohammedErritalib,Machine Learning and Semantic Sentiment Analysis based Algorithms for Suicide Sentiment Prediction in Social Networks, The 8th International Conference on Emerging Ubiquitous Systems and Pervasive Networks(EUSPN 2017), Procedia Computer Science 113 (2017) 65–72

[13] International Journal of Science and Research (IJSR) ISSN: 2319-7064 ResearchGate Impact Factor (2018): 0.28 | SJIF (2018): 7.426 Volume 9 Issue 1, January 2020 www.ijsr.net Licensed Under Creative Commons Attribution CC BY Machine Learning Algorithms - A Review Batta Mahesh

[14] Joachims T. Making large-scale SVM learning practical. SFB 475: Komplexitätsreduktion Multivariaten Datenstrukturen, Univ. Dortmund, Dortmund, Tech. Rep. 1998. p. 28.

[15]. Quinlan JR. Induction of decision trees. Mach Learn. 1986;1(1):81–106.

[16] Ihab Moudhich, Abdelhadi Fennani,A Comparison of ML approaches on sentiment analysis, based on ontologies, sarcasm and subjectivity detections in the case study of us elections.Journal of Theoretical and Applied Information Technology A COMPARISON OF ML 15th March 2021. Vol.99. No 5, ISSN: 1992-8645,

[17] D. Ramana Kumar, S. Krishna Mohan Rao, A self-attention layer mechanism based modified bi-directionel long short term memory for Twitter sentiment classification, Journal of Theoretical and Applied Information Technology, 30th April 2021. Vol.99. No 8, ISSN: 1992-8645

[18] Sarah M. Alsubaie , Kholoud M. Almutairi , Najla A. Alnuaim , Automatic semantic sentiment analysis on Twitter Tweets using machine learning: a comparative study, Journal of Theoretical and Applied Information Technology, 15th December 2019. Vol.97. No 23, ISSN: 1992-8645