

Sentiment Analysis Techniques and Application-Survey and Taxonomy

Mahmood Umar¹, Abdul-Azeez Abdullahi Bena², Buhari Wadata³

^{1,3}Department of Computer Science, Faculty of Science, Sokoto State University, Sokoto- Nigeria.

²Department of Computer Science, Waziri Umaru Federal Polytechnic, Birnin Kebbi, Kebbi State-Nigeria.

Corresponding Author: Mahmood Umar

ABSTRACT: Nowadays, social media platforms, blogs, and e-commerce are commonly use to express opinion on politics, movies, products, education respectively; for election forecasting, business boosting and improvement of teaching and learning. As a result, data generation becomes easier; producing big data which requires appropriate techniques and tools to analyse easily, accurately and timely. Thus, making sentiment analysis very demanding research area. This study will investigate on what basis (sentiment classification level) or area of application (data source) do supervised machine learning approaches particularly Support Vector Machine (SVM), Naïve Bayes, and Maximum Entropy algorithms, and other technique-lexicon-based approach give the best result in sentiment analysis. Based on the review of the literature there is a contradiction on the point that SVM generated the best result in analyzing student sentiment on document level. This study also discovers that sentiment analysis differs from system to system based on polarity (types of the classes to predict: positive or negative, subjective or objective), different levels of classification (sentence, phrase, or document level) and language that is processed. This research produces a taxonomy which serves as a guide for the choice of techniques in sentiment analysis. The taxonomy explores the sentiment classification levels and data preprocessing stages. It also explores that sentiment analysis techniques were organised in to three (3) groups; Machine learning, Lexicon and hybrid or combination. The machine learning techniques were sub-grouped in to two (2) namely; supervised and unsupervised. The supervised were organized in to two (2): Classification and Regression. un-supervised machine learning techniques includes clustering and association. The clustering technique consist of k-means. Decision tree which is a classification based under supervised type of machine learning technique consist of random forest,(Akinkunmi, 2019) while the ruled-based classifiers consist of confidence criterion and support criterion. The commonly used tools are Weka, Python compiler, and R programming tool.

KEYWORDS: Sentiment analysis, Taxonomy, Tools and Techniques, Support Vector Machine, Naïve Bayes, sentiment classification level, polarity, lexicon-based approach, framework.

1. INTRODUCTION

In a study by Bose et, al (2018) sentiment analysis can be expressed as a tool used to analyze opinions or moods on a specific matter-services, products or subject written in text or extracted from social media platforms, blog post, comments or web reviews and so on.

Sentiments analysis deals classification of opinions or based on its polarity-label or bearing of the opinion according to the three(3) classes: negative, positive or neutral (Alkubaisi et al. 2018). Neutral sentiment is the point at which a learner does not express a negative or positive feeling. The author further reported that, diverse techniques particularly, machine learning algorithm have been utilized in sentiment analysis and a couple has demonstrated to give the efficient and accurate result, these are Support Vector Machines (SVM), Naïve Bayes and Max Entropy (MaxEnt). Machine learning approaches performs better than other approaches (Ozturk et al 2017) on Large labelled data in English Language. In another study, lexicons perform better than machine learning

approaches on large data in Urdu Language based on Accuracy, Precision, Recall, F-measure and even economy of time and efforts used (Neelam et al 2018).

In a study by Munir et al, (2017), sentiment analysis techniques are grouped into three (3) namely: Lexicon based techniques, Machine Learning based techniques and hybrid techniques (combination of lexicon and machine learning algorithms). The author had identified Machine Learning approach as the most effective and reliable in the field of sentiment analysis and opinion mining for classification purposes. The survey will answer the following question:

Can taxonomy of all the sentiment analysis techniques enhances future survey in the same area? How can the survey promote choice of algorithm and analytical and data collection tool suitable in sentiment analysis?

Contribution of this research will produce a taxonomy of sentiment analysis techniques as well as guide for the choice of techniques and analytical and data collection tools in

sentiment analysis. The study is a literature for future survey in the related field.

There are several surveys on sentiment analysis techniques but mostly focused on one method; machine learning or lexicon-based, hardly to find all the techniques put together. And in most cases without identifying the tools used in the sentiment analysis. This study aimed at conducting a review on techniques or approaches, tools used in sentiment analysis for enhancing future surveys. A taxonomy will be developed also.

Serrano-Guerrero et al., (2015) reports in the paper title “Sentiment Analysis: A review and comparative analysis web services” that sentiment analysis techniques are grouped into two; Machine learning approach and lexicon-based approach. The former requires adequate labelled data sets for the training and comprises of decision tree classifiers, rule-based classifiers, and probabilistic classifiers (subdivided into Naïve Bayes, Bayesian Networks and Maximum Entropy). linear-based classifiers (subdivided into support vector machine and neural networks). The latter approach uses a bag of words for the token matching and it consists of dictionary-based approach and corpus-based approach. The corpus-based approach is subdivided in to semantic and statistical. In

my own opinion, the author investigated the algorithm using large data from twitter and movies.

The author (Lajis et al., 2018) reported that SA is applied on e-commerce, politics, health, Media, education etc. Education is the target area because of data access from the students.

2. LITERATURE REVIEW

This part presents the related works on the sentiment analysis frameworks and architecture, techniques and tools used in data analysis.

Regarding the sentiment analysis framework and architecture, Figure 1 and 2 present some of the technique used by some authors in the sentiment analysis. Figure 1 is the overall sentiment analysis framework for machine learning algorithm developed by (Atif, 2018) to classify sentiment as either positive or negative for the improvement of teaching and learning in an Open Arab University Business Program Courses Case Study. The framework is made up of five (5) stages which include: Data (i.e., sentence: + or -), training set, testing set, learn classifier(subjectivity) and evaluate.

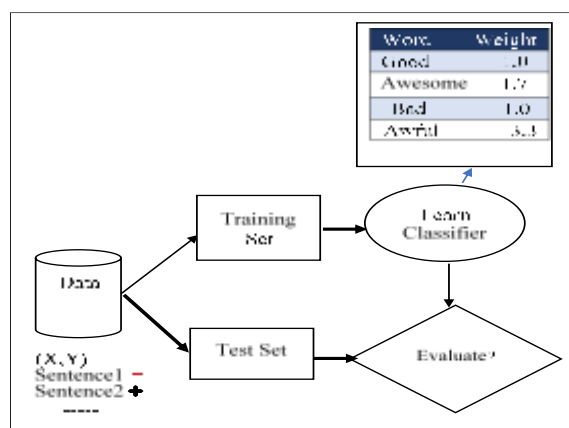


Figure 1. Overall Sentiment Analysis Framework Atif 2018

Figure 2 is the system architecture developed by Jadav and Scholar, (2016) for the hybridization of Naïve Bayes and SVM for the sentiment analysis of review data set. It is made up of four (4) phases namely: input Data, preprocessing, classification and result. This architecture used SentiWordNet package to calculate the sentiment of the review data.

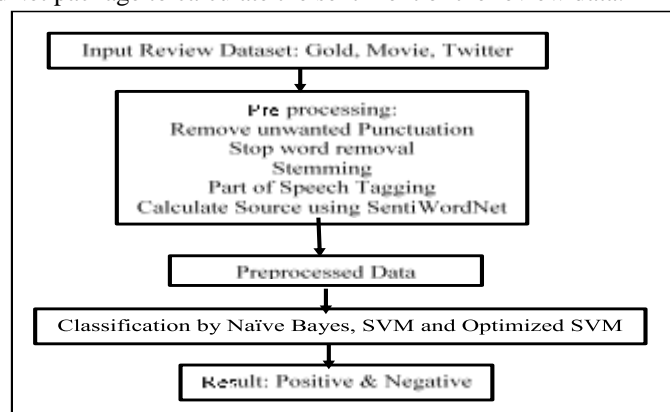


Figure 2. Proposed System Architecture (Jadav and Scholar, 2016)

In addition to the architecture in figure 2, (Archana & Kishore, 2017) designed a sentiment analysis architecture for NRC lexicon which consist of five phases viz: data collection, data pre-processing, sentiment and emotion identification, satisfaction and dissatisfaction computation and result visualization. The data collected from students via formal source (survey) and informal sources (blogs and forums) were pre-processed. The pre-processed data was passed to the next phase for sentiment and emotion identification. The sentiments are classified as either positive or negative. The emotions were classified based on eight categories (anger, anticipation, disgust, fear, joy, sadness, surprise, and trust). The proposed system computes satisfaction or dissatisfaction based on these labelled emotions. The issues with this architecture are the scope. The system processed the collected data in multilingual based on two entities. This is a motivational factor for this study to focus on one entity (Programming assessment) and one language (English).

For the sentiment analysis techniques, there are several studies that use both supervised and unsupervised machine learning techniques. (Saif, He, Fernandez, and Alani, 2016) reports that most of the supervised learning techniques depend on training classifiers, for example, Naïve Bayes (NB), Maximum Entropy (MaxEnt), and Support Vector Machines (SVMs), from different mixes of functionalities- word n-grams, Part-Of-Speech (POS) tags, with or without: words' earlier sentiment, words' semantic concept, sentiment topic features, semantic examples and tweets syntax features (e.g., hashtags, retweets, accentuations, and so forth. These strategies have accomplished generally great outcomes with correctness's revealed in the scope of 80– 84% as in (Saif et al., 2016). Rani and Kumar, (2017) report on sentiment analysis that performed opinion polarity of e-learning sites and discussions platforms utilizing a supervised half and hybrid method that joined hidden Markov models with support vector machines (SVMs). Rani and Kumar, (2017) additionally, write on research conducted on SA of students responses utilizing Bayes (NB), complement Naïve Bayes (CNB), SVM, and maximum entropy classifiers utilizing unigrams as features. They reasoned that an SVM with radial bias function kernel and the NBC method accomplished great outcomes for real-time sentiment analysis. They likewise experience better output without including the neutral class. The author (Rufai, S., & Umar, 2018) used artificial neural networks to diagnose heart disease. In a study by Bose et al., (2018) neurons are calculated using vectors which refers to the frequencies of words in a document across the lines-set of weights for each neuron used to calculate the function of the inputs.

Still on the sentiment analysis techniques, Kolchyna et al., (2015) writes- the performance of the lexicon-based approach depends on the quality and scope of the lexicon dictionary. As a result, many researchers have created their own sentiment lexicons or vocabularies of words. The commonly

used ones are Opinion lexicon, sentiWordnet, AFFINN, Loughran Mc Donald Lexicon, NRC-Hashtag and Harvard Liquierer lexicon. The procedure for the opinion-lexicon includes; Create or load opinion lexicon, Create a bag of words from data, pre-processing, sentiment score or classification.

In education, Rani and Kumar (2017) write regarding research on SA on responses from student assessment overview of Middle East College in Oman. They utilized the RapidMiner instrument to group sentiments into positive and negative based on attributes like instructor, test, model content, and materials. The author investigated the efficiency of their methodology utilizing Naïve Bayes classifier (NBC), SVM, k-nearest neighbours, and neural-network classifiers. The entity focused by the author are teacher, exam, module content, and resources Nevertheless, the data was collected in Arabic language and analysed using RapidMiner. In a study by Archana and Kishore, (2017); Rani and Kumar, (2017), NRC lexicon sentiment analysis architecture was designed for the improvement of teaching and learning . The authors collected students' feedback on teaching methodology and course satisfaction.

In movie review, B and M (2016) reports that SVM has a high efficiency on document level classification with data from movies review and twitter with respective accuracy of 82.9% and 85.5 % compared to NBC and MaxEnt. This author further reports that SVM on educational data operates and consist of two features classier and trainer. The function of the trainer is to analyze the extracted text as classified by the category. The classifier examines the document SVM decision on the result and declares the nearest one in that class. To implement SVM in R, caret and text mining packages are the suitable libraries in R compiler. Equation 3 is mathematical formula for the SVM algorithm as used by Jadav and Scholar, (2016).

$$f(x) = \sum_{i=1}^n a_i k(x, x_i) + b \tag{3}$$

$$\begin{aligned} & \text{where } k(x, x_i) \\ & = \exp\left(-\frac{\|x - x'\|^2}{2\sigma}\right) \end{aligned} \tag{4}$$

In election, (Ozturk, Cicek, & Ergul, 2017) reports that SA was used to estimate the votes in Singapore presidential election 2011. This author further reports that Naïve Bayes classifier (NBC) was used for SA in the study of 2012 U.S. presidential election cycle. The author also reports that NBC, random forest and support vector machine (SVM) were used to classify Twitter data and the result comparison of the three classification methods were presented. SA was conducted on electoral campaigns and have shown that Twitter has a remarkable ability to forecast electoral results.

3. RESEARCH METHOD

This section describes the procedure followed in order to carry out the research. This includes review of literatures to

form the topic of the research, procedure followed in order to form the taxonomy. Several surveys have described sentiment analysis tools and techniques depending on their relevance to the research in question. This study will make a survey of the available sentiment analysis tools, techniques (and their area of applications), and recent taxonomy. Based on the review of the literature, a taxonomy of sentiment analysis techniques would be design in a form of tree architecture as contribution to this study.

3.1 Taxonomy

Taxonomy is a tree structure showing the classification of sentiment analysis techniques based on their application in sentiment analysis. This section reports some of the recent taxonomy designed by the researchers in sentiment analysis and the proposed taxonomy for the sentiment analysis techniques based on the review of the literatures.

Figure 3 shows the taxonomy for the sentiment analysis techniques used by Serrano et al., (2015) in sentiment classification. The techniques cover almost all the areas of application of sentiment analysis ranging from education,

movie review, e-commerce, politics, marketing, etc. Based on Figure 3 it can be seen that the sentiment analysis techniques are classified in to two (2) namely: Machine learning approaches and lexicon-based approaches. The machine learning approach is characterized as supervised learning algorithms and they are subdivided into four (4) viz; Decision Tree classifiers, Linear classifiers, Rule-based classifiers, and probabilistic classifiers. Linear classifiers are subdivided into support vector machines and Neural networks. The probabilistic classifiers are subdivided into Naïve Bayes, Bayesian network and maximum entropy. All these algorithms are machine learning approach under supervised learning category. The lexicon-based approaches are subdivided in to two (2): dictionary-based approach and corpus-based approach-consisting of semantic and statistical. Supervised learning algorithm requires a lot of labelled data in order to give better result accuracy. The techniques under lexicon-based approach are unsupervised therefore do not need labelled data to work effectively in sentiment classification.

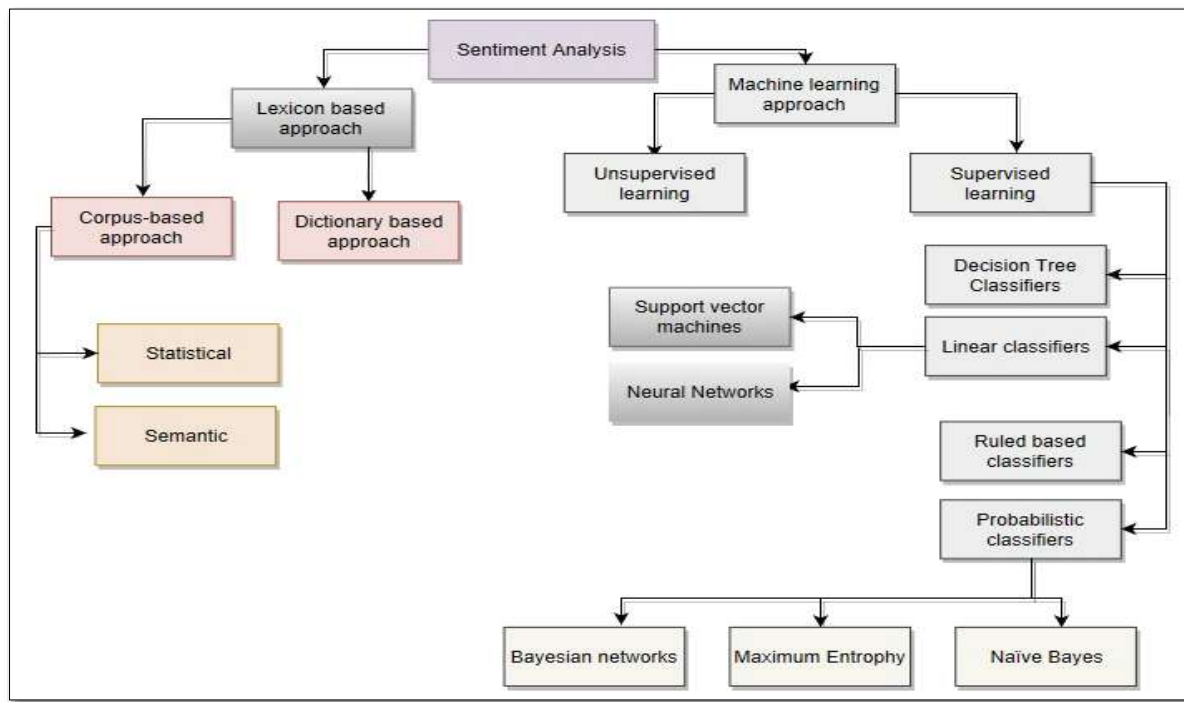
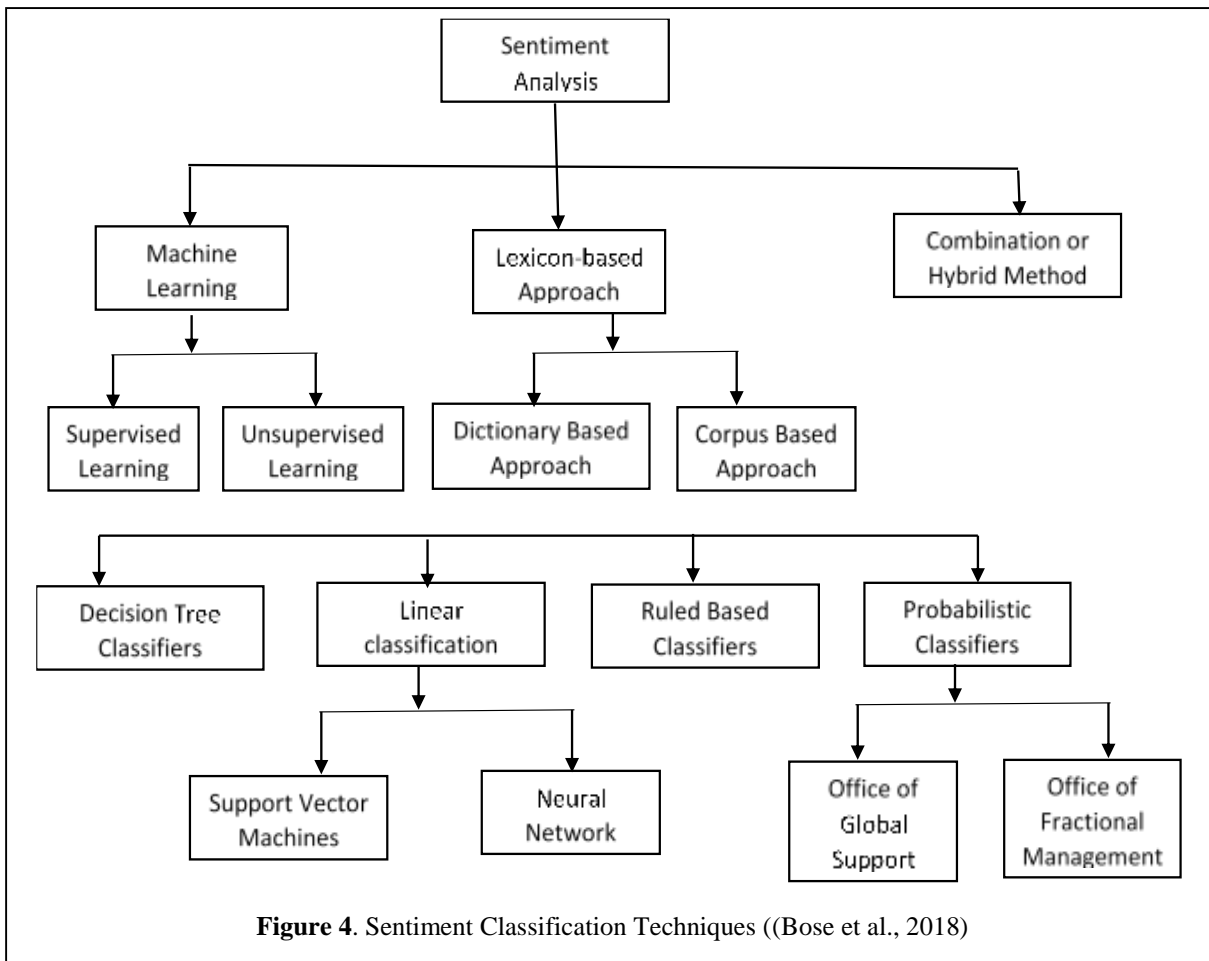


Figure 3. Sentiment Analysis Classification Techniques Taxonomy((Serrano-Guerrero et al., 2015))

Figure 4 is another taxonomy of sentiment analysis techniques developed by (Bose et al., 2018). The taxonomy categorized sentiment classification techniques in to three (3) namely: Machine learning, lexicon-based approach and hybrid approach or combination of both. The author introduced a new technique called combination or hybrid approach which is not available in the taxonomy in Figure 3. The machine learning techniques are classified as supervised and unsupervised. The supervised are subdivided in to three

(3) namely, Decision tree classifiers, Linear classifiers, Ruled-based classifiers, and probabilistic classifiers. This author also categorised probabilistic classifiers in to two (2) as office of the global support and office of the fractional management, unlike the one in Figure 3: probabilistic classifier consists of three (3) algorithm: Naïve Bayes, Bayesian network and maximum entropy. Thus, there is need for harmonization of these taxonomies in to one for easy referring and choice of algorithm among researchers.



3.2 Proposed Taxonomy for Sentiment Analysis Techniques

The proposed sentiment analysis techniques is developed as a result of harmonizing the two (2) taxonomies in Figure 3 and 4 and review of literatures. The proposed taxonomy is showed in figure 5. It can be seen that the taxonomy is a modification of the ones in figure 3 and 4 and the techniques were organised in to three groups; Machine learning, Lexicon and hybrid or combination. The machine learning is divided in to two (2) supervised and unsupervised. The new development here is the organization of un-supervised machine learning algorithm as clustering and association(Ahmad, Aftab, Bashir, Hameed, et al., 2018). For the clustering we have k-means. In addition to that, introduction of random forest as a technique under decision tree (Akinkunmi, 2019) and ruled-based classifiers;

consisting of confidence criterion and support criterion as reported by (Bose et al., 2018).

3.3 Machine Learning approach

Machine learning techniques are categorized in to three (3) namely: supervised and unsupervised or semi-supervised(Ahmad, Aftab, Bashir, & Hameed, 2018). The supervised one requires adequate, label training data set while unsupervised do not need label data set. Lexicon-based approach is also called unsupervised approach. According to Mukhtar, Khan, & Chiragh, (2018), machine learning approach can either be supervised explicitly classification or unsupervised- clustering.

3.3.1 Probabilistic Classifier

The probabilistic classifiers are supervised type of machine learning algorithms subdivided into Naïve Bayes, Bayesian network, and maximum entropy.

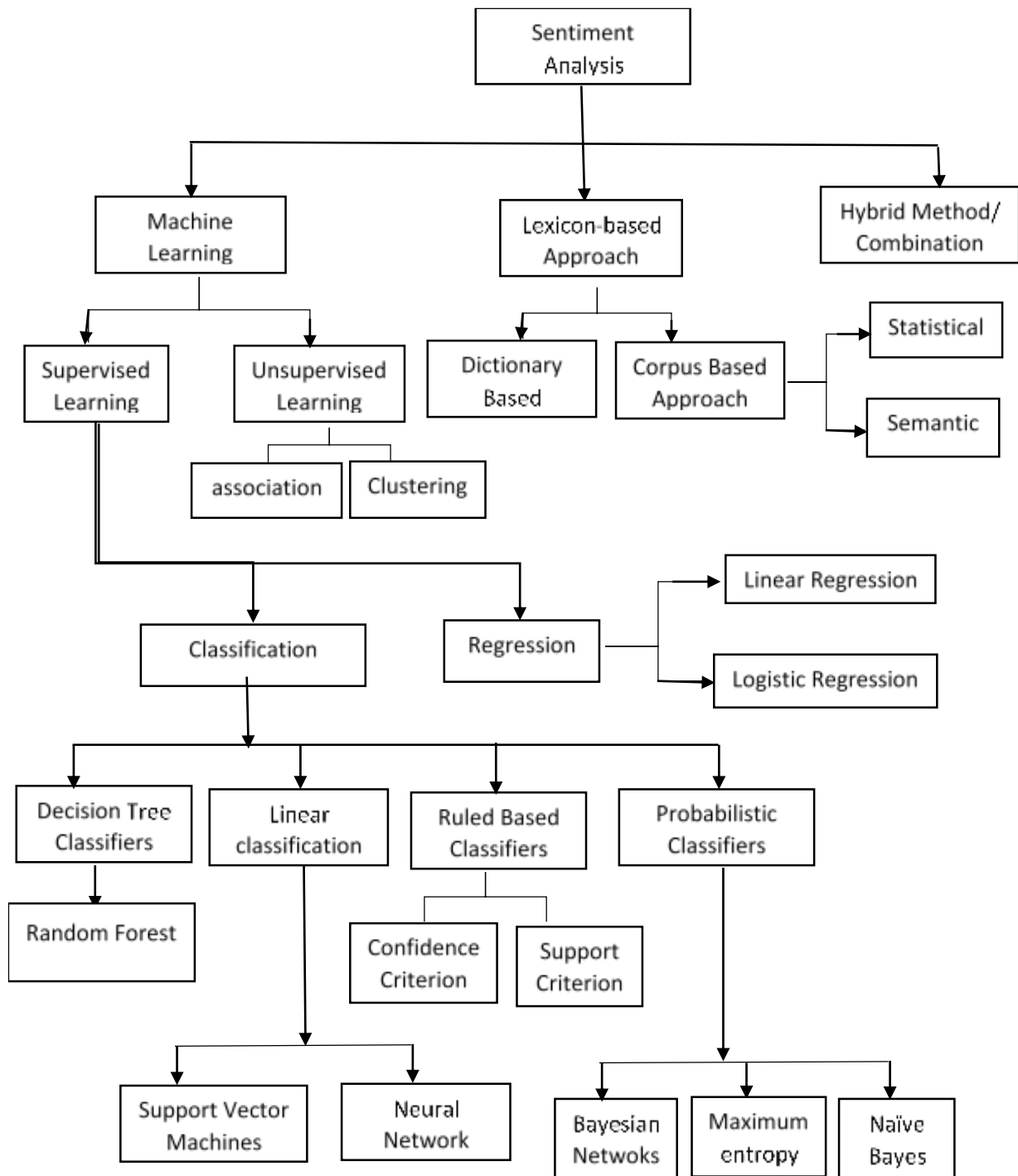


Figure 5. Proposed Taxonomy for Sentiment Analysis Techniques

3.3.1.1 Maximum Entropy

Maximum entropy (MaxEnt) is supervised type sentiment analysis technique that is characterized as feature-based and does not accept assumptions, Features like phrases and bigram can be added to MaxEnt without any overloading (Ahmad et al., 2017). It is used to estimate probability distribution using logistic regression. NBC takes preprocessed data as input for training model so that it can produce positive or negative sentiment (Srividya and Sowjanya, 2017). The Bayes theorem is displayed in equation 2 below.

$$P(c/d) = \frac{p(c)p(d/c)}{p(c)} \quad (2)$$

Where d is the predictor and c is the classes of the data. P(c) is the probability of the classified data set. P(d/c) is the probability of the predictor per class such that p(d) selects c. This functionality is the independence assumption. It may not be real. equation (1) below ensures that overloading does not hold in maximum entropy algorithm.

$$(c/d, \lambda) = \frac{\exp[\sum \lambda_i f_i(c, d)]}{\sum_c \exp[\sum_i \lambda_i \sum f_i(c, d)]} \quad (1)$$

3.3.1.2 Naïve Bayes Classifier (NBC)

NBC is a probabilistic type of supervised machine learning algorithm that is used to predict the probability for a given tuple to belong to another class. It is not like the maximum entropy; the assumption is a factor to consider. This is because it assumes that features in the documents are independent of the other features. This technique is based on Bayes theory and commonly used for its simplicity during training and classification.

3.3.2 Linear Classifiers

Linear-based classifiers are supervised type of machine learning algorithm. They are subdivided into two (2) support vector machine (SVM) and neural networks (NN).

3.3.2.1 Support Vector Machine

SVM is a linear based classifier type of machine learning algorithm efficiently used for document classification. SVM examine data and try to find hyper-plane that classify data into two classes with maximum margin. It supports classification and regression in statistical learning.

3.3.2.2 Neural Networks: This type of machine learning algorithm consists of neurons (similar to biological neurons in brain) otherwise called inputs that are connected to one another via weights linking the neurons for easy passage of signals to yield a single output. The neurons computes activation level as output signal whenever it receives new inputs (Rufai et al., 2018). This author used artificial neural networks to diagnose heart disease. In a study by Bose et al., (2018) neurons are calculated using vectors which refers to the frequencies of words in a document across the lines-set of weights for each neuron used to calculate the function of the inputs.

3.3.3 Decision Tree: This is another type of supervised machine learning algorithm where training data space is broken down in hierarchical order such that the values of attributes are used in data separation. This method predict data based on presence and absence of one or more words, and that will be done repeatedly until a minimum number of records that are registered with the lead nodes used in the classification (Bose et al., 2018).

3.3.4 Ruled-Based Classifiers: This is another type of supervised machine learning algorithm as mention in the taxonomy in figure 4. The rule-based finds the sentiment in text and classify it based on the number of positive and negative words it contains. The classification level can be document, sentence, word and even aspect level (Archana & Kishore, 2017). In a study by (Bose et al., 2018), ruled-based is based on rules which can be further classified in to two(3) as confidence criterion and support criterion (see in Figure 5).

3.4 Lexicon Based Approach

Literature has shown that lexicon-based techniques are unsupervised techniques as most of them don't require labelled and sometimes prepared data. Rather, they use vocabularies of words weighted with their estimation meaning to decide the general notion of given content. These

methodologies have appeared to work adequately on ordinary content (Terán and Mancera, 2019)

4. SENTIMENT LEVEL CLASSIFICATION

Sentiments are classified based on their contents. The contents may be document, sentence, phrase or word depending on the analysis to conduct. This gives the respective levels which categorize the sentiment analysis techniques (Srividya and Sowjanya, 2017). In a situation where by the whole file containing one person's feedback is considered for the analysis, it is termed as document-level sentiment classification. This kind of classification usually undergoes two stages. The first one is, a sentence is classified as either subjective or objective and it is termed as Subjectivity classification. The subjective sentence here refers to the factual information while the objective one means the emotions, believes, personal feelings contained in the sentence. (Saif et al., 2016) reports that techniques of the Naïve Bayesian classification are used to identify subjective sentence. The second one is the Classification of subjective sentences into positive or negative is known as Sentiment Classification. (Srividya and Sowjanya, 2017) reports that there are three (3) sentiment classes viz: *Positive*, *Negative* or *Neutral*. This author concludes that applying sentiment at the document and sentence level is helpful but people's actual opinion on like and dislike cannot be discovered, hence, it can't distinguish sentiment on focus. (Hussein, 2018) writes, Aspect/Feature Level Classification distinguishes and concentrates object including that have been remarked on by the sentiment holder and after that decides if the sentiment is positive, negative, or neutral. Here, similar features are gathered, and an attribute-based review of the various survey is delivered.

5. RESULT

Based on the review of the literature there is a contradiction on the point that SVM generated the best result in analysing student sentiment on document level but according to (Archana and Kishore, 2017), the processes of sentiment analysis differ from system to system based on types of the classes to predict (positive or negative, subjective or objective), different levels of classification (sentence, phrase, or document level and language that is processed. In a study by (Mika et al., 2017) in the paper titled “The evolution of sentiment analysis—A review of research topics, venues, and top-cited papers”, shows that most of the researches in the area of sentiment analysis focused on products reviews in e-commerce and movie reviews. In most cases, centred on document-level sentiment analysis classification specifically on Twitter data. Machine learning algorithms are considered better than other approaches on Large labelled data collected in English Language (Ozturk et al., 2017). Lexicons are better than the machine learning algorithm on large data in the Urdu Language based on Accuracy, Precision, Recall, F-measure

and even economy of time and efforts used (Mukhtar et al., 2018)

According to Srividya and Sowjanya,(2017) most of the studies in sentiment analysis that used machine learning algorithm focused on document-level classification as it gives the best result when the document is composed by a single person. Researchers’ attention focused on sentiment analysis of student evaluation of teachers (SET), twitter, and Facebook compared to an online.

Review of the literature shows that sentiment analysis techniques were organised in to three groups; Machine learning, Lexicon and hybrid or combination. The machine learning techniques are grouped in to two (2) namely; supervised and unsupervised. The un-supervised machine learning algorithm is sub-divided in to clustering and association(Ahmad, Aftab, Bashir, Hameed, et al., 2018). The clustering technique consist of k-means. The decision tree technique consist of random forest technique (Akinkunmi, 2019) while the ruled-based classifiers consist of confidence criterion and support criterion as reported by (Bose et al., 2018).

6. DISCUSSIONS

From the literature survey, it was found that a lot of researchers have dealt with the sentiment analysis tools and techniques. But hardly to find all the techniques and tools mention in a study. Some of the sentiment analysis techniques commonly used are Naive Bayes, Support Vector Machine, Max entropy. The review additionally recognizes three (3) normally utilized methodologies in sentiment analysis to be specific; Machine Learning, Ruled-based and Lexicon-based approaches. Machine learning algorithms are considered better than other approaches on Large labelled data collected in English Language (Ozturk et al., 2017). Lexicons are better than the machine learning algorithm on large data in the Urdu Language based on Accuracy, Precision, Recall, F-measure and even economy of time and efforts used (Mukhtar et al., 2018). The machine learning algorithm specifically Naïve Bayes is chosen for this study because it overcomes the lexicon-based approach generally in terms of accuracy in the area in which it is trained(Mukhtar et al., 2018). SVM outperforms NB on large data but the case may not be the same on small data.

The review of the literature had also reported that sentiment analysis involved a series of steps which include preparation, pre-processing, sentiment identification. And the commonly used tools are weka, python compiler, and R programming tool (Archana and Kishore, 2017). In R programming, text mining packages like NLP and “stringr” are used in sentiment analysis. The review also mentions that the result of sentiment analysis is either positive or negative which imply good emotions or bad emotions on the entity in question respectively.

7. CONCLUSION

It is observed that researchers developed their own frameworks and architectures for the implementation of machine learning algorithm in sentiment analysis in different case studies. Most of the frameworks incorporate at least four phases namely; data collection, data preprocessing, sentiment classification (training and testing set) and result evaluation (Ahmad et al., 2017, Atif et al., 2018, Abbas et al., 2019). For the lexicon-based approach the frameworks commonly include the following phases: data collection, preprocessing, token matching (Bag of words and lexicon) and result evaluation (Duwairi, Ahmed, & Al-Rifai, 2015).

The survey done by (Serrano-Guerrero et al., 2015) has shown that sentiment analysis is conducted on three levels based on the data; document level, sentence level, and word level. The survey had also identified that document level SA gives accurate result if the document is written by one person. Base on the taxonomy developed, we can say that sentiment analysis techniques are organised in to three groups; Machine learning, Lexicon and hybrid or combination. The machine learning techniques are grouped in to two (2) namely; supervised and unsupervised. The un-supervised machine learning algorithm is sub-divided in to clustering and association(Ahmad, Aftab, Bashir, Hameed, et al., 2018). The clustering technique consist of k-means. The decision tree technique consist of random forest technique (Akinkunmi, 2019) while the ruled-based classifiers consist of confidence criterion and support criterion as reported by (Bose et al., 2018). This research has produced a taxonomy which serves as a guide for the choice of techniques in sentiment analysis. Based on the result of this study, we conclude that the commonly used tools in sentiment analysis are Weka, Python compiler, and R programming tool.

REFERENCES

1. Ahmad, M., Aftab, S., Bashir, M. S., & Hameed, N. (2018). *Sentiment Analysis using SVM: A Systematic Literature Review*. 9(2), 182–188.
2. Ahmad, M., Aftab, S., Bashir, M. S., Hameed, N., Ali, I., & Nawaz, Z. (2018). *SVM Optimization for Sentiment Analysis*. 9(4).
3. Ahmad, M., Aftab, S., Muhammad, S., & Ahmad, S. (2017). Machine Learning Techniques for Sentiment Analysis: A Review. *Int. J. Multidiscip. Sci. Eng*, 8(3), 27–32.
4. Akinkunmi, M. (2019). Introduction to Statistics Using R. In *Synthesis Lectures on Mathematics and Statistics* (Vol. 11).
5. Alkubaisi, G. A. A. J., Kamaruddin, S. S., & Husni, H. (2018). Conceptual framework for stock market classification model using sentiment analysis on twitter based on Hybrid Naïve Bayes Classifiers. *International Journal of Engineering and Technology(UAE)*, 7(2), 57–61.
6. Archana, R., & Kishore, B. (2017). Role of

- Sentiment Analysis in Education Sector in the Era of Big Data: a Survey. *International Journal of Latest Trends in Engineering and Technology*, 022–024.
7. Atif, M. (2018). *An Enhanced Framework for Sentiment Analysis of Students' Surveys: Arab Open University Business Program Courses Case Study*. 9(1), 9–11.
 8. B., V., & M., B. (2016). Analysis of Various Sentiment Classification Techniques. *International Journal of Computer Applications*, 140(3), 22–27.
 9. Bose, R., Dey, R. K., Roy, S., & Sarddar, D. (2018). *Sentiment Analysis on Online Product Reviews*. (August).
 10. Duwairi, R. M., Ahmed, N. A., & Al-Rifai, S. Y. (2015). Detecting sentiment embedded in Arabic social media - A lexicon-based approach. *Journal of Intelligent and Fuzzy Systems*, 29(1), 107–117.
 11. Hussein, D. M. E. D. M. (2018). A survey on sentiment analysis challenges. *Journal of King Saud University - Engineering Sciences*, 30(4), 330–338.
 12. Jadav, B. M., & Scholar, M. E. (2016). *Sentiment Analysis using Support Vector Machine based on Feature Selection and Semantic Analysis*. 146(13), 26–30.
 13. Kolchyna, O., Treleaven, P., & Aste, T. (2015). *Twitter Sentiment Analysis: Lexicon Method, Machine Learning Method and Their Combination*. (January 2016).
 14. Lajis, A., Baharudin, S. A., Kadir, D. A., Ralim, N. M., Nasir, H. M., & Aziz, N. A. (2018). A review of techniques in automatic programming assessment for practical skill test. *Journal of Telecommunication, Electronic and Computer Engineering*, 10(2–5), 109–113.
 15. Mukhtar, N., Khan, M. A., & Chiragh, N. (2018). Lexicon-based approach outperforms Supervised Machine Learning approach for Urdu Sentiment Analysis in multiple domains. *Telematics and Informatics*, 35(8), 2173–2183.
 16. Ozturk, Z. K., Cicek, Z. İ. E., & Ergul, Z. (2017). *Sentiment Analysis: an Application to Anadolu University Sentiment Analysis: an Application to Anadolu University*. (October).
 17. Rani, S., & Kumar, P. (2017). A Sentiment Analysis System to Improve Teaching and Learning. *Computer*, 50(5), 36–43.
 18. Rufai, A., S., U., & Umar, M. (2018). Using Artificial Neural Networks to Diagnose Heart Disease. *International Journal of Computer Applications*, 182(19), 1–6.
 19. Saif, H., He, Y., Fernandez, M., & Alani, H. (2016). Contextual semantics for sentiment analysis of Twitter. *Information Processing and Management*, 52(1), 5–19.
 20. Serrano-Guerrero, J., Olivas, J. A., Romero, F. P., & Herrera-Viedma, E. (2015). Sentiment analysis: A review and comparative analysis of web services. *Information Sciences*, 311, 18–38.
 21. Srividya, K., & Sowjanya, A. M. (2017). *Sentiment analysis of facebook data using naïve bayes classifier Assistant Professor , Department of Computer Science and Engineering , Assistant Professor , Department of Computer Science and systems engineering , AU College of Engineering (A), Andhra . 15(1), 179–186.*
 22. Terán, L., & Mancera, J. (2019). Dynamic profiles using sentiment analysis and twitter data for voting advice applications. *Government Information Quarterly*, (March), 1–16.