UNIVERSITY OF NAIROBI

School of Computing and Informatics

PUBLIC SENTIMENTS ANALYSIS ON JUDICIAL OUTCOMES AGAINST EXPECTED OUTCOMES IN KENYA

By

ROBERT SABUNI WEKESA
REGISTRATON NUMBER: P53/64977/2013

SUPERVISOR: DR. STEPHEN MBURU

Project submitted in partial fulfillment of the requirements for the award of a degree in Master of Science in Distributed Computing Technology

DECEMBER 2018

DECLARATION

I, the undersigned, hereby confirm that the research project that is presented on this report is my original work and that to the best of my knowledge it has not been presented anywhere else for any other University Award

Signature: ____________________ Date: ___________________

Name: Robert Sabuni Wekesa- P53/64977/2013

This research has been submitted in partial fulfillment of the requirement for award of a degree in Master of Science in Distributed Computing Technology with my approval as university supervisor**.**

Signature: ____________________ Date: _________________________

Supervisor: Dr. Stephen Mburu

School of Computing and Informatics

University of Nairobi

ACKNOWLEDGEMENTS

My deep and sincere gratitude goes to my supervisor Dr. Stephen Mburu, University of Nairobi, for his guidance that helped me understand what a research project entails, more so a masters research project. He took his valuable time in guiding me and was always available for consultation whenever needed; this enabled me to progress with my project on schedule thus fulfilling the requirements of a master's degree award. I would also like to extend my gratitude to the entire panel of Dr. Miriti Evans, University of Nairobi, Dr. Chris Chepken, University of Nairobi, and Dr. Stephen Mburu for their constructive criticism and guidance accorded to me during my presentations.

I would also like to pass my heartfelt appreciation to my family, my wife Cate, my children Nasaka and Dixon, my brother Eric for their constant encouragement. Special appreciation goes to a dear friend, Christine Murunga for being the positive voice that picked me up whenever I felt like giving up. I would like to appreciate my late Uncle, Dixon Sabuni, whose love and constant reminding of my academic progress gave me energy to push on.

Lastly but most importantly, I'm very grateful to the almighty God for guiding and enabling me financially, health-wise and the providing the necessary energy and intellect required in completing this project.

Abstract

The Kenyan judiciary has for a long time struggled to gain the confidence of the public. Judicial decisions in some cases do conflict with the expectations of the public and this serves to erode the public's confidence in the institution. The judiciary needs a means to gauge the public's opinions on ongoing cases in order to measure the deviation of its decisions from the expected outcomes from preliminary hearings on ongoing litigations. This should facilitate the judiciary to get in touch with the public's expectation and the effect of its decisions on the public. This should in the long run guide the judiciary in bettering its service delivery procedures to effectively serve the public. A means to mining of sentiments of the public to analyze them and understand their opinions is important in achieving this. A number of sentiment analysis algorithms exist. This research investigated the suitability of these algorithms by reviewing literature on their performance in similar problem domains i.e. text classification. The chosen algorithm was trained by help of Weka, starting with 70 initial of instances, of these instances 53% were correctly classified while 17 were incorrectly classified. This gave a 75% classification. Subsequent training of the model with the same number of instances gave 81% classification accuracy. This trained model was applied to public sentiments on three public cases. From the results, it's clear to note that in some cases public's opinions were not aligned with the judiciary's decisions, an indication of public's dissatisfaction in such decisions. In some cases, there was agreement. In one of the cases, the 2017 presidential petition, the percentage of those who approved the handling of the petition stood at 54% but that number drops to 34.5% after the judicial decision denoting disapproval of the decision by the public though they supported the process. The research used two models, a mathematical model of the algorithm used and a structural model to illustrate interaction of various components of the system prototype.

Table of Contents

List of Abbreviations and Acronyms

SVM - Support Vector Machine

NB - Naïve Bayes

DT - Technique, Decision Tree

KNN - K-Nearest Neighbor

Weka -Waikato Environment for Knowledge Analysis

BR -Building Resource

FS -Feature selection

SC -Sentiment classification

List Figures

List Tables

CHAPTER ONE: INTRODUCTION

1.1Background

The ability to gauge the public's standing on various aspects and topics especially on issues that affect the well-being of society is an important area of concern. Computer science as a discipline has contributed so much on the mining of the public's opinions on various topics of interest. To effectively mine and analyze the opinions of various groups on various topics is important to improve the experience of these groups. From companies seeking to understand the performance of their various products and services to organizations seeking to understand the public's take on a number of topics, opinion mining or sentiment analysis becomes an important area of study. In Kenya, following the 2018 general elections, a number of cases were lodged in various courts across the country, the decisions to these cases determined major events in the country's history, for instance, the decision to annul the 2018 general elections led to various activities that affected the country and its populace in many ways, both negatively and positively. Various quarters of the population expected varied outcomes, as much as it is almost impossible to satisfy all parties in any judicial proceeding, it is important though, especially for public interest cases to understand the public's opinion on the matter. Sentiment analysis is a useful technique for mining opinions from online social platforms, and generally any form of text to understand the polarity of such text. Application of these techniques to specific domains is not as straightforward and requires prior evaluation, training of the classifiers involved and then performing a final test to confirm their effectiveness on the same. This study sort to identify an appropriate classifier for the analysis of public sentiments on judicial cases in order to help understand the disparity between the judicial decisions and the expected outcomes from the public. This should help the judiciary understand, re-evaluate its decisions and guide future policies on fostering the judicial neutrality, since it's important that it's not enough that justice is done but it should also be seen to be done. This study used the various court cases as sample public interest cases and analyzed the opinions of the public on the cases during the proceeding, capturing the sentiments at various stages of the proceedings and comparing these opinions to those after the decision by the Supreme Court of Kenya. This should offer a guide to understand the public's satisfaction or otherwise in the decisions, and by extension apply the same to other cases of public interest cases. Information from the public social platforms can be classified as either fact or opinion. An objective and sometimes accurate expression on certain entities or products can be termed as fact. On the other hand, subjective expressions made by people regarding products or entities mostly based on their emotions are classified as opinions. .Opinion mining or sentiment analysis is an emerging but very important field that offers preview into the public's opinions through sentiment analysis, a very important factor in decision making. Opinion mining has gained application in various fields, but this has largely been applied in marketing where companies seek to understand the opinions of their clients on their products in seeking to improve the client's experience. With careful selection of the right model, sentiment analysis can find application in other domains such as judicial sentiment analysis problems. This study will therefore evaluate the existing models in sentiment analysis most suitable for text classification and apply it to the problem at hand, judicial sentiment classification.

1.2 Problem Statement

The judiciary has faced a number of criticisms regarding its decisions, in most cases it being accused of being biased. As a branch of government charged with resolving conflicts, its neutrality is an important aspect. Even though it is hard to determine the neutrality or otherwise of the judiciary and the cases it deals with, efforts should be seen in working towards changing the perception of the public regarding various judicial proceedings. It is said that justice should not just be done but also seen to be done. This further enforces the concept of neutrality, that is, it is not enough for courts to be neutral but they should make every effort to appear neutral from the public's perspective. Understanding the public's take on various high octane cases can help the judiciary understand the disparity of its decisions from the expected outcome by the public. This may guide the judiciary in formulating policies to improve its standing in the public's eye. Sentiment analysis is an important emerging field that can effectively and efficiently used to evaluate the sentiments of the public from social media platforms. A number of algorithms are available for performing this task, thus finding an appropriate one and demonstrating its performance and suitability is imperative.

1.3 Objectives of the study

1.3.1 General Objective

The main of the objective of this study is to review sentiment analysis algorithms and implement a prototype of the best algorithm for analysis of public opinions and sentiments on active cases and the sentiments on the outcomes of the cases in order `to determine the disparity between sentiments on ongoing cases and sentiments on case outcomes

1.3.2 Specific Objectives

i. To review sentiment analysis algorithms and their performance on sentiment analysis
ii. To build a prototype to demonstrate the application of sentiment analysis algorithms in determining the public opinion on active cases and sentiments on the outcomes of the same cases
iii. To do a comparative evaluation of the results of the sentiment analysis during case progression against the sentiments on the actual rulings to determine the disparity between the public expectation and the actual judicial decisions

1.4 Research Questions

1. What are some of the sentiment analysis algorithms and how is their performance on sentiment analysis?
2. Can a prototype be built to demonstrate the application of sentiment analysis algorithms on judicial sentiment analysis?
3. Is there a disparity between decisions on ongoing cases and their outcomes from the public perspective based on the judicial sentiment analysis?

1.5 Significance

The findings of this study will help the judicial system in Kenya and policy makers understand the effect of the various judgments on the public by understanding the sentiments of the public on the various judicial pronouncements. The trend of the public's satisfaction and dissatisfaction regarding judicial decisions can help shape and align judicial policies to make sure justice isn't done but seen to be done in the eyes of the public

1.6 Scope and Limitations of the study

This study will be limited to a few cases of public interest, the cases that draw huge public interest are in most cases election petitions and those cases that touch on serious social dilemmas such as abortion, rape, divorce and other spheres of social life. The study will consider two election petitions and one case touching on the social sphere of the society. In determining the right algorithm to apply to the job, the study will constrain itself only to the study of existing literature and to experimental evaluation will be conducted. The experiment will be applied only during the training of the chosen algorithm. A prototype will be constructed to demonstrate the constructed prototype in classifying public sentiments on the judicial decisions.

This study may be limited by data on various cases, as the main platform for harvesting user opinions, twitter, does not give access to users on tweets older than seven days. Another limitation is the computing power that will be required to train the chosen algorithms. Linear classifiers tend to take huge computing power hence this may lead to longer training hours.

CHAPTER TWO: LITERATURE REVIEW

2.1 Categories of sentiment analysis methods

Classification methods for sentiment analysis can loosely be categorized in two broad categories i.e. machine learning approaches and Lexicon based approaches. The Lexicon based approaches can further be sub-divided to corpus approach and dictionary based approach that utilize semantic or statistical methods determining the polarity of sentiments while machine learning approaches can be divided (Walaa Medhat, Ahmed Hassan, Hoda Korashy, 2015) in two basic categories i.e. supervised and the unsupervised machine learning groups. Supervised machine learning utilize the availability of readily existing data to be used for the training of the algorithms while unsupervised learning algorithms are used mostly where there exist a large set of unlabeled data.

2.2 Lexicon based Classifiers

Lexicon approaches rely on finding the opinion lexicon and then using this to analyze a lexicon. We have two techniques that are used in this, the first approach is corpus based and the other is uses the dictionary approach. Dictionary approach relies on getting words to use for seeding i.e. seed words and opinions and thereafter utilizing this in searching the dictionary to see if their synonyms or antonyms exist. In the corpus approach, it starts with seed opinions which proceed to search for further opinion words in the rest of the corpus to find the particular opinion phrases that have orientation that is specific to the context.

2.3 Machine learning Classifiers

This uses machine learning algorithms for text classification problems, text classification problem is defined as the set D = (A_1, A_2, A_3….An) where A is a training record, each of this records is assigned to a particular class. The model of classification relates the underlying features of the class label. Hard classification, a situation that is encountered if just a single instance is assigned to one class label; in the other case, soft classification problem occurs if an instance is assigned to a probabilistic list of values of the labels.

2.3.1 Supervised machine learning

These learning algorithms usually rely on training sets of documents labelled appropriately and only find application where such labeled data exists. Most machine learning problems do utilize machine learning classifiers. Supervised machine learning algorithms can further be categorized depending on the type of the problem; we have classification problems where the output of the classifier is a category e.g. well or unwell, agree or disagree. The other category of supervised machine learning problems is regression; the output of these classifiers is a real value e.g. meters, age etc. There are number of classifiers under this category

2.3.1.1 Probabilistic Classifiers

Probabilistic Classifier utilizes a collection of various classification models. These models assume every class to part of the collection. These models are sometimes referred to as generative models. This research discusses three of those most commonly used probabilistic models.

2.3.1.1.1 Naïve Bayes Classifier

The naïve Bayes is based on Bayes theorem with the independent assumption between predictors. Naïve bays model is quite to build without complicated iterative parameter estimations. This makes it one of the most simple and also one of the commonly used algorithms. This is one of the models most suitable in the calculating a class's posterior probability is computed by this model according to the word distribution in a document.

This model utilizes the bag of words model, this model ignores the order of words in a document, a feature of extraction that comes in handy when the order of words is not an important factor. This classifier uses Bayes theorem for calculating the posterior probability, $P(c|x)$, from $P(c)$, $P(x)$, and $P(x|c)$. Naive Bayes classifier assumes that the effect of the value of a predictor (x) on a given class (c) is independent of the values of other predictors. This assumption is referred to as class conditional independence. (Sayad, 2017)

Figure 1 Naïve Bayesian Mathematical Model

Posterior Probability

Class Prior Probability

Likelihood

$$P(c|x) = P(x|c)P(c)/P(x)$$

Predictor Prior Probability

$$P(c|X) = P(x_1\,|c) * P(x_2\,|c) *\cdots\cdots* P(x_n\,|\,c) * P(c)$$

Source: (Sayad, 2017)

$P(c|x)$ is posterior probability of target class given predictor *attribute*

$P(c)$ gives the prior probability of the class.

$P(x|c)$ gives the likelihood that is the probability of the predictor given the class.

$P(x)$ gives the prior probability of the predictor.

2.3.1.1.2 Bayesian Network

This is another probabilistic model based approach that differs with Bayesian classifier on the assumption of the independence of features. While the main assumption of naïve Bayes is the based on features independence, the Bayesian Network utilizes some different but somehow extreme assumption that all the features actually completely dependent. The Bayesian Network is usually directed acyclic graph with nodes representing random variables; on the other hand, edges do represent conditional dependencies. In the case ,text mining, the of Bayesian network's computational complexity is believed to be too expensive thus it does not find very wide application (Walaa Medhat, Ahmed Hassan, Hoda Korashy, 2015)

2.3.1.1.3 Maximum Entropy Classifier

Maximum entropy classifier falls under the probabilistic classifiers belonging to the class of exponential models. The maximum entropy does not assume that the features are conditionally independent. The algorithm is based on the principle of maximum entropy and from all models that fit the training data, and selects the model with the largest entropy (Vryniotis, 2015). This algorithm is useful in solving a variety of text classification jobs e.g. language detection, topic classification, language detection and sentiment analysis. Due to the minimum assumptions this algorithm makes, it is mostly preferred when no knowledge exists on prior distributions hence no useful assumptions can then be made prior. It also finds application where it is not possible to make assumptions on conditional independence of features. Its weakness is that it requires more time to train compared to naïve Bayes due to majorly optimization problems that need to be worked out in order to estimate parameters of the model, but once these parameters have been computed, then it provides very robust results and its CPU and memory consumption is fairly competitive.

2.3.1.2 Linear Classifiers

Linear classifiers are one of the most practical classification ways. Linear classifiers will associate a coefficient with the counts of each word in a sentence. Logistic regression classifier, an example of linear classifier, allows one to predict a class and provides a probability associated with the prediction. These probabilities are very useful, since they provide a degree of confidence in the predictions. It can allow one to construct features from categorical inputs, and to tackle classification problems with more than two classes (multiclass problems). Linear classifiers make their classification based on a linear predictor function combining a set of weights with a feature vector

$$y = f\left(\underset{w}{\rightarrow}.\underset{x}{\rightarrow}\right) = f(\sum_{j}^{n} \omega_j x_j \ )$$

Taking $\underline{X}$ = {x_1,…….x_n} as the normalized word frequency in a document, the vector

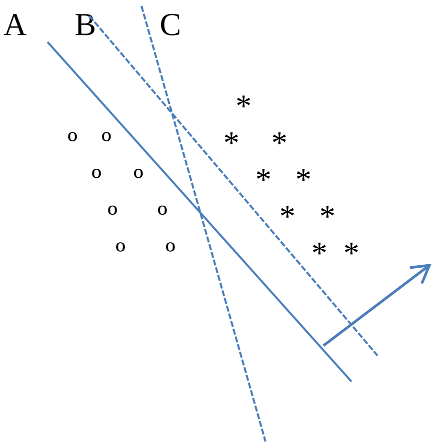
$\underline{w} = \{w_1, \ldots\ldots w_n\}$ taken as the linear coefficients vector. Their dimensionality is usually equal to feature space. Taking c as a scalar then the linear predictor's output then will defined as
$y = \underline{W}.\underline{X} + c$. The predictor p will separate the hyper-plane between the different classes.
There are various types of linear classifiers, the ones mostly used in sentiment analysis include

2.3.1.2.1 Support Vector Machines (SVM)

SVM in full support vector machines classifiers operate by determining good linear separators between classes. Support vector machines are most suited for text data , this is so because of the sparse nature on the text in most cases, this is done by correlating the features in a data set where the features are mostly irrelevant, the features are correlated the separated in linear separable classes . Support vector machines have the ability to build nonlinear decision surfaces in original feature space by mapping the data instances i.e. this is done no-linearly resulting in an inner product space, this then allows the classes now to be linearly separated along a hyper-plane. Support vector machines find application in a wide variety of areas but are better suited for classifying reviews, in most cases considering the quality of such reviews.

Support vector machines were part of the research conducted by Li and Li (Li Young-Ming, Li Tsung-Ying, 2013), as a classifier for polarity classification. Unlike in the problem of binary classification problem, they stressed that the credibility of the opinion giver and their subjectivity should be taken into consideration, they therefore put forward a framework that ensures production of a numeric summary of opinions in micro-blogging sites.
The identified the subjects discussed by the used and classified them with support vector machines. They also harvested posts twitter in the study. It was established that user credibility if, considered, then it follows that the subjectivity of opinions is very essential in aggregating micro-blogging site's opinion. It was established, from this study that the mechanism employed in this study be used in effectively in discovering market intelligence to offer support in decision making by decision makers by getting customer opinions on various areas of business for improvement in real-time.

Figure 2 Support Vector Machine Mathematical Model



Source: (Li Young-Ming, Li Tsung-Ying, 2013),

Figure 3: Support Vector machine Mathematical Model

$$K(X_i X_j) = \begin{cases} X_i . X_j & Linear \\ yX_i \ . X_j + C)^d & Polynomial \\ \exp(-y|X_i - X_i|^2 & RBF \\ \tanh\left(yX_i . X_j \ + C\right) & Sigmoid \end{cases}$$

Source: (Li Young-Ming, Li Tsung-Ying, 2013),

2.3.1.2.2 Neural Networks

The basic make up of this classifiers is a neuron, designed to emulate the human brain, thus a classifier may be composed of many neurons layered. The neurons receive their inputs as vectors that are the word frequencies within the document under consideration. Non-linear boundaries are better solved by multiplayer neural networks. The multiple layers find application in the induction of multiple piecewise linear boundaries that are used in enclosed regions of certain specific classes. The previous layers of the neurons will feed in the next layer of neurons in the network. The training process of neuro network is very intensive and a lot of work in involved. This is so since the errors from the layers ahead in the network are propagated back during the training process. Thus neural networks will not be appropriate for this study. This decision is largely informed by the empirical studies conducted by Moraes and Valiati (Moraes Rodrigo, Valiati Joao Fransisco, 2013) comparing support vector machines and artificial neural networks evaluating their performance on document-level sentiment analysis. The comparison was necessitated by the of huge success of support vector machines in sentiment analysis, on the other neural networks haven't got as much application in the field of sentiment analysis

In their results they discuss the requirements for both methods, the results from the models, and the prevailing circumstances or contexts under which those results are realized and the optimal conditions for better and accurate classification. They have, in their study utilized standard evaluation method applicable to most supervised methods of feature selection in the bag of words model. From their study, it was apparent that artificial neural networks gave better results compared with support vector machines save for in the cases of unbalanced data contexts. Three data sets were considered, that is movie rating, GPS and book reviews from the amazon online store. They established that artificial neural networks outperformed support vector machines on movie reviews by a big margin, statistically speaking. In their study, they encountered some limitations present in both models that were rarely discussed in sentiment analysis literature; these included the computational complexity of support vector machines at runtime and computational complexity of artificial neural networks at time of training. Within this study, it was established that the use of information gain, a feature selection technique that is cheap computationally, will significantly reduce the computational complexity for both artificial neural networks and also for the support vector machines but this will not significantly impact on the accuracy of the classification. Support vector machines and artificial neural networks perform significantly well in personal relationships classification in biographical texts as illustrated by Van de Camp and van den Bosch. Van de and van den used relations between two individuals; the first individual was designated as the subject of a biography and the second person being mentioned in the biography. The relations mapped to positive, neural or unknown. The research drew its data from historical biographical data describing persons in specific domains, regions and time periods. It was demonstrated that the classifiers classified the relations above the majority baseline score. It was established that using training sets with relations of multiple people yields better results that using training sets that focus only on particular entity, therefore showing that support vector machines and one layered neural networks classifiers result highest scores.

2.3.1.2.3 Decision tree classifiers

In decision tree classifiers, a condition on the value's attribute is used in dividing the data, providing hierarchal decomposition of training data space. The predicate in this case is the presence or absence on one or more words. Division will be carried out recursively till leaf nodes only are left with a particular record number that is then used in the classification. Other predicate kinds exist that rely on the similarity on document-text in correlating term sets that can also be utilized in further portioning on the documents. There are single attribute split that utilize either presence or absence of particular words in a specific node in the tree to allow it to do the split and the other split is the similarity based multi attribute that works by using word cluster frequency and the similarity of these words with the documents to do a split. Fischer discriminant is used when it comes to the discriminant based multi attribute split. When it comes to classification of text, the decision tree implementation is varies minimally from the standard packages in use e.g. C4.5 and ID3. C5, a successor to C4.5 is the algorithm that was used by Li and Jain. An approach proposed by Li (Li Young-Ming, Li Tsung-Ying, 2013) allows one in mining topical terms content structures in sentence level contexts by the structure of maximum spanning tree in discovering linkages among topical terms and their context words

2.3.1.3 Rule-based classifiers

In these classifiers modelling of data space is achieved by a set of rules. A condition on the set of features expressed in disjunctive normal form is represented by the left hand side whereas the class label is represented on the right side.

Rule: (Condition) $\rightarrow$ X

Condition : conjunctions of attributes

X : the class label

Left Hand Side : rule antecedent or pre condition

Right Hand Side: rule consequent

Conditions are specified based on presence of terms. Use of absence of terms is very rare due to its poor performance when considering sparse data. The generation of rules follows certain criterion, which the training phase also depend on during the training time. Confidence and support are the widely used criteria; support is an absolute number of instances in the data set being used in the training relevant to the rule while confidence is the conditional probability which the right hand rule will be satisfied when the left side is also satisfied

2.4 Lexicon-based approaches

Sentiment classification is a discipline that heavily relies on opinion words, where various desired states are expressed using positive opinions while negative opinions usually express undesired states. Opinion lexicons are opinion phrases and idioms put together. The approaches most employed in collecting in collecting the opinion word list can be categorized into; manual approaches, which are mostly very time consuming and thus it is rarely used. This approach is usually complemented by automated approaches in order to avoid any mistakes that might have resulted from the manual process. The automated approaches fall into two categories.

2.4.1 Dictionary-based approach

A small set of opinion words is collected manually with known orientations. Here we manually collect a small set of opinion words which is thereafter grown by a search in the various corpora e.g. WordNet or even the thesaurus to find the synonyms and their antonyms. The initial word list will grow by the addition of the newly found words, this is then followed by a manual inspection in order to either correct or remove errors. This method has a downside in that it is not able to trace opinion words that with domain and context specific orientation

2.4.2 Corpus-based approach

The other approach is the corpus based approach; this approach solves the question of getting opinion words with context specific orientation, the downside with the dictionary based approach. This approach relies on syntactic patterns or patterns that do. Corpus-based approach helps to solve the problem of finding opinion words with context specific orientations. This method depends on syntactic patterns or patterns that appear together with an initial list of words, these are the words that will be used as seed words in finding the rest of the opinion words in the larger corpus. Hatzivassiloglou and McKeown (Hatzivassiloglou V, McKeown K, 1997), Illustrated one of these methods by starting with a list of adjectives as their initial opinion words together with their orientations. The constraints applied included OR, AND, EITHER-OR, BUT, for example the AND conjunction implies that the adjectives conjoined are of the same orientation. Although this is supposed to achieve sentiment consistency, the reality is that practically it’s not possible for it always to be consistent

2.5 Sentiment analysis classification algorithms comparison

With the different approaches evaluated above, each method does have strengths and weaknesses and none of the approaches or algorithms stands out as the best. These, however should guide the choice of the algorithm. The choice of a learning algorithm should be based on the evaluation of the performance of the algorithm, be it in the context of itself, that is, in absolute terms or even in comparison to other algorithms. Search a choice should, according to (Nathaniel Japkowicz, Mohak Shah, 2013), address four components; performance measures, error estimation, statistical significance testing and test benchmark selection. The specific performance however relied on a number of factors including training data sets, sentiment type, domain orientation and other factors. Ayman Mohamed (Mustafa, 2017) provides a summary of the sentiment classification algorithms performance based on the later in the table below.

Table 1 Learning Algorithms Comparison

Sentiment Type	**Data Scope**	**Data Source**	**Algorithm Used**	**Best Algorithm Result**
BR	Social Media Data	Twitter	Support Vector Machine (SVM)	Support Vector Machine (SVM)
SA	Social Media Data	reviews social media websites	Naïve Bayes (NB) Technique, Decision Tree (DT), K-Nearest Neighbor (KNN), Support Vector Machine (SVM)	Naïve Bayes (NB)

SA	tweets	Twitter	(NB)Naïve Bayes, Support Vector Machine (SVM) , K-Nearest Neighbor (KNN),	Naïve Bayes (NB)
SC	Movie Reviews	OCA Corpus Book Review	Naïve Bayes (NB), Support Vector Machine (SVM)	Naïve Bayes (NB)
SA	Tweets and Facebook Comments	Twitter Facebook	The Naïve Bayes (NB), Support Vector Machine (SVM) and K-Nearest Neighbor (KNN)	Support Vector Machine (SVM)
SA	Comments	Maktoob Twitter	Manual, Dictionary-based, Corpus based, and Integrated Lexicons	Integrated Lexicons
SC	Tweets	Twitter	Johnson Reducer , Genetic-based reducer	Genetic-based reducer
SC	Comments	Facebook Blogs	Support Vector Machine (SVM), Naïve Bayes (NB)	Support Vector Machine (SVM)
SC	Facebook Comments	Facebook	(SVM), Naive Bayes (NB), K-Nearest Neighbor (KNN) and Decision Trees (DT)	(SVM) Support Vector Machines
SC	Tweets	Twitter	K-Means Clustering Algorithm	K-Means Clustering Algorithm
SC	Status Updates	Facebook Twitter	Support Vector Machine (SVM), Naïve Bayes (NB)	Support Vector Machine(SVM)
SA	Social Media Data	Twitter	Support Vector Machine (SVM), Naïve Bayes (NB), Maximum, Bayes Net, J48 decision tree (DT).	NB better for presence vector. SVM better for frequency vector

SA	Social Media Data	Twitter	Support Vector Machine (SVM)	Support Vector Machine
SC	Slang Comments	Aljazeera.net Facebook Youm7.com Alarabiya.net	Support Vector Machine (SVM)	Support Vector Machine (SVM)
SA	Politics and Arts	Twitter	Support Vector Machine (SVM) ,Naïve Bayes (NB), Decision-Tree (DT), K-Nearest Neighbor (KNN)	Support Vector Machine (SVM)
SA	Economic, Sport, News, Health, and Education	Facebook Twitter	Support Vector Machine (SVM) ,Naïve Bayes (NB), Decision –Tree (DT), K-Nearest Neighbor (KNN)	Support Vector Machine
SA	Arts, Politics, Science and Technolog y	Yahoo Maktoob	Support Vector Machine (SVM),Naïve Bayes (NB)	Support Vector Machine (SVM)
FS	Social Media	Twitter	Pattern Matching	Pattern Matching
SC	Social Media Data	Twitter	Naive Bayes (NB), and Support Vector Machines (SVM)	Support Vector Machines (SVM) in Unigrams and Bigrams

SC	Online Forums	Aljazeera	Naive Bayes (NB) Classifier, and Support Vector Machines (SVM)	Support Vector Machines (SVM)
SC	Education, Technology, Sports	Corpus	Naïve Bayes (NB), K-nearest neighbors (KNN) and Support Vector Machine (SVM)	K-nearest neighbors (KNN)
SC	Posts	Facebook	Naïve Bayes (NB), Naïve Search	Naïve Bayes (NB)

Source: (Mustafa, 2017)

The results obtained in the above are varied depending on the algorithm used and the type of the data set applied, and also the polarity in use. Each algorithm may have better performance based on all these aspect but the results of the performance of each algorithm visualized in the table below gives the Support vector machine as the best algorithm of social media sentiment analysis tasks.

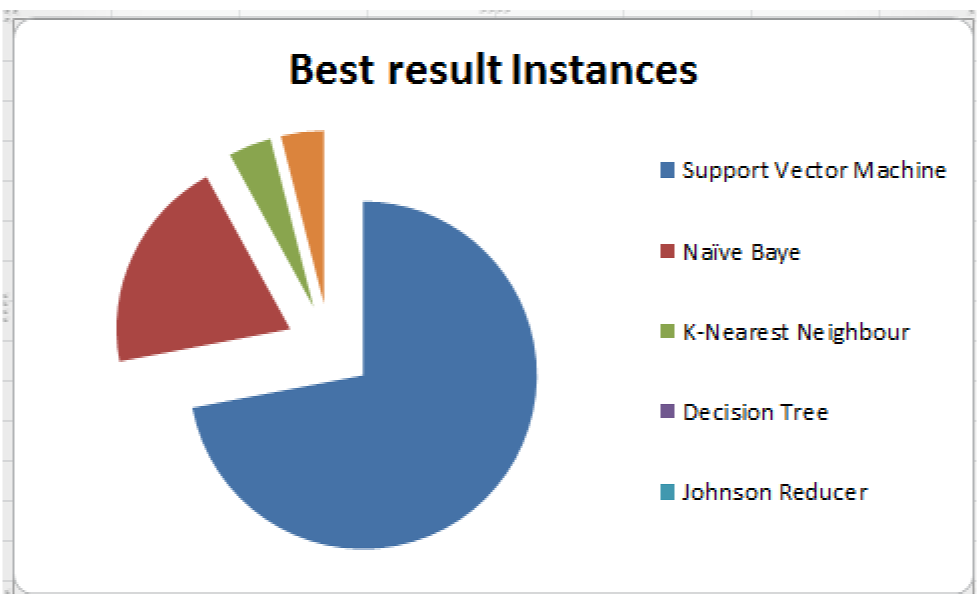


Figure 4: Summary of Evaluation of the different Classifiers

Source: (Author)

Considering an evaluation of the classifiers based on the performance metrics, Rich Caruana and Alexandru Niculescu-Mizil (Rich Caruana, Alexandru Niculescu-Mizil, 2015)give a very compressive Empirical Comparison of Supervised Learning Algorithms based on different problems and data sets, among the classifiers discussed in this research, Support vector machines perform better than the others. In their conclusion, they note that Learning methods such as boosting, random forests, bagging, and SVMs achieve excellent performance. Calibration improves the performance of boosted trees, SVMs, boosted stumps, and Naive Bayes, and provides a small, but noticeable improvement for random forests. Better performing algorithms include SVMs, Random forests and calibrated neural nets. The models that performed poorest were naive Bayes, logistic regression, decision trees, and boosted stumps. Although some methods clearly perform better or worse than other methods on average, there is significant variability across the problems and metrics.

Another study conducted by (M.Vohra, Prof. B. Teraiya, 2015) looks at the performance of various algorithms in sentiment analysis, their findings go on to illustrate the weaknesses of some of the algorithms and their performance based on the various tests. They cite a study conducted by (Pang, L. Lee, 2008)who compared the performance of three classifiers i.e. the support vector machines, Maximum entropy and Naïve Bayes at the document level sentiment classification. They based their evaluations on various feature including but not limited to considering only bigrams, unigrams, and a combination of both in most cases. The outcome show that feature presence is more significant than feature frequency. When the feature is small, Naïve Bayes performs better than support vector machine, but the Support vector machine will outperform the naïve Bayes when the feature space is increased. On the other hand, when feature space gets to be increased, then maximum entropy will come out as better that Naive Bayes, though with one problem, it may suffer from over fitting.

Abbas (A.Abbas, H. Chen, A. Salem, 2008) proposed techniques for sentiment analysis in classification of web forums in classifying hate based forums. This was undertaken in both the English language and Arabic language. They used syntactic and stylistic features. The researchers introduced another algorithm, a generic hybrid, which uses information gain heuristics to improve feature selection. They utilized Support Vector Machine with 10-fold cross-validation and bootstrapping to classify sentiments in all of the experiments. During this process, while using both syntactic and stylistic features, they achieved 95.55% accuracy in 10 crosses validation. This still point to the fact that SVM comes across as being better than the others as far as sentiment classification is concerned, though the variance of the experimental parameters may have a significant impact on the overall results, but averagely SVM will outperform other classifiers in the task of sentiment classification.

Supervised machine learning techniques demonstrate relatively improved performance in comparison to their unsupervised counterparts. It is wrong to write foo unsupervised methods since supervised classifiers require huge datasets of labelled data, this in most cases is very difficult to obtain and very expensive, on the other hand, unlabeled data is easy to acquire hence maybe very fundamental in the implementation of various solutions.

Support vector machines tend to emerge as one of the algorithms with good accuracy levels, but just like its supervised counterparts, it requires large corpora of labelled data sets for training and eventual application.

Movassate and Parikh (R. Parikh, M. Movassate, 2013) classified data from twitter (tweets) by the use of two Naive Bayes models, that is, unigram and bigram model and a also Maximum Entropy model. They established that Naive Bayes classifier was better in accuracy and efficiency as opposed to the Maximum Entropy model.

In a separate study, Go (A. Go, R.Bhayani, L.Huang, 2009), proposed a solution that consisted of tweets with emoticons in the training set. Since some tweet texts are not easy to classify, emoticons come in as noisy labels. They build models using Naive Bayes, Support Vector Machines and Maximum Entropy. The feature space composed of unigrams, bigrams and POS. They observed that SVM emerged with better performance and accuracy compared to the rest of the models and they also observed that unigram had more efficient features.

In other studies, captured by (P Dinkar Shinde, Dr. S Rathod, 2018), they review a number of works done to evaluate the performance of various algorithms The datasets utilized in this studies include movie reviews, product reviews and data from various social media platforms e.g. twitter and Facebook. They utilized pattern based approaches, Natural processing and also machine learning. The table below summarizes some of the findings obtained from the studies

Table 2: Summary of findings

	Machine Learning Classifier	Strength	Weakness
1	KNN	Simple and can be used in multiclass document categorization.	It takes long to categorize huge data sets. Utilizes a lot of memory for processing
2	Decision Tree	Fast during training period.	Difficult in handling datasets with noisy data. Has a problem in over fitting of data.
3	Naïve Bayesian	It is simple and is easy to use with both textual and numerical data. Easy to implement. It is computationally cheap and easy to implement	When feature set is highly correlated, its performance deteriorates very much and results in relatively low classification performance when using large data sets.
4	Support Vector Machine	High accuracy even when using large dataset. Works well with numerous dimensions. No over fitting.	Problems in representing document into numerical vector.

Source: (P Dinkar Shinde, Dr. S Rathod, 2018),

In the study carried out by (Chuanming Yu, 2009) , (R Varghese, M Jayasree, 2010) and (A Gupte, S Joshi, P Gadgul, A Kadam, 2014) propose aspect based sentiment analysis. Chuanming used four datasets to test the Support vector machine model, he compared Maximum Entropy classifier method for feature extraction with SVM method and he has concluded that SVM Method was superior as far as recall and precision rates were concerned. Raisa Varghese and Jayasree proposed different approach which combined the benefits of Senti-WordNet, dependency parsing, and co reference resolutions are well organized for the purpose of sentimental analysis. This was achieved through the use of SVM classifier.

Amit Gupte and Joshi (A Gupte, S Joshi, P Gadgul, A Kadam, 2014) compared between the more likely approaches used like Maximum Entropy, Naïve Bayes, Random Forest and Boosted trees algorithms When Random Forest algorithm is used in sentiment classification, it resulted in greater accuracy and better performance and was easy to understand. Its performance also improved with time and as a result of aggregation of decision trees, the accuracy was improved with a higher rate. The downside of this algorithm is that it required high processing power and training time. The researches therefore conclude that if accuracy is the first consideration as it is with the judicial sentiment analysis, then Random Forest classifier will be the most preferable though it consumes so much time during training, Naïve Bayes on the other hand utilizes lesser memory and less processors intensive. Also, Maximum Entropy requires shorter training period but a large memory and processor intensive. From these reviews, (P Dinkar Shinde, Dr. S Rathod, 2018) conclude that support vector machine yields higher accuracy in classification of product reviews, movie reviews and social media sentiments. It is to be noted that the authors have not dealt with sarcastic sentences and comparative sentences.

2.6. Conclusion

Narrowing down to sentiment analysis, the best algorithm used is support vector machine, this is in line with the previous discussion where the SVM came out as one of the best classifiers for sentiment analysis hence this research utilized the support vector machine in sentiment analysis on judicial procedures i.e. the sentiments from the public during the case prosecution to understand the expectation of the public on the outcome of cases compared to the actual rulings by the courts. This is important as it can guide the judiciary to evaluate its decisions on various public interest cases there by formulating policies to guide its procedures in line with the public's interest. Below is the SVM model illustrated graphically and mathematically

CHAPTER THREE: METHODOLOGY

This section describes in detail the process of training the model that will be used in the classification. It will start with describing what happens under feature selection, the evaluating the various feature selection methods. It goes on to propose a model that is then applied in the training of the model. Based on this model the study recounts the training process, the building of the prototype and eventual deployment of the model. It also covers the classification process of the model and presents the results of the training process.

3.1 Feature selection

Normally the first step as far as sentiment classification is concerned is extraction of text features. The features to be extracted are discussed below

3.1.1 Terms presence and frequency:

In terms of frequency feature selection, we have extraction of individual words and their frequency count, extraction of word n-grams and their frequency counts. A binary weighting is allocated to every time a word of interest is found, a zero is indicated if the word does not occur, also the relative importance of the words maybe be realized by use of frequency weights.

3.1.2 Parts of speech (POS):

The other feature selection method is getting the adjectives, adjectives are important as they best opinions expressed in the text

3.1.3 Opinion words and phrases:

In expression of opinions like good or bad happy or sad, opinion words will most be suitable, although it should be noted that this may not be sufficient as some texts will express opinions in sentences devoid of opinion words. An example would be it rained cats and dogs

3.1.4 Negations:

Another feature worth taking note of in feature selection is the use of negative words, this is important as it may alter the opinion orientation of words for example, not feeling healthy could in essence mean feeling sick

3.2 Feature selection methods

A number of feature selection methods exist, these methods are categorized into two mafor classes, and these are;

3.2.1 Lexicon-based methods:

Lexicon methods will normally depend on human annotation. This method will begin with a seed list of word; this list can then grow by searching the larger corpora for synonyms thus growing the lexicon. This method is in most cases faced with problems as observed by Whitelaw et al (Whitelaw Casey, Garg Navendu, Argamon Shlomo, 2005).

3.2.2 Statistical methods:

On the other hand, statistical methods are frequently used as they are fully automatic

Under statistical methods, the following are the common ones used in feature selection.

3.2.2.1 Point-wise Mutual Information (PMI)

Point-wise mutual information models mutual information between the features and the classes in which those features fall. The definition of mutual information is given by the ratio between the two values and is shown by the equation below:

$$M_i(w) = log\left(\frac{F_w . P_i(w)}{F(w) . P_i}\right) = \log\left(\frac{P_i(w)}{P_i}\right)$$

In the equation w correlates positively to i, w being the word and i the class. If $M_i(w) > 0$ then w greater than 0, then w correlates negatively to i when $M_i(w) < 0$ where w is the word and i is the class.

3.2.2.2 Chi-square (χ^2)

If we take n as the number of documents in the collection and take, $p_j(w)$ as the conditional probability of j with documents that have w, j being the class, and w words, then the global fraction of documents with i as P_j with F(w) being global fraction of documents containing w, then it follows that χ^2-statistic of w and j will be defined as

$$x_i^2 = n.F(w)^2.\frac{(p_i(w) - p_i)^2}{F(w).\big(1 - F(w)\big).p_i|(1 - p_i)}$$

PMI and Chi-square both are used in determining the correlation between terms and categories. χ^2 has better performance compared to PMI since it is a normalized value; hence their comparison across terms in the same category is much easier and optimized. This research used Chi-square in its feature selection

3.3 Bag of Words Model

The above feature selection methods treat the documents either as group of words (Bag of Words (BOWs)), or as a string which retains the sequence of words in the document. BOW is used more often because of its simplicity for the classification process. The most common feature selection step is the removal of stop-words and stemming (returning the word to its stem or root i.e. cries to cry). This was the model that was employed in this study

3.4 Data Preparation, Algorithm training and Evaluation

This study employed a framework proposed by (G Angiani, L Ferrari, T Fontanini, 2015) illustrated in the figure below

Model Training Model

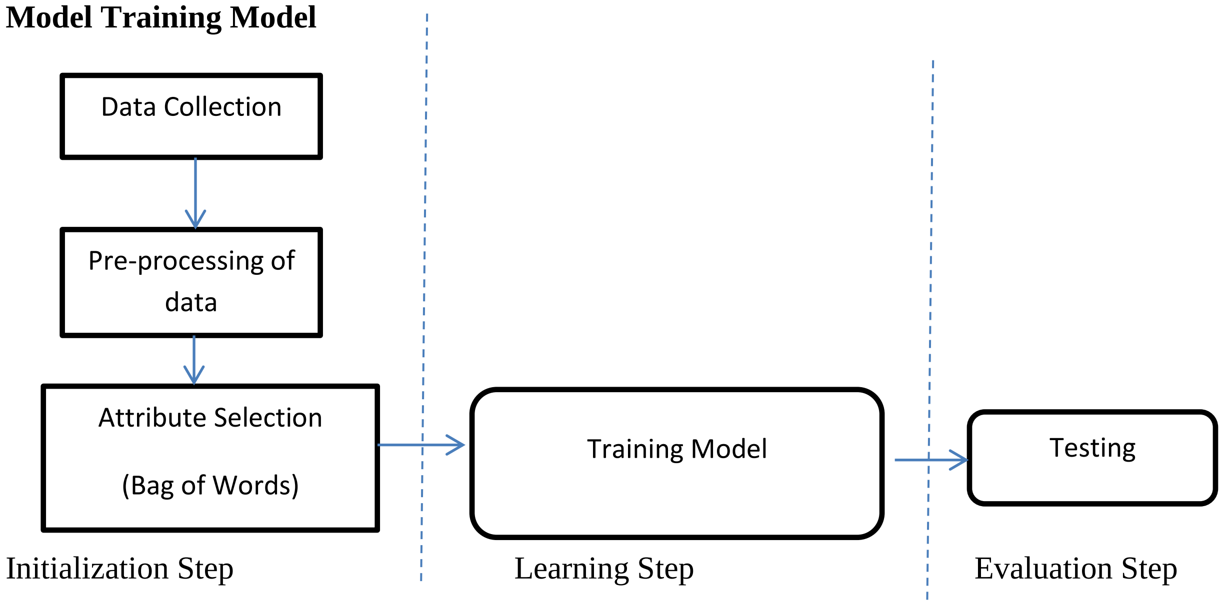


Figure 5 Model Training Model

Source: (G Angiani, L Ferrari, T Fontanini, 2015)

3.4.1 Data Collection

The training data sets were fetched from twitter, the chosen social network platform for this particular study. Since twitter restricts access to historical twits through the twitter API, the study depended on other online sources that harvest tweets and store them offline. To accomplish this task, the study deployed an open source tool, a python implementation GetOldTweets python master to harvest historical tweets that provided both the training set and the test set. The sentiments for the various tweets concerning the Supreme Court ruling of the presidential petition was analyzed and categorized either as agree with the ruling or don't agree with the ruling.

3.4.2 Preprocessing of Data

This section describes the preprocessing modules that have been used in this study. These modules have been implemented in python. The training sets were subjected to a number preprocessing techniques.

This first step involves the operations to clean the tweets harvested this involved removal of all elements that are not important, this is then forwarded for analysisand checking for words not spelled well during the process of normalization. Some of the elements removed at this level include URLs and hash-tags e.g. (i.e.#pissed) or mentions (i.e. @RaialBaba). Other elements that need to be removed at this level are line breaks and hashtags and this get replaced by a blank and quotation marks. This is useful in order to obtain a correct elaboration by our classifier in Weka this is so since not closing a quotation mark causes a wrong read by the data mining software causing a fatal error in the elaboration. In the next there is the removal of vowels that appear repeated over three times sequentially, this leads to those words' normalization, an example would be a word written as wroooooong instead of wrong such a word will be normalized to wrong. The second replacement will done on the words that express laughter, such words could be a string of the letters 'h' and 'a' and 'e' for example hahahhaaaaa or heehhhehe. These words are the substituted with the tag "laugh". The last procedure would the removal of or rather conversion of emoticons to words that express their emotion or sentiment, an example would be (i.e. :) →smile happy). The emoticons list in this study came from wikipedia5. Lastly all upper case text will be transformed to lower case letters and also there will be removal of any extra blank-spaces.

The expectation of this cleaning phase is to ensure that the text is uniform. This is to measure that only appropriate features are chosen at classification time will give an indicative frequency count.

After the text of each instance of a set has been preprocessed, the resulting sentences (the cleaned tweets) become the instances of a new training set. These cleaned tweets then are used as the data set for training the support vector machine classifier. This was accomplished via the online platform Weka.

3.4.2.1 Emoticon

This module reduces the number of emoticons to only two categories: smile_positive
and smile_negative, as shown in the figure below

List of substituted emoticons

Table 3: List of substituted emoticons

Positive emoticons	negative emoticons
:'-) :')	:'-(:'(
:-) :)	',:-|
:-] :]	',:-l
:-3 :3	<_<
:-> :>	>_>
:-> :>	:-< :<
8-) 8)	:-[:[

Source: (Researcher)

3.4.2.2 Negation

The knowledge of how negations can alter the classification results is an important step, since one instance of a negation word can change the tone of an entire sentence. The danger of ignoring the negations could mean miss-classification of a major part of the text under evaluation, thus it's important that negative constructs like won't, don't hasn't and never are properly replaced with not. This step will allow the model of the classifier to be much enriched enough negation bigram constructs. This constructs will have been excluded if the replacement will not be done thus leading to serious misclassification because of the low frequency of such negations

3.4.2.3 Correction of misspelled words

In order to correct misspelled words, there are a number of tools in existence that can help in the correction, this study utilized the PyEnchant6, a python library that provided functions that were very instrumental in the detection and subsequently correction of those words by the help of a dictionary, this tool also allowed for the substitution of the various slang words e.g. l8 to mean late and also the replacement of insults like the f-word with the tag of -a bad word. This is expected to greatly enhance the quality of the classified results of the tweets.

3.4.2.4 Stemming

Stemming techniques put word variations like "great", "greatly", "greatest", and "greater" all into one bucket, effectively decreasing entropy and increasing the relevance of the concept of "great". In other words, Stemming allows us to consider in the same way nouns, verbs and adverbs that have the same radix

Preprocessing of data as seen in Weka

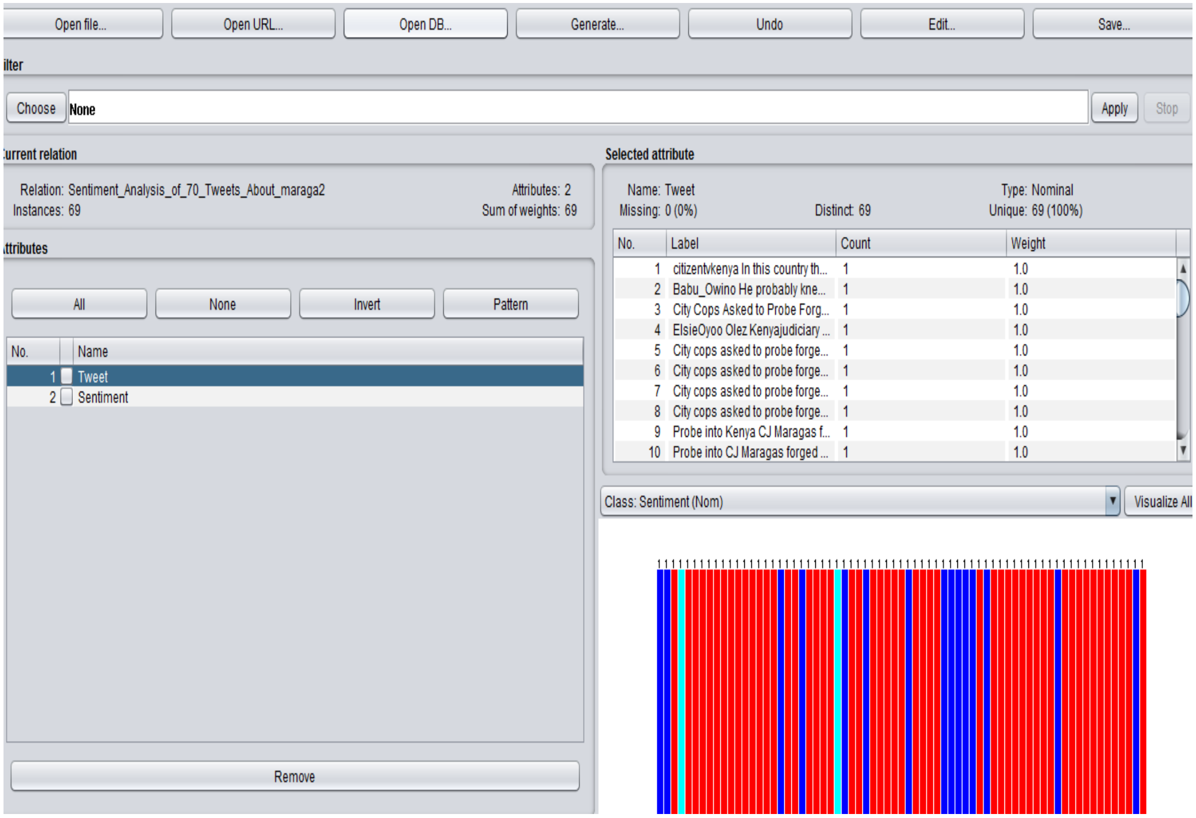


Figure 6:Pre-processing as seen in weka

Source: (Researcher)

3.4.2.5 Stop words

Stop words are words which are filtered out in the preprocessing step. These words are, for example, pronouns, articles, etc. It very important that such words are excluded from the model, this is so because they may compromise the accuracy of the results of the classification.

3.4.3 Attribute Selection using Bag of Words

A bag-of-words model, or BoW for short, is a way of extracting features from text for use in modeling, such as with machine learning algorithms. The approach is very simple and flexible, and can be used in a myriad of ways for extracting features from documents. A bag-of-words is a representation of text that describes the occurrence of words within a document. It involves two things, a vocabulary of known words, and measure of the presence of known words.

It is called a “bag” of words, because any information about the order or structure of words in the document is discarded. The model is only concerned with whether known words occur in the document, not where in the document.

3.5 Training Model

The training data was obtained using the twitter API using a python script. The data was captured into a CSV file. This file format is one of the file formats that WEKA utilizes when it comes machine learning.

The labeling of the training data was achieved through the use of the twitter API which classifies sentiments based on their polarity; this facility eliminated the bottleneck of having to label the data which would have been very time consuming

Using labelled tweets harvested from an active stream using the tweeter API, the SVM classifier was trained, the test was conducted using the split percentage, that is, 80% of the data was used for training and the remaining 20% used as test data, the results are summarized in the tables above

Classifier output

```
Time taken to build model: 0.19 seconds

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances          52               75.3623 %
Incorrectly Classified Instances        17               24.6377 %
Kappa statistic                          0
Mean absolute error                      0.2834
Root mean squared error                  0.3678
Relative absolute error                106.7644 %
Root relative squared error            102.3698 %
Total Number of Instances               69

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 0.000    0.000    ?          0.000    ?          ?        0.500     0.217     negative
                 1.000    1.000    0.754      1.000    0.860      ?        0.500     0.754     neutral
                 0.000    0.000    ?          0.000    ?          ?        0.500     0.029     positive
Weighted Avg.    0.754    0.754    ?          0.754    ?          ?        0.500     0.616

=== Confusion Matrix ===

  a  b  c   <-- classified as
  0 15  0 |  a = negative
  0 52  0 |  b = neutral
  0  2  0 |  c = positive
```

Figure 7 First results on Model Training

Source: (Researcher)

Of the total supplied data, 75.3623% instances were correctly classified while 24.6377% were incorrectly classified, below is a detailed accuracy table and the confusion matrix

```
=== Summary ===

Correctly Classified Instances          53               75.7143 %
Incorrectly Classified Instances        17               24.2857 %
Kappa statistic                          0
Mean absolute error                      0.2825
Root mean squared error                  0.3666
Total Number of Instances               70

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 1.000    1.000    0.757      1.000    0.862      ?        0.500     0.757     neutral
                 0.000    0.000    ?          0.000    ?          ?        0.500     0.214     negative
                 0.000    0.000    ?          0.000    ?          ?        0.500     0.029     positive
Weighted Avg.    0.757    0.757    ?          0.757    ?          ?        0.500     0.620

=== Confusion Matrix ===

  a  b  c   <-- classified as
 53  0  0 |  a = neutral
 15  0  0 |  b = negative
  2  0  0 |  c = positive
```

Figure 8: Second results on Model training

Source: (Researcher)

Below is the visualization of the results of the results obtained during the testing process, this results prove that a natural language classifier can effectively be used to gauge the judicial sentiments on ongoing cases.

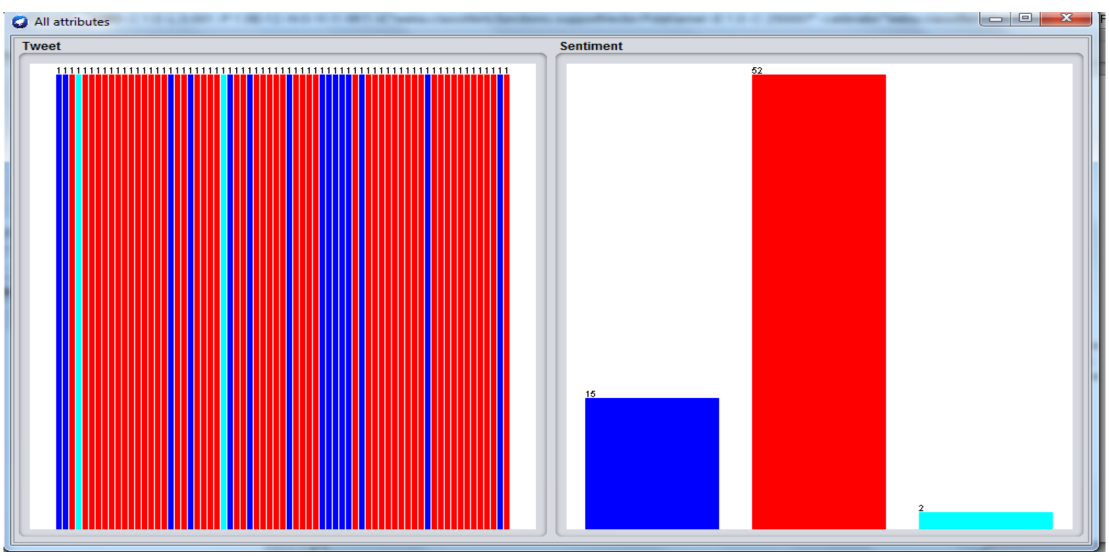


Figure 9: Graphical illustration of test results

Source: (Researcher)

With repeated training the model was able to improve its accuracy to 81.4815% of instances being correctly classified while only 18.52% of instances being incorrectly classified

```
=== Summary ===

Correctly Classified Instances          53               81.4815 %
Incorrectly Classified Instances        17               18.5185 %
Kappa statistic                          0
Mean absolute error                      0.2825
Root mean squared error                  0.3666
Total Number of Instances               70

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 1.000    1.000    0.757      1.000    0.862      ?        0.500     0.757     neutral
                 0.000    0.000    ?          0.000    ?          ?        0.500     0.214     negative
                 0.000    0.000    ?          0.000    ?          ?        0.500     0.029     positive
Weighted Avg.    0.757    0.757    ?          0.757    ?          ?        0.500     0.620

=== Confusion Matrix ===

  a  b  c   <-- classified as
 53  0  0 |  a = neutral
 15  0  0 |  b = negative
  2  0  0 |  c = positive
```

Figure 10 Third results on Model training

Source: (Researcher)

The SVM model was implemented and used to analyze sentiments of the public on cases that have been concluded using historical data (tweets) on those cases, both during and after the pronouncements on the various cases. The findings are discussed in the following chapter.

3.6 The Judicial Sentiment analysis prototype design

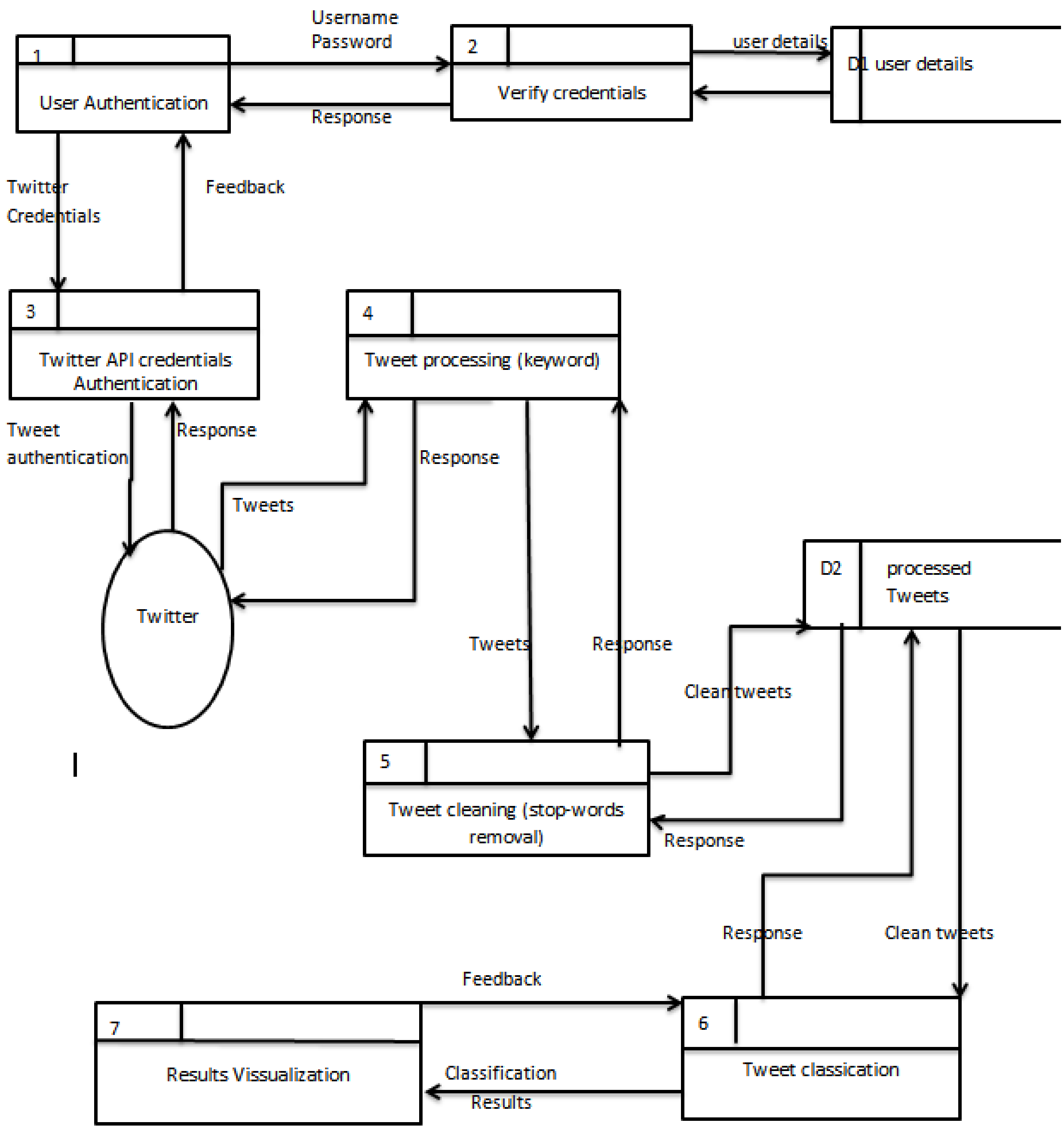


Figure 11: The Judicial Sentiment analysis prototype design

Source: (Researcher)

3.6.1 The Judicial Sentiment analysis prototype design Description

3.6.1.1 User authentication

Authenticates the user who can request to see the results, set different configuration for the twitter filter and provide the twitter API credentials which include the username, consumer key and secret key obtained from the twitter developer platform.

3.6.1.2 Twitter API authentication

Authenticates the twitter API authentication details (part of the twitter system). In case of an error, the user gets a response pointing out the printed error.

3.6.1.3 Tweet pre-processing

Once authentication has been successful, then tweets are harvested from twitter, represented as an external entity, those tweets are the pre-processed by subjecting them to a specific keyword filter i.e. a hashtag. The tweets are then committed to database and wait further processing.

3.6.1.4 Tweet cleaning and parsing

Removal of stop words, first step involves basic cleaning operations, which consist in removing None-important elements for the next phases of analysis and also the normalization of many misspelled words. In order to provide only significant information, in general a clean tweet should not contain URLs, hashtags (i.e.#pissed) or mentions (i.e. @RaialBaba). Furthermore, tabs and line breaks should be replaced with a blank and quotation marks with apexes. After this step, all the punctuation is removed, except for apexes, because they are part of grammar constructs such as the genitive.

3.6.1.5 Tweet classification

Using the trained model in the previous stage during model training, the clean tweets are then classified into three categories, positive, negative. For the research purposes the most important classes are the positive and negative and are the ones visualized in the results portal.

3.7 Judicial Sentiment Analysis Architectural Model

The algorithm can be implemented via prototype using the above architectural model, where by there is a tool, that is, a user interface to facilitate the visualization of the various analysis results of coming out of the Sentiment Analysis Engine

Sentiment analysis Engine was used to analyze the sentiments of the primary data, for this study twitter was used as the primary data source, this study utilized the twitter API, which gave developers access to the live tweets from the twitter platform.

3.7.1 Judicial Sentiment Analysis System Architectural Model

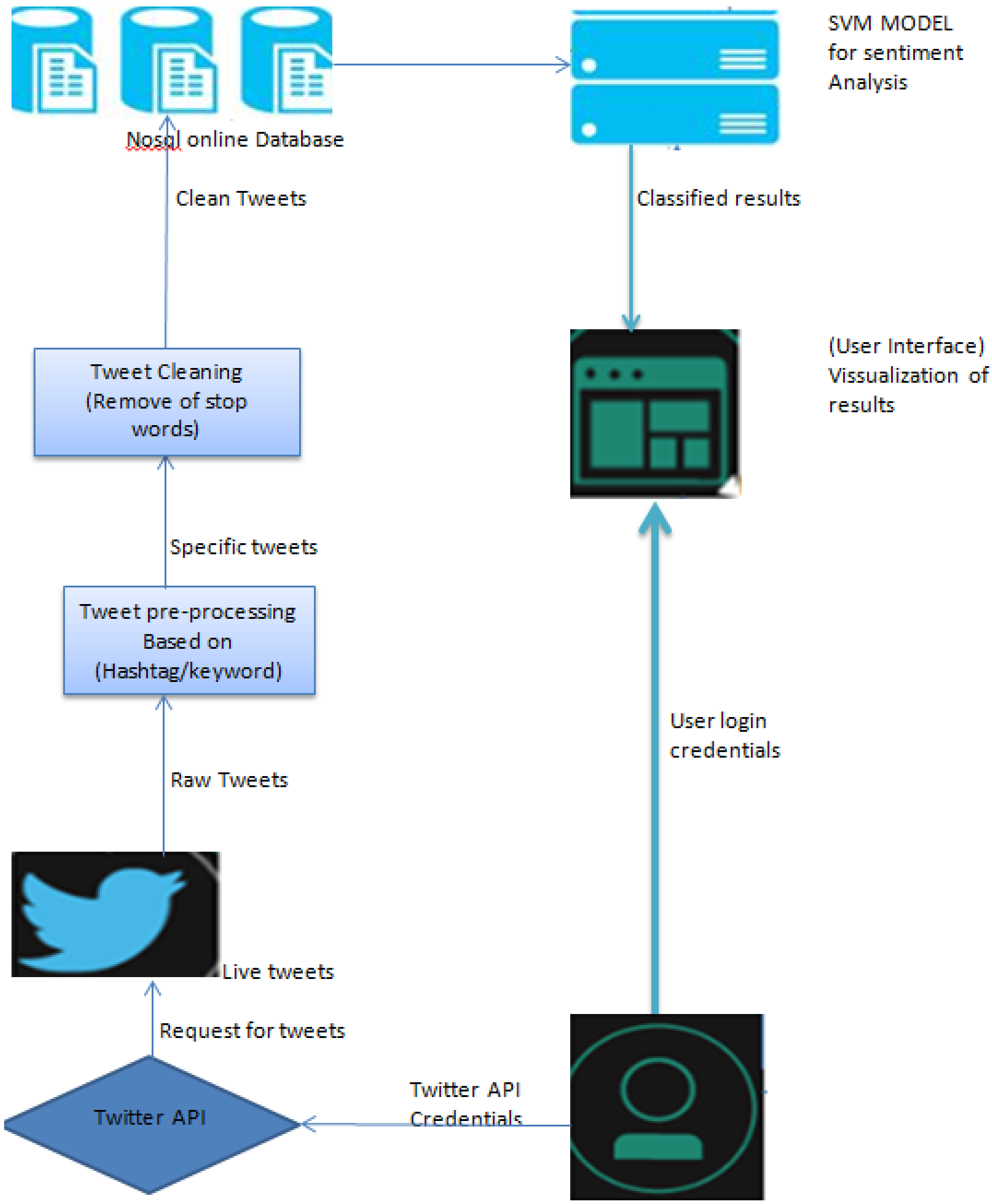


Figure 12: Judicial Sentiment Analysis System Architectural Model

Source: (Researcher)

3.7.2 Judicial Sentiment Analysis System Architectural Model Components

3.7.2.1 User Portal

User portal is used to login in order to input the necessary credentials for authentication, the credentials are twitter API credentials and for the user results visualization portal.

3.7.2.2 Twitter API

Though not part of the system developed, it is very important as it gives access to the twitter platform for the tweets required. It uses a keyword and a hashtag to filter the required tweets which are then processed further

3.7.2.3 Tweet cleaning and processing

The collected tweets are cleaned by removal of stop-words and other phrases that are not significant as far as the classification of the tweets is concerned, i.e. determining the class of each tweet, positive or negative.

3.7.2.4 NoSQL online database

The clean tweets are stored on the versatile online nosql database, offered as a service from the cloud, from here the tweets are then forwarded to the SVM model for classification.

3.7.2.5 SVM tweet classification

Here the trained model classifies the tweets based on their polarities and forwards the results to the user visualization

3.7.2.6 User visualization

This component represents the data analyzed from the trained model and the user can log in the analyze the results from the portal.

CHAPTER FOUR: SUMMARY, DISCUSSIONS, CONCLUSIONS AND RECOMMENDATIONS

4.1 Results discussion

Using historical data from twitter on the 2017 Kenyan judicial decision on the presidential petition, the sentiments during the progress of the case shows 54% that is 2546 of tweets as positive while 46% that is 2252 tweets as negative. For a process perceived to be by the public to be fair then results during the process should not greatly vary to those after the decision. This is not the case though from the figure below. The figure shows results after of tweets analyzed from the Elections petitions after the ruling in 2017, it shows that 2541 tweets were of negative sentiment on the ruling of the case, while only 1344 tweets were of positive sentiments, in percentage 65.4% of the sentiments were negative while 34.5% of the outcome was positive.

This shows that the public sentiment during the process is significantly shifted after the process, thus demonstrating that the public does disagree on the process. Whether the disagreement is positive or negative, that is to mean the public was not initially impressed with the process but are impressed with the outcome, or were impressed with the process but disappointed with the outcome, is up to the judiciary to re-evaluate its way of doing things if they understand the sentiments held by the public concerning their procedures, these then in return they can formulate better policies in-line with the new evidence on their procedures

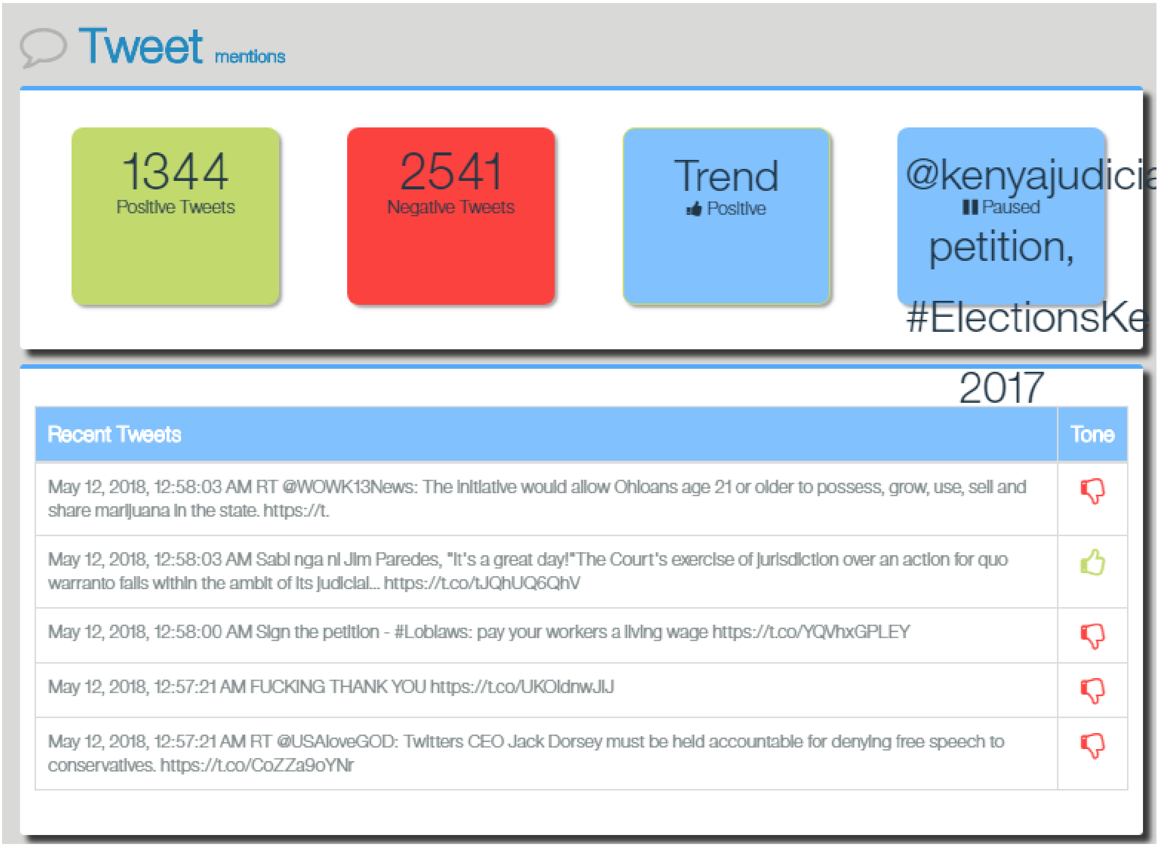


Figure 15 Sentiment Analysis

Source: (Researcher)

The same model can be effectively be applied on active ongoing cases to evaluate sentiment of a case prior to decisions and after the decision,

The following model can be used to evaluate the semantics of the comparative study from the sentimental analysis of public opinions on judicial cases in Kenya,

4.1.1 Summary of results analysis model

Table 4: results analysis model

Sentiments on the case during case progression	Sentiments on the case after Judicial decision	Expected Case outcome
NEGATIVE	POSITIVE	NO
POSITIVE	NEGATIVE	NO
POSITIVE	POSITIVE	YES
NEGATIVE	NEGATIVE	YES

The results of the presidential petition are summarized as below is the table below, there has been other comparative studies on other two of cases namely Wavinya Ndeti vs Alfred Mutua Gubernatorial Petition and Judith Wandera case of defiling 16 year old boy. The respective results are illustrated in the tables below

4.1.2 Study 1: 2017 General election petition

In the case of the general election petition, it's keen to note that there was contention after the ruling as opposed to the period during the handling of the petition. But this may be an opinion of the researcher in this context, to get a true picture of the perspective of the public's opinion on the matter; the table below summarizes the results. From the results it shows that 2546 of the sentiments or opinions analyzed represent a 54% of those in the support of the petition, after the conclusion of the case and the ruling being carried, the percentage drops to 34% of people still in support with the petition, this is a sharp dip in the numbers, an indication that the public's opinion has shifted on the matter.

Table 5 :2017 General election petition

Number of Tweets	**Sentiment before the ruling**		**Sentiment after the ruling**		**Expected Case outcome**
	Positive	**Negative**	**Positive**	**Negative**	**Expectation Met?**
Number	2546	2252	1344	2541	**NO**
Percentage	54%	46%	34.5%	65.4%	**NO**

The same results are here below extrapolated in a bar graph showing the sharp decline in the number of people supporting the petition in the y-axis. The first bar chart looks at the negative opinions before the judicial decision, where its clear that people in support of the petition was as high as 54%

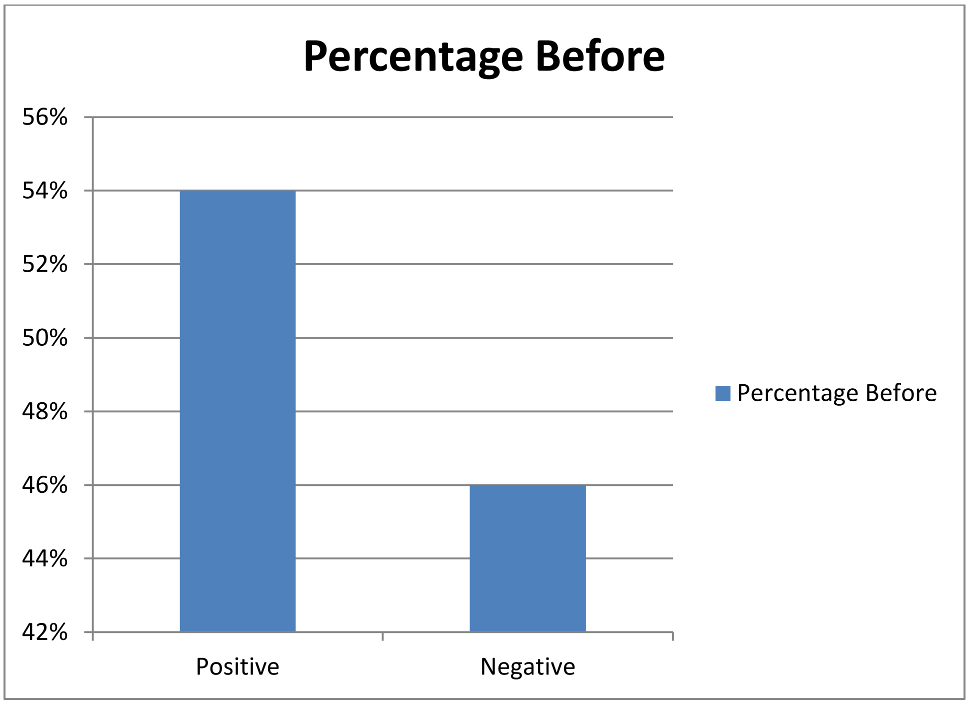


Figure 13: General election petition graph before ruling

The following bar chart below shows the opinions after the ruling on the petition, in comparison with the first above bar graph, the positive sentiments have dipped to about 34% while the negative sentiments have risen to about 65%, this shows people initially who had held the process in high regard are currently noting their apparent disapproval.

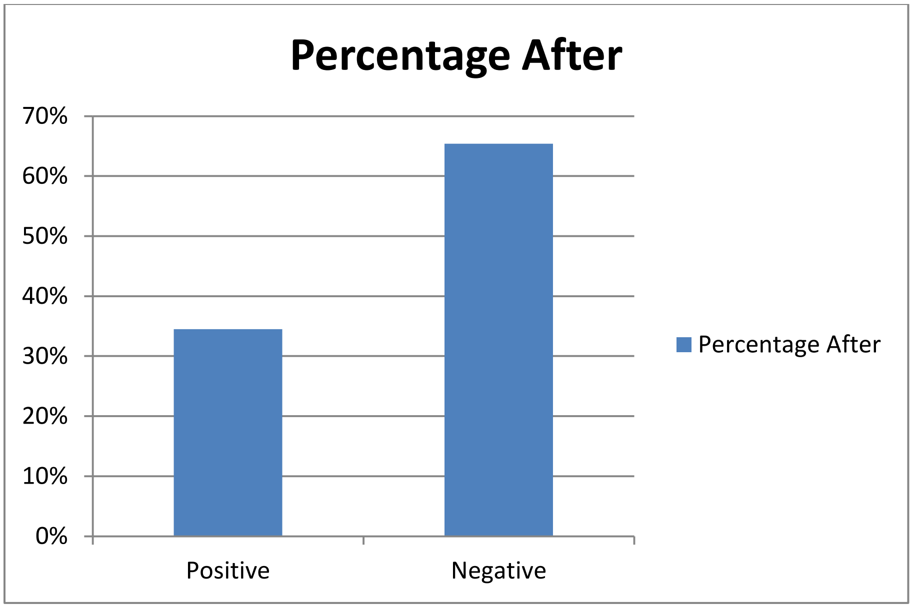


Figure 14: 2017 General election petition graph

Source: (Researcher)

From the graphical illustrations, it is apparent that the expectations of the public sharply varies from the initial perception, so that means the ruling goes contrary to the public's expectation

4.1.3 Study 2: 2018 Machakos Gubernatorial Petition

The second case study in the case of the Machakos gubernatorial petition, this case was selected as the stakes were high thus generating intense interest from the public on the ruling of the petition. It's a good case scenario that evaluates the public's expectation in line with the judicial ruling. As soon as the case was filed, the research collected the sentiments from twitter and analyzed the sentiments and opinions of the public concerning the case. A total number of 2362 sentiments analyzed, representing a 53% of the total sentiments evaluated. This shows still that a high number of people approved the resolution of the conflict within the corridors of justice. The only question is, were they satisfied with the resolution? According to the study, that number of people who approved the process drops to 52.5%, representing a significant disapproval of the final ruling by the court.

Table 6: 2018 Machakos Gubernatorial Petition

Number of Tweets	Sentiment before the ruling		Sentiment after the ruling		Expected Case outcome
	Positive	**Negative**	**Positive**	**Negative**	**Expectation Met?**
Number	**2362**	**2093**	**2487**	**2242**	**YES**
Percentage	**53%**	**46.98%**	**52.5%**	**47.4%**	**YES**

The bar graph below visually illustrates the results showing the significant dip in the number of people who supported the judicial process compared to those who were still in support at the time when the judicial decision was arrived. This points to some issues that affect how justice is dispensed and understanding how those decisions affect the public is paramount in achieving the necessary trust in the judicial processes.

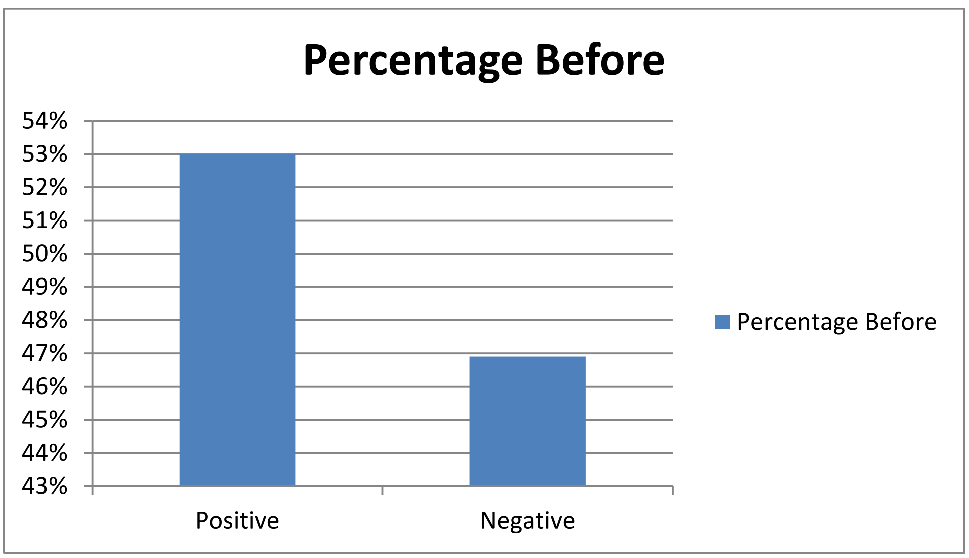


Figure 15: Machakos Gubernatorial Petition Graph before ruling

Source: (Researcher)

The following bar graph illustrates the number of positive and negative sentiments after the ruling, it's important to note that there was a drop in the number of people who were in approval of the process and a rise in the number of people who disagreed, a negative trend this was.

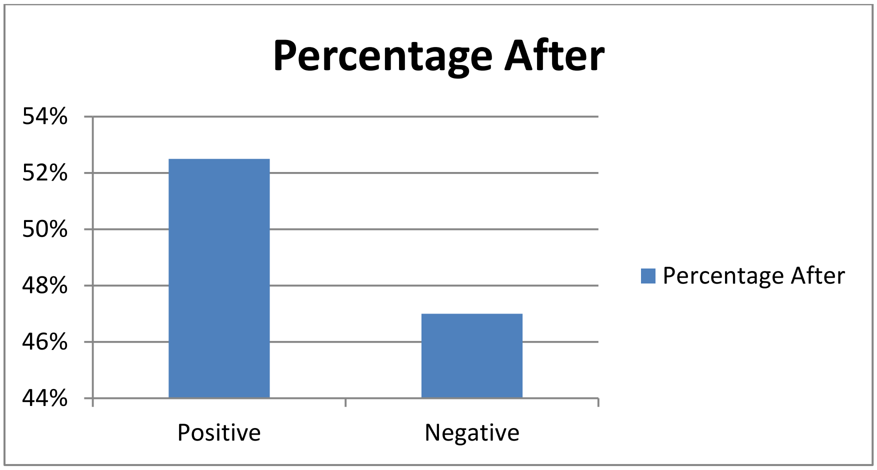


Figure 16 2018 Machakos Gubernatorial Petition Graph after ruling

Source: (Researcher)

This case on the other hand seems to adequately go by the public's expectation that's both during the case progression and after the case pronouncement.

4.1.4 Study 3: Judith Wandera case of defiling 16 year old boy

This case was selected from a different social dilemma equally affecting the publics as will election petition. The case generated a lot of interest from the public and thus providing a good chance to evaluate the sentiments of the public on the proceedings of the case and its final outcome. When the lady in the case was accused of defiling a 16 year old boy, there were questions concerning the circumstances surrounding the alleged rape, from the results the public jury had a different opinion. 54% of the sentiments analyzed showed public support for the resolution of the issue in the court. But their support changes when the ruling is made, the support goes from 54% to 43% denoting their disapproval of the ruling.

Table 7: Kisumu 16 year old boy defilement case

Number of Tweets	Sentiment before the ruling		Sentiment after the ruling		Expected Case outcome
	Positive	**Negative**	**Positive**	**Negative**	**Expectation Met**
Number	**2687**	**2288**	**2809**	**3644**	**NO**
Percentage	**54%**	**45.98%**	**43.53%**	**56.46%**	**NO**

Figure 17: Kisumu 16 year old boy defilement case Graph

The graphs represent that swift change in opinion. Given that in many jurisdictions, judicial systems constitute juries to decide on cases, its equivalent to this of the public's opinion as the jury's decision on the cases as opposed to one person's decision though schooled in law.

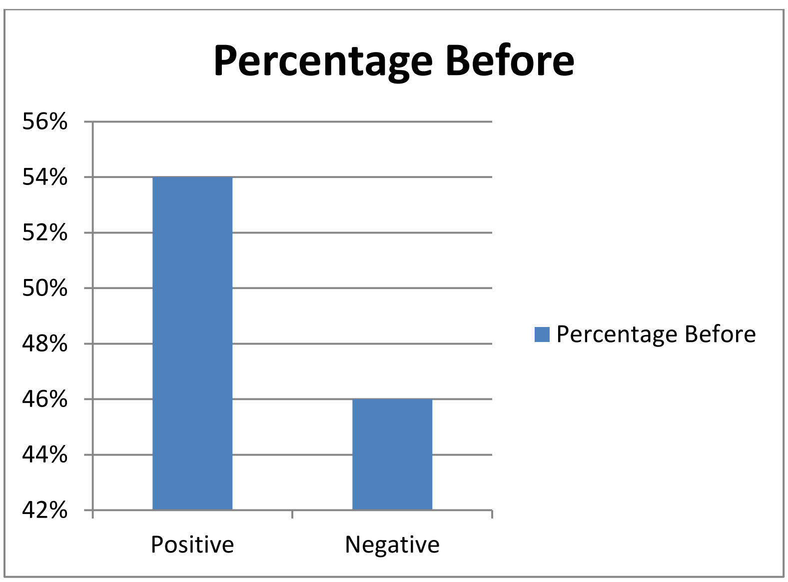


Figure 18: Kisumu 16 year old boy defilement case Graph before ruling

Source: (Researcher)

The second graph puts in perspective the sentiments of the public after the ruling, showing an increase in the number of those who disagree with the decision and a drop in number of those who still support the process bases on the results. This shows the trend of the public's take on most judicial decisions in the country. This is important in understanding the public's take on such decisions in order to improve service delivery.

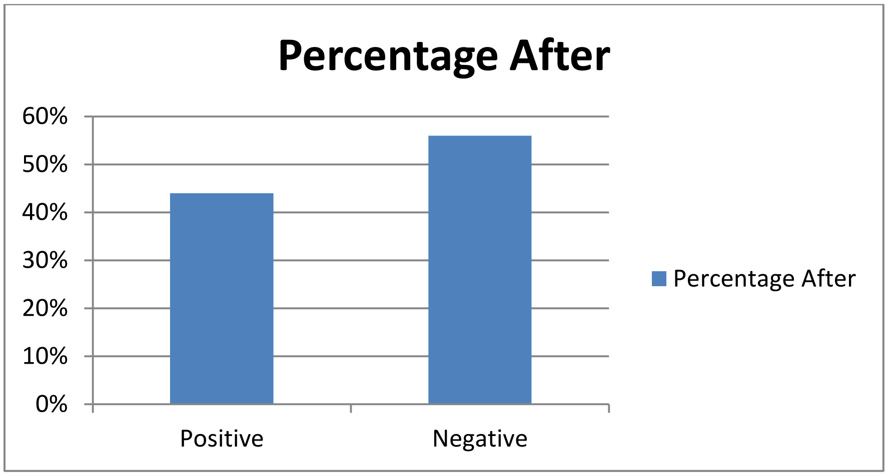


Figure 19 : Kisumu 16 year old boy defilement case Graph after ruling

Source: (Researcher)

This case in particular illustrates the disagreement of the case outcome with the public's expectation, and the disagreement seems to be very significant.

4.2 Conclusions

The model utilized in the analysis of the sentiments in this study was trained to a high degree of accuracy. Its accuracy stands at 81% hence it's able to classify the sentiments with a very minimal error rate. This model was applied in the problem of judicial sentiment classification and this leads as to draw very important conclusions.

In the first case study evaluated, the 2017 presidential election petition, the number of people who held positive sentiments during the filing and handling of the case stood at 54%, representing about 2546 sentiments of the total 3885 sentiments analyzed. This affirms the public support of the amicable resolution of the election petition through a judicial process. After the presidential petition ruling, the numbers significantly shift, the number of people supporting the ruling drops from 54% to 34.5%. This denotes a strong disagreement from the public towards the judicial decision. Given that opinions are mostly subjective in nature, this should not overshadow the overriding importance; that of understanding that justice should not just be done but seen to be done, a key aspect in achieving public confidence in the judicial processes. This should eventually culminate into citizens abiding with judicial decisions whether it's in their favor or not. Another case scenario where the public seems to disagree with the judicial decision is that of Judith Wandera case of defiling a 15 year old boy. From sentiments analyzed, the trend shows a 54% agreement with the judicial resolution of the matter, and the processes therein. Like the first case study, the sentiments move from predominantly positive to negative, that is from 54% to 43.53%. This ultimately indicates a strong disagreement of the public with the judicial decision.

One last case study though seems to vindicate the judiciary, a scenario where the judicial decision agrees with the public expectation. In the case of Machakos Gubernatorial petition, of the sentiments analyzed during case progression, 53% appear to agree with the process, and after the ruling, 52.5% still agree with the ruling. This implies that the judiciary in some cases gets the public's expectation, but this research offers a chance for it to re-evaluate in cases where its decisions seems to go completely contrary to the public expectations.

The important part is to note very significant deviations of the public's expectations and the judicial decisions; this may in turn serve as a good pointer into whether the public conceive judicial processes as serving justice or serving other interests. This in turn should guide the judiciary in aligning its procedures and formulating policies intended to further instill confidence in judicial processes.

This study has to some extent relied on historical data due to time constrains, though a real time mechanism that harvests current tweets going on the subject at hand, i.e. the case of interest. Twitter API allows only access to seven days old tweets; the rest of the tweets access was not possible via twitter API. The study therefore hugely depended on secondary sources for the historical tweets other twitter, the primary source. In the platform, though a real-time harvest of tweets is built such that the tweets are analyzed in real time. The rest of the two cases use real time twitter data harvested pre-processed and classified in real time.

4.3 Recommendations

This study utilized twitter as its primary source, as much as it is very rich as a platform for mining public opinions, the future work may consider incorporating other platforms such as Facebook and other popular social platforms in order to have better coverage of the public's sentiments on cases of interest.

In selecting the best model for sentiment analysis, this research heavily relied on existing literature due to time limitation; a better approach would be to carry out performance evaluations for each model considered before settling on the most appropriate model.

4.3.1 Contribution to body of Knowledge

It is apparent from the findings in this research that sentiment analysis as a field of study can find application in various domains. This study has demonstrated application of text classification in judicial domains. Text classification can also find application in other aspects of research where the public's interest is of benefit to the entities concerned. A number of models exist. After reviewing various models that are most suitable for task of text classification, established that support vector machines perform better to the task of text classification. Though it was noted in the research that support vector machines perform better, they are very resource intensive. It is best that their application should be done only when using smaller data sets unless there is enough computing power during the training period. The application of sentiment classification in the judicial domain in this study provides one important lesson; that public institutions in service of the public ought to be concerned with the opinions of the public in the various decisions they make and the services they provide.

4.3.2 Future Work

This research utilized most cases that had generated a huge interest within the public; these cases were mostly election petitions and one criminal case (sexual abuse). This was due to time limitations and also the available computing power. Availability of data on the cases was also a limitation since twitter only allows access to seven days into the past of its twitter data. Future research can evaluate a large spectrum of cases cutting across different spheres of life including public litigations and those cases that don't garner much of public attention, these cases still can be used to point out whether judicial ruling are in-line with expectations of the litigants. Future work can utilize a bigger dataset in the training of the model, although the accuracy of the model utilized in this study was 81%, classification accuracy can be made better if the dataset was large. Future studies should consider using distributed or parallel computing architecture to boost its computing power during the pre-processing stage to overcome the resource intensive nature of support vector machine model building task.

Bibliography

A. Go, R.Bhayani, L.Huang, 2009. *Twitter sentiment classification using distant Supervision*. Stanford University.

A.Abbas, H. Chen, A. Salem, 2008. Sentiment Analysis in multiple languages: Feature selection for opinion classification in web forums. *In ACM Transactions on information Systems*, 26(3), pp.1-34.

Amit Gupte, Sourabh Joshi, Pratik Gadgul, Akshay Kadam, 2014. Comparative Study of Classification Algorithms used in Sentiment Analysis. *International Journal of Computer Science and Information Technologies*, 5.

Chuanming Yu, 2009. *"Mining Product Features from Free-Text Customer Reviews: An SVM-based Approach*. Nanjing.

Giulio Angiani, Laura Ferrari, Tomaso Fontanini, 2015. *A Comparison between Preprocessing Techniques for Sentiment Analysis in Twitter*. Parma: Dipartimento di Ingegneria dell'Informazione.

Hatzivassiloglou V, McKeown K, 1997. Predicting the semantic orientation of adjectives. *Proceedings of annual meeting of the Association for Computational Linguistics*.

Li Young-Ming, Li Tsung-Ying, 2013. Deriving market intelligence. In *Decis Support Systems*. pp.8-12.

M.Vohra, Prof. B. Teraiya, 2015. A comparative study of sentiment analysis Techniques. *JOURNAL OF INFORMATION, KNOWLEDGE AND RESEARCH IN COMPUTER ENGINEERING*, 02(02), pp.2-5.

Moraes Rodrigo, Valiati Joao Fransisco, 2013. Document-level sentiment classification: an empirical comparison. In *Expert Syst Appl*. pp.621 - 633.

Mustafa, A.M., 2017. An Evaluation of Sentiment Analysis and Classification. *International Journal of Computer Applications* , 158(3), p.33.

Nathaniel Japkowicz, Mohak Shah, 2013. *Evaluating learning algorithms, A classification Perspective*. 1st ed. New York,: Cambridge University Press.

Pang, L. Lee, 2008. Opinion mining and sentiment analysis. *Foundations and trends in information Retrieval*, 2(2), pp.1-135.

Pooja Dinkar Shinde, Dr. Sunil Rathod, 2018. A Comparative Study of Sentiment Analysis Techniques. *International Journal of Innovations & Advancement in Computer Science*, 7(3), pp.3-7.

R. Parikh, M. Movassate, 2013. *Sentiment Analysis of UserGenerated Twitter Updates using Various Classification Techniques*. India: Finolex Academy of Management and Technology.

Raisa Varghese, Jayasree M, 2010. Aspect Based Sentiment Analysis using Support Vector Machine Classifier. *IEEE*.

Rich Caruana, Alexandru Niculescu-Mizil, 2015. An Empirical Comparison of Supervised Learning Algorithms. p.6.

Sayad, D.S., 2017. Real Time Data Mining. *ResearchGate*, 1(1), pp.24-56.

Vryniotis, V., 2015. *http://blog.datumbox.com*. [Online] (1) Available at: http://blog.datumbox.com/machine-learning-tutorial-the-max-entropy-text-classifier/ [Accessed 03 April 2018].

Walaa Medhat, Ahmed Hassan, Hoda Korashy, 2015. Sentiment analysis algorithms and applications. *Ains Shams Engineering Journal*, (1), pp.6-21.

Whitelaw Casey, Garg Navendu, Argamon Shlomo, 2005. Using appraisal groups for sentiment analysis. In *Proceedings of the ACM SIGIR Conference on Information and Knowledge*., 2005.

A. Go, R.Bhayani, L.Huang, 2009. *Twitter sentiment classification using distant Supervision,* s.l.: s.n.

A.Abbas, H. Chen, A. Salem, 2008. Sentiment Analysis in multiple languages: Feature selection for opinion classification in web forums. *In ACM Transactions on information Systems*, 26(3), pp. 1-34.

Giulio Angiani, Laura Ferrari, Tomaso Fontanini, 2015. *A Comparison between Preprocessing Techniques for Sentiment Analysis in Twitter,* Parma: Dipartimento di Ingegneria dell'Informazione.

Hatzivassiloglou V, McKeown K, 1997. Predicting the semantic orientation of adjectives. *Proceedings of annual meeting of the Association for Computational Linguistics*.

Li Young-Ming, Li Tsung-Ying, 2013. Deriving market intelligence. In: *Decis Support Systems.* s.l.:s.n., pp. 8-12.

M.Vohra, Prof. B. Teraiya, 2015. A comparative study of sentiment analysis Techniques. *JOURNAL OF INFORMATION, KNOWLEDGE AND RESEARCH IN COMPUTER ENGINEERING*, 02(02), pp. 2-5.

Moraes Rodrigo, Valiati Joao Fransisco, 2013. Document-level sentiment classification: an empirical comparison. In: *Expert Syst Appl.* s.l.:s.n., pp. 621 - 633.

Mustafa, A. M., 2017. An Evaluation of Sentiment Analysis and Classification. *International Journal of Computer Applications* , 158(3), p. 33.

Nathaniel Japkowicz, Mohak Shah, 2013. *Evaluating learning algorithms, A classification Perspective.* 1st ed. New York,: Cambridge University Press.

Pang, L. Lee, 2008. Opinion mining and sentiment analysis. *Foundations and trends in information Retrieval,* 2(2), pp. 1-135.

R. Parikh, M. Movassate, 2013. *Sentiment Analysis of UserGenerated Twitter Updates using Various Classification Techniques,* India: s.n.

Rich Caruana, Alexandru Niculescu-Mizil, 2015. An Empirical Comparison of Supervised Learning Algorithms. p. 6.

Sayad, D. S., 2017. Real Time Data Mining. *ResearchGate,* 1(1), pp. 24-56.

Vryniotis,V.,2015.*http://blog.datumbox.com.*[Online]Availableat:
http://blog.datumbox.com/machine-learning-tutorial-the-max-entropy-text-classifier/
[Accessed 03 April 2018].

Walaa Medhat, Ahmed Hassan, Hoda Korashy, 2015. Sentiment analysis algorithms and applications. *Ains Shams Engineering Journal,* Issue 1, pp. 6-21.

Whitelaw Casey, Garg Navendu, Argamon Shlomo, 2005. *Using appraisal groups for sentiment analysis.* s.l., s.n., pp. 625-631.

Liu Bing, Hsu Wynne, Ma Yiming. Integrating classification and association rule mining. In: Presented at the ACM KDD conference; 1998.

Ko Youngjoong, Seo Jungyun. Automatic text categorization by unsupervised learning. In: Proceedings of COLING-00, the 18th international conference on computational linguistics; 2000.

Turney P. Thumbs up or thumbs down? : semantic orientation applied to unsupervised classification of reviews. In: Proceedings of annual meeting of the Association for Computational Linguistics (ACL'02); 2002.

Read J, Carroll J. Weakly supervised techniques for domain in dependent sentiment classification. In: Proceeding of the 1st international CIKM workshop on topic-sentiment ana

lysis for mass opinion; 2009. p. 45–52.

Somasundaran S, Wiebe J. Recognizing stances in online debates. In: Proceedings of the joint conference of the 47th annual meeting of the ACL and the 4th international joint conference on natural language processing of the AFNLP; 2009. p. 226–34.

Hu Minging, Liu Bing. Mining and summarizing customer reviews. In: Proceedings of ACM SIGKDD international conference on Knowledge Discovery and Data Mining (KDD'04);2004.

Kim S, Hovy E. Determining the sentiment of opinions. In: Proceedings of international conference on Computational Linguistics (COLING'04); 2004.