RUFORUM Working Document Series (ISSN 1607-9345), 2018, No. 17 (3): 736-749
Available from http://repository.ruforum.org

Research Application Summary

Towards the creation of an ensemble model for sentiment analysis based on naïve bayes and support vector machine for product review classification: A Literature Survey

Ogutu, R.V., Otieno, C., Rimiru, R. & Ochieng, P.

Jomo Kenyatta University of Agriculture and Technology, CBD Mombasa Campus, Institute of Computer Science and Information Technology, P.O.Box 81310-80100, Mombasa, Kenya
Jomo Kenyatta University of Agriculture and Technology, Juja Main Campus, Institute of Computer science and information Technology, P.O.Box 81310-80100, Mombasa, Kenya
Corresponding author: achiengrh@gmail.com

Abstract

Sentiment analysis has demonstrated that the automation and computational recognition of sentiments is possible and evolving, due to factors such as; emergence of new technological trends and the continued dynamic state of the human language. Sentiment analysis is therefore an Information extraction task that aims at obtaining private sentiments that can either be expressed as positive or negative, toward a specific object or subject. However, social media platforms are marred with informal texts that make extraction and parsing of relevant information a problem for most systems and models. This can pose a challenge to companies, individuals or organizations that need to make specific business decisions based on the available data. To overcome such inefficiencies, this research proposes an ensemble model on the basis of performance evaluation on sentiment classification of product reviews. The research will explore the use of a detailed pre-processing technique with the integration two classifiers, Naïve Bayes and SVM as an ensemble. The effect (in terms of performance measure and evaluation) of such a computational model, and how the model can be implemented within machine learning approaches to sentiment analysis, has formed grounds for this research.

Key words: Feature extraction, machine learning, opinion, social listening, supervised learning

Résumé

L'analyse des sentiments a démontré que l'automatisation et la reconnaissance informatique des sentiments sont possibles et évolutives, en raison de facteurs tels que l'émergence de nouvelles tendances technologiques et l'état dynamique continu du langage humain. L'analyse des sentiments est donc une tâche d'extraction d'informations qui vise à obtenir des sentiments privés qui peuvent être positifs ou négatifs, envers un objet ou un sujet spécifique. Cependant, les plateformes sociales de médias sont entachées de textes informels qui font de l'extraction et de l'analyse des informations pertinentes, un problème pour la plupart des systèmes et modèles. Cela peut poser un défi aux entreprises, aux particuliers ou aux organisations qui doivent prendre des décisions commerciales spécifiques en fonction des données disponibles. Pour pallier ces inefficacités, cette recherche propose un modèle d'ensemble sur la base d'une évaluation des performances sur la classification des sentiments des avis produits. La recherche explorera l'utilisation d'une technique de prétraitement détaillée avec l'intégration de deux classificateurs, Naïve Bayes et SVM en tant qu'ensemble. L'effet (en termes de mesure et d'évaluation des performances) d'un tel modèle de calcul, et la manière dont le modèle peut être mis en œuvre dans les approches d'apprentissage automatique de l'analyse des sentiments, a servi de fondement à cette recherche.

Mots clés: extraction de fonctionnalités, apprentissage automatique, opinion, écoute sociale, apprentissage supervisé

Introduction

The rise and popularity of informal language and social media platforms, especially Twitter has made Sentiment analysis of tweets an important area of research (Vosoughi *et al.,* 2015), while at the same time given web users a venue for expressing and sharing their thoughts, opinions or sentiments on all kind of topics and objects. As expressed by Saif *et al.* (2012), Twitter has millions of users worldwide that constantly tweet, making it a gold mine for communities, organizations as well as individuals to monitor their reputation and brands by extracting and analyzing the sentiment of tweets posted by the public about them, their market or competitors. These monitoring processes can be referred to as Social Listening. According to Yang *et al.* (2014) these online social interactions can be used to reveal a 'digital footprint' of individuals or groups of individuals' behavior or a community dynamics, by mining the digital traces left by users while interacting with cyber-physical space, such as microblogs, for profitable reasons. This has led to the emergence of research on Social and Community Intelligence that present opportunities to compile these digital footprints into a comprehensive picture of individuals' daily life facets, transform the understanding of our needs, organizations and societies; while also enabling innovative improvement on products, public safety, resource management and environmental monitoring.

Such information submitted to the online services are a form of data sources that can be used in Sentiment analysis. Models in Sentiment analysis over twitter data and other microblogs faces several new challenges due to the short length and irregular structure of textual data (high dimensions) . These may include informal or colloquial content and use of various languages, voluminous data, among other challenges (Kiritchenko *et al.,* 2014; Chandni *et al.,* 2015; Awachate and Kshirsagar, 2016; Shirbhate and Deshmukh, 2016). As such, a problem of focusing on the most relevant information from the voluminously complex data, during Sentiment Analysis, may arise. In addition, relevant feature extraction is significant for Sentiment classification as the opinionated texts may have high dimensions, which can affect classifier performance (Sharma Dey, 2012).

The aim of this research is to develop an Ensemble Machine learning model with a bag-of-word approach, for accurate classification purposes so as to achieve and evaluate a better predictive model, as the research problem originates as a combination of a Machine Learning and Information Retrieval challenge. By combining these fields, a problem of extraction of relevant Features in Sentiment analysis exists.

Theoretical literature review

As indicated by Lo *et al.* (2016), in the field of Sentiment analysis, an increased attention is now focused on analysis of social media content, that is, the user-generated contents. This is to facilitate the understanding of the different social aspects and confidence level measure of products; or the perceived image of a company or its product. These social media contents also, often contain informal or mixed linguistic languages (colloquial) that are textual in nature. While, it is no longer sufficient to consider only standard/formal languages such as English in Sentiment analysis processes, these short informal textual messages bring in new challenges to Sentiment Analysis. They are limited in length, tend to have many misspellings, slang terms, shortened form of words or non-standard expressions (e.g. 'gr8' instead of using the word 'great'). They also have special markers such as hashtags and other characters that express special meaning (Agarwal *et al.,* 2011; Kiritchenko *et al.,* 2014; Awachate and Kshirsagar, 2016). These can be referred to as high dimensions in textual data (Gaikwad *et al.,* 2014).

For this reason, in the field of Artificial Intelligence, Machine learning approaches have been widely applied for the automation of Sentiment analysis to provide computers the ability to learn without being explicitly programmed, and also improve efficiency for classifiers and models. According to Gaikwad *et al.* (2014) Sentiment analysis is an Information Extraction task that aims to obtain a writer's feelings expressed in positive or negative comments by analyzing a large number of documents. It is therefore the computational technique for extracting, classifying, understanding and determining opinions expressed in various contents. Sentiment analysis attempts to identify a sentiment held towards an object and helps in the automation of extraction or classification of sentiment from unstructured text. Further, in their research, the researcher describe Sentiment analysis as an aims to determine the state of mind of a speaker/ writer with respect to some topic or the overall tonality of a document.

As Routray *et al.* (2013) clearly puts it in their research, there are challenges in sentiment analysis such as subjectivity classification, word sentiment classification, document sentiment classification and opinion extraction. These challenges can be solved through various computational approaches for sentiment analysis such as Linguistic approaches and Machine learning approaches. Linguistic approach relies on disambiguation using background information such as a set of rules and vocabularies. Thus, such a system normally contains lexicons, which consist of words and their polarity values (positive/ negative, bad/good etc). There are also a set of rules that help produce more accurate results as an integral part of such a system. The machine learning approaches however, are used for automatic sentiment classification and are approved by many researchers as the efficient way to analyze sentiment laden term in a document. The researchers also recommend that improving the quality of the system is an area for future work.

With this regard, feature selection and extraction has been exhibited as an important area geared towards the improvement of quality and efficiency in Sentiment analysis by many researchers, and therefore is the basis of this study. According to Akba *et al.* (2014) employing various feature reduction and extraction techniques decreases the running time of learning while increasing success rate of algorithms. Consequently, to reach an optimal performance level, and improve efficiency of classifiers during Sentiment analysis, it is advisable to include important features in prediction and extraction of Sentiment information. These important features can be referred to as relevant features.

Agarwal *et al.* (2011) carried out a research that contributed to the introduction of POS-specific prior polarity features. They therefore build models for two classification tasks: a binary task of classifying sentiments into positive and negative classes, and a 3-way task of classifying sentiments into positive, negative and neutral classes. They experimented with three types of models, unigram, feature based and a tree kernel based model. On the feature based model, they proposed new features from some of the features proposed in past literature and carried out an extensive analysis of 100 new proposed features. Information gain is used as the attribute evaluation metric for feature selection. Their experiments showed that features that have to do with Twitter specific features such as emoticons and hashtags, only add marginal value to the classifier, while, features that combine prior polarity of words with their part-of-speech tags are most important for the classification tasks. Combining unigrams with Senti-features outperformed the combination of kernels with Senti-features by 0.78%, and turned out to be the best performing system for the positive versus negative task. Important aspects in Sentiment analysis such as topic modeling was also not considered in this paper, therefore, the researchers advocate for further research on these aspects.

Abbasi *et al.* (2008) proposed a Sentiment analysis methodology for classification of Web forum opinions in multiple languages (hate/ extremist group forum postings in English and Arabic languages). They evaluated utility of stylistic and syntactic features and integrated specific feature extraction components to account for linguistic characteristics of Arabic. A hybrid genetic algorithm that incorporates information-gain heuristics for feature selection and extraction, Entropy Weighted Genetic Algorithm (EWGA), was developed to improve performance and get a better assessment of key features from a benchmark movie review datasets and Web forum postings. Experimental results using EWGA with SVM (Support Vector Machine) indicated high performance levels, with accuracies of over 91%. Stylistic features significantly enhanced performance across all datasets, while EWGA also outperformed other feature selection methods. Their identified research gaps were work on Sentiment analysis on Web forums in multiple languages, research on usage of stylistic features and feature reduction/ selection technique.

Sentiment Analysis. Sentiment Analysis refer to the application on Natural language processing, Computational linguistics, Text analytics to identify and extract subjective information from source materials (Hovy, 2015) such as internet text like documents, Product reviews, Tweets and other social media materials (Tejwani, 2014). Sentiment Analysis can be used interchangeably with Opinion Mining, and involves the process of computational treatment of opinions, sentiments, and subjectivity in texts (Pang and Lee, 2008). Its basis is to ascertain the attitude of a writer or speaker, in reference to certain topic or specific targeted object/ entity or the overall contextual polarity of a document (Hovy, 2015). This ‘attitude’ that changes and changes again, reflects the speaker’s or writer’s appraisal, judgment, opinion, or evaluation; his/her affective state (personal feeling at the time of writing) and intended emotional communication (effect to the reader).

Sentiments on the other hand are emotions, judgments, opinions or ideas (Boiy *et al.,* 2007). Gaikwad *et al.* (2014) also states in their research that a Sentiment is basically a thought or view based on emotion instead of reason. It is a kind of subjective impression and not fact, and can be termed as the expression of sensitive feeling in art and literature.

Human beings are subjective creatures by nature (Tejwani, 2014). Our decisions are based on or influenced by personal feelings, stimuli, tastes or opinions. According to Kim and Hovy (2004), as much as sentiments involve the Holder's emotions or desires, people however, express their sentiments in complex ways. These can be in terms of explicit or implicit expressions. For example; *'I think that buying an iphone for grandma is quite unreasonable'* is an implicit sentiment, while, *'Buying an iphone for grandma is a waste!'* is an explicit sentiment. Explicit aspects are aspects explicitly mentioned as nouns or noun phrases in a sentence, while, implicit aspects are those aspects not explicitly mentioned in a sentiment or sentence but are implied.

Approaches used for informal or mixed linguistic languages' sentiment analysis. According to Lo *et al.* (2016) there are two main approaches that are fundamentally used in for Informal or mixed linguistic language Sentiment analysis, these are Subjectivity and Polarity detection.

Subjectivity detection. Subjectivity detection involves procedures that help in understanding if contents contain personal views and opinions in 'private state'. Subjective expressions may arise from cultures or various experiences of an individual or community, and hence can be localized and specific to individuals or society as a whole. According to Kaya *et al.* (2012), most previous works focus on Sentiment analysis of highly subjective texts, such as product or movie reviews.

Polarity detection. Polarity detection is about studying subjective expressions in terms of different polarities, intensities or rankings. These categories can also be referred to as classes, such as 'positive', 'negative' and 'neutral' or 'joy', 'sadness', 'anger' and 'fear' (Rashid *et al.,* 2013).

Aspects in Sentiment Analysis

Text Mining. According to Vijayarani and Janani (2016) Text Mining (also referred to as Text Data mining) is a domain in Data Mining that is used to extract interesting information, knowledge or pattern from both unstructured and semi-structured data in databases. It is a process that involves processes such as document gathering, pre-processing, text transformation, attribute selection, pattern selection and finally interpretation and evaluation. Data mining however, is the process of extracting hidden predictive information from databases and transforming it into meaningful and understandable formats for future use.
Text mining is used to analyze large quantities of Natural Language text (these include text in both Standardized languages and Informal languages), or used in Computational Linguistic, and it is assumed universally by researchers that writers or speakers have some affective value (sentiment) to entities (Hovy, 2015). Text mining helps detect lexical patterns that are used for the ultimate extraction of unseen data or useful information (Vijayarani and Janani, 2016). This unseen data is useful to organizations and individuals in various Social listening ways, for instance, Business Analytics, trend predictions among others.

Opinionated Information. Textual information can be categorized into Factual and Opinionated information (Schrauwen, 2010). Factual information (facts) are objective expressions that describe entities, events and their properties, while, Opinionated information (opinions) are commonly subjective expressions that describe individual's sentiment, opinions or feelings toward entities, events and their properties. These kind of information are mostly available on user-generated content platforms such as Social networks, internet forums, discussion groups and blogs (Liu, 2010).

Sentiments and opinions can be described in terms of: Topic, Holder, Claim and Sentiment (Kim and Hovy, 2004). The Holder believes a Claim about a Topic, and associates a Sentiment, for example 'good' or 'bad', with his/her belief. Generally, human beings love to express their sentiments/opinions and emotions over entities or objects in both their physical and online interactions. For example, 'this lipstick shade is better than that one' or 'I hate this food!' As a result, it is evident that there is a great deal of importance to opinions especially in decision making. Sometimes, we need to hear other people's opinion in order to make decisions (Liu, 2010). Large volumes of data are therefore collected from the World Wide Web, while User-generated content forums are also provided to collect such opinions and sentiments (an online word of mouth) which are later presented to the world. To help navigate through these large volumes of data and address challenges that are posed during analysis and mining of the said data, processes such as Sentiment Analysis and Opinion Mining have risen to provide automatic and semi-automatic methods for generalization and interpretation of Sentiments/Opinions in form of texts.

Social Networking

Social Networking is a social infrastructure technology that allows microblogging. It is the grouping of individuals into specific groups such as cliques. These groups can be influenced on grounds based on friendship, political, religious, business or peer preferences; all together sharing information of their interest online. Social Media however, is the content that a user uploads, share, like and comment on. This could be a status update, post, images, blog, video, podcast, newsletter or links to content (Burriss, 2016).

Social Networking (for example, Instagram, Facebook and Twitter) is a platform that is characterized by ease of access and is mostly free of charge to users. For this reason, as well as the increased dependence on technology globally, individuals and organizations indulge in these forms of interactions to make new friends, stay in touch with loved ones, give feedback/ opinions on product/ service and have fun while also do business. Consequently, a lot of opinionated information are generated from such interactions that are collectively referred to as User-generated content. These are textual information that can be inform of known standard languages e.g. English or Informal and mixed linguistic languages.

Microblogging. According to Ehrlich and Shami (2010) Microblogging is a form of online collaboration that has become increasingly popular with time. It is a web service that allows subscribers to relay short or 'micro' textual messages, links, videos and images to other users of the service. These contents can be sent from either a computer or mobile device. 'Blog' is a short for 'web blog' which provides a platform where users write about their feelings and activities in form of sentiments/ opinions, so other people can read them. Blogs are therefore generally website that are periodically updated by authors and writers about and around some common theme or topic of interest. Commercial microblogs exist to promote service and/or products, websites and collaboration within an organization or externally with consumers. Microblogging tools include Twitter, Facebook, Instagram among others.

The major appeal of microblogging is for both its immediacy and profitability (ease of use and accessibility). Blogs or posts are typically brief and short, and can be relayed with various computing devices including phones. According to Hornak (2009) with microblogging, one can get quick tidbits of information to many or hundreds of potential user/ followers in seconds. The following benefits can be achieved by individuals or companies;
-Build brand awareness
-Grow a business network
-Give the client base important announcements
-Give and get feedback
-Quickly syndicate content.

Machine Learning. According to Mitchell (1998), Machine Learning is a field concerned with the question of how to construct computer programs that automatically improve with experience, that is, Machine Learning explores the study and construction of algorithms that can learn from and make predictions on data. Precisely put, a computer program is said to learn from experience 'E' with respect to some class of tasks 'T' and performance measure 'P', if its performance at tasks in 'T', as measured by 'P', improves with experience 'E'. The purpose of Machine Learning is to learn from training data so as to make predictions on new or unseen data. These algorithms operate by building specific models (as demonstrated in Figure 1) from example inputs in order to make data driven predictions and decisions, rather than follow a strict static program instruction.

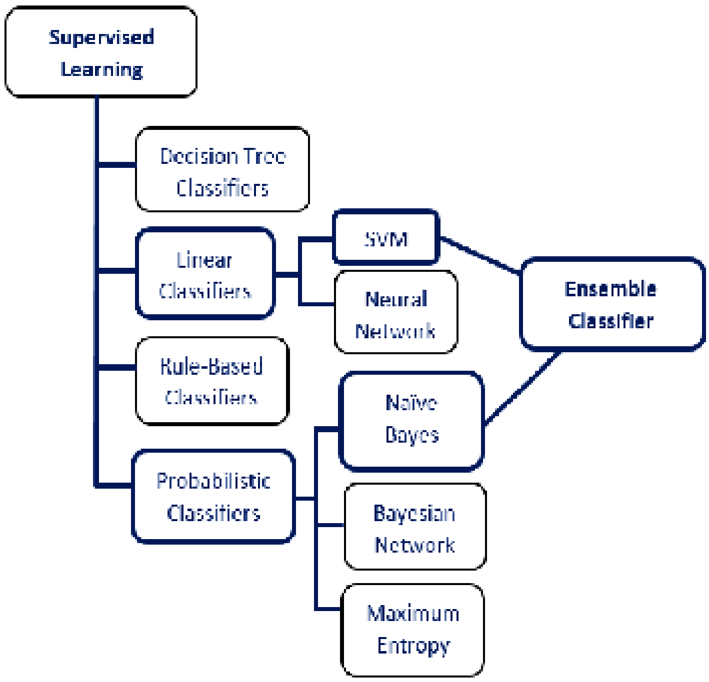

Figure 1. Supervised machine learning

Proposed Classifiers. This research will be regarded as a qualitative research. According to Kumar (2011) a qualitative research usually involves studying perceptions, beliefs or feelings and a researcher does not make any attempt to establish uniformity in them across respondents and hence measurements and variables do not carry much significance.

The use of algorithms is usually dependent on the type of problem to be studied. In addition, a researcher can look at a number of dimensions to give a sense of what will be a reasonable algorithm or classifier. Some of these could include;
-Does the researcher expect the problem to be linearly separable?
-Are features independent?
-The number of training examples to be used in the study
-Dimensionality of the feature space
-Is over fitting expected to be a problem?
-What are the system's requirement in terms of speed/performance/memory usage...?

The question of identifying a reasonable set of algorithm will therefore be addressed by following the Occam's Razor principle: 'use the least complicated algorithm that can address your needs and only go for something more complicated if strictly necessary'. The proposed ensemble will therefore involve the use of SVM (Support Vector Machine) and Naïve Bayes, which are Linear and Probabilistic classifiers respectively (Figure 2.4). The significance and weakness of the algorithms are discussed below.

Naïve Bayes. Most language processing tasks are Classification tasks. According to Khan *et al.* (2016) Naïve Bayes algorithm is extensively used for text classification and also widely used in solving classification problems such as text categorization and is based on Bayes theorem. This study focuses on the problem of text categorization, the task of classifying an entire text by assigning it a label drawn from some set of labels.

Support Vector Machine. Support Vector Machine (SVM) is a supervised machine learning algorithm which is also widely used for classification problems, even though it can be employed for both classification and regression purposes. It is a non-probabilistic binary linear classifier that has the ability to linearly separate classes by a large margin. According to Balahur and Perea-Ortega (2015) in their research on Sentiment analysis system adaptation for multilingual processing, the case of tweets, SVM has been proven to be highly effective in traditional text categorization and has been applied successfully in many opinion mining tasks, performing better than other machine learning techniques. Process wise SVM is based on the idea of finding a hyper plane that best divide a dataset into two classes, that is, an optimal separating hyper plane, and a hyper plane, as far as possible from data points from each category.

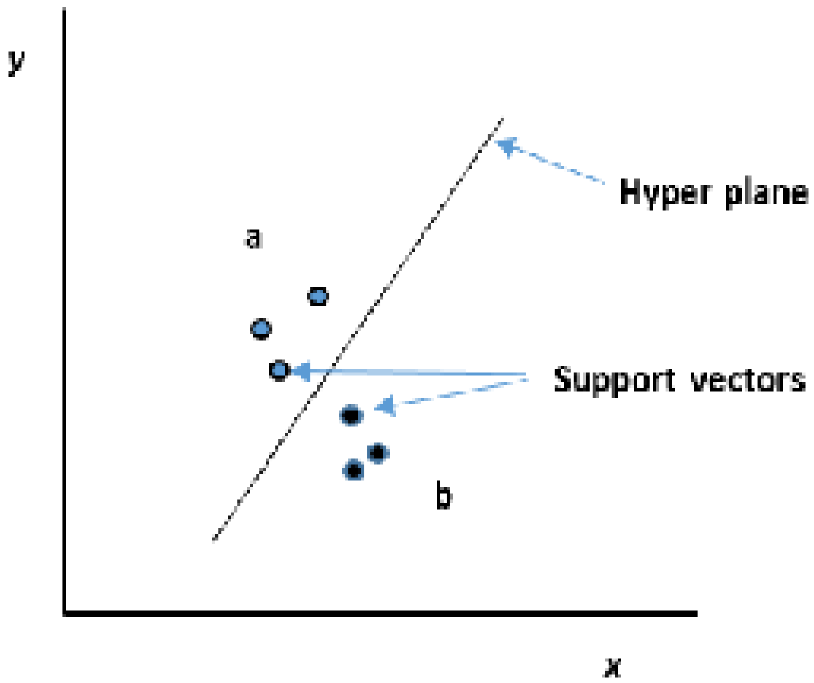

Figure 2. Given a particular data point (y and x) a classification a or b is made

Ensemble Learning and Approach. These are Machine Learning methods that construct a set of predictive models and combine their outputs into a single prediction. Basically, according to Džeroski *et al.* 2009), the purpose of learning is typically to achieve better predictive performance, as it has been exhibited in various works such as Dietterich, (2000) and Zhu *et al.,* 2016) that ensembles can be more accurate than single models. Dietterich (2000) gave three fundamental reasons for why ensemble methods are able to outperform any single classifier within the ensemble in terms of statistical, computational and representational issues. The concept is actually quite simple: train several models from the same data set, or from samples of the same data set, and combine the output predictions, typically through a stacking ensemble approach. This will be done after a systematic pre-processing approach that will be used to eliminate noise from the dataset.

Currently, ensemble methods represent a standard Machine Learning method which has to be considered whenever a good predictive model is demanded (Džeroski *et al.,* 2009). This study proposes to train the two models singularly from the same data set by partitioning the data set into a training set and a test set, and finally creating a stacked ensemble.

Related works

According to Vilares *et al.* (2017) research on Supervised Sentiment analysis in multilingual environments, automatically understanding all the information shared on the Web and transforming it into knowledge is one of the main challenges in this age of Big Data. In terms of NLP (Natural Processing Language), this usually involves comprehending different human languages, which are implicitly related with relevant human aspects such as cultures, countries or regions.

Kataria and Shah (2015) research acknowledged that considering modern writing styles (like misspelled words, abbreviations, concatenated words and emoticons) can increase the accuracy of sentiment analysis. In addition, they state that, people tend to use different words for a particular feature of a product. Thus identifying frequent nouns and noun phrases automatically will classify more number of reviews, while also helping in identifying any new feature of the product that is being talked about.

Most focus currently in sentiment analysis is on online reviews. People express their opinion on social media which consist of product review sites, social networks or blogs (examples include Facebook, Amazon, Twitter, Flicker, LinkedIn etc.) Information from these sources are very helpful for the customers to make purchase decisions. According to researchers such as Bhaskar *et al.* (2014), Sentiment analysis has become a new knowledge resource after the onset of the Internet and the World Wide Web. Its main purpose is to automatically predict the sentiment polarity of users' opinions on the web. Opinions or preferably, sentiments, play an important role in the understanding of collective sentiments and help to make better decisions. The sentiments can be positive, negative or even neutral. Positive sentiments encourage the prospective customer to make positive decisions toward a product; negative sentiments may usually result to negative decisions. None the less, Sentiment analysis of textual communication extracts subjective information in the text.

Abirami *et al*. (2016) proposed an approach to sentiment analysis by addressing three major issues in analyzing social media content; computing the intensity of a polarity, identifying the more effective sentiment in such cases where various sentiments are mentioned, context dependency of opinion words and deliberate spelling errors. The researchers acknowledge that there is an increase in interest in developing improved opinion mining algorithms for accuracy, and developing a more efficient understanding of the dynamics of the human sentiment. Their task was a classification ranking problem with a goal of producing mathematical score for the sentiments observed. Using Twitter, the idea took into consideration every aspect of the text and used that information to compute final score. The aspects included social media tendencies such as use of emoticons, punctuations, casing, vowel stretches, abbreviations, interjections and the context in which the opinions are mentioned. Evaluation of the algorithm was then done by comparing the result to a baseline produced by manual scoring.

Vilares *et al.* (2017) tackled the problem of performing multilingual polarity classification on Twitter, comparing three techniques: (1) a multilingual model trained on a multilingual dataset, obtained by fusing existing monolingual resources, that did not need any language recognition step, (2) a dual monolingual model with perfect language detection on monolingual texts, and (3) a monolingual model that acted based on the decision provided by a language identification tool. The techniques were evaluated on monolingual, synthetic multilingual and code-switching corpora of English and Spanish tweets, their goal being to compare the performance of supervised monolingual models based on bag-of-words, with respect to their corresponding multilingual version (i.e., a model that is a collection of weights from English and Spanish features). To do this, they relied on standard sets of features. The aim was to show how current state of the art supervised approaches can successfully address (or not) situations where monolingual, multilingual and code-switching texts appear. They also relied on an L2-regularized logistic regression.

Feature selection and Extraction can be used interchangeably in Feature engineering. Sabri and Saad (2016) acknowledged that Feature engineering is a very important task in the domain of sentiment analysis and generally in text categorization, and converting original documents to feature vectors is critical. The study proposed a supervised learning approach for sentiment analysis in Arabic language, while empirical study was done to evaluate various Feature Selection Methods (Information gain, Chi-square and Gini Index) and also showcases that selection of the right feature set determines the overall performance of classifiers. Three classification approaches (association rule mining, n-gram model and meta-classifier approach) are implemented. It was proven from their results that the use of Feature selection methods especially as a combination, can enhance and increase performance of sentiment classification models in comparison to results obtained using original classifiers.

Saraee and Bagheri (2013) researched and proposed a model, using n-gram features, stemming and feature selection to overcome some Persian language challenges in Sentiment classification. The researchers acknowledged, according to their findings, that feature selection in Sentiment analysis can improve classifier performance. The proposed Modified version of Mutual Information (MMI) method considers all possible combinations of co-occurrences of a feature and class label, and is concluded to improve performance. However, it was also proven through experiments that other feature selection approaches such as Document Frequency (DF), Mutual Information (MI) and Term Frequency Variance (TFV) does not measure the co-occurrences of other features and classes.

Duwairi and Qarqaz (2014) researched on Sentiment analysis in Arabic reviews from a machine learning perspective, and employed three classifiers, Naïve Bayes, SVM and K-Nearest neighbor in a parallel experimentations to detect polarity of reviews. These classifiers were run on the dataset of tweets and comments that were collected from Twitter and Facebook. The tweets and comments addressed general topics such as education, sports and political news. Precision and recall parameters were used as measures of accuracy and evaluation for each classifier.

Conclusion

To overcome problems such as classification inefficiencies and overfitting in models, Feature Extraction techniques or Single Models have been used to address such issues by means of reducing the number of features in consideration for a model. This has been evidenced by research done by Akba *et al.* (2014), Prusa *et al.* (2015) and Savita and Gore (2016). It is clear that some researchers, however, do not explore the possibility of creating an Ensemble so as to improve performance. It is evident that every classification model or technique has its own benefit and drawbacks. Selection of a classification model is normally decided on the basis of resources available, accuracy requirement, training time available among other factors (Gupte *et al.,* 2014), for this reason no model can either be referred to as superior or inferior to another.

This study will to create and use Ensemble Machine Learning approach, that is, a combination of Support Vector Machine (SVM) and Naïve Bayes for Classification of Sentiments, and evaluate the results. As Zhou (2012) described in his book, Ensemble methods, also referred to as Multiple classifier systems, try to construct a set of learners or algorithms and combine them for generalization or, combine their outputs into a single prediction. According to Zhou (2012)the generalization of ensembles is often much stronger than the use of single learners or algorithms, as ensembles are able to boost weak learners and are unlikely to over fit and improve performance, among other various advantages. According to Džeroski *et al.* (2009), the purpose of learning is typically to achieve better predictive performance as ensembles can be typically more accurate than single learners. Our aim is to evaluate the performance of the proposed classifiers for Sentiment Classification in terms of robustness, precision and recall.

Reference

Abbasi, A., Chen, H. and Salem, A. 2008. Sentiment analysis in multiple languages: Feature selection for opinion classification in Web forums. *ACM Trans. Inform. Syst.* 26 (3): 34 http://doi.acm.org/10.1145/1361684.1361685

Abirami, M. S., Uma, M. and Prakash, M. 2016. Sentiment analysis of informal text using a Rule based Model. *Journal of Chemical and Pharmaceutical Sciences (JCPS)* 9 (4): 2854 – 2858.

Agarwal, A., Xie, B., Vovsha, I., Rambow, O. and Passonneau, R. 2011. Sentiment Analysis of Twitter Data. pp.1-9. In: Proceedings of the Workshop on Languages in Social Media, LSM'11.

Akba, F., Uçan, A., Sezer, E.A. and Sever, H. 2014. Assessment of Feature Selection Metrics For Sentiment Analyses: Turkish Movie Reviews

Awachate, P. B. and Kshirsagar, V. P. 2016. Improved twitter sentiment analysis using n-gram feature selection and combinations. *International Journal of Advanced Research in Computer and Communication Engineering* 5 (9): 154-157.

Balahur, A. and Perea-Ortega, J.M. 2015. Sentiment analysis system adaptation for multilingual processing: The case of tweets. *Information Processing and Management* 51 : 547–556

Bhaskar, J., Sruthi, K. and Nedungadi, P. 2014. Enhanced sentiment analysis of informal textual communication in social media by considering objective words and intensifiers. IEEE International Conference on Recent Advances and Innovations in Engineering (ICRAIE-2014), May 09-11, 2014, Jaipur, India.

Boiy, E., Hens, P., Deschacht, K. and Moens, M. F. 2007. Automatic sentiment analysis of on-line text. pp. 349–360. In: Chan, L. and Martens, B. (Eds.), Openness in digital publishing: Awareness, discovery and access. Vienna: IRIS-ISIS.

Chandni, C., N., Gupta, S. and Pahade, R. 2015. Sentiment analysis and its challenges. *International Journal of Engineering Research and Technology (IJERT)* 4 (3): 968-970.

Chris Hornak, C. 2009. Micro-blogging for businesses: Benefits and tips. Retrieved from http://eyeflow.com/micro-blogging-for-businesses-benefits-and-tips/

Dietterich, T.G. 2000. Ensemble methods in machine learning. pp. 1-15. In: International workshop on multiple classifier systems. Springer, Berlin, Heidelberg.

Duwairi, R. M. and Qarqaz, I. 2014. Arabic sentiment analysis using supervised classification. The 1st International Workshop on Social Networks Analysis, Management and Security (SNAMS - 2014), Barcelona, Spain.

Džeroski, S., Panov, P. and Ženko, B. 2009. Machine learning ensemble methods. pp. 5317-5325. In: Encyclopedia of Complexity and Systems Science, Springer.

Ehrlich, K. and Shami, N.S. 2010. Microblogging inside and outside the workplace. Proceedings of the Fourth International AAAI Conference on Weblogs and Social Media

Gaikwad, S.V., Chaugule, A. and Patil, P. 2014. Text mining methods and techniques. *International Journal of Computer Applications* 85 (17): 42-45.

Hovy, E. H. 2015. What are sentiment, affect, and emotion? Applying the methodology of Michael Zock to sentiment analysis. Paper presented at Springer International Publishing, Switzerland.

Kataria, M. and Shah, S. M. 2015. Extended comprehensive sentiment analysis for informal opinion Text. *International Journal of Engineering Research and Technology (IJERT)* 4: 481-483.

Kaya, M., Fidan, G. and Toroslu, I. H. 2012. Sentiment analysis of Turkish Political News. 2012 IEEE/WIC/ACM International Conferences on Web Intelligence and Intelligent Agent Technology.

Khan, M. T., Durrani, M., Ali, A., Inayat, I., Khalid, S. and Khan, K. H. 2016. Sentiment analysis and the complex natural language. Complex Adaptive Systems Modeling.

Kim, S.-M. and Hovy, E. 2004. Determining the Sentiment of Opinions. pp. 1-8. In: Proceedings of the COLING conference, Geneva.

Kiritchenko, S., Zhu, X. and Mohammand, S. M. 2014. Sentiment analysis of short informal texts. *Journal of Artificial Intelligence Research* 50: 723-762.

Kumar, R. 2011. Research Methodology. 3rd Edition: A step-by-step guide for beginners. SAGE Publications Ltd, London.

Liu, B. 2010. Sentiment Analysis and Subjectivity. Handbook of Natural Language Processing, Second Edition.

Lo, S.L., Cambria, E., Chiong, R. and Cornforth, D. 2017. Multilingual sentiment analysis: from formal to informal and scarce resource languages. *Artificial Intelligence Review* 48 (4): 499-527.

Mitchell, T. M. 1998. Machine Learning. McGraw-Hill Science/Engineering/Math. 432pp.

Pang, B. and Lee, L. L. 2008. Opinion mining and sentiment analysis. *Foundations and Trends in Information Retrieval* 2 (1-2): 1-135.

Rashid, A., Anwer, N., Iqbal, M. and Sher, M. 2013. A survey paper: areas, techniques and challenges of opinion mining. *International Journal of Computer Science Issues (IJCSI)* 10 (6): 18-31.

Routray, P., Swain, C. K. and Mishra, S. P. 2013. A Survey on Sentiment Analysis. *International Journal of Computer Applications* 76 (10): 0975 – 8887.

Sabri, B. and Saad, S. 2016. Arabic sentiment analysis with optimal combination of features selection and machine. *Research Journal of Applied Sciences, Engineering and Technology* 13 (5): 386-393.

Saif, H., He, Y. and Alani, H. 2012. Semantic sentiment analysis of twitter. pp. 508-524. In: International semantic web conference. Springer, Berlin, Heidelberg.

Saraee, M. and Bagheri, A. 2013, June. Feature selection methods in Persian sentiment analysis. pp. 303-308. In: International Conference on Application of Natural Language to Information Systems. Springer, Berlin, Heidelberg..

Schrauwen, S. 2010. Machine Learning Approaches to Sentiment Analysis using the Dutch Netlog Corpus. Computational Linguistics and Psycholinguistics Research Center, pp.30-34.

Sharma, A. and Dey, S., 2012. Performance investigation of feature selection methods and sentiment lexicons for sentiment analysis. *IJCA Special Issue on Advanced Computing and Communication Technologies for HPC Applications* 3: 15-20.

Shirbhate, A. G. and Deshmukh, S. N. 2016. Feature extraction for Sentiment classification on Twitter Data. *International Journal of Science and Research (IJSR)* 5 (2): 2183 – 2189.

Teddy, B. 2016. What is the difference between social media and social networking? [Msg 2]. Retrieved from https://www.quora.com/What-is-the-difference-between-social-media-and-social-networking#MoreAnswers.

Tejwani, R. 2014. Sentiment Analysis: A Survey. arXiv preprint arXiv:1405.2584.

Vijayarani, S. and Janani, R. 2016. Text Mining: Open Source Tokenization Tool – An Analysis. *Advanced Computational Intelligence: An International Journal (ACII)* 3 (1): 37-47.

Vilares, D., Alonso, M.A. and Gomez-Rodriguez, G. 2017. Supervised sentiment analysis in multilingual environments. *Information Processing and Management* 53 (3): 595-607.

Vosoughi, S., Zhou, H. and Roy, D. 2015. Enhanced twitter sentiment classification using contextual information. arXiv preprint arXiv:1605.05195.

Yang, D., Zhang, D., Yu, Z. and Zeghlache, D. 2014. Mining user digital footprints for fine-grained preference-aware social media search. *ACM Transactions on Internet Technology (TOIT)* 14 (4): 1-24.

Zhou, Z. 2012. Ensemble Methods. Foundation and Algorithms. Chapman & Hall/ CRC Machine Learning and Pattern Recognition on Series.

Zhu, Y., Xie, C., Wang, G.J. and Yan, X.G. 2017. Comparison of individual, ensemble and integrated ensemble machine learning methods to predict China's SME credit risk in supply chain finance. *Neural Computing and Applications* 28 (1): 41-50.