



EXPLORING MULTINOMIAL NAÏVE BAYES FOR YORÙBÁ TEXT DOCUMENT CLASSIFICATION

I. I. Ayogu

DEPARTMENT OF COMPUTER SCIENCE, FEDERAL POLYTECHNIC, IDAH, KOGI STATE, NIGERIA.

E-mail addresses: ig.ayogu@gmail.com

ABSTRACT

The recent increase in the emergence of Nigerian language text online motivates this paper in which the problem of classifying text documents written in Yorùbá language into one of a few pre-designated classes is considered. Text document classification/categorization research is well established for English language and many other languages; this is not so for Nigerian languages. This paper evaluated the performance of a multinomial Naive Bayes model learned on a research dataset consisting of 100 samples of text each from business, sporting, entertainment, technology and political domains, separately on unigram, bigram and trigram features obtained using the bag of words representation approach. Results show that the performance of the model over unigram and bigram features is comparable but significantly better than a model learned on trigram features. The results generally indicate a possibility for the practical application of NB algorithm to the classification of text documents written in Yorùbá language.

Keywords: Supervised learning, text classification, Yorùbá language, text mining, BoW Representation

1. INTRODUCTION

Document or text classification is the task of separating or grouping natural language text matter according to their subject matter [1,2]. Rather than having to always depend on humans to hand-classify text documents into the categories of interest, learning algorithms are used to automatically learn the classification rules [3] from carefully hand-made document-class instances (in the case of supervised learning) or from a set of documents with which no class information is provided (in the unsupervised case). The learned model is then used to automatically classify previously unseen instances. Although the accuracy of the classification model has been found to depend on a number of factors, ranging from the quality of data, suitability of the algorithm, the kind of features and the experience of the model designer, automatic document classification is a more affordable, easily attainable solution. The cost of using humans far outweighs its benefit in terms accuracy or model performance; more still, situations that involves millions of documents and multiplicity of classes can

overstretch human capacity for consistency of judgement beyond limits.

Research activities in text classification and associated text mining applications are wide-spread and have increased in scale and scope in the two preceding decades. This has been linked by various researchers to the unprecedented growth in the multiplicity, type and quantity of text, and the rate at which they emerge online as well as the increase in the organizational and societal awareness and the attendant need for more effective and efficient use of information without incurring unbearable expenses [4,5]. Research has mostly focused on English and many other languages; the same cannot be said of Nigerian and most other African languages due to the lack of online or electronic presence of these languages. The recent advances in language-independent natural language processing technology is increasingly creating and sustaining the hope for changing the narrative as there is a noticeable increase in the rate and the scale of the emergence of Nigerian language text online, especially Hausa, Igbo and Yorùbá languages. This fact actually motivates this research.

* Corresponding author, tel: +234-803-417-8787

Today, natural language processing (NLP) research, such as Machine Translation, has largely achieved state-of-the-art results that have now found useful adaptation and extension to other languages of the world. Nigerian languages have begun benefiting from these successes and as a result, Nigeria language texts are now increasingly emerging online.

Given that research in the domain of text mining is not well established for the Nigerian language and given the growing availability of these texts in the electronic space, the time is right to begin experimenting on the possibility of extending the use of highly successful text mining techniques to the development of applications for Nigerian language texts. It is interesting to study the efficacy of successful machine learning methods on the classification of texts documents written in Nigerian languages. Other motivating factors for this research includes the fact that text classification research for other world languages have been applied to various fields and problems – its application to Nigerian domain is lacking and since text classification is useful for realizing organizational economic potentials [4,5], many of the good techniques and algorithms for text classification needs to be investigated on Nigerian languages. An acceptable outcome of text classification research of this kind would stimulate the development of text classification and systems and adoption for use by relevant stakeholders.

The problem of document classification has been defined by [4,5] and accordingly, this paper re-states the problem of document classification, with the assumption that each document in the collection belongs to only one class, as follows:

Given a document d , and a set of predefined classes $C_1, C_{i+1}, \dots, C_n: i \geq 2$, to which d is supposed to belong, determine the most appropriate class c and assign it to d .

The task of determining the most appropriate class for d is performed by a classification model learned from input documents, represented in the vector space model as feature vectors and some parameters. Text classification task, with increasing complexity, can be binary, multi-class or multi-label depending on the number of possible classes a document in the collection can belong. In the binary case, every document in the set can belong to only one of two classes and in the multi-class case, a document could belong to one of possibly many

classes – classes are usually more than two. In the multi-label scenario, there is the possibility of a given document belonging to more than one class at the same time.

Text classification has applications in spam email filtering, sentiment analysis and classification, topic modelling, business intelligence, routing and filtering systems, improvement of search and general performance of information retrieval systems[6]. The growing interest in text classification in organizational circles is attributed to the increasing awareness of the benefits derivable from text analysis especially as it concerns knowledge discovery in voluminous data, effective information use and improved understanding of organizational process with a reduced effort and reliability of outcomes [4 - 7].

This research studied the performance of the Multinomial Naive Bayes (MNB) classifier on a multi-class Yorùbá text document classification task. The model was learned on a dataset consisting of 100 samples each from business, sporting, entertainment, technology and politics domains. Experimental results showed that the performance of the algorithm over unigram and bigram features is comparable and significantly better than trigram features. Although Naïve Bayes classifier makes an extremely simplistic assumption of conditional independence of the discriminating features, it has been found competitive for many complex tasks, including text classification and so it is for our results. The main reason for its good performance is because of the dependence distribution of the features [8]. This paper suggests that conferences in the manner organized for the development of text classification and other NLP applications for English, European languages, Chinese, and so on, be established for Nigerian and African languages in order to foster the development and application of systems that promote language use and propagation.

2. LITERATURE REVIEW

Text classification is notably challenging, especially in situations where the documents is composed of short text segments, data is insufficient or there exist other subtlety such as noise in the text. Further, natural languages embody high-level semantic and abstract concepts that are not easy to represent and capture [9]; there is always an alternative way to express a given concept or

meaning in natural languages. Text classification as a problem has stirred up much research efforts, leading to the exploration and development of many interesting methodologies for developing text classification tools.

Among the approaches, regularized linear classifiers have been found to be of more practical relevance. This is due to their high computational efficiency and ability to scale with increasing dataset size. More complex, non-linear approaches have been found to give superior performance but are often not appealing for common practical applications, especially where the luxury of powerful processing nodes are not affordable. The literature is quite enormous for text classification and associated techniques but the review of text classification techniques by [10] is a good starting point. A number of researches have compared the performance of classifiers on the task of text classification; some of these are highlighted next.

In [9], the text documents were represented using the bag of words approach. The authors compared Naïve Bayes (NB), Nearest Neighbour, Decision trees (DT) and a subspace method on the Yahoo news group dataset comprising seven document classes. The authors highlighted that notwithstanding the complexity imposed by the overlap of words in the documents in the dataset, the NB and the subspace model individually performed better than the others as well as in combination. Thangaraj and Sivakami [11] compared NB to decision tree, Neural Networks (NN), and Support Vector Machines (SVM) and found it better in terms of accuracy and computational efficiency. Pawar and Gawande [12] also compared k-Nearest Neighbour (KNN), Rocchio's algorithm, decision tree, NB, NN, SVM on document classification task. The study observed that the performance characteristics of decision tree and NN were complimentary while NB was found to be weaker in comparison to SVM. The authors suggested that high dimensionality was the reason for NB's weaker performance. Similarly, [13] studied SVM, NB, KNN, DT on short text segments from social media in a cross-lingual setting. The motivation for this particular research was that short text segments are known to degrade the performance of text classification algorithms as well as the need to study Arabic.

In other works, [14] devised a partially supervised classification methodology aimed at reducing the efforts needed to create labelled data sets for text

classification problems. Their approach involves a two-stage strategy that attempts to reduce the efforts needed to create labelled documents for classification by using a small set of positive examples. Their experiments revealed that using Expectation-Maximization (EM) based methods does produce better classification models than SVM for document classification. Colace, *et al* [15], also proposed a method that uses a single label for text classification and reported a better performance than baseline methods when the size of the training set is small. The approach used features composed of weighted pair of words which were automatically learned and extracted from text using latent Dirichlet allocation. Apart from these, others include [16] which modelled text categorization as a graph classification problem, [17] which dwelt deeply on the selection of highly distinguishing features to improve on classifier performance, [18] where the use of terms-based discriminative information space for text classification was explored. Other noted research include the use of linear discriminant analysis [19], character-level document classification by combining convolution and recurrent neural networks [20], Recurrent convolutional NN [21], the use of heterogeneous information network kernels [22], representation of text documents using sentence-vector-space model and unigram representation models and fusing the scores [23]. Arras, *et al*. [24], in the quest to explain the relevant items in a text document experimented layer-wise relevance propagation technique on convolutional NN and bag-of-word SVM classifier. Although both performed similarly based on accuracy, convolutional neural networks (CNN) showed a better explainability and thus more comprehensible and suitable for more useful applications.

Kusumaningrum, *et al* [25], similar in principle to this paper, considered classification of Indonesian news articles using latent Dirichlet allocation. Also [26] dwelt on Arabic text classification. The motivation to explore MNB for the current problem comes from the findings presented by [3]. In [3], Multinomial, Bernouli and Gaussian event Naive Bayes models were compared on the 20 Newsgroup dataset. MNB was found to perform better than the other two variants. Although the 20 Newsgroup data set is many folds larger than our research data, the results will contribute towards further confirming the efficacy of MNB or otherwise.

3. METHODOLOGY

This research uses a dataset obtained from online sources and the bag-of-words (BoW), a widely used approach for feature engineering [21,22], to generate features from the documents. Unigram, bigram, and trigram features were selected after tokenization and stop word removal. The term frequency - inverse document frequency (tf-idf) pre-processing technique was employed to take care of normalization and issues arising from zero frequency or low frequency lexical items in the documents. The NB classifier was investigated on the data set.

3.1 Datasets, Representation and Pre-processing

Contrary to situations obtainable for languages like English, where a variety of standard, benchmark datasets exist, like the 20 Newsgroup data set for text classification (approximately 20,000 data points), no such data set is available for Yorùbá or any other Nigerian language. The data set for this research is custom-made and it is composed of electronic texts obtained from online news sites. There are five document classes - sports, politics, entertainment, business and technology, each with 100 instances, giving a total of 500 documents in the data set. Each news article was designated a document.

Pre-processing is a required step in text classification; it removes the least relevant, noisy textual elements from the input. This is known to improve classification performance [5,27]. Each document in the collection was subjected to a pre-processing pipeline that cleans, tokenizes, removes punctuations; numbers; tags and symbols, lowercases the text, and removes stop-words using a lexicon of pre-designated stopwords.

The raw text documents naturally lack any form of structure that is suitable for machine learning algorithm to learn from. Hence there is always the need to impose a structure since most of the learning algorithms require vectorized documents. The vector space model provides a straight-forward way of representing text documents as vectors of weights. To represent a document in this model, the basic features are the individual terms, further to this, this paper uses bigram and trigram features of terms/words in each document after pre-processing. This paper utilized the corpus-based term frequency

- inverse document frequency (tf-idf) feature weighting technique [4].

3.2 Learning Algorithms and Evaluation Measures

The NB classifier first appeared in machine learning literature in the later part of 1980s; [28] reported that Cestnik and colleagues were the first to apply NB classifier in the machine learning community. The NB classifier ignores statistical dependence that often exists among features. This is often considered naïve since this is hardly the case in reality. Nonetheless, the classifier has proven successful in many well-known practical applications, especially text classification. It profits from its computational simplicity and scalability which is a direct result of the independence assumption.

The model derives from the Bayesian theorem which expresses the posterior probability distribution of the class/category as a product of the likelihood of the data given the class and the prior probability distribution of the class; it is normalized by the probability of the sample as given in equation (1).

$$P(Y|X) = \frac{P(X|Y)P(Y)}{P(X)} \quad (1)$$

$$\equiv \frac{P(X=x_1, x_2, \dots, x_n|Y)P(Y)}{P(X=x_1, x_2, \dots, x_n)} \quad (2)$$

Using the chain rule, the likelihood probability $P(X = x_1, x_2, \dots, x_n|Y)$ is decomposed into equation (3)

$$P(X = x_1|x_2, \dots, x_n, Y)P(X = x_2|x_3, \dots, x_n, Y) \dots P(X = x_{n-1}|x_n, Y)P(X = x_n|Y) \quad (3)$$

The NB classifier is popular for its simplifying, conditional independence assumption, which despite being simplistic, has been found to be practically successful. The NB assumes that the input features are conditionally independent and using the naïve conditional independence assumption, $P(X = x_1|x_2, \dots, x_n, Y)$ can be restated as $P(X = x_1|Y)$ and this then implies that $P(X = x_1|x_2, \dots, x_n, Y) = \prod_i^n P(X = x_i|Y)$. The $P(X = x_1|Y)$ is usually modelled using the same distribution: binomial or Gaussian. Thus, since x_i, x_{i+1}, \dots, x_n are conditionally independent, given a class Y , then the number of parameters to be estimated reduces to just $2n + 1$ parameters, making it time efficient and robust.

NB is trained using the Maximum Likelihood Estimation (MLE) and Maximum a Posteriori (MAP)

estimates from the data. The class priors are obtained by computing for each class, Y in the data, the probability $P(Y = y_i)$ as given in equation (4).

$$\hat{\pi} = P(Y = y_i) = \frac{\text{count}(Y=y_i)}{\text{numer of classes in the data}} \quad (4)$$

The likelihood probabilities are also obtained from the data. For this, on each values $x_{i1}, x_{i2}, \dots, x_{ik}$ and for each Y_i and each X_i compute an estimate of $P(X_i = x_{ij}|Y = y_k)$ as:

$$\theta_{ijk} = P(X_i = x_{ij}|Y = y_k) = \frac{\text{count}(X_i = x_{ij}|Y = y_k)}{\text{count}(Y = y_k)} \quad (5)$$

3.2.1 Training and Applying the NB Algorithm

The training steps and application of then Naïve Bayes algorithm is summarized in the following listings.

Given X training instances with n -dimensional features and K classes, NB algorithm has a training complexity of $O(nX)$ and a prediction complexity of $O(nK)$. These imply that the algorithm is time-efficient and thus suitable for the bag of words approach employed in this paper.

Begin Train:

Input: Labelled data X with n features: x_1, x_2, \dots, x_n ; A vector of class labels y_1, y_2, \dots, y_k .

Output: Parameter estimates for the model: π_k and θ_{ijk}

// compute prior probability

For each class, y_k :

$$\pi_k = P(Y = y_k) \equiv \frac{\text{count}(Y = y_k)}{|X|}$$

End for

// compute likelihood probability

For each feature, x_i :

For each feature value, x_{ij} :

$$\theta_{ijk} = P(X_i = x_{ij}|Y = y_k)$$

Begin Classify:

Input: test data, X_i^t , without the associated class vector.

Output: the predicted classes for each instance in the test data

For each new instance, x_i^t

// compute a predicted class

$$y_{pred} = \underset{y_k}{\operatorname{argmax}} P(Y = y_k) \prod_i P(x_i^t | Y = y_k)$$

$$\Rightarrow y_{pred} = \underset{y_k}{\operatorname{argmax}} \pi_k \prod_i \theta_{ijk}$$

End Classify

4. EXPERIMENTAL SETUP AND RESULTS

4.1 Experimental Setup

The experimental results reported were obtained from three separate experimental sessions; each based on unigram, bigram or trigram features obtained from the texts and represented using the bag of words approach respectively. The training data comprises 500 instances, evenly distributed across the five categories of documents in the data set. After pre-processing, the data instances were normalized using the tf-idf transformation technique. The experiments were performed using 10-fold cross validation setup since the size of the data set is considered small. In order to evaluate the performance of the model, three standard metrics – precision, recall and F₁-measure were used. These are usually stated as follows:

$$P = \frac{TP}{TP+FP} \tag{7}$$

$$R = \frac{TP}{TP+FN} \tag{8}$$

$$F_1 = \frac{2PR}{P+R} \tag{9}$$

where:

TP = the number of correct positive prediction

FP = the number of incorrect positive prediction

FN = the number of incorrect negative prediction

4.2. Results and Discussion

The experimental results presented in Table 1 were obtained for three separate multinomial Naïve Bayes models respectively learned on unigram, bigram and trigram features. The results show that the performance of the learning algorithm over unigram and bigram features are comparable and significantly better than trigram features. Although Naive Bayes classifier makes an extremely simplistic assumption of conditional independence of the discriminating features, it has been found competitive for many complex tasks, including text classification and so it is for our results.

Table 1 shows the result of the experiments in terms of precision, recall and f-scores. It is not difficult to see that the general performance of the model across the three feature categories is not generally impressive. It is notable, however, that the unigram features performed significantly better overall. The mean unigram accuracy ($\mu \approx 0.55$) is not significantly better than that of bigram features ($\mu \approx 0.51$) at 95% confidence interval ($p < 0.5$) but is significantly better than trigram features ($\mu \approx 0.43$) with p value of 0.667.

Again, the performance due to bigram features categories is significantly better (p value = 0.287) than that due to trigram features at 95% confidence interval.

Table 1: Experimental results in terms of precision, recall and f₁-measure by feature categories.

Feature category	Class	P (%)	R (%)	F ₁ (%)	Overall Accuracy (%)
Unigrams	Technology	0.566	0.600	0.583	54.8
	Sport	0.62	0.620	0.62	
	Entertainment	0.511	0.460	0.484	
	Business	0.521	0.620	0.566	
	Politics	0.518	0.440	0.477	
	Average	0.5472	0.548	0.546	
Bigrams	Technology	0.532	0.433	0.477	51.0
	Sport	0.559	0.576	0.567	
	Entertainment	0.500	0.480	0.490	
	Business	0.450	0.540	0.491	
	Politics	0.505	0.380	0.503	
	Average	0.5092	0.4818	0.5056	
Trigrams	Technology	0.463	0.380	0.418	42.6
	Sport	0.529	0.540	0.535	
	Entertainment	0.370	0.400	0.385	
	Business	0.362	0.420	0.389	
	Politics	0.424	0.390	0.406	
	Average	0.4296	0.426	0.4266	

These observations are not quite surprising. The experimental dataset comprises just 500 instances. The size of data compares poorly to the 20 Newsgroup benchmark dataset used in [3] with mean precision, recall and f-scores values of 0.826, 0.828 and 0.827 respectively. In terms of overall accuracy, the trigram features also scored the lowest at just 42.6%. This is expectedly so because of sparsity that the lean size of the dataset naturally imposes. The imperative therefore, is to increase the size of the data and include more algorithms in subsequent study.

5. CONCLUSION

Text classification is a significant, supporting task in many information retrieval [3] and natural language processing problems. The wave of increasing presence of texts written in Nigerian languages online motivates the desire to develop tools and systems for processing, propagation and utilization. Such tools would also accelerate the production and publication of more text matter online, leading to the creation of standard, benchmark datasets as had been done for many languages of the world. This paper explored the use of Naïve Bayes algorithm, a popular, well established text classification algorithm for the classification of text written in Yorùbá language. The results, while not so impressive, demonstrate the potentials of NB being a suitable algorithm for the task.

6. REFERENCES

- [1] Pang, B., Lee, L., and Vaithyanathan, S. "Thumbs up?: sentiment classification using machine learning techniques", *Association for Computational Linguistics Conference on Empirical methods in natural language processing*, Pennsylvania USA, July 7-12, 2002, pp.79-86.
- [2] Sebastiani, F. "Machine learning in automated text categorization", *ACM computing surveys (CSUR)*, Vol. 34, Number 1, 2002, pp. 1-47.
- [3] Xu, S. "Bayesian Naïve Bayes classifiers to text classification", *Journal of Information Science*, Vol. 44, Number 1, 2018, pp. 48-59.
- [4] Kobayashi, V. B., Mol, S. T., Berkers, H. A., Kismihók, G., and Den Hartog, D. N. "Text classification for organizational researchers: A tutorial", *Organizational Research Methods*, Vol. 21, Number 3, 2018, pp. 766-799.
- [5] Kobayashi, V. B., Mol, S. T., Berkers, H. A., Kismihók, G., and Den Hartog, D. N. "Text mining in organizational research", *Organizational Research Methods*, Vol. 21, Number 3, 2018, pp. 733-765.
- [6] Hull, D., Pedersen, J., and Schutze, H. "Document routing as statistical classification", *AAAI Spring Symposium on Machine Learning in Information Access*, Palo Alto, USA, March 25-27, 1996, pp. 49-54.
- [7] Adeva, J. G., Atxa, J. P., Carrillo, M. U., and Zengotibengoa, E. A. "Automatic text classification to support systematic reviews in medicine" *Expert Systems with Applications*, Vol. 41, Number 4, 2014, pp. 1498-1508.
- [8] Zhang, H. "The optimality of naive Bayes", *Seventeenth International Florida Artificial Intelligence Research Society Conference*, Miami Beach, USA, May 17 -19, 2004, pp. 562 – 567.
- [9] Li, Y. H., and Jain, A. K. "Classification of text documents", *The Computer Journal*, Vol. 41, Number 8, 1998, pp. 537-546.
- [10] Ting, S. L., Ip, W. H., and Tsang, A. H. "Is Naive Bayes a good classifier for document classification", *International Journal of Software Engineering and its Applications*, Vol. 5, Number 3, 2011, pp. 37-46.
- [11] Thangaraj, M., and Sivakami, M. "Text classification techniques: A literature review", *Interdisciplinary Journal of Information, Knowledge and Management*, Vol. 13, 2018, pp. 117-136.
- [12] Pawar, P. Y., and Gawande, S. H. "A comparative study on different types of approaches to text categorization", *International Journal of Machine Learning and Computing*, Vol. 2, Number 4, 2012, pp. 423-426.
- [13] Faqeeh, M., Abdulla, N., Al-Ayyoub, M., Jararweh, Y., and Quwaidar, M. "Cross-lingual short-text document classification for facebook comments", *2014 IEEE International Conference on Future Internet of Things and Cloud*, Barcelona Spain, August 27-29, 2015, pp. 573-578.
- [14] Saad, F. H., de la Iglesia, B., and Bell, G. D. "Comparison of documents classification techniques to classify medical reports", *Lecture Notes in Computer Science*, Vol. 3918, 2006, pp. 285-291.
- [15] Colace, F., De Santo, M., Greco, L., and Napoletano, P. "Text classification using a few

- labeled examples", *Computers in Human Behaviour*, Vol. 30, 2014, pp. 689-697.
- [16] Rousseau, F., Kiagias, E., and Vazirgiannis, M. "Text categorization as a graph classification problem", *The 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing*, Beijing China, July 26-31, 2015, pp.1702-1712.
- [17] Rehman, A., Javed, K., and Babri, H. A. "Feature selection based on a normalized difference measure for text classification", *Information Processing and Management*, Vol. 53, Number 2, 2017, pp 473-489.
- [18] Junejo, K. N., Karim, A., Hassan, M. T., and Jeon, M. "Terms-based discriminative information space for robust text classification", *Information Sciences*, Vol. 372, 2016, pp. 518-538.
- [19] Torkkola, K. "Linear discriminant analysis in document classification". *IEEE ICDM Workshop on Text Mining*, San Jose USA, November 29, 2001, pp. 800-806.
- [20] Xiao, Y., and Cho, K. "Efficient character-level document classification by combining convolution and recurrent layers" 2016, <https://arxiv.org/abs/1602.00367>, Accessed on January, 15th, 2019.
- [21] Lai, S., Xu, L., Liu, K., and Zhao, J. "Recurrent convolutional neural networks for text classification", *Twenty-ninth AAAI conference on artificial intelligence*, Austin Texas USA, January, 25-30, 2015, pp. 2267-2273.
- [22] Wang, C., Song, Y., Li, H., Zhang, M., and Han, J. "Text classification with heterogeneous information network kernels", *Thirtieth AAAI Conference on Artificial Intelligence*, Phoenix, Arizona USA, February 12-17, 2016, pp. 2130-2136.
- [23] Bhushan, S. B., and Danti, A. "Classification of text documents based on score level fusion approach", *Pattern Recognition Letters*, Vol. 94, 2017, pp. 118-126.
- [24] Arras, L., Horn, F., Montavon, G., Müller, K. R., and Samek, W. "What is relevant in a text document?: An interpretable machine learning approach", <https://doi.org/10.1371/journal.pone.0181142>, Accessed on January, 15, 2019.
- [25] Kusumaningrum, R., Wiedjayanto, M. I. A., and Adhy, S. "Classification of Indonesian news articles based on Latent Dirichlet Allocation", *2016 IEEE International Conference on Data and Software Engineering (ICoDSE)*, Bali Indonesia, October 26-27, 2016, pp.1-5.
- [26] Ahmed, N. A., Shehab, M. A., Al-Ayyoub, M., and Hmeidi, I. "Scalable multi-label Arabic text classification", *The 6th IEEE International Conference on Information and Communication Systems*, Amman Jordan, April 7-9, 2015, pp. 212-217.
- [27] Uysal, A. K., and Gunal, S. "The impact of pre-processing on text classification", *Information Processing and Management*, Vol. 50, Number 1, 2014, pp. 104-112.
- [28] Kotsiantis, S. B. "Supervised machine learning: A review of classification techniques". *Informatica*, Vol. 31, 2007, pp. 249-268.