

## NEWS ARTICLE CLASSIFICATION USING KOLMOGOROV COMPLEXITY DISTANCE MEASURE AND ARTIFICIAL NEURAL NETWORK

Temitayo Matthew Fagbola<sup>1,2\*</sup>, Colin Surendra Thakur<sup>1,2</sup>, Oludayo Olugbara<sup>2</sup>

<sup>1</sup>*KZN eSkills CoLab, Durban University of Technology, Durban 4000, South Africa*

<sup>2</sup>*Department of Information Technology, Durban University of Technology, Durban, South Africa*

(Received: September 2018 / Revised: November 2018 / Accepted: January 2019)

### ABSTRACT

News article classification is a recently growing area of interest in text classification because of its associated multiple matching categories. However, the weak reliability indices and ambiguities associated with state-of-the-art classifiers often employed make success in this domain very limited. Also, the high sensitivity and large disparity in performance results of classifiers to the varying nature of real-world datasets make the need for comparative evaluation inevitable. In this paper, the accuracy and computational time efficiency of the Kolmogorov Complexity Distance Measure (KCDM) and Artificial Neural Network (ANN) were experimentally evaluated for a prototype large dimensional news article classification problem. 2000 News articles from a dataset of 2225 British Broadcasting Corporation (BBC) news documents (including examples from sport, politics, entertainment, education and technology, and business) were used for categorical testing purposes. Porter's algorithm was used for word stemming after tokenization and stop-words removal, and a Normalized Term Frequency–Inverse Document Frequency (NTF-IDF) technique was adopted for feature extraction. Experimental results revealed that ANN performs better in terms of accuracy while the KCDM produced better results than ANN in terms of computational time efficiency.

*Keywords:* Artificial neural network; Kolmogorov complexity distance measure; News article dataset; Text classification

### 1. INTRODUCTION

In the domain of text classification problems, news article classification has become an area of significant interest due to the overwhelmingly growing volume of news corpus on the World Wide Web (WWW). However, during classification, news articles often suffer from deep ambiguity because of their various matching categories and the weak reliability performance of most classification systems being used. These often resulted in low efficiency and poor performance evident in many current approaches (Kaur & Bajaj, 2016; Birabadar & Raikar, 2017). In recent times, learning systems based on ANN and KCDM for classification tasks in high-dimensional problem space, including intrinsic plagiarism detection, image and speech recognitions, identity and non-parametric testing, risk assessment, cellular automata classification, spam filtering, malicious URL detection, text and music classifications, DNA analysis, radar signal classification, EEG classification, e-commerce product classification, etc, are becoming more evident (Revolle et al., 2016; Oyewole & Olugbara, 2017; Abdalkafor, 2017; Haris et al., 2018). Actually, ANN has previously been identified as a good approach for

---

\*Corresponding author's email: [temitayo.fagbola@gmail.com](mailto:temitayo.fagbola@gmail.com), Tel. +27-78-7246995; +234-70-30513010  
Permalink/DOI: <https://doi.org/10.14716/ijtech.v10i4.2339>

dealing with large text classification problems (Lai et al., 2015). However, due to the complex structure of news article datasets, the identification of an efficient and accurate classifier that is a best fit for their classification remains an open problem. This makes it a highly challenging process to annotate topical news, based on different categories in an accurate and time efficient manner.

Recently, a growing trend of emerging, user-aware, big data analytic concepts with tags like “user-assisted classification”, “interactive classification”, “user-aware classification” and “user-centred classification” is becoming more evident. This reflects learning systems and/or big data analytic techniques that incorporate users’ feedbacks, reviews, ratings and personalized opinions into their classification process to augment the quality of classification decisions in an automated/semi-automated fashion (Donkers et al., 2018). For example, this type of approach includes personalized and sentiment-enhanced recommender systems (Yibo et al., 2018). However, this approach is best suited for unstructured data analysis (Donkers et al., 2018).

In the present work, the general aim was to conduct a performance comparison of the accuracies and time efficiency of the Kolmogorov Complexity Distance Measure (KCDM) and an Artificial Neural Network (ANN) for solving a prototype news article classification problem. In the experiments conducted, ANN and the KCDM were implemented using Microsoft Visual C# language. 2000 news articles were obtained from the publicly available BBC News article dataset. These were pre-processed using Porter’s algorithm after tokenization and stop-words removal. An NTF-IDF technique was used to extract and select relevant features before training and classification with the KCDM and ANN. The rest of this paper is summarized as follows: in Section 2, relevant literature on multi-labelled text classification, ANN and the KCDM are discussed. In Section 3, materials and method are presented. Section 4 discusses the results obtained, while Section 5 presents the conclusions with future directions. The major contributions of this work include:

- a) Development of a classification method for a large corpus of news articles using ANN and the KCDM by combining Porter’s algorithm with an NTF-IDF technique.
- b) Experimental comparison of the performance of ANN and the KCDM on news article classification using accuracy and computational time efficiency as evaluation metrics.

## 2. LITERATURE REVIEW

In this section, multi-labelled text classification problems and algorithms used in this study are discussed.

### 2.1. Text Classification Problem

News article classification is an example of a multi-label text classification problem. Assuming a hyperplane function  $\Phi: \mathcal{O} \rightarrow \mathcal{P}(C)$  and an unknown target function  $\check{\Phi}: \mathcal{O} \rightarrow \mathcal{p}(C)$ , then text classification is illustrated by estimating  $\check{\Phi}$  using  $\Phi$ , such that each instance  $x \in \mathcal{O}$  is mapped to 0 or more predetermined classes  $C = \{c_1, \dots, c_{|C|}\}$  (Skjennum, 2016). Given an arbitrary problem with function  $\gamma: \mathbb{X} \rightarrow \mathbb{C}$ , where  $\mathbb{X}$  is the feature space and  $\mathbb{C} = \{C_1, C_2, \dots, C_n\}$  is a unique distribution of classes having  $n$  as the number of unique classes, then the class,  $C_i, i \leq n$ , of an unknown test input  $\hat{\mathbb{X}}$  is determined by  $\gamma$ . An observation  $Z$  could be presented in the input-output form  $(x, y)$ , where  $x$  is the input into the learning procedure and  $y$  is the output estimated using  $y = f(x)$ . That is, with a training dataset  $T_D$  defined as:

$$T_D = \{z_n = (x_n, y_n)\}_{n=1}^N \quad (1)$$

where  $N = |T_D|$ , that is, the size of  $T_D$ , the expected error rate of a hypothesis  $h$ , is defined as (Van Meeuwen, 2013):

$$\pi(h) = \int_{x \sim P_x} \mathbb{I}[h(x) \neq f(x)] \quad (2)$$

where the Boolean evaluation  $\mathbb{I}[h(x) \neq f(x)]$  is 1 if the relation is true and 0 otherwise, while  $P_x$  depicts an unknown probability distribution  $x$ . Basically, the main goal of learning algorithms is to select a  $h$  with a low  $\pi(h)$  in an attempt to enhance the accuracy of a classification process. Hence, a learning algorithm is used to determine a  $h$  with the least number of errors in the training data,  $T_D$ , in such a way that (Li & Mostafa, 2006):

$$e_D(h) = \sum_{n=1}^N \mathbb{I}[h(x_n) \neq y_n], \quad (3)$$

where  $e_D(h)$  is the hypothesis with the least number of learning errors.

(Zhang et al., 2016) introduced a convolutional network-based on character-level coding for classifying news dataset using “small and large convolution”, “bag of means”, “bag of words” and “small and large full convolution”. Gurmeet and Karan (2016) classified a 1000 article subset of BBC news with categories in business, health, entertainment and sports using ANN and a Support Vector Machine (SVM). However, specific methodological detail on steps, including the network topology used, data pre-processing, tokenization and stop-word removal and feature selection was not included in the published article. Chan et al. (2001) developed a classical online news categorizer system with personalization for Channel Asia’s financial news classification using an SVM with 10 distinct categories from the Reuters-21578 collections. Skjennum (2016) designed and implemented a robust and scalable system, based on an ensemble of feedforward multilayer perceptron network, n-binary multinomial Naïve Bayes and a Long Short-Term memory Recurrent Neural Network, to classify 1.8 million news articles originating from a *New York Times annotated corpus*. However, the reliability and efficiency of the developed multilingual classification system could not be verified.

## 2.2. Classification Algorithms

### 2.2.1. Artificial Neural Networks

An ANN is made up of atomic units called artificial neurons. Each neuron is composed of summation, weighting, transfer, error, learning, scaling and limiting, as well as output functions (Huang, 2009). A calculated weight adaptively depicts how strong the connection and magnitude of an input signal to a neuron is and is often associated with the summation function. May et al. (2011) established a functional dependency between input variables, weight and the optimal performance of an ANN model. While choosing input variables, it is very important to consider the factors that closely influence the final output of the classification process. These include relevance, comprehensibility, training difficulty, computational effort and dimensionality of features. To solve a prototype classification problem, an array of neural networks is created and trained with some labelled data to generate a model that explicitly represents the physical classification system (Huang, 2009). New data are then introduced to this model for classification.

To achieve this, a weighted sum of the inputs must be computed. This is done by summing the dot products of the respective input elements and their associated weights. Given inputs  $i_1, i_2, \dots, i_n$  and their corresponding weights  $w_1, w_2, \dots, w_n$ : input 1 =  $i_1 * w_1$ , input 2 =  $i_2 * w_2$  and input  $n = i_n * w_n$ ; then using the summation function, the final input is computed as  $input1 + input2 + \dots + input n$  (May et al., 2011).

In this paper, probabilistic neural nets, a type of ANN, were used to model the hierarchical relationship among tokens generated from the training data by computing a vector space that interprets into a model capable of minimizing the distance among the tokens, as shown in Figure 1. The input to the ANN units is a vector  $X_i$ , which represents the term frequencies in the

*ith* document. For each neuron  $A$ , representing a set of weights,  $(w_1, w_2, \dots, w_n)$ , is computed using the relation  $pi = A \cdot X_i$  to learn the internal structure of the training data  $(X_i, y_i) \forall i \in \{1 \dots n\}$ . Random weights are used to initiate the learning process and updated via the relation  $pi$  that was introduced to the training data. It is important to monitor the progress of the update by introducing a learning rate,  $\mu$ , which ensures that the direction of the neuron's error is well regulated.

### 2.2.2. Kolmogorov Complexity Distance Measure

Kolmogorov complexity (KC) is best used for apt representation of data elements or simply documents (Haynes et al., 2010). Assuming  $D$  is the formal depiction of an object  $obj$  defined as  $D = [S_1, S_2, \dots, S_n]$  where  $S_1, S_2$  and  $S_n$  are texts contained in  $D$  and  $l_i$  depicts the length of string  $S_i$  in digital bits for all  $i \leq n$ ,  $L = [l_1, l_2, \dots, l_n]$ .

```

Input: Training Data  $(X_i, y_i) \forall i \in \{1 \dots n\}$  and the learning rate,  $\mu$ ;
weight vector,  $A = (w_1, w_2, \dots, w_n)$ ;
  initialize  $A$ ; (set  $A \leftarrow 0$ ) ||  $A \leftarrow$  small random numbers
  for each category of BBC news
  repeat
  neural network  $\leftarrow$  training data,  $T_D = (x_n, y_n) \}_{n=1}^N$ 
  if (sign of  $A \cdot X_i \neq y_i$ ) then
    update weights in  $A$  using the learning rate,  $\mu$ 
  else end;
  until weights in  $A$  converge.
end for

```

Figure 1 The ANN algorithm

By definition, the KC of a string is the size of a string  $S_i$  with the smallest length  $l_i$ , such that  $K(obj) = \min[|S_1|, |S_2|, \dots, |S_n|]$  or  $\min[l_1, l_2, \dots, l_n]$ . Therefore, given any two objects,  $obj_1$  and  $obj_2$ , with  $K$  complexities  $K_1$  and  $K_2$ , respectively,

$$K(obj_2|obj_1) = K(obj_1 \text{ c } obj_2) - K(obj_1) \quad (4)$$

where 'c' is a concatenation operator and  $K(obj_1 \text{ c } obj_2)$  is the  $K$  complexity when  $obj_1$  and  $obj_2$  are concatenated (Kolmogorov, 1965). Kolmogorov distance can help to determine the underlying similarities among disparate data. Similarly, the normalized Kolmogorov distance,  $K_D$ , between  $obj_1$  and  $obj_2$  can be determined by using (Cilibrasi & Vitanyi, 2007):

$$K_D = [L(l_1 \text{ c } l_2) - \min(L(l_1), L(l_2))] / [\max(L(l_1), L(l_2))]; \quad (5)$$

$$K_D(obj_1, obj_2) = (K(obj_1) + K(obj_2) - K(obj_1 \text{ c } obj_2) - \min(K(obj_1), K(obj_2))) / \quad (6)$$

$\max(K(obj_1), K(obj_2))$ ). In reduced form,

$$K_D(obj_1, obj_2) = (1 - K(obj_1) + K(obj_2) - K(obj_1 \text{ c } obj_2)) / \max(K(obj_1), K(obj_2)) \quad (7)$$

## 3. METHOD

A 4-staged architecture comprising of the news article dataset collection, news article dataset pre-processing, feature extraction and classification using ANN and the KCDM is presented in Figure 2. After the classification stage was completed, the performance of ANN and the KCDM were evaluated using True Positive Rate and classification time.

### 3.1. News Article Dataset Collection

The British Broadcasting Corporation (BBC) news dataset prepared by Greene and Cunningham (2006) was used for the experiments conducted in this study. The dataset contains 2225 news

articles in five topical categories including politics (417), sport (511), entertainment (386), education and technology (401) and business (510). These articles were collected between 2004 and 2005 from the BBC news website and made publicly available through <http://mlg.ucd.ie/datasets/bbc.html> for research purposes.

### 3.2. News Article Dataset Pre-processing

Data pre-processing is a major requirement to ensure that all missing values, outliers and incomplete data are cleaned or fixed before the actual feature extraction process. This is due to the strong connection between classification accuracy and how well the documents have been pre-processed (Wang and Wang, 2005). In this paper, tokenization, removal of stop-words and stemming were performed. Tokenization ensures that sentences are morphologically segmented into tokens delimited by white spaces (Khan *et al.*, 2010). Stop words that frequently occur but do not add up relevant information to the news classification process were filtered out from the corpus of terms using the BBC News stop word lists. Similarly, the non-descriptive tokens with less than three characters were removed through a term-filtering process before the final classification was conducted. Stemming is heuristically base/stem/root words from words that are inflected by removing the affixes and reducing inflectional forms (Greene and Cunningham, 2006). Porter's stemmer was used due to its higher efficiency than other baseline approaches like Paice/Husk, *S* and Lovins (Gurmeet and Karan, 2016).

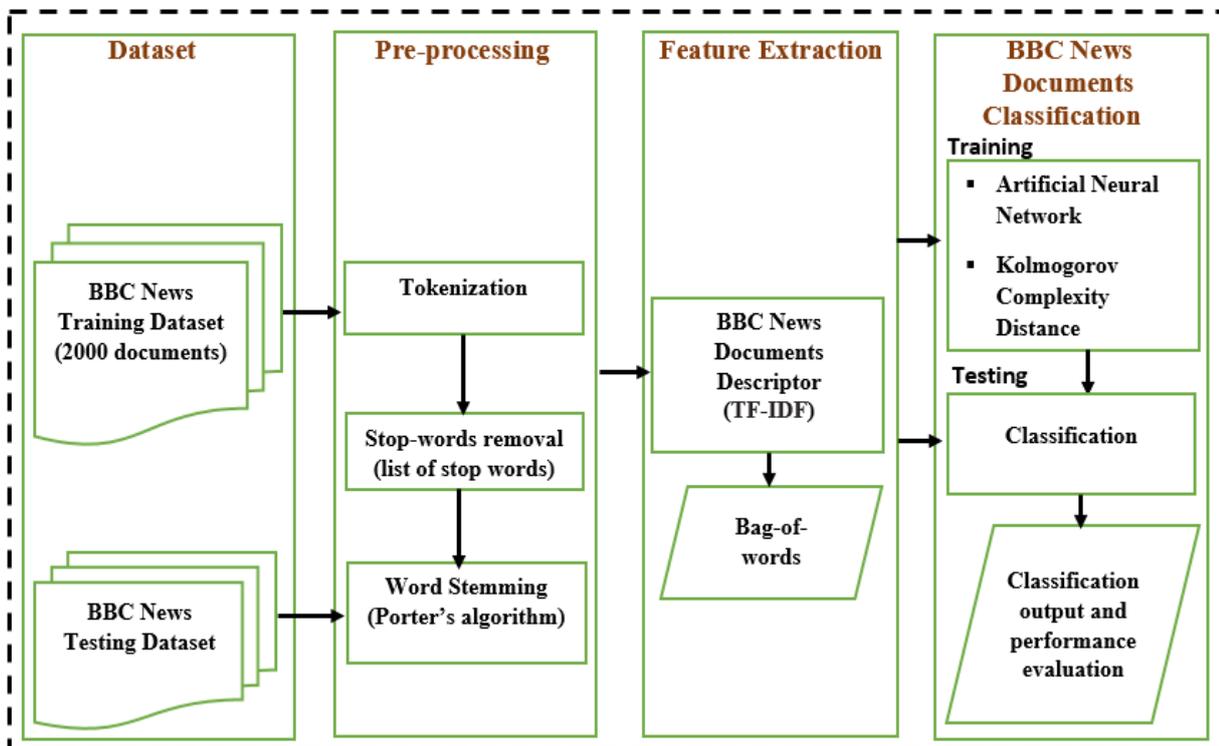


Figure 2 The classification architecture used with the BBC News dataset.

### 3.3. Feature Extraction

Feature extraction is a major dimensionality reduction step to ensure the accuracy of the classification and improved time efficiency (Fagbola *et al.*, 2012; Fagbola *et al.*, 2017). In this paper, relevant features were extracted from the terms returned via the pre-processing step using a normalized TF-IDF technique. Normalized TF-IDF is a vector space technique that easily renders the weight of terms numerically (Joho and Sanderson, 2007). It was adopted due to its highly accurate performance when compared to most other statistical approaches. For each term  $i$ , the weight is given as (Mandal and Sen, 2014):

$$W_i = \frac{(TF_i \times \log(\frac{N}{n_i}))}{\sqrt{\sum_{i=1}^n (TF_i \times \log(\frac{N}{n_i}))^2}} \tag{8}$$

where  $n_i$  is the number of documents containing term  $i$  and  $N$  is the total number of documents. TF defines the number of occurrence of each term in a document, while IDF provides length normalization. A weight-term matrix with documents forming the rows and TF-IDF weights forming the columns was generated by calculating the TF-IDF for each feature, where  $w_{ij}$  is the weight of term  $i$  in document  $j$ ,  $D$  is a sample document,  $T$  represents a term and  $c_i$  is the class

$$\begin{bmatrix} & T_1 & T_2 & \dots & T_i & \\ D_1 & w_{11} & w_{12} & \dots & w_{1i} & c_1 \\ D_2 & w_{21} & w_{22} & \dots & w_{2i} & c_2 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ D_j & w_{j1} & w_{j2} & \dots & w_{ji} & c_k \end{bmatrix} \tag{9}$$

of document  $D_j$ . The pseudocode for the feature extraction step is presented in Figure 3. The relevant-term matrix features obtained at this stage were introduced into the classifiers (ANN and KCDM) for term similarity estimation and final classification.

```

Input
  T: the unique terms in all documents
  D: the documents of the training set
Output: weight Matrix
Step:
i. for each term  $t_i \in T$  do
ii. for each document  $d_j \in D$  do
       $w_{ij} =$  frequency of term  $t_i$  in document  $d_j$ 
iii. end for
iv. end for
    
```

Figure 3 The pseudocode for the normalized TF-IDF feature extraction step.

### 3.4. BBC News Article Classification

The resulting discriminating features of the dataset obtained from the feature extraction phase were repeatedly split into training and testing sets for different k-fold values from 10 through 70 for the purpose of classification in a supervised learning mode (Mandal and Sen, 2014). A total of 2000 news articles were used for experimental purposes in this study. The categories, actual number of documents and the training size extracted from the BBC News dataset are presented in Table 1. Similarly, the description of the training and test datasets for each distinct k-fold value are presented in Table 2. The training dataset depicts the past observations used to generate learned models of ANN and the KCDM, while the test dataset are the set of unknown observations that were classified into various categories.

Table 1 BBC News data articles by category

Category	Actual number of Documents	Number of Documents used
Education & Technology	401	350
Sport	511	500
Entertainment	386	300
Business	510	500
Politics	417	350
Total	2225	2000

### 3.5. Performance Evaluation Metrics

The number of true positives and classification time are the basic evaluation metrics considered in this work. These metrics assisted in determining the best-fit algorithm between ANN and the KCDM for news article classification. *True positives* were the actual number of documents or topics that were correctly placed into their classes by a decision plane, a model or a classifier. The higher the true positive rate, the more reliable the classifier (Skjennum, 2016). *Classification time* was best defined as the total finite time expended by a classifier or a decision plane to identify and categorize a given set of documents into their respective classes.

Table 2 Training/testing subset size of BBC News dataset by K-fold size.

K-fold	Size of training set for all categories	Size of testing set for all categories
10	1800	200
20	1600	400
30	1400	600
40	1200	800
50	1000	1000
60	800	1200
65	700	1300

## 4. EXPERIMENTAL RESULTS

The experiment was conducted on a 64-bit Microsoft Windows 7 system with AMD Athlon X2 CPU running at 2.2 Ghz, 4 GB of random access memory and a 350 GB hard disk drive. The Java programming language was used to implement the algorithms. Each technique (ANN and KCDM) was applied to the seven distinct training sets corresponding to the different k-fold values as shown in Table 2. The average classification time for a news article and the total number of true positives were computed and are presented in Table 3. The graph of the true positives and the training documents obtained by ANN and the KCDM are presented in Figure 4.

Table 3 Results of implementing the ANN and KCDM algorithms on the test data.

Number of Test Documents	True Positives (ANN)	True Positives (KCDM)	Classification time (s) (ANN)	Classification time (s) (KCDM)
200	189	168	431	52
400	335	307	571	67
600	537	411	802	96
800	751	628	1025	119
1000	932	663	1429	166
1200	1001	699	1466	179
1300	1026	716	1600	280

Based on the evaluation results obtained, ANN showed better performance in terms of the number of true positives obtained than KCDM, which indicates that it is more accurate. However, as the training data size decreased, the classification error generated by ANN and the KCDM grew as shown in Figure 5 and Table 4. This result corroborates Oloyede et al. (2016) who asserted that the higher the training data size, the lower the misclassification error rate. In Figure 6, a graph of BBC News testing set and classification time is presented depicting a corresponding increase in classification time as the size of the testing set grows. However, KCDM was more computationally efficient than ANN in all the experiments.

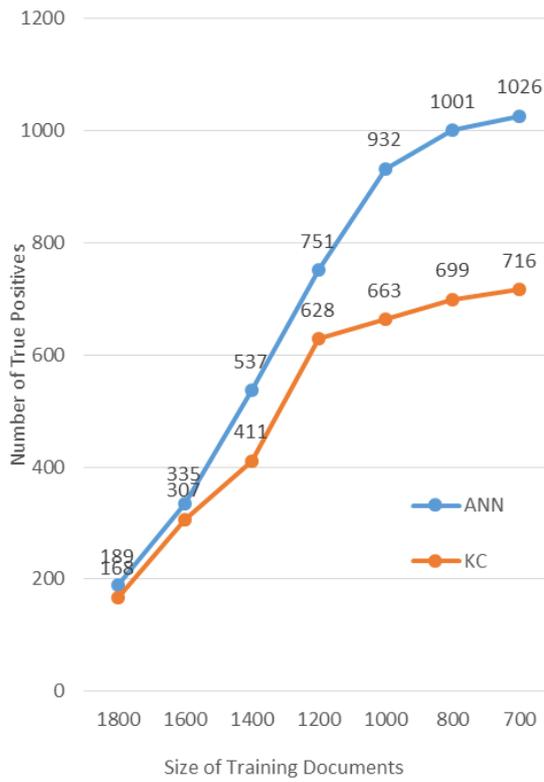


Figure 4 Plot of training set size against number of true positives.

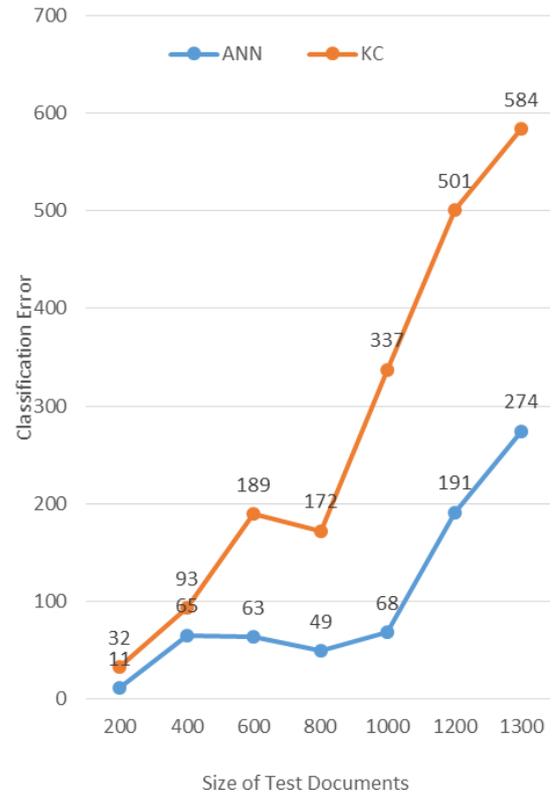


Figure 5 Plot of testing set size against classification error.

Table 4 Classification error for ANN and KCDM

Number of test document for all categories	Error (ANN)	Error (KCDM)
200	11	32
400	65	93
600	63	189
800	49	172
1000	68	337
1200	191	501
1300	274	584

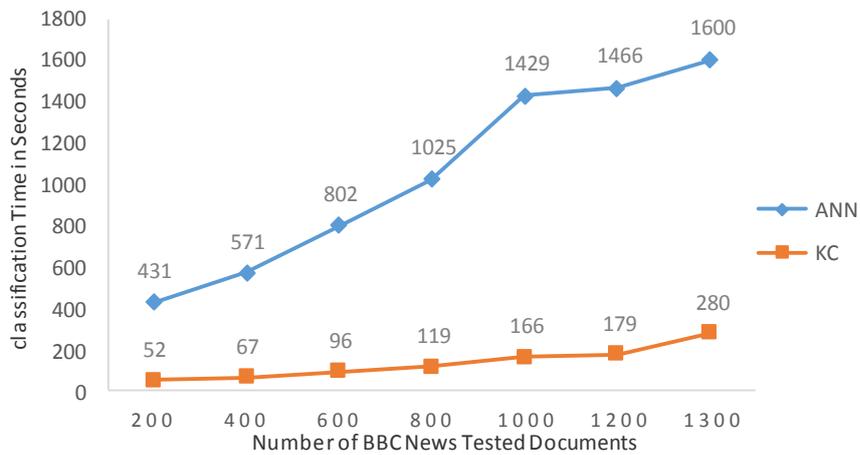


Figure 6 Plot of testing set size against total classification time

## 5. CONCLUSION

In this paper, a performance comparison between two methods (ANN and the KCDM) for addressing the news article classification problem was conducted. The experimental results revealed that ANN was better in terms of accuracy, while KCDM was better for developing time-efficient applications. Summarily, this paper establishes the relative importance of conducting performance evaluation as a core part of choosing the best test routines during development to ensure an overall high reliability for deployed applications. Furthermore, this process can help identify certain trade-offs associated with each algorithm and synergize decision making on what algorithm to apply to a specific problem domain of interest, especially when developing fault-tolerant systems.

The major findings of our experiments are: (1) ANN can produce higher classification accuracy for large datasets than KCDM. In all the experiments conducted, ANN yielded the more true positives than the KCDM; (2) The time efficiency of ANN was very low when compared to the KCDM. In the experiments, as the size of the testing set grew, its time complexity also increased. With a testing set containing 1300 features, the classification time spent by ANN was approximately eight times more than that of the KCDM.

In future works, evaluation of some emerging and other baseline classifiers like Adaboost, SVM, naïve-Bayes and  $k$ -nearest neighbour could also be conducted for news article classification in large multi-dimensional features space. In addition, an ensemble of ANN and KCDM can also be developed to realize an algorithm with improved classification accuracy and time efficiency.

## 6. REFERENCES

- Abdalkafor, A.S., 2017. Designing Offline Arabic Handwritten Isolated Character Recognition System using Artificial Neural Network Approach. *International Journal of Technology*, Volume 8(3), pp. 528–538
- Cilibrasi, R.L., Vitanyi, P.M.B., 2007. The Google Similarity Distance. *IEEE Transaction on Knowledge and Data Engineering*, Volume 19(3), pp. 370–383
- Chan, C., Sun, A., Lim, E.P., 2001. Automated Online News Classification with Personalization. *In: Proceedings of the 4<sup>th</sup> International Conference of Asian Digital Library (ICADL2001)*, pp. 320–329
- Donkers, T., Loepp, B., Ziegler, J., 2018. Explaining Recommendations by Means of User Reviews. *In: Joint Proceedings of the ACM IUI 2018 Workshops*, March 11, Tokyo, Japan, pp. 1–4
- Fagbola, T.M., Olabiyisi, S.O., Egbetola, F.I., Oloyede, A., 2017. Review of Technical Approaches to Face Recognition in Unconstrained Scenes with Varying Pose and Illumination. Federal University, Oye-Ekiti (FUOYE). *Journal of Engineering and Technology*, Volume 2(1), pp. 1–8
- Fagbola, T., Olabiyisi S., Adigun A., 2012. Hybrid GA-SVM for Efficient Feature Selection in E-Mail Classification. *Journal of Computer Engineering and Intelligent Systems*, Volume 3(3), pp. 17–28
- Greene, D., Cunningham, P., 2006. Practical Solutions to the Problem of Diagonal Dominance in Kernel Document Clustering. *In: Proceedings of the 23<sup>rd</sup> International Conference on Machine Learning*, pp. 377–384
- Gurmeet, K., Karan, B., 2016. News Classification and its Techniques: A Review. *IOSR Journal of Computer Engineering (IOSR-JCE)*. Volume 18 (1), pp. 22-26.
- Haris, A., Murdianto, B., Susatty, R., Riyanto, A., 2018. Transforming Seismic Data into Lateral Sonic Properties using Artificial Neural Network: A Case Study of Real Data Set. *International Journal of Technology*. Volume 9(3), pp. 472–478

- Huang, Y., 2009. Advances in Artificial Neural Networks-Methodological Development and Application. *Algorithms*, Volume 2(3), pp. 973–1007
- Joho, H., Sanderson, M., 2007. Document Frequency and Term Specificity. In Large Scale Semantic Access to Content (Text, Image, Video, and Sound), Paris, France: Le Centre de Hautes Etudes Internationales D'informatique Documentaire. pp. 350–359
- Kaur, G., Bajaj, K., 2016. News Classification and its Techniques: A Review. *IOSR Journal of Computer Engineering (IOSR-JCE)*. Volume 18(1), pp. 22–26
- Khan, A., Baharudin, B., Lee, L.H., Khan, K., 2010. A Review of Machine Learning Algorithms for Text-Documents Classification. *Journal of Advances in Information Technology*, Volume 1(1), pp. 4–20
- Haynes, K.E., Kulkarni, R., Stough, R.R., Laurie, S., 2010. *Exploring a Region Classifier based on Kolmogorov Complexity*. School of Public Policy, George Mason University, Available Online at <http://ssrn.com/abstract=1499218>
- Kolmogorov, A.N., 1965. Three Approaches to the Quantitative Definition of Information. *Problems of Information Transmission*, Volume 1(1), pp. 3–11
- Li, L., Mostafa, Y.S.A., 2006. *Data Complexity in Machine Learning*. Computer Science Technical Reports, 2006.003, California Institute of Technology, Pasadena, USA California Institute of Technology, Pasadena, USA
- May, R., Dandy, G., Maier, H., 2011. *Review of Input Variable Selection Methods for Artificial Neural Networks*. Suzuki, K. Edition, Artificial Neural Networks Methodological Advances and Biomedical Applications, University of Adelaide, Australia
- Mandal, A.K., Sen, R., 2014. Supervised Learning Methods for Bangla Web Document Categorization. *International Journal of Artificial Intelligence & Applications (IJAA)*, Volume 5(5), pp. 93–105
- Oloyede, A., Fagbola, T., Olabiyisi, S., Omidiora, E., Oladosu, J., 2016. Development of a Modified Local Binary Pattern-Gabor Wavelet Transform Aging Invariant Face Recognition System. In: Proceedings of ACM International Conference on Computing Research & Innovations, Nigeria, pp. 108–114
- Oyewole, S.A., Olugbara, O.O., 2017. Product Image Classification using Eigen Colour Feature with Ensemble Machine Learning. *Egyptian Informatics Journal*, Volume 19(2), pp. 83–100
- Revolle, M., le Bihan, N., Cayre, F., 2016. *Algorithmic Information Theory for Automatic Classification*. Grenoble, France: Gipsa-Lab, Universite Grenoble Alpes, France
- Birabadar, S., Raikar, M.M., 2017. Performance Analysis of Text Classifiers based on News Articles-A Survey. *Indian Journal of Scientific Research*, Volume 15(2), pp. 156–161
- Lai, S., Xu, Liheng, Liu, K., Zhao, J., 2015. Recurrent Convolutional Neural Networks for Text Classification. In: Proceedings of the Twenty-Ninth AAAI Conference on Artificial Intelligence, pp. 2267–2273
- Skjennum, P.L., 2016. Multilingual News Article Classification. *Master Thesis*, Department of Computer and Information Science, Norwegian University of Science and Technology, pp. 1–125
- Van Meeuwen, F., 2013. Multi-label Text Classification of News Articles for ASDMedia, *Master Thesis*, Department of Information and Computing Sciences, Utrecht University
- Wang, Y., Wang, X., 2005. A New Approach to Feature Selection in Text Classification. In: Proceedings of 4<sup>th</sup> International Conference on Machine Learning and Cybernetics, IEEE, Volume 6, pp. 3814–3819
- Yibo, W., Mingming W., Wei X., 2018. A Sentiment-Enhanced Hybrid Recommender System for Movie Recommendation: A Big Data Analytics Framework. *Wireless Communications and Mobile Computing*, Volume 2018, pp. 1–9

Zhang, X., Zhao, J., LeCun, Y., 2016. Character-level Convolutional Networks for Text Classification. Ithaca, NY: Cornell University, arXiv:1509.01626v3