ISSN 2278-2540 | DOI: 10.51583/IJLTEMAS | Volume XV, Issue VII, July 2026

Classifying News Articles on Nigerian Newspapers Using Machine Learning

Echezona Stephenson C., Chukwu Deborah Ogoamaechi, Uzo Blessing, Okereke George E, Ugwu Celestine, Uzo Izuchukwu

Department of Computer Science, University of Nigeria, Nsukka.

DOI: https://doi.org/10.51583/IJLTEMAS.2026.150700177

Received: 14 August 2026; Accepted: 19 August 2026; Published: 27 August 2026

ABSTRACT

With the escalation of news sources, it presents challenge for a reviewer to find the exact newspaper article(s) that would meet ones need(s); hence, the need to classify news articles so that users can efficiently get the information they need with minimal efforts. The classification of news presents many advantages, such as, determining which newspaper to look out for albeit the date of publication, as well as providing the user with personalized recommendations based on the user's previous areas of interest. Also, for research purposes, it can tell the class of news a particular string belongs to, in addition to finding trending news articles. Studies on classification of news for other countries' Newspapers have been ongoing, into a variety of groups/categories. Examples are: Arabic, Tigrigna, Uzbek, China, and Thai news headlines have all been classified into a variety of categories. These would help visitors to quickly find the news stories that interest them, little to no research has been done to categorize Nigerian news articles. Therefore, this research proposed the classification of Nigerian news articles into five different categories; such as, sports, entertainment, business, health, and politics using machine learning - to the extent of displaying the class of a string keyed in by the user - as a means of filling-in the gap in the existing body of literature. A local source was used to compile a dataset consisting of five different Nigerian news articles. Feature extraction employed Linear Discriminant Analysis (LDA), and Support Vector Machines (SVM). These two algorithms were used to train the model. The research will also evaluate the outcomes of several categorization algorithms and compare the efficacy of these algorithms according to some of performance metrics including accuracy, precision, and recall.

Keywords: News sources, newspaper articles, classify news articles, Personalized recommendation, Nigerian News articles

INTRODUCTION

Classification of news is a branch of natural language processing and information retrieval system. With the ever-increasing amount of news articles available online, it is becoming increasingly challenging for users to find relevant and trustworthy information. Machine learning techniques provide a promising solution to automate the classification process and assist users in efficiently finding news articles of interest.

Machine learning algorithms use a large dataset of labeled news articles to learn patterns and features that can distinguish different type of news. This algorithm can then classify new and unseen news articles into pre-defined categories such as Sport, Politics, Entertainment, or Technology, etc. The classification process involves analyzing the textual content of the news articles, extracting relevant information and making predictions based on learned pattern.

The use of Machine learning for news classification offers several benefits which include; firstly, enabling the automation of classification process, reducing the need for manual categorization and saving time and

ISSN 2278-2540 | DOI: 10.51583/IJLTEMAS | Volume XV, Issue VII, July 2026

effort for users. Secondly, it improves the accuracy and consistency of news classification, as machine learning algorithm are capable of handling large volume of data and identify subtle patterns that may be difficult for humans to detect, (Ahmed J. et al, 2021).

Additionally, machine learning can adapt and learn from new data, making it suitable for handling evolving news topics and trends.

In this era of information overload, news classification using machine learning plays a vital role in providing users with personalized and relevant news articles. By efficiently categorizing news based on their content, machine learning algorithms can help users filter through the vast amount of information available and focus on the topics that interest them the most. Furthermore, machine learning techniques can also be applied to identify and flag fake news, helping to combat the spread of misinformation (Babalola, E. A., 2002).

Related Works

To better understand the challenges of news article categorization and offer potential solutions, which are current with the state of the art, we survey prior research in the field. Qadi et al., (2019), classified Arabic news articles into four main categories: Business, Sports, Technology, and the Middle East, using a newly constructed dataset collected from various news websites which contain about 90,000 Arabic news articles. The authors used 10 different machine learning algorithms which include Logistic Regression, Nearest Centroid, Decision Tree (DT), Support Vector Machines (SVM), K-nearest neighbors (KNN), XGBoost Classifier, Random Forest Classifier, Multinomial Classifier, Ada-Boost Classifier, and Multi-Layer Perceptron (MLP) to investigate the effectiveness of the dataset. The result of their experiment shows that SVM performed better than all other classifiers and achieved an accuracy of 97%.

Noppakaow and Uchida, (2019), also, evaluated the performance of various machine learning models used in classifying news articles. The experiment was carried out on three classification algorithms which include Decision Tree, Support Vector Machine, and deep learning using a Thai news dataset. The authors measure the performance of the algorithms using precision, recall, and F-score measure. The results show that the accuracy of Decision Tree, SVM, and Deep Learning models are 86%, 94%, and 95%, respectively. This indicates that deep learning outperforms the other algorithm by about 1%. However, the study focused on Thai news article.

Winster and Kumar, (2020), presented a method for classifying the many different emotions that are expressed in reports of recent events. Fear, joy, and melancholy were the three sentiments that they classified into their respective groups. The performance of this strategy was evaluated with the use of the SVM classifier, which produced an accuracy score of 87.83% as a result of the evaluation.

Rabbimov and Kobilov, (2020), conducted a multi-class text categorization in Uzbek. The dataset for the study is based on 10 different types of news stories taken from the Uzbek "Daryo" online news edition. In order to classify the dataset's text samples into many categories, researchers employed six (6) distinct machine learning methods. Some examples of these algorithms include the Support Vector Machine (SVM), the Decision Tree Classifier (DTC), the Random Forest (RF), the Logistic Regression (LR), and the Multinomial Naive Bayes (MNB). Word-level and character-level n-gram models, as well as the (TF-IDF) Term Frequency Inverse Document Frequency algorithm, were employed for feature extraction. The hyperparameters for text categorization were defined using 5-fold cross-validation. The results of their experiments achieve the best accuracy on the SVM, which is 86.88%.

From several Tigrigna news sources, Fesseha et al. (2020) compiled a new dataset covering six topics: agriculture, sports, health, education, religion, and politics. To evaluate the efficacy of the datasets, the authors utilized several of ML techniques, such as LR, Nearest Centroid (NC), DT, SVM, KNN, RFC and MLP. Hybrid models, which combine the best classifiers based on a superior vote, are also used to improve accuracy. The experimental findings demonstrated dependable performance, with the closest centroid

ISSN 2278-2540 | DOI: 10.51583/IJLTEMAS | Volume XV, Issue VII, July 2026

achieving a minimum F1-score of 89.2% and SVM achieving a maximum performance of 96%. Precision, recall, and F1 scores were employed to describe the experimental outcomes

Ahmed and Ahmed, (2021), developed a platform for the automatic grouping of online news articles. In order to achieve the required accuracy, the authors tried different classifiers. The result of their experiment shows that the Bayesian classifier achieved the topmost accuracy of 93% presented in terms of confusion metrics.

Research Gap

Several studies have been carried out to classify different countries or languages such as Arabic (Qadi et al., 2019), Tigrigna (Fesseha et al. 2020), Uzbek (Rabbimov and Kobilov, 2020), Chinese (Huang and Jiang, 2019), Thai (Noppakaow and Uchida, 2019), and Bahiyah O. et al, (2020), into various news categories. However, no studies have considered classifying Nigerian news articles into various categories. As such, this research aims to develop a dataset of the Nigerian news article and classify them into various categories such as Education, Crime, Gorvernment, Drug Abuse, Businesses, and many more.

METHODOLOGY

The researcher will employ Linear Discriminant Analysis (LDA), and Support Vector Machines (SVM), Huang M. W., et al, (2017), and (Egwom et al., (2022). SVM will be used to train the model. The procedure that will be followed in the development of the model that we have proposed is shown in Figure 3.1. It demonstrates the progression of the model's creation from the gathering of the datasets (by compiling the articles from local newspapers) as is the case with some low resourced language like Tigrigna, (Fesseha, et al., (2020)), through their subsequent processing, preprocessing, then, configuration of the development tools, testing, and ultimately assessment of the news classifier machine learning model. This starts with getting the datasets together and ends with evaluating the machine learning model.

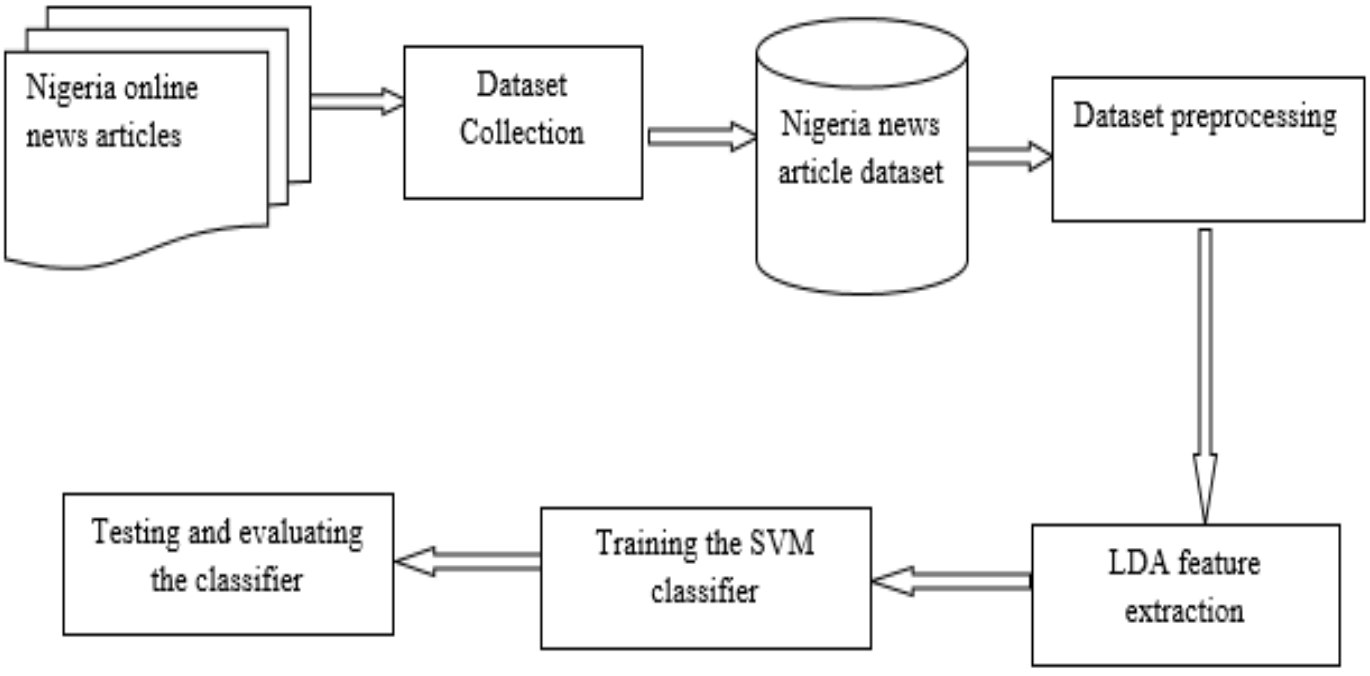

Figure 3.1 Diagrammatical representation of the research methodology

Figure 3.2: shows the details of the execution process from splitting the dataset to the evaluation of the model using a confusion matrix.

ISSN 2278-2540 | DOI: 10.51583/IJLTEMAS | Volume XV, Issue VII, July 2026

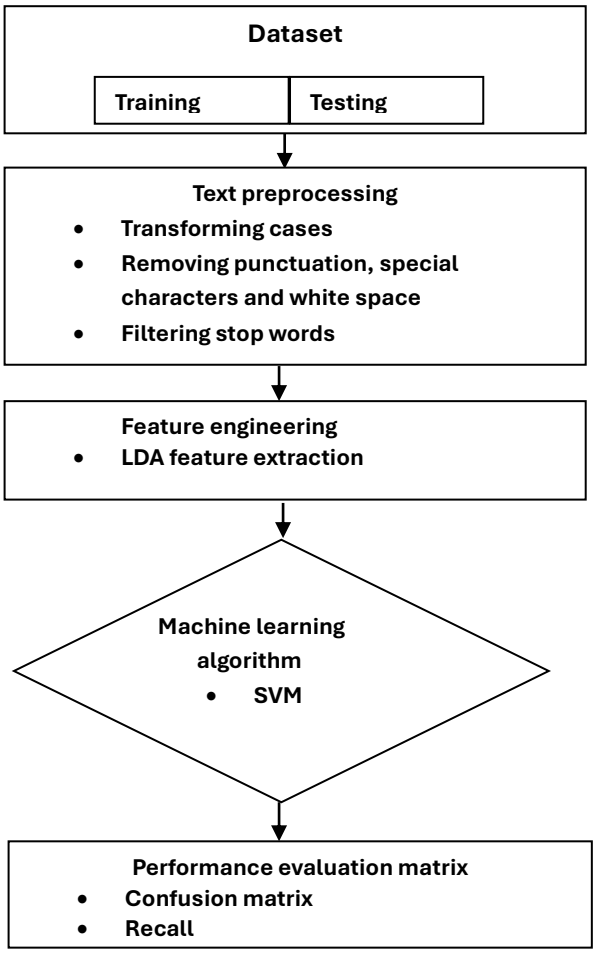

Figure 3.2 Implementation process

Dataset Collection

The table 1, sample of the dataset, which has five categories of news articles in CSV format, as shown below.

Table 1: Sample of the Dataset

S/N	ARTICLELD	TEXT	CATEGORY
1	1833	IPoB	Exbusiness
2	154	Nigeria-Chinese	Busibusiness
3	1101	Dolla appreciates	Business
4	1469	Electiong	Government
5	893	Rangers	Sport
6	2115	How the acae	Drug Abuse
7	915	Electricity	Business
8	1582	PDP govs	Government
9	651	Eto'o's silent	Sport
10	1797	Honour	Sport

Dataset Description

The provided dataset presents an extensive collection of news articles from reputable platforms, including Twitter, Jang, et al, (2019), and well-established Nigerian newspaper websites like Nigeria Tribune, Punch, Vanguard, and others. It comprises news articles categorized into five domains: Sport, Business, Politics, Entertainment, and Tech, with three essential columns - ArticleId, Text, and Category.

ISSN 2278-2540 | DOI: 10.51583/IJLTEMAS | Volume XV, Issue VII, July 2026

The ArticleId field uniquely identifies each news article, facilitating seamless referencing and meticulous tracking of individual entries. This identifier serves as a valuable asset for data analysts and machine learning practitioners, enabling efficient management and analysis of the entire collection.

The Text column contains the complete textual content of each news article, capturing a wealth of information that offers researchers an unbridled opportunity to explore and implement cutting-edge techniques in natural language processing and text analysis.

The dataset's categorization aspect empowers the algorithmic community to develop and fine-tune robust news classification models. Each news article is thoughtfully assigned to its relevant category, enabling the algorithm to discern the primary subject matter with accuracy and efficiency.

The dataset exhibits a well-balanced distribution of news articles across its five categories. The Sport category is the largest domain, comprising 346 news articles (23.2% of the dataset). Business follows with 336 articles (22.5%), Politics contributes 227 articles (17.5%), Entertainment encompasses 273 articles (18.3%), and Tech contains 261 articles (18.6%).

Dataset Preprocessing

The following preprocessing process will be applied to the dataset.
Transformation of Cases: we will transform the text to lower cases.
Removing unwanted text: Removing punctuation, special characters, and white space.
Filtering stop words.

Feature Engineering

Linear Discriminant Analysis

Figure 3.3 shows an overview of how LDA is been calculated.

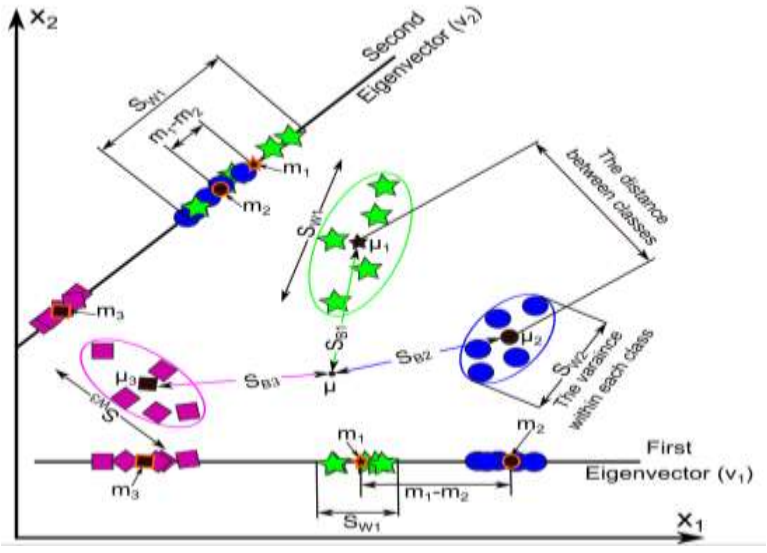

Figure 3.3 A visualized comparison between two lower-dimensional subspaces which are calculated using three different classes (Alaa Tharwat et al., 2017)

Steps for Computing Linear Discriminant Analysis

Strengths of Linear Discriminant Analysis

1. Improves Accuracy
2. Speeds up training

ISSN 2278-2540 | DOI: 10.51583/IJLTEMAS | Volume XV, Issue VII, July 2026

3. Maximize the means of each class and minimize the spreading within the class.

Modeling of the Classifier

A classification model will be adopted for this research. Support Vector Machines (SVM) classifiers that have shown great performance in previous research for classification will be adopted.

Performance evaluation metrics (Alexander Berkovich, (2025))

The model was evaluated using confusion matrix**.** A confusion matrix describes the performance parameters for the classifier as illustrated in Table 3.1,

Table 3: A confusion matrix for a binary classification task

PREDICTED CLASS		
	POSITIVE	NEGATIVE
TRUE	(TP)	(FN)
FALSE	(FP)	(TN)

Now to evaluate the accuracy, precision, and recall for a model using the formula

Accuracy measure: Addition of the correctly classified patterns divided by the addition of all patterns as seen in equation (3.16).

$$\frac{TP + TN}{TP + FP + FN + TN} \quad (3.16)$$

Precision: equation (3.17) shows the extent the classifier was able to correctly identify the positive class.

$$\frac{TP}{TP+FP} \qquad \textbf{(3.17)}$$

Recall: The number of extent the classifier predicted a negative class out of all the times the class is negative as seen in equation (3.18), Huang M. W. et al,(2017)

$$\frac{TP}{TP+FN} \qquad \textbf{(3. 18)}$$

RESULTS AND DISCUSSION

Choice of Development Environment

The Visual Studio Code also, known as the VSCode, became our primary Integrated Development Environment for both frontend and backend. In terms of the backend, we utilized the Python Flask, which is the minimum web framework for server organization for application routing and HTTP request handling. This is important because we were looking for a more flexible way to combine our machine learning model with a web interface and Flask was just the tool for the job. In the front-end section, we systematically employed HTML, CSS, Bootstrap, and JavaScript languages. HTML and CSS were used as basis for creating web pages while Bootstrap for the mobile responsiveness and the cool appearance of the web page. To use interactivity, the researcher chose JavaScript, where it was possible to implement the expected functionality, such as replacing the content of the web page with something else, or smoothly moving from one page to the other. When performing the task of creating the machine learning model, the programming tool that was used was Jupyter Notebook. This tool was beneficial due to the notebook format, as all the code could be written and executed in cells; data could also be viewed in various ways; and an extended description of the modeling process was possible as well. Jupiter Notebook was used to ensure that constant improvements and

ISSN 2278-2540 | DOI: 10.51583/IJLTEMAS | Volume XV, Issue VII, July 2026

assessments could be made to the model, and with its help, employment of the diverse algorithms and parameters was more effective.

Implementation Architecture

The news classification system focuses on its efficiency, flexibility, and ease of use through a well-planned implementation structure. This architecture leverages distinct layers, each crafted to handle specific functionalities: aggregation of the input data, model building, user engagement, and application running. The system's core is in the ML model, which was carefully implemented in Python and supported by libraries, such as scikit-learn, pandas, as well as NumPy. This has the role of a classifier, which properly categorizes news articles into the predetermined category list. The thoroughness of the data processing pipeline ensures that the raw text data is formatted in a way that can be delivered to downstream ML algorithms. It includes the steps in data collection, data cleaning, data preprocessing, feature selection, and vectorization processes. The backend of the system is built using Python Flask by serving as a connecting link between the model and the interface. The Flask framework effectively coordinates the HTTP requests and calls the ML model and returns the classification outcome to the requesting client. This framework is especially designed for creating RESTful APIs which will enhance the interaction of the components of this system. Figure 4.1 depicts the System Implementation Architecture

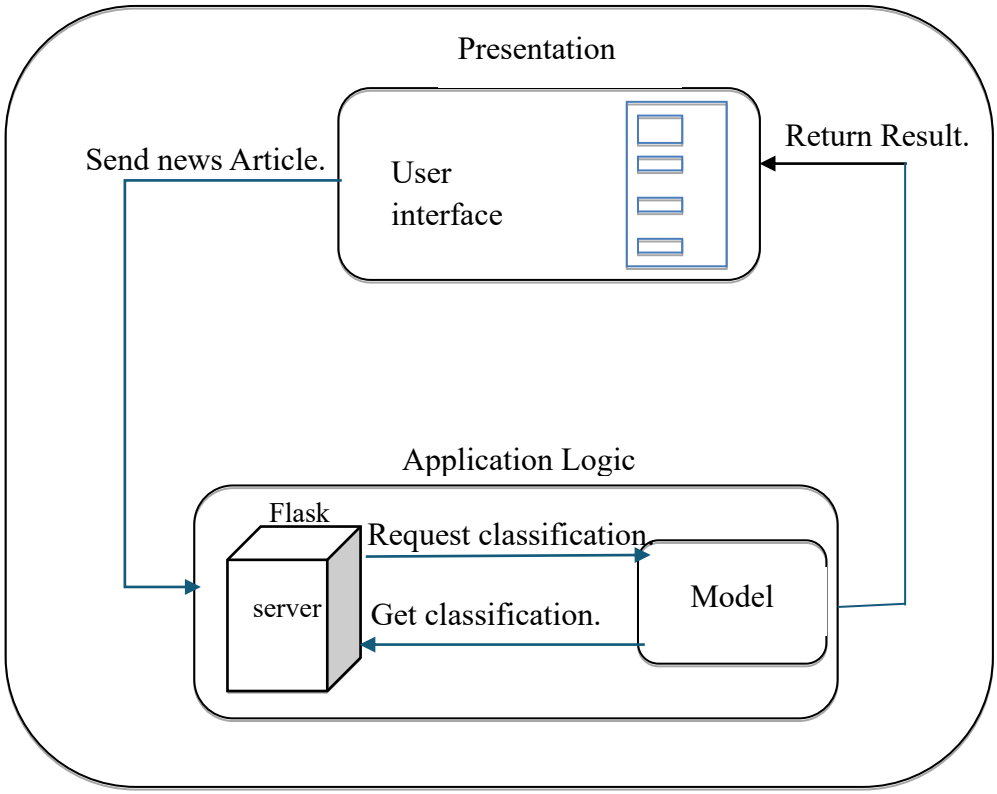

Figure 4.1: System Implementation Architecture

Algorithm Implementation

Several stages of the news classification algorithm are data preprocessing, the feature selection, the model learning, and assessment of its performance. In this section, steps taken throughout the entire process of building classification model, and the methods and libraries used to create same are discussed. The first step in the process of preprocessing is the loading of the dataset that consists of the Nigerian news articles into the DataFrame. This is essentially the first step when working with textual data towards building a model: data pre-processing. To begin with, any non-alphabetical characters are omitted, and all characters are transformed to lower case to cover all letter case differences. After that, Preprocessing is done by excluding the common words which don't play the vital role in creating prediction model and these words referred as stop-words. Also, words are stemmed where applicable by removing prefixes and suffixes to transform words into their

ISSN 2278-2540 | DOI: 10.51583/IJLTEMAS | Volume XV, Issue VII, July 2026

stem or root form in preparation for analysis. These are the preprocessing techniques that are used to ready data from text for use in feature descriptions.

In the process of text analytics, the next step to be performed after preprocessing of text data is feature extraction. This process is relevant because it transforms the text-based data generated during data cleaning into numerical formats that are usable for feeding in a machine learning algorithm. This is done using the CountVectorizer, a tool from the scikit-learn toolkit for python that transforms text documents into a matrix of token counts through a bag-of-words model. This model is capturing the counts of each word in the dataset and hence provides the necessary representation of the textual value. To successfully input, train and apply the model on the specified task, textual data have to be preprocessed and converted into numerical values. After performing feature extraction, the dataset is divided into two; for training and testing. Different classification techniques are then used for the training set. The following models were also used: LR, RF, MNB, SVC, DTC, KNN Classifier, GNB Classifier. On the same basis, each model is trained with the training data and then checked with the help of testing data to determine how well it has worked. The specific evaluation metrics used include accuracy, precision, recall and F1 score which together give a complete picture of the performance of the models and their ability to accurately and confidently predict instances of cyber security threats.

In evaluating the models, each of them is given standard parameters which are evaluated to determine the performance of each model. In trained models, the Support Vector Machines (SVM) hclassifier was found to be the best model, showing highest accuracy among other models. However, due to the Random Forest model's advantage in accuracy, the model was taken for implementation. As a final step of preparing the model for real-life application, the Python's pickle module was used to save the model, which quickly and easily allows the model to be loaded and used for making further predictions from new data. Moreover, the pipeline included the feature extraction model (CountVectorizer) for further reusing the same preprocessing steps with new input.

Exploratory Data Analysis

Exploratory data analysis (EDA) serves to get a first look at the main properties of the used dataset and an understanding of what can be observed by making use of graphical visualization. It is useful for visualization and further analysis of the data, that is required for further preprocessing for building models. In order to get a better understanding to what is in this dataset, a pie chart is built to show the percentage of the articles in each category. This chart makes a similar input to the one made by the bar chart, creating the impression of a relatively balanced number of categories with minuscule fluctuations. The Percentage of the articles in each category is depicted in chart 4.1

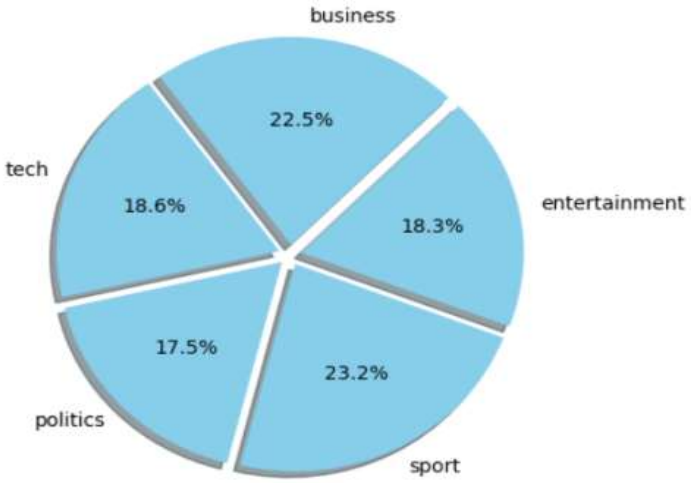

Figure 4.2: Percentage of the articles in each category

ISSN 2278-2540 | DOI: 10.51583/IJLTEMAS | Volume XV, Issue VII, July 2026

Moreover, the word cloud for each of the category is also plotted. Word clouds Figure 4.3, presents the most often used words in the news articles of each category, demonstrating what it is more continuously argued and which terms are more frequent. The following are word-clouds derived from the texts after having eliminated stop words and other words with low semantic content, so as not to distort the word-clouds.

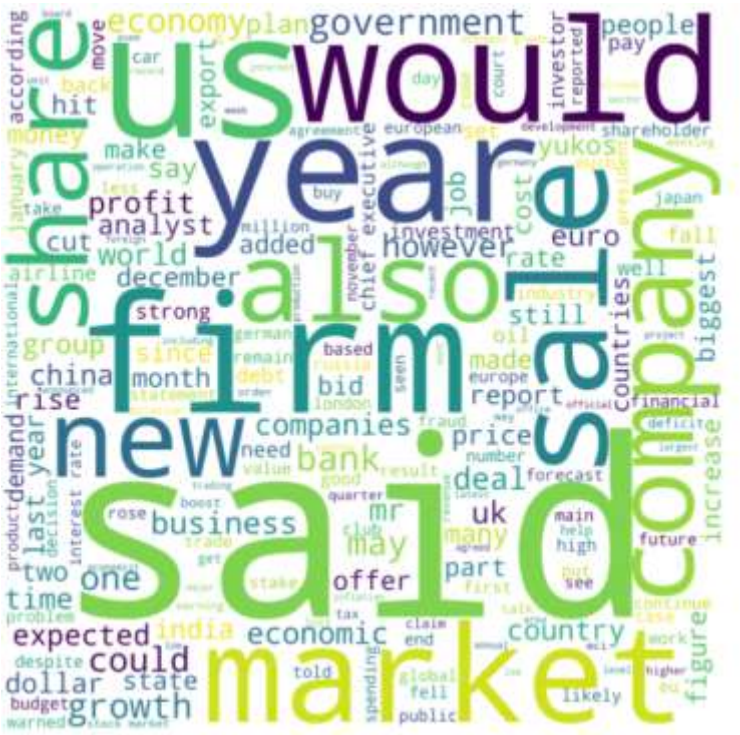

Figure 4.3: word cloud of text

Model Development and Testing

This step involved the development, training and selection of various models which can work best in this classification task. The models implemented in the research investigation were: Logistic Regression, Random Forest, Multinomial NB, SVC, DTC, KNN, and Gaussian Naive Bayes. The dataset was split into the training set and the test set with the ratio of 70:30 so it became possible to gauge the models on new data. The training set contained data used for model training, and the test set was utilized in the assessment of the models' performance. The performance of each model was assessed using several metrics: It includes the following metrics: accuracy, which is the ratio between the number of correctly forecasted instances and the total number of instances; precision, which is the ratio between the number of instance that predicted to be actually positive and all the instances that involves with positive prediction; recall, which is the ratio between correct positive prediction to the all instances of the actual class; and F1 score, which is the weighted average of precision and recall.

Table 2: Results of percentage Accuracy, Precision, Recall and F1 Score for the various models tested.

Model	**Test Accuracy (%)**	**Precision**	**Recall**	**F1 Score**
Logistic Regression	95.76	0.96	0.96	0.96
Random Forest	94.42	0.96	0.96	0.96
Multinomial Naive Bayes	95.98	0.96	0.96	0.96
Support Vector Classifier	96.21	0.94	0.94	0.94
Decision Tree Classifier	79.91	0.80	0.80	0.80

ISSN 2278-2540 | DOI: 10.51583/IJLTEMAS | Volume XV, Issue VII, July 2026

K Nearest Neighbour	72.10	0.72	0.72	0.72
Gaussian Naive Bayes	79.69	0.80	0.80	0.80

Table 2 presents the highest accuracy of 96.21 percent, and the model used was the Support vector Machines (SVM). On this basis, the program can be deemed to be more accurate than the current commonly used system in classifying the news articles.

Software Testing

The system was tested using unit testing, system testing and integration testing

Unit Testing

Testing that is done at this level involves testing of the individual whole parts of a software system

System Testing

System testing is an assessment of the developed software as a whole system with the view to determining level of conformity with the set specifications.

Integration Testing

Integration testing targets making sure that the components or the subsystems of the system interact in an acceptable manner. integration tests also verified the communications between the classification models and the performance measures to confirm that measured the predicted results correctly. When the integration testing was conducted, various problems concerning module interfaces and data translation and compatibility were discovered before the final integrated system was established.

Sample Screen Shots

This section aims to give the students the sample graphic illustrations on the ways the news classification system

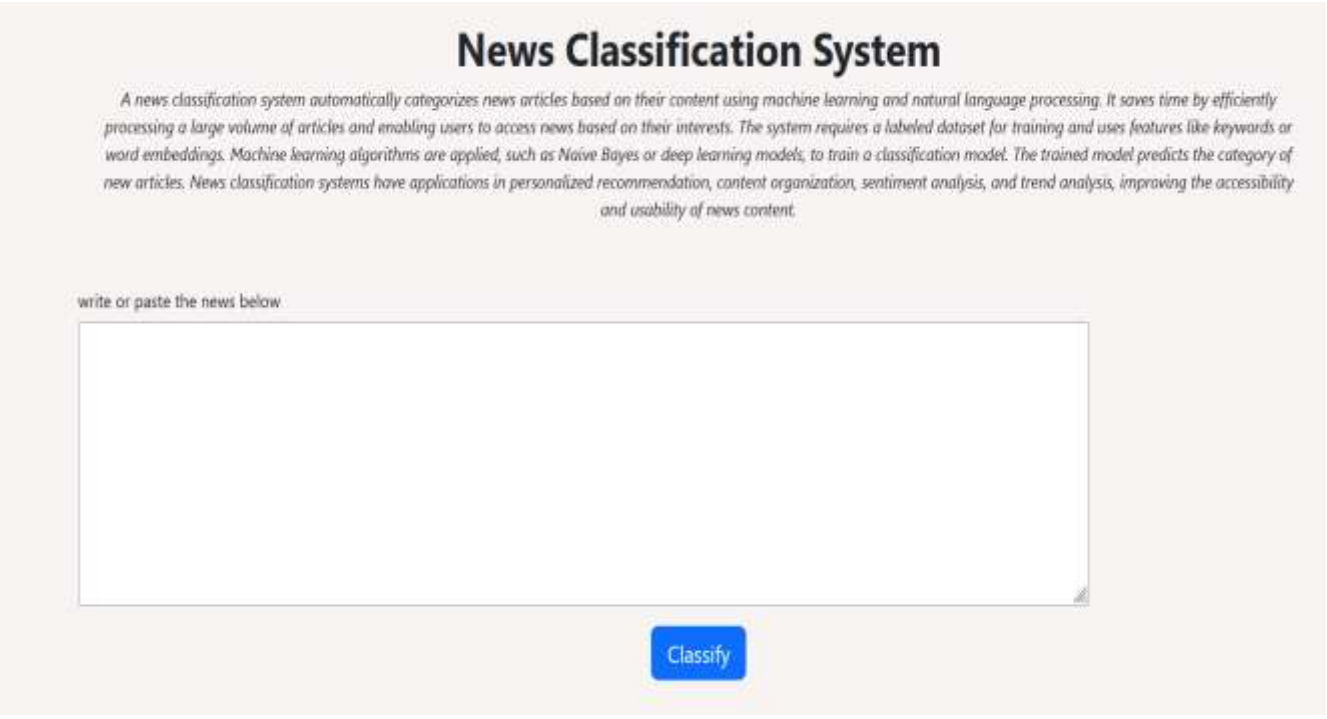

Figure 4.4: main interface

ISSN 2278-2540 | DOI: 10.51583/IJLTEMAS | Volume XV, Issue VII, July 2026

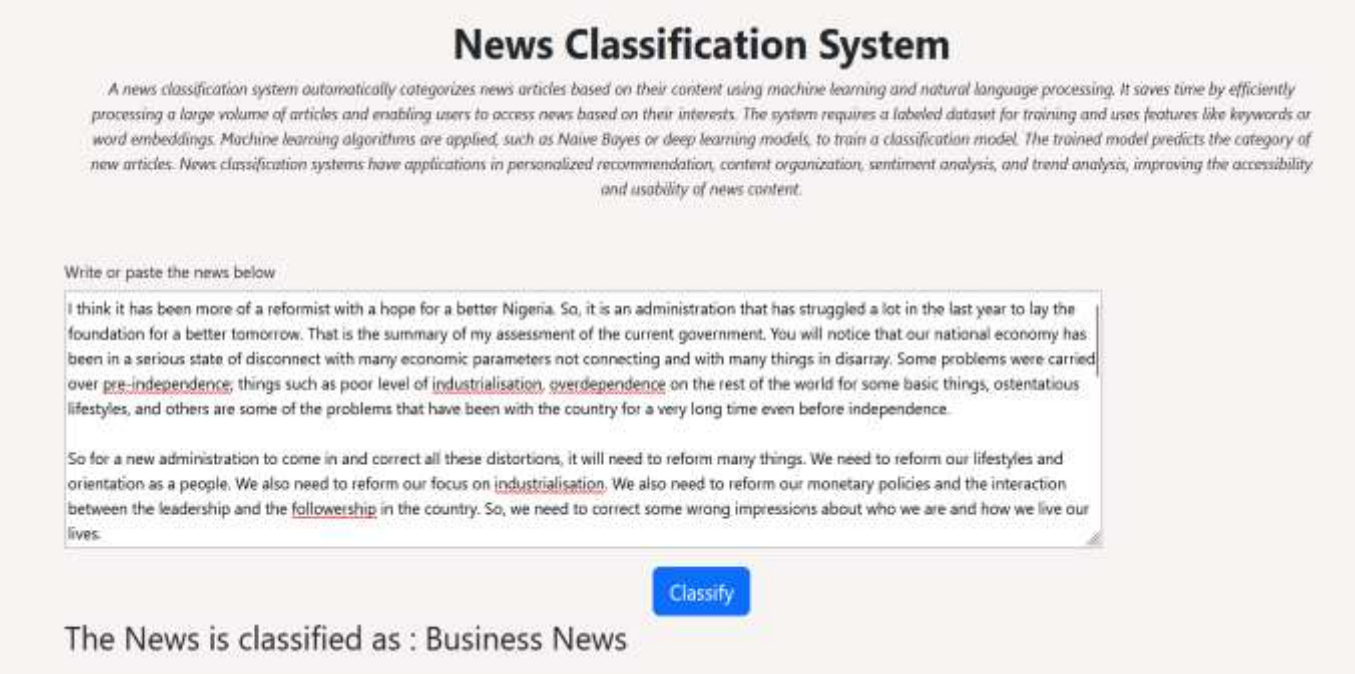

Figure 4.5: main interface with news classified.

Figure 4.5 shows a sample snippet of news article pasted for classification.

SUMMARY

This research presents the development and evaluation of a news classification system aimed at categorizing news articles into predefined categories using various machine learning algorithms. The process was structured, encompassing data collection, preprocessing, model training, evaluation, system testing, and documentation. The dataset utilized consisted of 1,493 news articles classified into different categories, such as, business, politics, tech, sport, entertainment etc. Data cleaning and preprocessing involved the removal of tags, special characters, and stop-words, conversion of text to lowercase, and lemmatization to standardize the words. Exploratory Data Analysis (EDA) provided insights into the dataset's structure and distribution through visualizations like pie charts, and word clouds, illustrating the most frequent words in each

CONCLUSION

In conclusion, this research has successfully classified Nigeria newspapers. By employing a detailed preprocessing pipeline that included text cleaning, tokenization, and lemmatization, we ensured that the input data was of high quality. Rigorous model evaluation was conducted, and among several models tested, the Support Vector Machines (SVM) classifier emerged as the most effective, achieving an impressive accuracy of 96.21%. This high performance underscores the importance of thorough data preparation and the selection of appropriate algorithms in machine learning projects.

RECOMMENDATIONS

Based on the findings and conclusions of this study, several recommendations are proposed for future work and practical implementation. Firstly, future research should explore the integration of advanced natural language processing techniques, such as, BERT or GPT, to potentially enhance the classification accuracy and capture more nuanced contextual information within news articles. These models have shown great promise in various text classification tasks and could further improve the performance of the system developed in this study. Secondly, expanding the dataset to include a more various news sources and categories could provide a broad gauge of the model's generalizability and robustness. This would ensure that the classification system performs well across different types of news content and is not biased towards the specific dataset used in this research.

ISSN 2278-2540 | DOI: 10.51583/IJLTEMAS | Volume XV, Issue VII, July 2026

REFERENCES

1. Ahmed J. and Muqeem A. (2021) Online News Classification Using Machine Learning Techniques.
2. Alaa Tharwat, Tarek Gaberc, Abdelhameed Ibrahim, and Aboul Ella Hassaniene, (2017), Linear discriminant analysis: A detailed Tutorial a Department of Computer Science and Engineering, Frankfurt University of Applied Sciences, Frankfurt Main, Germany, A research paper, AI Communications00 (20xx) 1–22 1 DOI 10.3233/AIC-170729 IOS Press, pp1 – 22
3. Alexander Berkovich, (2025), Understanding Model Evaluation Metrics for Image Classification, 175 S San Antonio Rd SUITE 102, Los Altos, CA 94022, United States, https://akridata.ai/blog/understanding-model-evaluation-metrics-for-image-classification/
4. Babalola, E.A. (2002). Newspapers as instruments for building literate communities: The Nigerian experience. *Nordic Journal of African Studies* 11(3), 403-410.
5. Bahiyah O. and Oberiri D. A. (2020) Newspaper Readership Pattern Among Nigerian University Students: Perspectives from Mass Communication Students; Library Philosophy and Practice (e-journal)
6. Dadgar S. M. H., Araghi M. S., and Farahani M. M. (2016) "A novel text mining approach based on TF-IDF and Support Vector Machine for news classification," in *2016 IEEE International Conference on Engineering and Technology (ICETECH)*, Coimbatore, India, Mar. 2016, pp. 112–116, doi: 10.1109/ICETECH.2016.7569223.
7. Egwom, O.J., Hassan, M., Tanimu J.J., Hamada, M., Ogar, O.M. (2022) an LDA–SVM Machine Learning Model for Breast Cancer Classification. Biomedinformatics 2022,2, 345–358. https://doi.org/10.3390/biomedinformatics2030022
8. Fesseha A., Xiong S,. Emiru E. D., Dahou A. (2020) Text classification of news articles using machine learning on low-resourced language: Tigrigna. In Proceedings of the 3rd International Conference on Artificial Intelligence and Big Data (ICAIBD), Chengdu, China, pp 34-38. https://doi.org/10.1109/ICAIBD49809.2020.9137443.
9. Hartmann J., Huppertz J., Schamp C., and Heitmann M., (2019) "Comparing automated text classification methods," *Int. J. Res. Mark.*, vol. 36, no. 1, pp. 20–38, 2019.
10. Huang C. M, Jiang Y. J. (2019) An empirical study on the classification of Chinese news articles by machine learning and deep learning techniques. In Proceedings of the International Conference on Machine Learning and Cybernetics (ICMLC), Kobe, Japan, pp 1-6. https://doi.org/10.1109/icmlc48188.2019.8949309
11. Huang M. W., Chen, C. W., Lin, W. C., Ke, S. W., & Tsai, C. F. (2017). SVM and SVM ensembles in breast cancer prediction. *PLoS ONE*, *12*(1), 1–14. https://doi.org/10.1371/journal.pone.0161501
12. Jang B., Kim I., Kim J. W. (2019) Word2vec convolutional neural networks for classification of news articles and tweets. PLOS ONE, 14(8): e0220976. https://doi.org/10.1371/journal.pone.0220976
13. Miao F., Zhang P., Jin L., and Wu H. (2018) "Chinese News Text Classification Based on Machine Learning Algorithm," in *2018 10th International Conference on Intelligent Human-Machine Systems and Cybernetics (IHMSC)*, Hangzhou, Aug. 2018, pp. 48–51, doi: 10.1109/IHMSC.2018.10117. Nigerian experience. *Nordic Journal of African Studies* 11(3), 403-410.
14. Noppakaow A, Uchida O. (2019) Examinations on the Performance of Classification Models for Thai News Articles. In Proceedings of the 2019 11th International Conference on Information Technology and Electrical Engineering (ICITEE), Pattaya, Thailand, 2019, pp. 1-4 https://doi.org/10.1109/iciteed.2019.8929959
15. Qadi L. A, Rifai H. E, Obaid S., Elnagar A. (2019) Arabic text classification of news articles using classical supervised classifiers. In Proceedings of the 2nd International Conference on new Trends in Computing Sciences (ICTCS), Amman, Jordan, pp 1-6. https://doi.org/10.1109/ICTCS.2019.8923073.
16. Rabbimov I. M, Kobilov S. S. (2020) Multi-class text classification of Uzbek news articles using machine learning. Journal of Physics: Conference Series, 1546(1): 012-097.
17. Rajbharath R., and Sankari L. (2017). *Predicting Breast Cancer using Random Forest and Logistic Regression*. *7*(4), 10708–10713

ISSN 2278-2540 | DOI: 10.51583/IJLTEMAS | Volume XV, Issue VII, July 2026

18. Winster S. G, Kumar M. N. (2020) Automatic classification of emotions in news articles through ensemble decision tree classification techniques. Journal of Ambient Intelligence and Humanized Computing, 1–12. https://doi.org/10.1007/s12652-020-02373-5