Bima Journal of Science and Technology, Vol. 9(1A) Mar, 2025 ISSN: 2536-6041

DOI: 10.56892/bima.v9i1A.1233

Social Media Fake News Detection Model Using Support Vector Machine

Samson Isaac[1]*, Galadima Mihasdyenum Peter[1], Christopher Habu[2], Saratu Habu[3], Juliet David Shekarau[3] and Thomas Lass Barna[3]

[1]Department of Computer Science, Kaduna State University, Nigeria
[2]Department of Industrial Chemistry, Mewar International University, Abuja, Nigeria
[3]Department of Computer Science, Kaduna State College of Education, Gidan-Waya

Corresponding Author: samson.isaac@kasu.edu.ng

ABSTRACT

News is information on societal events currently taking place, and it can be distributed via various social media channels. Fake news is information that has been deliberately manufactured to deceive readers or listeners. It is a form of propaganda that is disseminated under the guise of real news, which can occasionally be hard to distinguish from false news. Social media and conventional news media are both used to promote fake news. The research adopts Support Vector Machine (SVM) algorithm to detect fake news. The SVM algorithm, is a tool for guided learning which can be applied to a variety of tasks based on the categorization of the data. Support Vector Machine is a fast and reliable algorithm that enhances the process for the detection of fake reviews The evaluation metric of the proposed model is accuracy and the proposed model achieved 99.67% accuracy

Keywords: Fake news, Real News, Support Vector Machine, Recurrent Neural Network and Long Short Time Memory

INTRODUCTION

For a very long time, fake news has been a concern. It has been more challenging to distinguish between real news and fake news since the advent of social media, which has amplified the propagation of fake news. Concern should be expressed about the propagation of false information that sways public opinion. By the end of the 2016 American presidential election, it was projected that over 1 million tweets had been tied to "Pizza gate" fake news. The widespread dissemination of false information may have a very harmful effect on both people and society at large. Claire (2017) identifies seven types of fake news they include Satire or parody, false connection, misleading content, false context, Impostor content, Manipulated content and Fabricated content. Social media is a digital tool that enables users to produce and disseminate material to the general audience. Numerous websites and applications fall under the category of social media, such as Twitter, Facebook, Instagram, and WhatsApp. Social media is widely used and mostly unrestricted.(Hudson, 2020). The study of computer algorithms that evolve spontaneously with use is known as machine learning (ML).(Mitchell, 1997). It is seen as a subset of artificial intelligence. Without being expressly taught to do so, machine learning algorithms create a model using sample data, also referred to as "training data", in order to make predictions or judgments. In a wide range of applications, including email filtering and computer vision, when the development of traditional algorithms to carry out the required tasks is challenging or impractical, machine learning techniques are used. While computational statistics, which focuses on using computers to make predictions, is closely related to a subset of machine learning,

DOI: 10.56892/bima.v9i1A.1233

not all machine learning is statistical learning. The field of machine learning benefits from the tools, theory, and application domains that come from the study of mathematical optimization. Data mining is a related area of study that focuses on unsupervised learning for exploratory data analysis.(Friedman, 2019). Supervised learning is a machine learning task of developing a function that translates an input to an output based on sample input-output pairs. From labeled training data made up of a collection of training instances, it infers a function. Each example in supervised learning is a pair, consisting of the intended output value and an input object (usually a vector) (also called the supervisory signal). Brownlee (2019) defined Natural language processing is the computer-controlled use of everyday language, much like how programming controls discourse and writing. With the development of computers, the study of natural processing, which has been around for more than 50 years, has outgrown the study of linguistics. A platform called the fake news detection system uses machine learning to identify bogus news and is regarded as a change-making force. This will help to eliminate fake news to the bare minimum with the aid of machine learning

RELATED WORKS

Feng et al. (2017) proposed a novel Two-level convolutional neural network with user response generator where the TCNN captured the semantic information from article text by representing it at the sentence and word model of user response to article text from historical user responses. Eugenio et al. (2017) to new articles in order to assist fake news detection. They made use of one dataset for the detection of fake news. The experimental result showed that TCNN-URG outperforms the baselines based on prior approaches that detect fake news from article text alone. presented two classification techniques, the first based on the logistic regression and the other on a novel adaptation of boolean crowdsourcing algorithms. Using a data set consisting of 15,500 Facebook posts and 909,236 users an accuracy exceeding 99% was obtained (Eugenio et al 2017). William (2017) presented a data set titled liar for fake news detection. Automatic Fake news detection was investigated based on surface level linguistic patterns. In the article a novel and convolutional neural network was designed to investigate data with text. The article showed that when combining meta-data with text significant improvements can be achieved for fine-grained fake news detection. The data set can be used for stance classification, argument mining, topic modelling and rumor detection.

In the article, various models were compared using text features only. The data set gives about 0.204 and 0.208 accuracy on the validation and test sets. SVMS models obtained significant improvements. In Thota et al (2018) neural network architecture was presented to accurately predict the stance between a given pair of headline and article body. In the article the focus of the study was to make use of stance detection for the detection of fake news. The dataset used in the article was (FNC-1) data the data set is split into train, validation and test splits. In the paper the following word vector representations were used to experiment Bag of words, TF-IDF, GloVe and Word2Vec. Using the TF-IDF-dense neural network the article achieved an accuracy of 94.21% was achieved on test data. The bag of words was the second performing model in the article. Kelly (2018) discussed on linguistic cue and network analysis approaches and proposed a three-part method using Naïve Bayes classifier, support vector machines and semantic analysis as a way of detecting fake news accurately on social media.

DOI: 10.56892/bima.v9i1A.1233

Okoro et al (2018) proposed a hybrid model for social media fake news detection using both the human based and machine based detection approaches. The human based social media news literacy education tool was combined with the machine based approaches for linguistic and network analysis. Two parallel approaches of detection were put to work each helping to provide a balance for the other. Karimi et al. (2018) proposed a Multi-Source Multi-class Fake News Detection framework MMFD which combines automated feature ex- traction, multi-source fusion and automated degrees of fakeness detection into a logical and interpretable model. Multi-source Multi-class Fake News Detection framework (MMFD). MMFD incorporates three coherent components into an end- to-end way automated feature extraction, interpretable multi-source fusion, and fakeness discrimination.

Monther & Ali (2018) proposed a simple but effective approach that allows users to install a tool into their browser which is used to detect click baits. They made use of simple and carefully selected features of the title and post to identify fake posts accurately. They tested their approach on Bayes net, logistic, random tree and Naïve Bayes. Their result showed a 99.4% accuracy using logistic classifier. In Georgios et al. (2019), an implementation of enhanced linguistic feature set with word embedding along with ensemble algorithms and support vector machines was used to classify fake news. The article made use of content based features and machine learning algorithms. Several feature sets proposed were evaluated for deception detection and word embedding. The most popular machine learning classifiers were tested and ensemble ML methods such as Adaboost and bagging were investigated to see the possible improvements reached. The article introduced a new text body the “Unbiased dataset” which merges and fulfills various news sources and standards and rules to avoid biased results in classification task.

In the article the features combined with machine learning algorithms obtained accuracy up to 95% over all datasets used with the Adaboost to be first in rank and SVM and bagging algorithms to be next in ranking. In the research in both datasets the fake news class is the same while what they are vary only in the real news sources, it is believed that this difference is an indicator of bias risk concerning the use of only one source either for real or fake news articles. A possible improvement in the future would be to employ several meta-data about the source and the author of news, along with social media information diffusion features and deep learning methods with larger datasets should be used. Subhadra et al. (2019) proposed a new model for the detection of fake news using Naïve Bayes classifier. The model takes news events as input and based on reviews on twitter and classification algorithms it predicts if the news is fake or real.

Vasu et al. (2019) examined the approach of natural language processing and machine learning in solving the problems of fake news they made use of bag of words, n-grams, count vectorizer and TF-IDF and five classifiers to see which of them works well for the dataset used. Their results showed that SVM and logistic regression classifier have the best performance on the dataset. Limeng et al. (2019) presented a system for fake news detection called defend, which can detect authentic news by identifying user comments that can explain why the news is fake or real. Neural networks were used as the computing methodology. The system provides check-worthy scores and explainable results at the same time. Xinyi and Reza (2019) proposed a network based pattern-driven fake news detection approach. different patterns of fake

DOI: 10.56892/bima.v9i1A.1233

news in social networks were studied and represented into various network levels for being further utilized to detect fake news. The experiments were conducted on two public benchmark datasets of fake news detection.

Olivieri et al. (2019) proposed a methodology to create task-generic features that are paired with textual features in order to detect fake news. Task-generic features were created by elaborating on metadata attached to answers from Google's search engine, and they made use of crowdsourcing for missing values. The method was tested on a dataset for detection of fake news based on the PolitiFact website. their results showed an improvement in F1-Score of 3% over the state of the art, which is significant for a 6-class task. Algorithms: Multinomial Naive-Bayes (MNB) with basics scikitlearn settings, Support Vector Machine (SVM) with polynomial kernel, and Neural Networks (NN) with 2 hidden layer of 25 neurons each and Adam as weight optimization solver. Devi et al. (2019) came up with the application of some NLP (Natural Language Processing) techniques for the detection of fake news which is news published in order to misguide the readers, they made use of Naïve Bayes classifier, logistic regression, support vector machine and random forest classifier. Dataset was collected from various legitimate news domains. Therefore, the research gives proper and appropriate results for the test data present in the dataset.

The future scope of the methodology is to associate it with internet news so as to give appropriate result even for test data not present in the dataset. Amjad et al. (2020) investigated whether machine translation at its present state could be successfully used as an automated technique for annotated corpora creation and augmentation for fake news detection focusing on the English-Urdu language pair. fake news classifier was trained for Urdu on the manually annotated dataset originally in Urdu and the machine-translated version of an existing annotated fake news dataset originally in English. Rutvik et al. (2020) developed a two stage automated pipeline for detection of COVID-19 fake news using machine learning models for natural language processing and transformer models The first model leverages a novel fact checking algorithm that retrieves the most relevant facts concerning user claims about particular COVID-19 claims. The second model verifies the level of "truth" in the claim by computing the textual entailment between the claim and the true facts retrieved from a manually curated COVID-19 dataset. Results showed that the transformer based models are significantly better than classical NLP models (Rutvik et al., 2020).

IFtikhar et al. (2020) proposed a machine learning ensemble approach for automated classification of news articles. Different machine algorithms were trained and tested on four real world datasets. The article made use of ensemble techniques with diverse linguistic feature sets to classify different articles from different domains as real or fake. In the article machine learning algorithms were trained with different hyperparameters to achieve maximum accuracy for a given dataset, with an optimal balance between variance and bias. Each model was trained multiple times with a set of different parameters using a grid search to find the best parameters is computationally expensive however the measure is taken to ensure the models do not overfit or underfit the data logistic regression, support vector machine, ensemble learners was used and performance metrics. In the article an accuracy of 97.67% was attained by ensemble learners. Bagging classifier and boosting classifier were the best performing algorithms achieving accuracy of 94%. Kasra et al. (2020) proposed a feature selection method with the integration

DOI: 10.56892/bima.v9i1A.1233

of K-means clustering and support vector machines (SVM) which work in four steps.

First of all, the similarities between all features are calculated. Then the features are divided in to several clusters. The final feature set is then selected from all clusters. Finally, fake news is classified based on the final feature subset using the SVM method. In the article the detection performance was improved in two aspects. First the detection runtime process decreased, secondly the classification accuracy increased because of the elimination of redundant features and the reduction of datasets dimensions. K-means and SVM were used for feature selection and SVM classifier for the fake news detection. The article focused on social media news, the methods can be extended or tested on other news sources like blogs and other online forums. Das, Basak, &

Dutta (2021) described their fake news detection system which automatically detects whether tweets related to COVID-19 are either real or fake. In this paper they made use of an ensemble model consisting of pre-trained models. They approached the task as a classification problem which means each news item to be classified as either fake or real. The method used consist of five main parts text processing, tokenization, backbone model, ensemble and heuristic processing. The system achieved an f1-score of 0.9883 by improving their system with a novel heuristic algorithm based on username handles and link domains in tweets. Gundapu & Mamidi (2021) observed that social networks such as twitter has led to a tendency of the spread of fake news and false information due to fast technological advancement. This spread of fake news is especially frequent in the on-going COVID-19 pandemic. They made use of an approach based on an ensemble of three transformer models (BERT, ALBERT AND XLNET). The model was trained and evaluated in the context of the Constraint AI 2021 shared task COVID19 Fake News Detection in English. The system obtained a result of 0.9855 f1-score on test set and had a ranking of 5th among 160 teams.

Gautam & Masud (2021) did an extensive research due to the rapid dissemination of information and its easy access on the internet it is difficult to separate fake news from real news in this paper they made use of a novel method they introduced an approach to combine topical distributions from Latent Dirichlet Allocation(LDA) with contextualized representations from XLNet. By comparing the method used with existing baselines they were able to show that XLNet + topic distributions outperform other approaches by getting an f1-score of 0.967. Saleh, Alharbi, & Alsamhi (2021) proposed novel approaches based on machine learning and deep learning to address the issue of fake news. They proposed an optimized convolutional neural network model for the detection of fake news(OPCNN-FAKE). The performance of the OPCNN-FAKE was compared with recurrent neural network(RNN), long-short term memory(LSTM) and six regular other ML techniques. Accuracy, precision, recall and f1-measure were used to evaluate the performance of the OPCNN-FAKE. The OPCNN-FAKE achieved the best performance for each dataset compared with other models.

Mahlous & Al-laith (2021) addressed the problem of detecting fake news concerning COVID-19 in arabic tweets. More than seven million Arabic tweets using trending hashtags were collected relating to the COVID-19 pandemic from January 2020 to august 2020. A small corpus was extracted from the tweets and was manually annotated and was used as a baseline to build a system for the automatic detection of fake news. The manually annotated database was able to achieve an f1-

DOI: 10.56892/bima.v9i1A.1233

score of 87.8% using logistic regression(LR) and term frequency-inverse document frequency(TF-IDF) and an f1-score of 93.3% using same classifier with count vector feature Sanovsky, Maslej-kresnakova, & Invancova (2022) put their focus on the detection of fake news automatically using deep learning models in the slovak language. Data were collected from various online news sources related to the COVID-19 pandemic and those data were used to train and evaluate some deep learning models. With the combination of bidirectional long-short term memory network with one-dimensional convolutional layers an average macro F1-score on an independent test set of 94% was achieved.

Shetty et al. (2022) demonstrated a web application for the analyzation and the detection of fake news. (Isaac et al., 2024) the proposed web application uses LSTM, CNN and hybrid models for the classification, they made use of a web mining system to inculcate auxiliary facts for the user and an OCR module for text extraction for pieces of news transmitted through images. The neural network provided a satisfactory classification over a generalized news dataset and the web application provides useful features in addition to providing a classifier.

MATERIALS AND METHODS

Machine learning algorithms were used to develop all experimental configurations at this stage. The data description explains several fake new problems and associated attributes and visualizes their distribution using Python classification tools. This explanation describing the system architecture is represented in Figure 1.

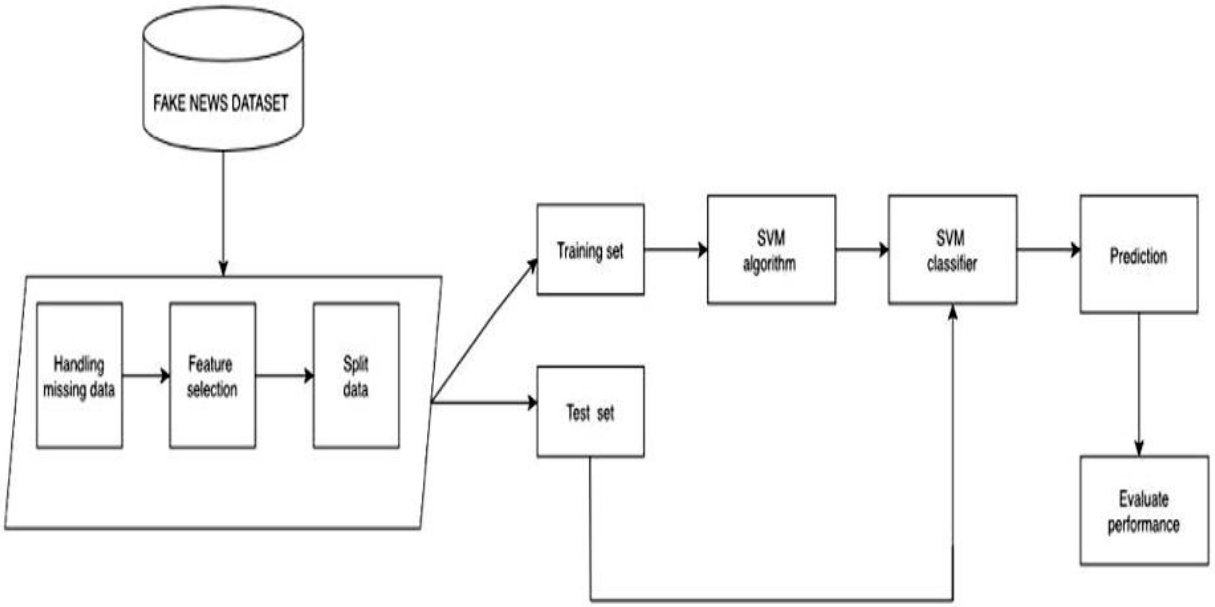

Figure 1: proposed system Architecture

Linear Support Vector Classification is the SVM library obtained from Scikit-learn and was utilized as the primary approach for training the model. LinearSVC is a variant of SVC with the kernel 'linear' parameter. However, it is implemented in terms of liblinear rather than libsvm, giving it more flexibility in terms of penalties and loss functions and the ability to scale to vast numbers of samples. Before the linguistic features were extracted using the Natural language toolkit, the data set underwent data

DOI: 10.56892/bima.v9i1A.1233

cleaning and investigation after it had been acquired. Eighty percent (80%) of the extracted linguistic features were used for the training data and twenty percent (20%) for the testing data and evaluate the performance. After the model requirement was met, the method finished; otherwise, the hyperparameters changed, and the model was retrained until the conditions were met. The support vector machine (SVM) is a supervised learning model with associated learning algorithms that can analyze and recognize patterns for classification and regression tasks. SVM is a binary classifier.

The basic idea of SVM was to find an optimal hyper-plane that can linearly separate instances of two given classes. This hyper-plane was assumed to be in the gap between some marginal models called support vectors. Introducing the kernel functions, the idea was extended for linearly inseparable data. The dot product of the projections of two data points in a high-dimensional space is represented by a kernel function. It is a transform that disperses data by mapping from the input space to a new (feature space) where the instances are more likely to be linearly separable. Kernels, such as radial basis function (RBF), can be used to learn complex input spaces. In classification tasks, given a set of training instances marked with the label of the associated class, the SVM training algorithm finds a hyper-plane that can assign new incoming models into one of two classes. The class prediction of each unique data point is based on which side of the hyper-plane it falls on feature space. SVM has been successfully applied to a broad range of applications.

RESULTS AND DISCUSSION

The model has now been trained to recognize fake news as well as real news. The proposed model is evaluated based on accuracy, precision, recall, F1-score, and support metrics. Additionally, a comparative analysis is carried out as shown in figure 4.

Splitting the data is the most essential step in machine learning. We train our model on the trainset and test our data on the testing set. We split our data in train and test using the train_test_split function from Scikit learn.

We split our 80% data for the training set and the remaining 20% data for the testing set as shown in figure 2.

```
In [16]: #splitting data for training and testing
         import sklearn
         from sklearn.model_selection import train_test_split
         x_train,x_test,y_train,y_test = train_test_split(data['text'],data['label'],test_size=0.2, random_state = 1)
```

Figure 2: Data Splitting.

Testing and training is a very essential part while building any model. The complete base of the model is relied on how far we train the model. Figure 3 shows confusion matrix obtained for the proposed model.

Bima Journal of Science and Technology, Vol. 9(1A) Mar, 2025 ISSN: 2536-6041

DOI: 10.56892/bima.v9i1A.1233

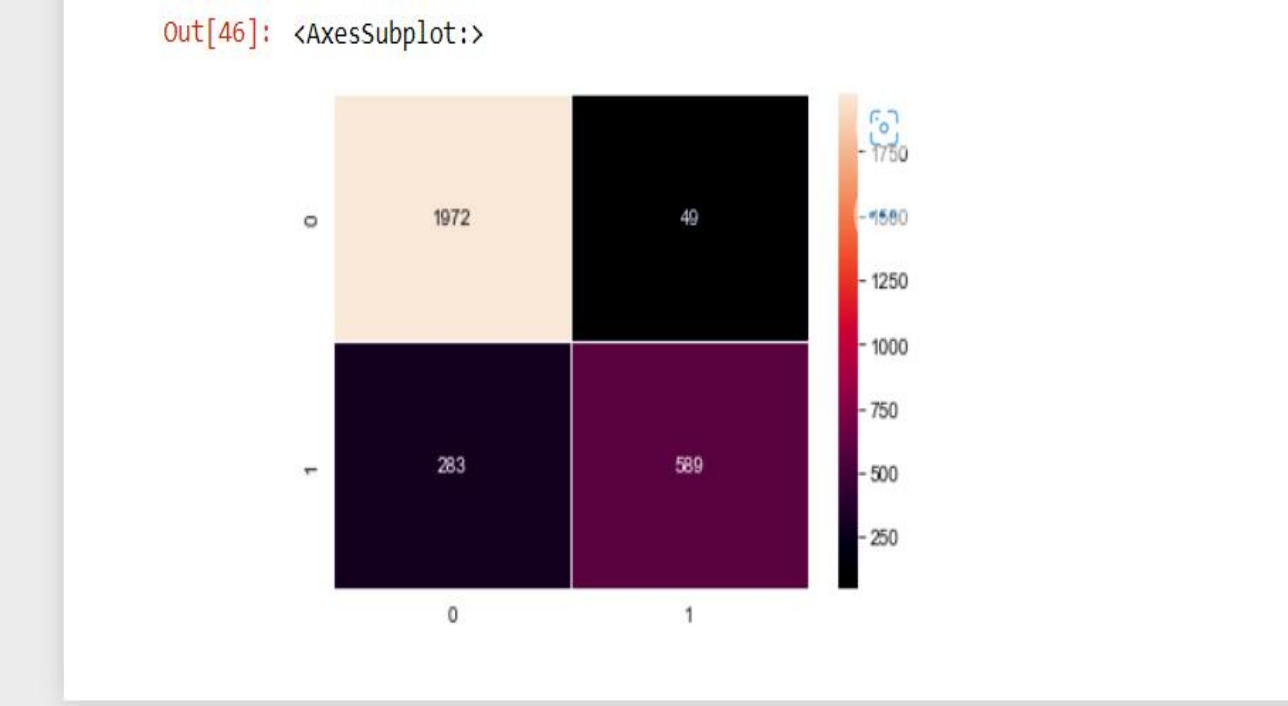

Figure 3: Confusion matrix obtained for the proposed model.

Table 1: Prediction of Text Data with the Fitted Model and Check the Accuracy (SVM)

Epoch	Precision	Recall	F1-Score	Accuracy
0.0	0.87	0.98	0.92	0.98
1.0	0.92	0.68	0.78	0.99
Macro Average	0.90	0.83	0.85	0.98
Weighted Average	0.89	0.89	0.88	0.98
Area Under the Curve = <function roc_curve at 0x0000023407D0A670>				

As shown in figure 4, The ROC (Receiver Operating Characteristic) curve is usually a good graph to summarize the quality of our classifier. The higher the curve is above the diagonal baseline, the better the predictions. Although the AUC ROC (Area Under the Curve ROC) is very good, we should not use here the ROC curve to assess the quality of our model.

The remind the False Positive Rate formula, which corresponds to the x axis of the ROC curve: FPR (False Positive Rate) = # False Positives / # Negatives.

Here the # Negatives corresponds to our number of good reviews which is very high because our dataset is imbalanced. This means that even with some False Positives, our FPR will tend to stay very low. Our model will be able to make a lot of false positives predictions and still have a low false positive rate, while increasing the true positive rate and therefore artificially increasing the AUC ROC metric.

DOI: 10.56892/bima.v9i1A.1233

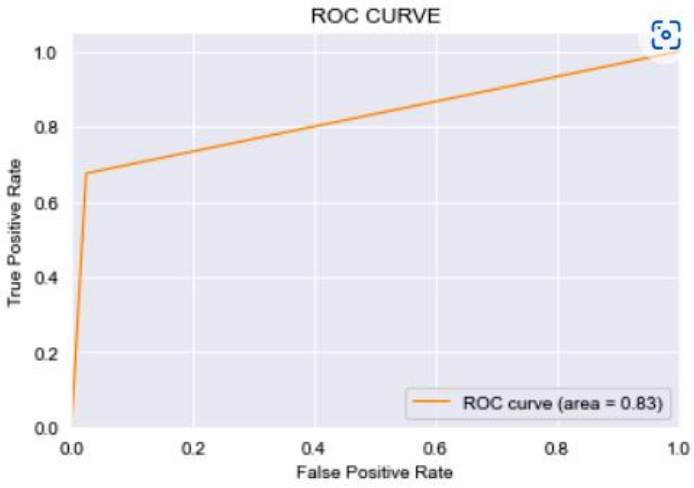

Figure 4: Receiver Operating Characteristics

CONCLUSION

The focal point of our research lies in differentiating and detecting fake news and original or real news. The goal of this study is to determine and comprehend the impact of fake news on the government and society. This research aims in helping society from the spread of fake news, making people aware of the fake news propagators and their expansion in today's world. As a result, the spread of fake news has been a hot topic for quite some time. The primary obligation for the betterment of society is to detect fake news. The proposed model used a deep learning Model, which makes use of memory architecture. This model has been presented to improve the detection of misleading news. Before training the model, the concept of stop words has been employed for data pre-processing, which enhances the accuracy of the model. A tokenization method has been proposed for feature extraction or vectorization, which assigns tokens to word embedding. The GloVe technique was used to represent each word in vector form for word embedding. The information then passes via the main SVM model, where it is routed through various levels in the architecture. The long short -term memory model is used for the construction of the model which mainly is based on the concept of classification of data for sequential prediction. Finally, the model is programmed to distinguish between true and false news. The evaluation metric of the proposed model is accuracy. Our proposed model achieved 99.67% accuracy. The future direction lies to improve and outstretch the existing work for implementation of existing work towards building an automated system for e-commerce websites, where detection of fake news has become equally important.

REFERENCES

Anita, S. K., & Mogalla, S. (2019). Vectorization of Text Documents for Identifying Unifiable Articles. jornal of advanced computer science apllications, 10(7), 6. Retrieved from http://www.ijaca.theai.org

Brian, W. K., & Sawyer, S. C. (2011). a practical introduction to computers & communication. New york: McGraw-hill.

Claire, W. (2017, February 16). Fake news it's complicated. First Draft, 1-11. Retrieved from firstdraftnews.org

DOI: 10.56892/bima.v9i1A.1233

Devi, C., Priyanka, R., Surendra, P. S., Priyanka, B. S., & Nikhila, C. (2019, April). Fake News Detection Using Machine Learning. Journal of Emerging Technologies and Innovative Research, 6(4), 1-6.

Dylan, B. d., & Machdel, M. (2020, May 5). Approaches to identify fake news: A systematic literature review. integrated science in digital age, 1-10. doi:10.1007/978-3-030-49264-9_2

Das, D. S., Basak, A., & Dutta, S. (2021, January 10). A Heuristic-driven Ensemble Framework for COVID-19 Fake News Detection. 1, 1-13.

Dey, A. (2016). Machine learning algorithms: A Review. International journal of computer science and information technologies., 7(3), 1174. ISSN (Online): 0975-9646

Eugenio, T., Gabriele, B., Marco, V. L., Stefano, M., & Luca, d. (2017, April 25). Some like it Hoax: Automated Fake News Detection in Social Networks. arXiv, 1, 1-12.

Ertel, W. (2017). Introduction to artificial intelligence Springer, Cham. 2(356), 267. 10.1007/978-3-319 58487-4

Feng, Q., Chengyue, G., Karishma, S., & Yan , L. (2018). Neural user response generator: Fake news detection with collective user intelligence. International Joint Conference on Artificial Intelligence, 18, pp. 1-7. california.

Friedman, J. H. (2019, November). data mining and statistics whats the connection. Computing Science and Statistics, 1, 1-8.

Georgios, G., Athena, V., Konstantinos, D., & Panagiotis, K. (2019, March 21). Behind the cues: A benchmarking study for fake news detection. Expert systems with applications, 1-13.

Gundapu, S., & Mamidi, R. (2021, January 21). Transformer Based Automatic COVID-19 Fake News Detection System. International Institute of Information Technology, Hyderadab, 3, 1-12.

Gautam, A., V, V., & Masud, S. (2021, January 12). Fake News Detection System using XLNET model with Topic Distributions. 1, 1-12.

Hudson, M. (2020). Definition and exampkes of social media. doi:https://www.thebalancemb.com

Isaac, S., Ayodeji, D. K., Luqman, Y., Karma, S. M., & Aminu, J. (2024). Cyber Security Attack Detection Model Using Semi-Supervised Learning. Fudma Journal of Sciences, 8(2), 92-100.

Jonathan, O. A., & Steve, W. (2015, July 22). Social Media Definition And The Governance Challenge: An introduction to the special issue. Special issue Telecommunications Policy, 39(9), 1-21. doi:10.1016

Kelly, S. (2018, May 15). Fake news detection in social media. 1-6. Retrieved from http://www.csustan.edu

Kot, J. (2020, October 6). A brief history of machine learning - Concise software. Retrieved from https://concisesoftware.com/history-of-machine-learning/ on March 3, 2020

Limeng, C., Kai , S., Dongwon , L., Huan, L., & Suhang, W. (2019, November 3). A system for explainable fake news detection. 28th ACM International Conference on Information and Knowledge management, (p. 4). Beijing. doi:10.1145/335

Lutins, E. (2017, August 2). Ensemble methods in machine learning: what are they and why use them? Retrieved from

DOI: 10.56892/bima.v9i1A.1233

https://towardsdatascience.com/ensemble-methods-in-machine-learning-what are-they-andwhy-use-them-68ec3f9fef5f on May 3, 2021

Mahlous, A. R., & Al-Laith, A. (2021, July 8). Fake News Detection in Arabic Tweets during the COVID-19 Pandemic. International Journal of Advanced Computer Science and Apllications, 12(6), 1-11.

Mitchell, T. M. (1997). Machine learning (illustrated ed.). McGraw-Hill Education. Retrieved from http://www.cs.cmu.edu

Monther, A., & Ali, A. (2018). Detecting Fake News in Social Media Networks. 9th International Conference on Emerging Ubiquitous Systems and Pervasive Networks (pp. 1-8). Abu dhabi: science direct.

Mwiti, D. (2021, April 10). Gradient boosted decision trees [guide] – a conceptual explanation. Retrieved from https://neptune.ai/blog/gradient-boosted-decision-trees-guide on May 3, 2021

Nolan Higdon. (2020). The anatomy of fake news: A critical news education. Oakland, california: university of california press. Retrieved from http://www.ucpress.edu.com

Rutvik, V., Prathyush, P., Siddharth, K., & Sundeep, T. (2020). Two stage transformer model for covid-19 fake news detection and fact checking. arXiv:2011.13253, 1, 1-10. doi:https://arxiv.org/abs/2011.13253

Ray, S. (2017, September 13). "Support Vector Machine" (SVM) is a supervised machine learning algorithm Retrieved from https://www.analyticsvidhya.com/blog/2017/09/understaing-support-vector machine-exam ple-code/ on May 3, 2021

Rahm, E., & Hong, D. H. (2000). Data cleaning problems and current approaches. data engineering, 23(4), 23-40.

Roger, S. (2021, Februrary 22). What are the branches of artificial intelligence? Retrieved from, https://www.h2kinfosys.com/blog/what-are-the-branches-of-artificial-intelligence/ on May 3,2021

Subhadra, G., Swati, S., Supriya, S., Prachi, W., & Sumit, H. (2019, January). survey on automated system for fake news detection using NLP & machine learning approach. international research journal of engineering and technology, 06(01), 2.

Shetty, A., Thawani, P., Rao, A., Uphade, A., & RL, P. (2022). News Check: A Fake News Detection and Analysis System. Soft Computing for Security Applications (pp. 2-16). ResearchGate.

Saleh, H., Alharbi, A., & Alsamhi, S. H. (2021, September 14). OPCNN-FAKE: Optimized Convolutional Neural Network for Fake News Detection. 9, 1-19.

Sarnovsky, M., Maslej-Kresnakova, V., & Ivancova, K. (2022). Fake News Detection Related to the COVID-19 in Slovak Language Using Deep Learning Methods. 19(2), 1-15.

Singh, A. K., & Mogalla, S. (2019). vectorization of text documents for identifying unifiable news articles. 10(7), 6. Retrieved from http://www.ijacsai.org

Singh, A. (2018, June 18). A comprehensive guide to ensemble learning. Retrieved from https://www.analyticsvidhya.com/blog/

DOI: 10.56892/bima.v9i1A.1233

2018/06/comprehensive-guide-for-ensemble-models / on May 3, 2021

Tanu, V., Renu, & Deepti, G. (2014). Tokenization and filtering process in rapidminer. international journal of applied information, 7(2), 3. doi:www.ijias.org

Vasu, A., Parveen, s. H., Srijan, M., & Amitrajit, S. (2019). Analysis of classifiers for fake news detection. trends in advanced computing (p. 7). New zealand: science direct.

Wood, T. (2019, June 1). Random forests. Random forests [Definition]. Retrieved from https://deepai.org/machine-learning-glossary-and-terms/random-forest on May 3, 2021

Xinyi, Z., & Reza, Z. (2019, june 10). Network based fake news detection: A pattern-driven approach. ACM SIGKDD, 1, 1-13.

Yu, J., & Posey, B. (2018, december). Data reduction. Retrieved from https://techtarget.com