***Table of Content***

[**1.0 Introduction** 1](#_Toc124194183)

[**4.0 Data Pre-processing** 3](#_Toc124194184)

[**5.0 Modelling** 4](#_Toc124194185)

[5.1 SVM 4](#_Toc124194186)

[5.2 Bayesian model 5](#_Toc124194187)

[**6.0 Results and Discussion** 6](#_Toc124194188)

[**7.0 Conclusion** 9](#_Toc124194189)

[**8.0 Reference** 10](#_Toc124194190)

# **1.0 Introduction**

With the rapid development of social networking media, rumors can easily spread on the internet. Tweets spread on platforms such as Weibo, Twitter, WeChat groups and friends' circles often contain misleading political tweets that affect public perception, and most social media users believe fake political news because they lack knowledge of the subject. Fake political news is also a great threat to people's safety, as some individuals or organizations make use of social media platforms’ high spread ability for their own benefit, Therefore, early and efficient detection of fake political news or rumors on social media platforms are crucial for the well-being of a society.

On a personal level, disinformation distorts people's judgement, disturbs their thought processes, and makes it harder for them to identify good from wrong. Some individuals are likely to listen to incorrect political information and even forward it to their friends and family, allowing it to spread extensively. What is frightening is not fully misleading information, but half-true, half-false information that confuses media consumers. Unconfirmed political intelligence should not be dismissed, and it requires challenging judgement to establish its reliability and trustworthiness. At the national level, political disinformation is sufficient to affect the outcome of an election, as it influences the electorate's decision. In addition, political disinformation has a negative effect on the national interest and even rips social fault lines apart. As can be shown, political deception causes considerable harm. Once misunderstanding and confusion have been produced, the government must issue corrections and clarifications and request that the author remove the published information to prevent additional damage. The scourge of falsehood can be as devastating as monetary loss, social discord, and national panic. False political news undermines not only the cyber security and financial safety of individuals, but also social stability and national security in a significant way.

Machine learning is dedicated to learning models from data which mainly consists of supervised, semi-supervised, and unsupervised learning. The purpose of unsupervised learning is to classify or aggregate similar groups into the same dataset, whereas supervised learning is used to train machine learning models with training sample data with corresponding target values, supervised learning is used to extract feature values and map relationships by making connections between data sample factors and known outcomes, and to continuously learn and train on new data with known outcomes. Machine learning uses computers and algorithms to learn and discover hidden patterns and insights from data faster and more accurately than what the human mind is capable of. The detection of false political news can be performed by the time of publication of an article, the title of the article, the text of the article and the subject of the article. The following research focuses on how machine learning can be implemented to detect fake political news on social media.

2.0 Research Questions and Objectives

Research Questions：

1.Is it possible to use machine learning for multi-class classification and classify fake news into levels of authenticity?

2.Does the multi-class classification model perform the same as the svm classifier on different evaluation metrics?

3.How to update the global language network disinformation corpus faster and more accurately?

Research Objectives ：

1.To Improve the currently proposed fake news svm classification model into a multi-classification model.

2.To Improve the performance of the classification model through different evaluation metrics.

3.To Optimize programmers who develop algorithms to detect fake news.

**.0 Related Work**

Hussain, Hasan, Rahman, Protim, and Al Hasan (2020) demonstrated an experimental investigation into the detection of fake news from Bangladeshi social media, as this area still requires a great deal of attention. Throughout the study, the authors have utilized two supervised machine learning techniques, Support Vector Machines (SVM) and Multinomial Bayesian (MNB) classifiers to identify fake news in Bangladesh. The term frequency - inverse document frequency vector quantizer and count vector quantizer have been used as feature extraction. The authors' proposed system identifies fake news based on the polarity of the relevant posts. The final study showed that the SVM with a linear kernel provided 96.64% accuracy outperforming the MNB's 93.32%.

Qalaja, Al-Haija, Tareef, and Al-Nabhan classified fake news about COVID-19 collected from Twitter using four machine learning-based models: a decision tree (DT), a simple Bayesian (NB), an artificial neural network (ANN), and a k-nearest neighbour (KNN) classifier. Furthermore, detection models were constructed and assessed in real time on their new Twitter dataset using conventional assessment measures such as detection accuracy (ACC), F1-score (FSC), under the curve (AUC), and Matthew correlation coefficient (MCC). The DT-based detection model scored 99.0% for ACC, 96.0% for FSC, 98.0% for AUC, and 90.0% for MCC in the first set of experimental assessments utilising the entire dataset. The DT-based detection model achieved the greatest detection performance scores of 89.5%, 89.5%, 93.0%, and 80.0% for ACC, FSC, AUC, and MCC in the second set of trials utilising small data sets. The best-selected features were used to derive the findings for all experiments.

Rahman, Hasan, Billah, and Sajuti (2022) employed four traditional machine learning (ML) algorithms as well as long and short-term memory (LSTM) methods in their research. Logistic regression (LR), decision tree (DT), k-nearest neighbour (KNN), and basic Bayesian (NB) classification are the four traditional approaches. To achieve the best optimal results, the dataset was trained using LSTM and Bi-LSTM (bi-directional long-term short-term memory). To determine the best model for detecting fake news, they used four traditional methods and two deep learning models. The logistic regression approach fared the best of the four traditional methods, with an accuracy of 96%, while the Bi-LSTM model had an accuracy of 99%.

Alameri and Mohd (2021) set out to find the highest performing machine learning model among two: a simple Bayesian (NB), a support vector machine (SVM), and three deep learning models: long short-term memory (LSTM), Neural Network with Keras (NN-Keras), and Neural Network with TensorFlow (NN-TF). The authors used two separate English news datasets to test the five models. The accuracy, precision, recall, and F1-score of the models were used to evaluate their performance. The results reveal that deep learning models outperform typical machine learning models in terms of accuracy. All other models tested were outperformed by the LSTM model. It had a 94.21% average accuracy. NN-Keras likewise performed admirably, with an average accuracy of 92.99%. The arrangement of words conveys essential information and is used to classify bogus news, on which the LSTM makes its predictions.

Z Xu et al. (2020) address the issue of using text syntactic structure to improve pretrained models like BERT and RoBERTa. In a dependency tree, predicts the syntactic distance between tokens. Injecting auto-generated text grammar into pre-trained models can help them improve. Second, when compared to the local center relation between consecutive tokens, the global syntactic distance between them yields greater performance gains.

JY Khan et al. (2021) proposed using hardware constraints to investigate many advanced pre-trained language models, as well as traditional and deep learning models, for fake news detection. Naive Bayes (with n-grams) is an excellent option. Deep learning models outperform traditional models in detecting fake news. Traditional and deep learning models are outperformed by BERT-based models.

KB Nelatoori (2022) presented the domain adaptation capabilities of RoBERTa and BERT models on HASOC and OLID datasets containing out-of-domain text from Twitter and found a 3% improvement in F1 scores over single-task models. Training ROBERTA on unlabeled data for each domain adaptation task (task-adaptive pre-training), according to S Gururangane et al. (2020), improves performance even after domain-adaptive pre-training.

By fine-tuning BERT, X Yang et al. (2022) train a classifier. To train the summarizer, fine-tune the T5 model. By comparing the performance of T5 and other summary models and use the PEGASUS model as one of our classifiers. To validate the TCS framework, use a small number of samples from the XSUM dataset at random. The results demonstrate that the TCS framework can generate text summarization in a variety of styles.

# **4.0 Data Pre-processing**

The dataset includes the files "Fake.csv" & "True.csv".

Fake.csv:

|  |  |  |  |
| --- | --- | --- | --- |
| The title of the article | The text of the article | The subject of the article | The date at which the article posted |
| 17903unique values | 22851 values | Include News (9250), politics (6878), Other (7590) | 2015/3/31-2018/2/19 |
| Example:  Donald Trump Sends Out Embarrassing New Year’s Eve Message; This is Disturbing etc. | Example:  Donald Trump just couldn’t t wish all Americans a Happy New Year and leave it at that. Instead, he had...etc. | Example:  News | Example: December 31, 2017 |

True.csv: This dataset contains a list of articles considered as "real" news, including "The title of the article", "The text of the article", "The subject of the article ", "The date at which the article was posted". This data includes the number of true news stories from 2016-2017, with a maximum value of 20826.

True.CSV:

|  |  |  |  |
| --- | --- | --- | --- |
| The title of the article | The text of the article | The subject of the article | The date at which the article posted |
| 20826  unique values | 21192  unique values | politicsNews 53%  Worldnews 47% | 2016/1/13-2017/12/31 |
| Example:  US. military to accept transgender recruits on Monday: Pentagon etc. | Example:  WASHINGTON (Reuters) - Transgender people will be allowed for the first time to enlist in the U.S. m... etc. | Example:  politicsNews | Example: December 29, 2017 |

1.Import NumPy, pandas, nltk (natural language toolkit), and matplotlib, seaborn plotting library, and string string library.

2.Set up a target tag feature to transform the false data into target tag 1 and the true data into target tag 0.

3.Merge the false data with the true data via pd.concat and reset the dataframe index to a dataframe named as combined\_df. 'http' to 'link', '\n' to ' ', and the text data by removing spaces. The resulting cleaned training data is 70%, the test data is 20% and the validation data is 10%.

4.The training set, test set and validation set are encapsulated into a dataset (dataset\_text).

# **5.0 Modelling**

## 5.1 SVM

SVM (support vector machines) is a two-category model that maps an instance's feature vector to some points in space. SVM's goal is to draw a line that "best" distinguishes the two class points. Data is then divided into two groups. SVM's goal is to draw a line that "best" distinguishes these two types of points, so that if new points appear in the future, this line can also make a good classification. The kernel function is referred to as SVM. When a sample is linearly inseparable in the original space, data can be mapped from the original space to a higher-dimensional feature space, where it is linearly separable. After introducing such a mapping, there is no need to solve the real mapping function to solve the dual problem, but only to know its kernel function. The kernel function is defined as K(x,y)=(x),(y)>, which means that the inner product in the feature space equals the result calculated by the kernel function K in the original sample space. On the one hand, the data in a high-dimensional space becomes linearly separable. On the other hand, there is no need to solve specific mapping functions. Instead, only specific kernel functions must be provided, greatly reducing the difficulty of solving.

Dataframe Text processing

1.Delete the unnecessary attributes ‘title’, ‘subject' and 'date' columns in combined\_df, then convert the text column in combined\_df to lowercase by x.lower(),

2.Delete the punctuation in the text column in combined\_df by “punctuation\_removal” self-defined function to remove punctuation from the text column in combined\_df.

3. Remove the stop words from the ‘text’ column in combined\_df to form the final data dataframe.

To validate the model performance, we performed train-test split method to the dataset.

1. The combined dataset is split into catogirories, which are training set, testing set and validation set.

2. The spliting for them are 70%; 20%; 10% respectively.

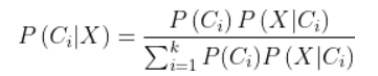
3. Then, the training dataset is used to train a few candidate models with each different parameter. The validation dataset is used to evaluate the candidate models and one of the candidates is chosen.

4. The chosen model (linear kernel function with c parameter equals to 1.0) is trained with a new training dataset, and evaluated with the test dataset

The text features in the training set are then extracted by the constructed Count Vectorizer model and fitted to the training and test sets to transform them into word frequency matrices, then the training word frequency matrix and training labels are fitted by the SVM model, i.e. svm.fit(x\_count\_train,y\_train), then the test word frequency matrix is predicted and the predicted labels are derived and the accuracy, precision, recall, f1-score and area under the curve of the test set are derived by the accuracy\_score, precision\_score, recall\_score, f1\_score and roc\_auc\_score functions.

## 5.2 Bayesian model

The core idea of the Bayesian model is to calculate the maximum probability of the object of study in each classification. Ci denotes one possibility of the research object, and for which specific category the research object is classified is to calculate the maximum possible outcome of Ci.



The 'subject' feature column in the combined\_df was encoded with unique heat (pd.get\_dummies) and then concat, the dataframe shape was (44898, 13). The title features in the training set were then extracted by the constructed CountVectorizer model and fitted to the training and test sets to transform them into a word frequency matrix, which was then fitted to the training word frequency matrix and training labels by a Bayesian model, i.e., nb = MultinomialNB(alpha=0.1) nb.fit(X\_count\_train, y\_train).

5.3 RoBERTa

The innovation of BERT lies in the Transformer Decoder (containing Masked Multi-Head Attention) as the extractor and the use of a mask training method that goes with it. Although the use of dual encoding makes BERT incapable of text generation, BERT utilizes all the contextual information of each word in the encoding of the input text, giving BERT a greater ability to extract semantic information than a one-way encoder that can only extract semantics using preorder information.

The RoBERTa model is an improved version of BERT. The next sentence prediction (NSP) task is removed, and dynamic masks and text encoding are used. Compared to BERT, Robert has a larger number of model parameters and a larger training data set. By calling the Roberta-based model from the automatic word splitter in transformers' library. The tokens, token\_span, token\_ids and token\_mask of the text data is obtained by tokenizing the dataset.

The tokenized\_functions are called to embedding text features to form tokenized\_datasets, and then the TrainingArguments function is used to encapsulate the training model parameters. The model and training (train\_args) parameters are evaluated using acc, pre, recall, f1, four indicators. Similarly, the training\_text, test\_ text,andvalid\_ text were cleaned, encapsulated, and segmented for feature extraction to form tokenized\_datases\_ text data. The trained model was then used to predict the test\_ text.

# **6.0 Results and Discussion**

For SVM, the accuracy, precision, recall, f1-score and area under the curve of the SVM model was obtained from the test set. The text features were extracted by the CountVectorizer model, and the title column could be transformed into a word frequency matrix with a shape of (44898, 23401). The results of the evaluation and visualisation of confusion matrix for SVM model in a heat map are as below:

|  |  |
| --- | --- |
| Accuracy: | 0.9962 |
| Precision: | 0.9974 |
| Recall: | 0.9953 |
| F1-score: | 0.9963 |
| Area Under the Curve: | 0.9963 |



Confusion Matrix for Support Vector Machine

“Fake” word cloud map



“True” word cloud map

We have generated Wordcloud using the text column data for both True and False political news dataset. First, we ignore the names of the popular politicians such as ‘Obama’, ‘Trump’, ‘Hillary’ and ‘Clinton’ as their names appeared the most and everywhere in True and False political news dataset. Then, by inspecting on the False political tweet news, we could observe that the words include but not limited to ‘Black’, ‘White’, ‘War’, and ‘Muslim’ that could be used in tweets to express messages that can spread racism and hate in the social media are being used relatively frequent in our dataset. Whereas by inspecting True political tweet news, words such as ‘Will’, ‘Say’, ‘House’, ‘Call’, ‘Official’ and ‘May’ that are often being used to quote or convey official statements are relatively more frequent in True political news in terms of usage compared to False political news. Besides, we could observe that the social media treats the information from other countries seriously as in the True political news dataset, Wordcloud produced ‘China’, ‘Korea’, ‘Russian’, ‘Syria’ and ‘Brexit’ as frequent words.

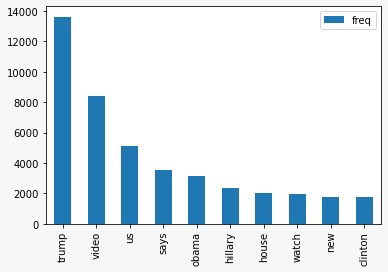
Bayesian models were fitted to the training word frequency matrix and training labels, and the accuracy of the test set was derived by the accuracy\_score function. The accuracy rate: 0.9748, precision: 0.9883, Recall: 0.9629, F1-Score: 0.9754, AUC: 0.9753. The accuracy rates did not differ significantly. By fitting a good Bayesian model, the test set is predicted, then the roc\_auc\_score and confusion\_matrix functions are called to generate the AUC values of the test set (0.9754), and the confusion matrix of the test set, and the precision,recall,f1-score of the test set is reported through classification\_report and pre=0.96, recall=0.93, f1-score=0.94, accuracy=0.95, and all four metrics were high, and the model predicted well.

A picture containing chart

Description automatically generated

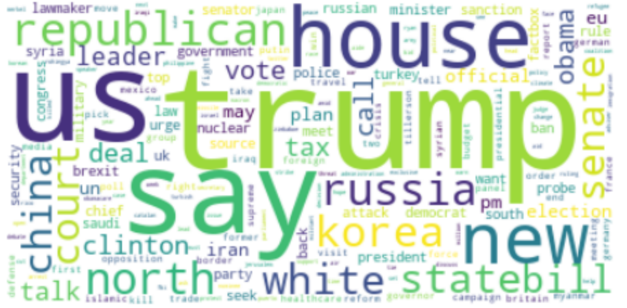
Confusion Matrix for Naïve Bayes

By constructing the histogram text word frequency, the top ten words, from high to low, were: trump, video, us, says, obama, hillary, house, watch, new, clinton.



Text Word Histogram

Bayes generation of cloud map is like generation of false data title word cloud. The text of false data is split, and then lower() is carried out. Finally, stop words are removed to generate false data word cloud map. people, said one, appeared most often. Similarly, to generate real data text word cloud, can get real data words in the text in the us, said, united, the state, trump, will appear the highest frequency.

 A picture containing text

Description automatically generated

“Fake” text word cloud “True” text word cloud

RoBERTa model trained trainer () through early stopping call back and got the best model score of 0.9983. To predict tokenized\_datasets[' test '] on the text attribute, we can get:

|  |  |
| --- | --- |
| test precision | 0.9983 |
| test recall | 0.9983 |
| testf1 | 0.9983 |
| test accuracy | 0.9983 |

# **7.0 Conclusion**

This paper studies the application of machine learning in identifying false political news, mainly focusing on the SVM, Bayesian and RoBERTa models to identify false political news. The first part introduces the hazards of false information and the background of the development of machine learning, as well as the methods used to identify false news. The second part summarizes many literatures, looking for research methods in the past literatures. The third part introduces the dataset used in the research and does simple cleaning and preprocessing to the dataset. Parts 4 and 5 describe the definitions of SVM, Bayesian and RoBERTa models and how to use these models for false message recognition. High scores of accuracy of 0.95 were calculated by Bayesian model. The SVM model achieved an accuracy of 0.9962. RoBERTa model improved Bert model to get the result of 0.9990968 in accuracy. Through the score comparison of the SVM model, Bayesian model, and RoBERTa model, the RoBERTa model has the highest accuracy. RoBERTa is based on BERT's language masking strategy, modifying key hyperparameters in BERT, including deleting BERT's next sentence pre-training target, and using a larger batch size and learning rate for training, so the accuracy rate is higher than SVM models and Bayesian models. If you only compare the SVM model from the accuracy rate, the Bayesian model and the RoBERTa model will also have problems. Because each element of the transformer can interact with global information like CNN, ignoring the distance. The attention head can learn to perform different tasks. The RoBERTa model is a black box and cannot populate a word cloud.

# **8.0 Reference**

1. Nelatoori, K. B., & Kommanti, H. B. (2022). Multi-task learning for toxic comment classification and rationale extraction. Journal of Intelligent news Systems, 1-25.
2. Gururangan, S., Marasović, A., Swayamdipta, S., Lo, K., Beltagy, I., Downey, D., & Smith, N. A. (2020). Don't stop pretraining: adapt language models to domains and tasks. arXiv preprint arXiv:2004.10964.
3. Khan, J. Y., Khondaker, M. T. I., Afroz, S., Uddin, G., & Iqbal, A. (2021). A benchmark study of machine learning models for online fake news detection. Machine Learning with Applications, 4, 100032.
4. Kovachevich, N. (2021). Nkovachevich at CheckThat! 2021: BERT fine-tuning approach to fake news detection. In CLEF (Working Notes) (pp. 537-544).
5. Yang, X., Wang, Y., & Li, X. (2022, March). TCS Framework: Generating Text Summaries of Specified Style Using Classifier and Summarizer. In 2022 4th International Conference on Natural Language Processing (ICNLP) (pp. 404-408). IEEE.
6. Xu, Z., Guo, D., Tang, D., Su, Q., Shou, L., Gong, M., ... & Jiang, D. (2020). Syntax-enhanced pre-trained model. arXiv preprint arXiv:2012.14116.
7. Alameri, S. A., & Mohd, M. (2021). *Comparison of fake news detection using machine learning and deep learning techniques.* Paper presented at the 2021 3rd International Cyber Resilience Conference (CRC).
8. Hussain, M. G., Hasan, M. R., Rahman, M., Protim, J., & Al Hasan, S. (2020). *Detection of bangla fake news using mnb and svm classifier.* Paper presented at the 2020 International Conference on Computing, Electronics & Communications Engineering (iCCECE).
9. Qalaja, E. K., Al-Haija, Q. A., Tareef, A., & Al-Nabhan, M. M. Inclusive Study of Fake News Detection for COVID-19 with New Dataset using Supervised Learning Algorithms.
10. Rahman, M., Hasan, M., Billah, M. M., & Sajuti, R. J. (2022). Political Fake News Detection from Different News Source on social media using Machine Learning Techniques. *AIUB Journal of Science and Engineering (AJSE), 21*(2), 110-117.