Advanced Approaches to Combatting Disinformation: A Comprehensive Analysis for Fake News Detection

1st Parshwa Paresh Bhavsar
Department of Multidisciplinary
Texas A&M University
College Station, TX
parshwa_2108@tamu.edu

2nd Rohan Dalvi
Department of Multidisciplinary
Texas A&M University
College Station, TX
rohan.dalvi@tamu.edu

***Abstract*—Politics worldwide has faced a significant setback due to the proliferation of false information. Deliberately crafted to mislead, fake news poses a challenge in identifying its deceptive nature solely through content analysis. This misinformation has influenced the perspectives of the general populace, highlighting the urgency to verify the authenticity of news, especially in the online domain. The widespread dissemination of fake news carries the potential for severe societal repercussions. To address this issue, we propose employing machine learning techniques for detection. Our approach involves vectorizing news titles and analyzing word tokens using a dataset containing a curated list of news items categorized as either authentic or fake. The objective is to develop a model capable of classifying given articles as either true or false. Addressing the global challenge of misinformation in politics, our proposed solution leverages machine learning techniques to detect fake news. By vectorizing news titles and analyzing word tokens within a curated dataset, our goal is to develop a robust model capable of classifying articles as either authentic or fake, thereby contributing to the ongoing efforts to ensure the accuracy and reliability of information in the digital age.**

Code: https://tinyurl.com/568v3pa2

Presentation: https://youtu.be/fjWGbnn9skM?feature=shared

***Index Terms*—Fake News, Self Learning, Pattern Matching, Response Generation, Artificial Intelligence, Natural Language Processing, Context Free Grammar, Term Frequency Inverse Document Frequency, Stochastic Gradient Decent, Word2Vec.**

I. Introduction

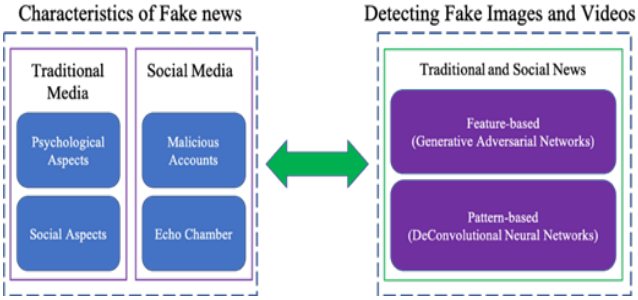
A. What is Fake News?

The term ”fake news” carries varied interpretations for different individuals. Fundamentally, we define ”fake news” as news stories that lack authenticity; these stories are entirely fabricated, lacking verifiable facts, sources, or quotes. At times, these narratives may serve as intentional propaganda aiming to mislead readers, or they may be crafted as ”click bait” designed for economic gains, with the writer profiting based on the number of clicks the story receives. The proliferation of fake news has been notable on social media platforms in recent years, facilitated by the ease and speed with which information can be shared online.

To effectively address fake news detection, it is crucial to comprehend the nature and characteristics of fake news. This involves two key aspects: characterization, which pertains to understanding what constitutes fake news, and detection. Building detection models necessitates initiating the process with characterization; indeed, comprehending what fake news entails is essential before attempting to identify it.

B. Fake News Characterization

The definition of fake news comprises two essential components: i. Authenticity & ii. Intent. Authenticity refers to the presence of false information in fake news that can be verified as such. This excludes conspiracy theories from the realm of fake news, as they are challenging to prove true or false in many instances. The second component, intent, underscores that the dissemination of false information is purposefully crafted to mislead the reader.



C. Fundamental Theories

Foundational theories in human cognition and behavior, developed across diverse disciplines such as social sciences and economics, offer valuable insights for the analysis of fake news. These theories present opportunities for both qualitative and quantitative studies of extensive fake news data sets. Moreover, they contribute to the creation of well-justified and explainable models for the detection and intervention of fake news—an area that has been relatively under explored thus far. Through a thorough literature survey spanning various disciplines, we have identified established theories suitable for investigating fake news.

*1) **News-related theories**:* Theories related to news highlight potential distinctions between the characteristics of fake news content and genuine news content. For instance, these theories suggest that fake news may differ from the truth in aspects such as writing style and quality, quantity (e.g., word counts), and the sentiments conveyed. It's important to note that these theories, originating from forensic psychology, primarily address deceptive statements or testimonies rather than specifically targeting fake news, although the concepts share similarities. Consequently, a research opportunity arises to examine whether these identified attributes are statistically distinguishable among disinformation, fake news, and truthful information, particularly through the analysis of extensive fake news data sets.

*2) **User-related theories**:* Theories pertaining to users delve into the attributes of individuals engaged in fake news activities, including actions such as posting, forwarding, liking, and commenting. Unlike information such as fake reviews, fake news has the capacity to draw both malicious and regular users. Malicious users deliberately disseminate fake news, motivated by potential gains. On the other hand, certain regular users, identified as vulnerable normal users, may unwittingly propagate fake news without realizing its falsity. This vulnerability is rooted in psychological factors, namely, (i) social impacts and (ii) self-impact.

D. Novelty

In our research and project focused on fake news detection in the CyberSecurity domain, we aim to make significant contributions to the field. Our work stands out in several ways: firstly, we carefully selected and integrated diverse Kaggle datasets relevant to CyberSecurity, providing a comprehensive perspective on fake news detection. Secondly, we adopt a multi-model approach, integrating various machine learning and deep learning models to thoroughly evaluate and compare performance. We prioritize ethical considerations by creating a transparent and explainable AI system that respects individual freedoms and privacy. Lastly, our research questions are tailored to the CyberSecurity domain, emphasizing linguistic, source, and contextual features specific to fake news in this field.

II. Related Work

In 2017, **Nguyen Vo, a student at Ho Chi Minh City University of Technology (HCMUT) in Cambodia**, conducted research on fake news detection and successfully implemented it. His project utilized the Bi-directional GRU with Attention mechanism, initially proposed by Yang et al., for fake news detection. Additionally, Nguyen Vo explored other deep learning models such as AutoEncoders, GAN, and CNN.

Stanford University's Samir Bajaj also delved into fake news detection in a research paper, employing NLP perspectives and various deep learning algorithms. Bajaj utilized an authentic dataset from Signal Media News.

Avinash Shakya, a student at ABES Engineering College in Lucknow, contributed to the field in 2019 by proposing a strategy that combined Naive Bayes classifier, Support Vector Machines, and semantic investigation. This approach achieved an impressive accuracy of 93.50%.

Parikh, S. B., & Atrey, P. K, in their April 2018 paper, introduced various detection techniques, including linguistic basis, deception modeling, clustering, predictive modeling, content cue-based methods, and non-text cue-based methods, with accuracy ranging from 63 to 70 percent.

In November 2015, **Conroy, N. J., Rubin, V. L., & Chen, Y**, focused on linguistic cue approaches, machine learning, Bag of Words approach, rhetorical structure and discourse analysis, network analysis approaches, and SVM classifiers.

Helmstetter, S., & Paulheim, H, in August 2018, classified tweets/posts as a binary classification problem based on the source. They used various algorithms such as Naive Bayes, Decision Trees, SVM, Neural Networks, Random Forest, and XG Boost, revealing 15 percent fake tweets, 45 percent real tweets, with the rest undecided.

Wang, W. Y., in 2017, proposed deception detection using the labeled benchmark dataset 'LIAR,' emphasizing improved efficiency in detecting fake posts/news.

Della Vedova, M. L., Tacchini, E., Moret, S., Ballarin, G., DiPierro, M., & de Alfaro, L, in May 2018, introduced the need for hoax detection, employing an ML approach with a mix of content and social content approaches.

Julio CS Reis and his team, in 2019, used machine learning techniques on BuzzFeed articles related to the US election, achieving accuracy through various algorithms and handcrafted features.

In 2017, **Natali Ruchansky, Sungyong Seo, and Yan Liu** utilized a hybrid network merging news content features and metadata for fake news detection, achieving improved performance.

In 2017, **James Thorne, Mingjie Chen, Giorgos Myrianthous, Jiashu Pu, Xiaoxuan Wang, and Andreas Vlachos** took on the Fake News Challenge by proposing a stack of different classifiers. Their approach involved using a multilayer perceptron, logistic regression, and gradient boosted trees, leveraging various features such as word embeddings, tf-idf vectors, and headline-article body concatenation.

In 2018, **Yang Yang, Lei Zheng, Jiawei Zhang, Qingcai Cui, Zhoujun Li, and Philip S. Yu** introduced a novel approach using a CNN with images contained in the articles for classification. They utilized the Kaggle fake news dataset and real news from trusted sources like the New York Times, incorporating both textual and image branches into their network.

Mykhailo Granik and Volodymyr Mesyura, in 2017, presented their fake news detection approach using a Naive Bayes classifier, achieving an accuracy of 74% on the test set.

Sohan Mone, Devyani Choudhary, and Ayush Singhania, also in 2017, proposed a system that calculates the probability of news being fake by applying NLP and using methods like Naive Bayes, SVM, and logistic regression.

In 2018, students from **Southern Methodist University (SMU), including Aswini Thota, Priyanka Tilak, Simrat Ahluwalia, and Nibrat Lohia**, concluded that their finely tuned Tf-IDF – Dense Neural Network (DNN) model outperformed existing architectures by 2.5%, achieving an impressive accuracy of 94.21% on test data.

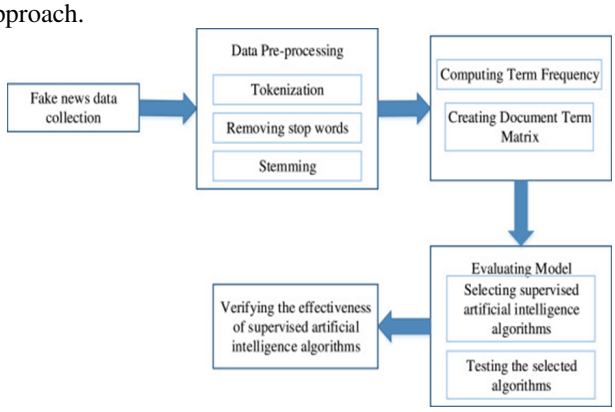
Xinyi Zhou and Reza Zafarani, in 2020, conducted a comprehensive survey reviewing current fake news research. They defined fake news, differentiated it from related concepts, evaluated interdisciplinary approaches, and highlighted methods for detection based on various perspectives, including writing style, propagation patterns, and source credibility.

These studies collectively demonstrate the diverse strategies researchers have employed to combat the challenges of fake news detection, incorporating machine learning, deep learning, NLP techniques, and even image analysis.

III. Methodology

A. Proposed Framework

In our proposed framework, we expand on the current literature by introducing ensemble techniques with various linguistic feature sets to classify news articles from multiple domains as true or fake. The ensemble techniques, along with the Linguistic Inquiry and Word Count (LIWC) feature set used in our research, constitute the novelty of our proposed approach.



B. Algorithm

We utilized the following learning algorithms in conjunction with our proposed methodology to evaluate the performance of fake news detection classifiers.

*1) **Naïve Bayes**:* We employed Naive Bayes as a probabilistic classifier inspired by the Bayes theorem under a simple assumption: the attributes are conditionally independent. The classification was conducted by deriving the maximum posterior, which is the maximal P(Ci—X), with the above assumption applying to Bayes theorem. This assumption greatly reduced the computational cost by only counting the class distribution. Even though the assumption is not valid in most cases, as the attributes are dependent, surprisingly, we found that Naive Bayes has been able to perform impressively.

$$P(\mathbf{X} \mid C_i) = \prod_{k=1}^{n} P(x_k \mid C_i) = P(x_1 \mid C_i) \times P(x_2 \mid C_i) \times \ldots \times P(x_n \mid C_i)$$

*2) **Logistic Regression**:* We named our approach Logistic Regression for the function used at its core, the logistic function. The logistic function, also called the sigmoid function, was developed by statisticians to describe properties of population growth in ecology, rising quickly and maxing out at the carrying capacity of the environment. It's an S-shaped curve that can take any real-valued number and map it into a value between 0 and 1, but never exactly at those limits.

$$\frac{1}{(1 + e^{-value})}$$

We combined input values (x) linearly using weights or coefficient values (referred to as the Greek capital letter Beta) to predict an output value (y). A key difference from linear regression is that the output value being modeled is a binary value (0 or 1) rather than a numeric value.

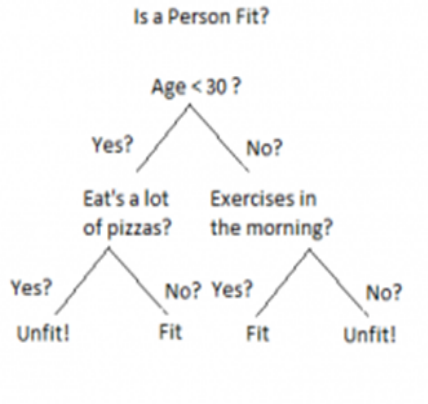
$$\frac{e^{(\beta_0 + \beta_1 x)}}{1 + e^{(\beta_0 + \beta_1 x)}}$$

*3) **Support Vector Machine (SVM)**:* Support Vector Machine, or SVM, is one of the most popular Supervised Learning algorithms, utilized for both Classification and Regression problems. However, our primary focus lies in solving Classification problems in Machine Learning. Our goal with the SVM algorithm is to create the best line or decision boundary that can effectively segregate n-dimensional space into classes. This enables us to easily categorize new data points in the correct category in the future. This optimal decision boundary is referred to as a hyperplane. As SVM advocates, we select the extreme points or vectors that aid in constructing the hyperplane. These exceptional cases are known as support vectors, hence earning the algorithm its name, Support Vector Machine.

$$h(\boldsymbol{x}_i) = sign\Big(\sum_{j=1}^{s} \alpha_j \; y_j \; K(\boldsymbol{x}_j, \boldsymbol{x}_i) + b \Big)$$

$$K(\boldsymbol{v}, \boldsymbol{v}') = \exp\Big(\frac{||\boldsymbol{v} - \boldsymbol{v}'||^2}{2\gamma^2}\Big)$$

*4) **Decision Tree Learning**:* Decision Trees belong to the category of Supervised Machine Learning, where the model learns from labeled training data, meaning it understands the relationship between input and corresponding output. In this method, data is systematically divided based on specific parameters. The structure of a Decision Tree is defined by two key elements: decision nodes and leaves. Decision nodes represent points in the tree where data is split, while leaves signify the ultimate decisions or final outcomes of the model. The process involves recursively branching through decision nodes until reaching the leaves, which hold the conclusive results or predictions.



*5) **Random Forest**:* Random Forest (RF) stands as an advanced iteration of decision trees (DT) within the realm of supervised learning models. In RF, a multitude of decision trees operate independently, collectively working to predict the outcome of a class. The final prediction is determined by aggregating most votes from the individual trees. One notable advantage of Random Forest is its lower error rate when compared to other models, attributed to the reduced correlation among the trees. The training process of our Random Forest model involved experimenting with various parameters. Through a grid search, different numbers of estimators were tested to identify the optimal model capable of achieving high-accuracy predictions. In the context of decision trees, multiple algorithms can be employed to determine a split, depending on whether the problem involves regression or classification. For our classification task, we utilized the Gini index as a cost function to guide the estimation of a split in the dataset.

$$Gini = 1 - \sum_{i=1}^{C} (p_i)^2$$

C. Datasets

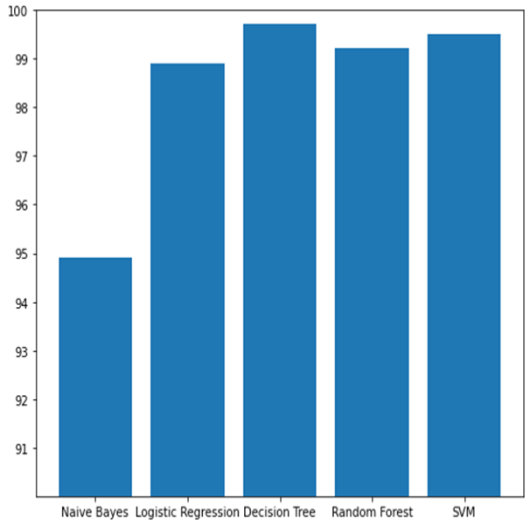
The datasets employed in this study are open source and accessible online without charge. They encompass both fake and truthful news articles spanning various domains. Truthful news articles provide accurate depictions of real-world events, whereas the fake news sources disseminate claims that deviate from information. In this study, three distinct datasets were utilized. The first comprises articles from the three datasets individually, which are henceforth denoted as True and Fake.

D. Performance Metrics

To assess the effectiveness of the algorithms, we employed a confusion matrix. A confusion matrix is a tabular representation of how well a classification model performs on a test set. It comprises four key parameters: true positive, false positive, true negative, and false negative. These elements provide a detailed snapshot of the model's accuracy and error rates, offering valuable insights into its performance across different categories.

IV. RESULT ANALYSIS

A. Result



The graph above provides a summary of the accuracy attained by each algorithm on the final dataset. Clearly, the highest accuracy is observed with the Decision Tree algorithm, reaching 99.65%. Following closely is the Support Vector Machine (SVM) with an accuracy of 99.57%. The next notable accuracy is achieved by the Random Forest algorithm, registering at 99.12%. The subsequent highest accuracy is attained with Logistic Regression, reaching 98.83%. The lowest accuracy is recorded for Naïve Bayes, which is 94.81%. The table below displays the names of the classifiers along with their corresponding accuracy scores.

Classifier	Accuracy
Decision Tree	99.65%
Support Vector Machine (SVM)	99.57%
Random Forest	99.12%
Logistic Regression	98.83%
Naïve Bayes	94.81%

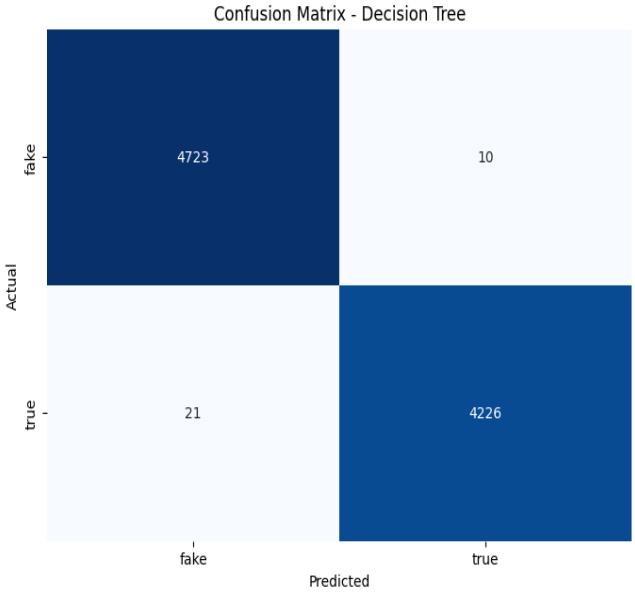
TABLE I: Classifier Accuracy

B. Confusion Matrix

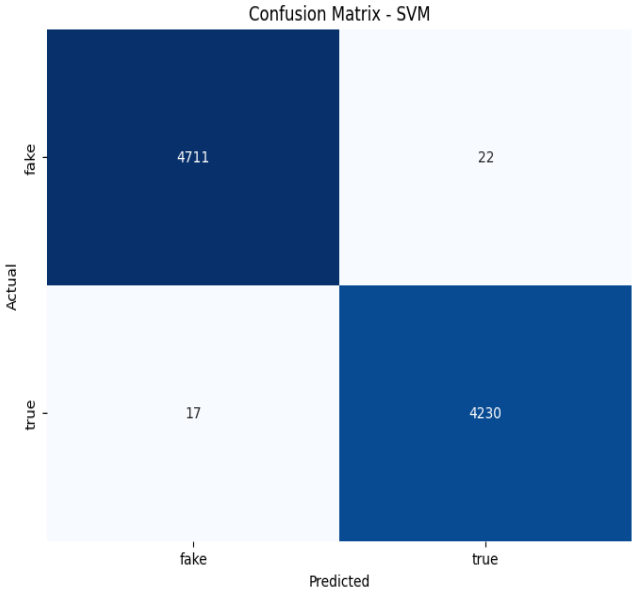
A confusion matrix is a table used in machine learning and statistics to evaluate the performance of a classification algorithm. It summarizes the counts of true positive, true negative, false positive, and false negative predictions, providing insights into the model's accuracy, precision, recall, and other performance metrics. The matrix is particularly useful for assessing how well a model distinguishes between different classes in a classification task.

1. The **Decision Tree classifier**, with specified parameters, achieved accurate predictions and displayed a clear confusion matrix. It effectively captured the relationships within the

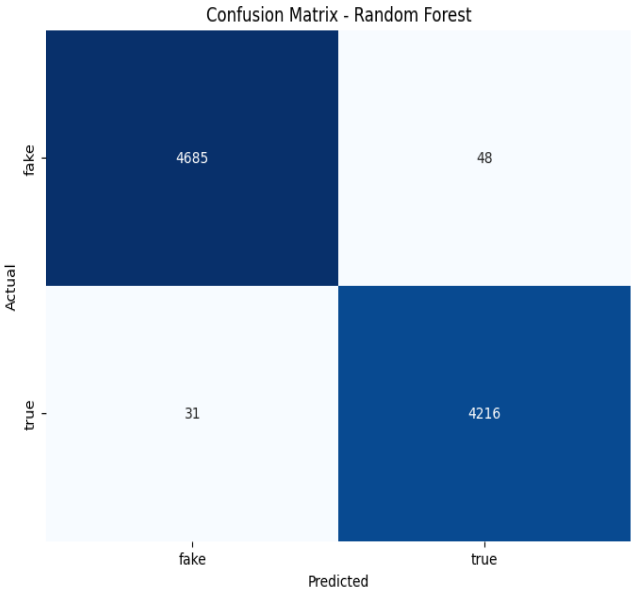
data, leading to accurate classification across different classes.



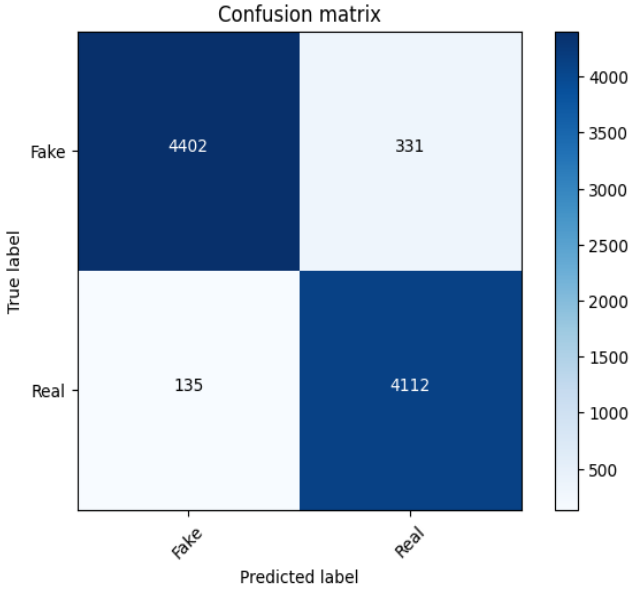
2. The **SVM classifier**, employing a linear kernel, delivered high accuracy and a well-structured confusion matrix. It effectively separated classes in the dataset, showcasing its strength in handling non-linear decision boundaries and achieving accurate predictions.



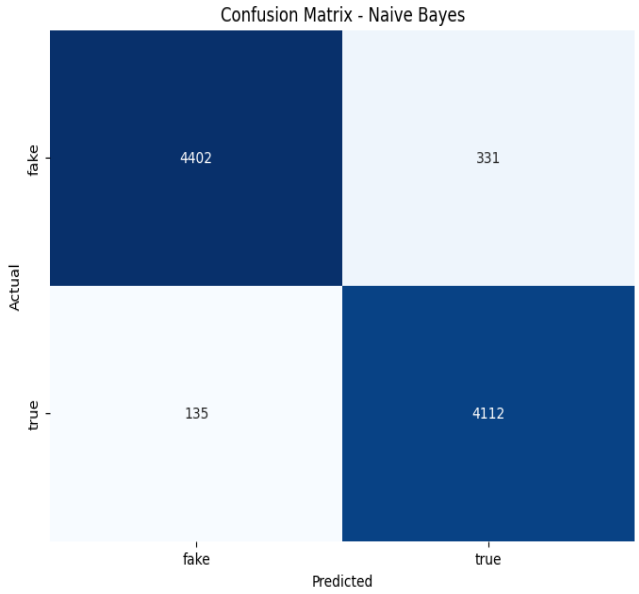
3. The **Random Forest classifier**, utilizing an ensemble of decision trees, demonstrated remarkable accuracy. The confusion matrix illustrates the model's ability to handle complex patterns, providing reliable predictions across multiple categories.



4. **Logistic Regression** exhibited outstanding accuracy, and its confusion matrix indicates precise predictions across various classes. The model shows robust performance in distinguishing between different categories in the dataset.



5. The **Naive Bayes classifier** demonstrated high accuracy with a well-balanced confusion matrix. It effectively classified instances across multiple classes, showing reliable performance in text classification.



V. CONCLUSION

The manual classification of news articles demands a profound understanding of the domain and expertise to discern anomalies in the text. This research addresses the challenge of classifying fake news articles through the application of machine learning models and ensemble techniques. The dataset utilized is sourced from KAGGLE and encompasses news articles spanning diverse domains, ensuring comprehensive coverage beyond the specific classification of political news. The primary objective is to discern text patterns that distinguish fake articles from genuine news.

The learning models underwent training and parameter tuning to achieve optimal accuracy. Some models demonstrated notably higher accuracy compared to others. Multiple performance metrics were employed to compare results across each algorithm. Ensemble learners exhibited an overall superior performance across all metrics when compared to individual learners.

Fake news detection presents numerous open issues warranting further research. For instance, to mitigate the dissemination of fake news, understanding the key elements in news propagation is crucial. Graph theory and machine learning techniques can be applied to pinpoint the pivotal sources involved in the spread of fake news. Additionally, real-time identification of fake news in videos presents a promising avenue for future exploration.

In conclusion, this application represents just one component of a broader toolkit essential for a highly accurate fake news classifier. Other tools, including a fact detector and a stance detector, would be integral additions. To integrate all these components effectively, a model capable of combining and weighing each tool in its final decision-making process would be necessary.

REFERENCES

[1] Soll, J., White, J. B., Sitrin, S. S. and C., &; Gerstein, B. M. and J. (2016, December 18). The long and brutal history of fake news. POLITICO Magazine. https://www.politico.com/magazine/story/2016/12/fake-news-history-long-violent-214535

[2] Rinehart, A. (2022, September 2). Fake news. it's complicated. First Draft. https://firstdraftnews.org/articles/fake-news-complicated/

[3] T. Ahmad, H. Akhtar, A. Chopra, and M. Waris Akhtar, "Satire detection from web documents using machine learning methods," pp. 102–105, 09 2014.

[4] Kang, C., &; Goldman, A. (2016, December 5). In Washington Pizzeria attack, fake news brought real guns. The New York Times. https://www.nytimes.com/2016/12/05/business/media/comet-ping-pong-pizza-shooting-fake-news-consequences.html

[5] S. Sedhai and A. Sun, "Semi-supervised spam detection in twitter stream," arXiv preprint arXiv:1702.01032, 2017.

[6] A. Bhowmick and S. M. Hazarika, "Machine learning for e-mail spam filtering: Review, techniques and trends," arXiv preprint arXiv:1606.01042, 2016.

[7] W. Y. Wang, "" liar, liar pants on fire": A new benchmark dataset for fake news detection," arXiv preprint arXiv:1705.00648, 2017.

[8] James W Pennebaker, Martha E Francis, and Roger J Booth. Linguistic inquiry and word count: Liwc 2001. Mahway: Lawrence Erlbaum Associates, 71(2001):2001, 2001.

[9] Natali Ruchansky, Sungyong Seo, and Yan Liu. Csi: A hybrid deep model for fake news detection. In Proceedings of the 2017 ACM on Conference on Information and Knowledge Management, pages 797–806. ACM, 2017.

[10] Eugenio Tacchini, Gabriele Ballarin, Marco L. Della Vedova, Stefano Moret, and Luca de Alfaro. Some like it hoax: Automated fake news detection in social networks.

[11] James Thorne, Mingjie Chen, Giorgos Myrianthous, Jiashu Pu, Xiaoxuan Wang, and Andreas Vlachos. Fake news stance detection using stacked ensemble of classifiers. In Proceedings of the 2017 EMNLP Workshop: Natural Language Processing meets Journalism, pages 80–83, 2017.

[12] Yaqing Wang, Fenglong Ma, Zhiwei Jin, Ye Yuan, Guangxu Xun, Kishlay Jha, Lu Su, and Jing Gao. Eann: Event adversarial neural networks for multi-modal fake news detection. In Proceedings of the 24th acm sigkdd international conference on knowledge discovery & data mining, pages 849–857. ACM, 2018.

[13] Mykhailo Granik and Volodymyr Mesyura. Fake news detection using naive bayes classifier. In 2017 IEEE First Ukraine Conference on Electrical and Computer Engineering (UKRCON), pages 900–903. IEEE, 2017.

[14] Zichao Yang, Diyi Yang, Chris Dyer, Xiaodong He, Alex Smola, and Eduard Hovy. Hierarchical Attention Networks for Document Classification. In Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, pages 1480–1489, San Diego, California, 2016. Association for Computational Linguistics.

[15] Kamran Kowsari, Mojtaba Heidarysafa, Donald E. Brown, Kiana Jafari Meimandi, and Laura E. Barnes. RMDL: Random Multimodel Deep Learning for Classification. Proceedings of the 2nd International Conference on Information System and Data Mining - ICISDM '18, pages 19–28, 2018. arXiv: 1805.01890.

[16] David R. Karger, Sewoong Oh, and Devavrat Shah. Iterative learning for reliable crowdsourcing systems. In J. Shawe-Taylor, R. S. Zemel, P. L. Bartlett, F. Pereira, and K. Q. Weinberger, editors, Advances in Neural Information Processing Systems 24, pages 1953–1961. Curran Associates, Inc., 2011.

[17] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, and Illia Polosukhin. Attention is all you need. In NIPS, 2017.

[18] Bisaillon, C. (2020, March 26). Fake and real news dataset. Kaggle. https://www.kaggle.com/datasets/clmentbisaillon/fake-and-real-news-dataset

[19] Fake news. Kaggle. (n.d.). https://www.kaggle.com/c/fake-news/data

[20] S. H. Kong, L. M. Tan, K. H. Gan and N. H. Samsudin, "Fake News Detection using Deep Learning," 2020 IEEE 10th Symposium on Computer Applications & Industrial Electronics (ISCAIE), Malaysia, 2020, pp. 102-107, doi: 10.1109/ISCAIE47305.2020.9108841.

[21] N. F. Baarir and A. Djeffal, "Fake News detection Using Machine Learning," 2020 2nd International Workshop on Human-Centric Smart Environments for Health and Well-being (IHSH), Boumerdes, Algeria, 2021, pp. 125-130, doi: 10.1109/IHSH51661.2021.9378748.