Fake News Detection with News Headlines

Md Mahmud Ferdous

***Abstract*—The bubble of information that is now travelling fast in a short span of time over online news sites is a major threat to society and trust. This project discusses the possibility of classification of news headlines as either fake or real news. Based on the dataset of the Fake and Real News presented at Kaggle, we used the TF-IDF approach to vectorize the text of headlines and assess the functionality of three traditional machine learning algorithms: Logistic Regression (LR), Support Vector Machine (SVM), and Random Forest (RF). The results of the experiments show that all of three models produced high classification accuracy (SVM, 0.949; LR, 0.944; RF, 0.944). One can see that SVM produced the best F1-score (0.948) compared with both LR (0.942) and RF (0.942). Confusion matrices show that classification is uniform and effective through fake and real labels. Though the main concentration reflected the lexical features, the project is the preparatory step towards further extensions relying on deep learning structures, sentiment analysis, and explainability procedures. All in all, scalable and interpretable headline-level fake news detection is possible with lightweight NLP-based models, as demonstrated by the findings.**

***Index Terms*—Fake News Detection, Natural Language Processing, Machine Learning, TF-IDF, Support Vector Machine, Logistic Regression, Random Forest, News Headline Classification, Text Classification, Misinformation.**

I. Introduction

A. Problem Definition

Deliberately deceptive news, referred to as fake news, has spread using digital platforms, causing a decline in the level of trust people have in institutions and affecting domains such as politics, health, and finance. Especially important are headlines, which are all that most users will ever see: researchers indicate that readers are unlikely to read further, but headlines can be misleading by themselves. Training automatic detection on headlines is thus possible and practically meaningful.

B. Research Questions

This study is driven by the following research questions:

1) Can we accurately classify news headlines as fake or real using NLP techniques?
2) What words or phrases are most indicative of fake news?
3) Does sentiment polarity vary significantly between real and fake news?

II. Literature Review

Machine learning and natural language processing methods have been spreading in the field of fake news detection, with each approach trying to cope with the problems of detecting deceptive or misleading information. This literature review unites existing solutions in three large groups by methodological orientation: (1) traditional machine learning techniques and feature engineering, (2) hybrid representations combining multiple embedding strategies, and (3) deep learning architectures leveraging neural networks. In the review, the author not only points out the changes in the modelling methods but also philosophizes about the advantages and shortcomings of each method.

A. Traditional Machine Learning Models and Feature Engineering

A couple of researchers comes up with a simple approach for fake news detection using a naive Bayes classifier [2]. Data were collected from three large Facebook pages as well as three large mainstream political news pages (Politico, CNN, and ABC News). The developed system was tested on a relatively new data set of Facebook news posts each of which represents a news article allowing evaluating of its performance on recent data, and achieved a classification accuracy of 74%. During implementation, only a few of the fields are used as the link to the Facebook post with the text of the news article. 1771 cleaned news article was obtained after data cleanings, such as ignoring the null texts and news articles with mixed labels. Dataset was then randomly shuffled and divided into training, validation, and test datasets. The validation dataset was used for tuning some global parameters of the classifier. The test dataset was used to get an unbiased estimation of how well the classifier performs on new data. The Precision achieved for the given classifier equals 0.71; recall equals 0.13. Such a low value of the Recall is caused by the skewness of the data in the test dataset. The researcher's goal was to show that even quite a simple artificial intelligence algorithm, such as a naive Bayes classifie,r may show a good result on such an important problem as fake news classification. They have mentioned several ways to improve the classifier, such as getting more training data, greater the length of the news article, removing stop words, using stemming, treating rare words separately and using a group of words instead of separate words for calculating probabilities.

This paper proposed a news text classification architecture based on Latent Dirichlet Allocation (LDA) [3]. This model uses a topic model to make text dimension reduced and get features because the dimension of the news texts is too high. The news text information of some websites has a variety of categories, such as entertainment, sports, politic, and so on. For this reason, the paper also makes a research on the Softmax regression algorithm to solve multi-class of text problems in our life and make it a model classifier. The proposed LDA is a kind of topic model algorithm based on a probability model and thinks that each article is comprised of a plurality of topic mixture thus it can detect potential hidden knowledge topics in a larger document set. The algorithm assumes that each

word in the corpus in an article is through by "with a certain probability to choose a topic, and then from this subject with a certain probability to select a word". A topic model is a kind of unsupervised learning method, but text classification is usually used in supervised learning. So, they used Latent Dirichlet Allocation Supervised (SLDA), and the class labels are from experimental data. The paper uses Precision, Recall, and F1-Measure as a measure of predictive performance evaluation index. The dataset contains 20 different news groups written in the past few years. The dataset covers a wide range of topics, including religion, medicine, art, and so on. The model chooses three kinds of categories out of these 20-news groups, and the data was divided into training and test groups for performance measurement. This paper does an excellent job in two dimensions- the first one is to propose a news text classification based on the LDA topic model, which is mainly based on the latent topic information of the text, and the second one is it use multi class classification using Softmax regression. The researchers also discussed some of the shortcomings which can be improved in the future, such as the choice of the parameters of the topic model, the size of the news text, etc.

Nikam & Dalvi [4] proposed and developed a method on Twitter for the binary classification of fake news. They created a web-based GUI for this system to classify tweets into two binary classes as genuine or fake. They have used passive-aggressive ML algorithms and naïve Bayes, and the TF-IDF vectorization method and compared tweets with real news sources. In the paper, they did some literature review and mentioned some works done using numerous classification algorithms such as k-means clustering, n-gram classifier, k-nearest neighbor, support vector machine (SVM), logistic regression, decision tree, random forests, and feature extraction methods like TF, TF-IDF, Doc2Vec, Bag of Words, etc. They also mentioned some shortcomings of this research, such as lower Accuracy considering domain-specific features, punctuation errors not considered, and a lack of use of real data. For their project work, after collecting data from Kaggle, which contains 4048 samples with a balanced class, having around 50% samples from each class they have done some preprocessing work, such as discarding duplicate data and stop words, tokenization, and stemming, and got 3983 cleaned samples. Then, for feature extraction, they used term frequency (TF) and inverse document frequency (IDF), and for the classifier, they used naïve Bayes and passive-aggressive algorithms. To compare the performance of these two algorithms, they used Accuracy, Precision, Recall, and F1-score metrics and achieved F1-Score as slightly above 50% for both classifiers and an accuracy score of 73% for naïve Bayes and 78% for the passive-aggressive classifier. They mentioned their scope is limited to text datasets only, and for future work, they suggested using image data as well.

Kesarwani et al. [5] come up with a simple approach for classifying fake news on social media, the Facebook news posts dataset, using a k-nearest neighbour classification algorithm. In this paper the reviewed some past research work with different approaches such as N-gram, non-negative factorization, RST-SVM, LIWC, and SVM classifier and decided to experiment with the k-nearest neighbour classification algorithm. KNN evaluates the distance (e.g., Euclidean distance) among the data points to be classified, then picks the closest points in terms of the lowest distance and then considers the majority of the points. For training and testing dataset from Buzz Feed News organization has been used, which collects information about Facebook posts which represent a news article independently. The dataset was kind of imbalanced because out of 2282 samples 1669 posts are true, 104 are false, 245 are mixed, and 264 undetermined and contains a small number of articles, such as 2000. They split the datasets into training and testing using 80% to use as training purpose of the k-nearest neighbour algorithm and 20% for testing purposes to see how well the algorithm works for unseen data. The researchers considered only five features to train the model. The hyperparameter of this model, K, is carefully chosen so that KNN performs well for this dataset. They measured classification outcomes such as Precision, Recall, Accuracy, and F1-score. This model managed to get Accuracy of 79% with a Precision of 0.75 and a Recall of 0.79 after testing the algorithm with different value setting of K. Future work in this paper suggested testing this with different other classification algorithms and compare the results.

Researchers from the School of Computer Science and Engineering, Vellore Institute of Technology, suggested Support Vector Machine (SVM) with the feature extraction method TF-IDF as the most accurate one after simulating in various scenarios with classification algorithms such as Naïve Bayes, Support Vector Machine (SVM), Logistic regression, and decision tree. Moreover, they compared all these classifiers and score against two feature extraction techniques as count vectorizer and TF-IDF vectorizer [9]. The researchers have used a Kaggle dataset where content has been extracted from 244 websites and contains 13000 posts for one month. The parent dataset contains articles from the top 15 United states publications, with 150000 articles from news websites and RSS feeds. However, researchers chose randomly selected 13000 articles from this and merged them with the fake news dataset for avoiding skewness as well as accurate prediction results. After data collection, they have done some preprocessing to make the data clean, such as tokenization and stop word removal. Then they have used the CountVectorizer and TF-IDF vectorizer to generate vectors of the document. Count vectorizer only considers the frequency of each word appearing in the document while TF-IDF not only considers frequency but also the importance of the words in the documents. After vectorization, the researchers put that in several algorithms such as Naïve Bayes, Logistic Regression, Decision Trees, Support Vector Machine, and Artificial Neural Networks. For the Artificial Neural Network architecture, they have used ReLU for the hidden layers and the sigmoid function for the output layer as an activation function. They have simulated all these algorithms for both vectorizers, which are the count vectorizer and the TF-IDF vectorizer. From the result analysis, researchers have found that SVM, along with TF-IDF vectorizer, give the highest Accuracy although both Logistic Regression and SVM tend to give better scores with larger datasets. Artificial Neural Network performs worst

for both vectorizers, although the researchers had expected better results from the neural network architecture. As a future direction, they suggested making use of multiple attributes such as content along with the title, source of news, and other statistics for better results.

Two researchers from KLS Gogte Institute of Technology experimented and compared various machine learning techniques, such as Logistic Regression, Naïve Bayes, Support Vector Machine, Random Forest, and Deep Neural Network for detecting fake news in binary classification. They have used a news dataset that is collected online and consists of news from different sites. They have preprocessed the data by stemming and removing stop words. They have built their architecture on the Java system. For evaluation metrics, they have used an Accuracy score to see the performance of each algorithm. They have expanded their analysis further by adding the time and memory consumption of each algorithm during training [10]. From the result analysis, it has been found that Deep Neural Network consumes more memory than other algorithms and for training time, Deep Neural Network is much faster than others, where Logistic Regression takes lots of time to train. Deep Neural Network outperformed all other algorithms, achieving an Accuracy score of 0.91, where Logistic Regression has done the worst. The purpose of the researchers in this paper was to compare different machine learning algorithms for detecting fake news. They have considered three different aspects to compare – memory consumption, training time, and Accuracy score. Although Deep Neural Network consumes more memory, the training time is faster than other algorithms, and the Accuracy score is also far better. So, the researchers have concluded as Deep Neural Network is the most convenient algorithm for this case.

B. Hybrid and Advanced Embedding Techniques

With the help of machine learning and NLP, a couple of researchers have come up with several models and methodologies, including seven different ML classification algorithms, along with several NLP vectorization techniques, such as count vector, word embedding, and TF-IDF, used to detect fake news. They have used the Kaggle dataset that is separate for real and fake news, and they combined the dataset for training purposes. After doing a literature review, they come up with methodology steps as text collection, preprocessing, vectorization and classification [6]. They have collected a dataset from Kaggle that has 18574 posts as a training dataset and 9149 posts as a test dataset from 244 websites. For preprocessing, they have performed a lowercase operation; removal of numbers, punctuations, accent marks, white spaces, and stop words. For feature extraction, they have used Bag of Words and Word Embedding using Spacy. Count vectorizer and TF-IDF are used for the vectorization of text data. For classification purposes, researchers have used seven different classification algorithms: Support Vector Machine, Logistic Regression, Decision Trees, Random Forest, Gradient Boosting, XG-Boost, and Neural Network. Several performance evaluation metrics, Accuracy, Precision, Recall, and F1-score used to measure and compare the performance of different algorithms. While result analysis, it is observed that, Neural Network performed best using the count vectorizer and word embedding with Accuracy scores 0.94 and 0.90 respectively, whereas for the TF-IDF vectorizer, the Support Vector Machine did well with an Accuracy score 0.94. They have concluded with the statement that because the Accuracy score is quite close and the neural network is complex as well as takes longer to train so they have considered linear SVM, which is a simple model and takes less time to compute. The future direction provided by the researchers to use deep learning methods, as well as sentiment analysis, and more Accuracy can be achieved with a dataset with many articles.

Three researchers proposed a method to identify fake news using a Support Vector Machine (SVM) classifier along with TF-IDF and Word2vec. Word2vec is effective to calculate text similarity, and TF-IDF calculates feature weights evaluating the importance of a word in a document, and has a record of success for information extraction. Word2vec is calculated in terms of the context of the word, capturing full semantic information, and from the vectors, the similarity of two words can be calculated easily. Word2vec comprises two models named CBOW which predicts the probability of occurrence within a certain size of window, and Skip-gram predicts the context words from the current words [8]. A total of 5600 social news articles were collected from two different data sources such as from some websites and research institutes. After data cleaning, 5581, they managed to get news items of which 51.5% are real news and the rest are fake. The researchers segregated their method into three main parts- (1) Data preprocessing, (2) Feature extraction, and (3) Text classification. For data preprocessing they have done word segmentation as well as the removal of stop words and punctuation. In the feature extraction process, they have combined the two methods, Word2Vec and TF-IDF. They first calculated the word vector using Word2vec and TF-IDF value, and then concatenate these two to generate a document vector. Finally, they used Support Vector Machine (SVM) for binary classification purposes to find out whether it was real or fake. They have trained and experimented with an SVM classifier with three different scenarios- TF-IDF alone, Word2vec alone, and a combination of Word2vec and TF-IDF. For evaluation metrics, the researchers have used Precision, Recall, and F1-score. They have split the dataset into a 73:27 ratio as train and test to get 4079 samples for training and 1502 for testing. From the experiment, they found a good average F1-score of 0.87, and the combined method performed better than solo. So, they concluded that both Word2vec and TF-IDF performed well in their experiment, but the combined method performed better and thus most suitable and did not provide any future direction of work in this paper.

C. Deep Learning Architectures and Hybrid Neural Models

Girgis et al. [7] proposed a hybrid model comprising the GRU and CNN algorithms, considering a deep learning perspective rather than looking only at the content, and applied to the dataset called LIAR. They discussed that most researchers in the past used different deep learning algorithms such as

CNN, RNN, Bidirectional-LSTM, but they have mostly been unsuccessful because the dataset itself because of small and contains unrealistic news. So, in this paper, they mentioned dataset LIAR as the potential solution to this problem. This dataset has 12836 small statements labelled and with many features and natural context, and collected from different sources such as social media, TV's, interviews, political discussions, news channel, etc. For data preparation before feeding into the algorithms researchers have done sentence splitting, removal of stop words, and stemming. Word Embedding is used for feature extraction, which is the input of the RNN and LSTM's algorithms which have the output as binary classification as real or fake. Accuracy score is used as a metric for performance evaluation of seven different classifiers, where vanilla performed worst, and LSTM also is not promising compared to GRU and CNN. So, they have suggested a hybrid approach merging GRU and CNN as future work.

Umer et al. [11] proposed a hybrid approach building a Neural Network architecture by combining the capabilities of CNN and LSTM along with two different dimension reduction algorithms, Chi-square and Principal Component Analysis (PCA). They have used Accuracy and F1-score as evaluation metrics to compare the performance of PCA with Chi-square while classifying the fake news dataset, which has four different types of stances. They have used the benchmark dataset of Fake News Challenges, which has $75,385$ labelled samples with 2587 articles related to 300 headlines having four stances: 1) Agree-relation between headline and body, 2) Disagree-no relation at all, 3) Discuss-little match, and 4) Unrelated-totally different. The dataset was split as a train-test ratio of around $67:33$, the train data is for training the algorithm, while the test data is used to evaluate how the algorithm performs against unseen data, thus generalizing the system. Several preprocessing tasks were performed, such as lowercase conversion, removal of stop words, stemming, and tokenization. For stemming, they have used NLTK's Porter Stemmer algorithm. After preprocessing, they have used word2vec to convert text to a list of vectors and a dictionary of 5000 words created. They have constructed the hybrid architecture with CNN and LSTM and simulated for four different scenarios- a) all the features without preprocessing, b) preprocessed but with all the features, c) with reduced features by PCA, and d) with reduced features by Sci-square. For CNN, they have used ReLu as activation function fed to a 1-D max-pooling layer, and a dropout rate of 0.2, then used 100 units in the LSTM layer and finally a dense layer as fully connected, followed by a softmax activation function. They have set hyperparameters as batch size 32, the number of epochs 50, and the Adam optimizer. The researchers have used Accuracy, Precision, Recall, and F1-score as evaluation metrics. From the result analysis after training and testing, it is observed that the CNN-LSTM architecture with PCA outperformed all others, producing a promising result of 97.8% Accuracy, which is better than previous studies. The researchers have concluded their research with future direction as validation of performance with a larger dataset, introducing tree-based learning, and analyze of various textual features and their combination.

D. Summary of Literature Review

The presented literature allows tracing the actual trend of the shift of classical supervised learning algorithms with handcrafted features to the multifaceted hybrid and deep learning models. Classical methods are interpretable and computationally efficient, whereas hybrid methods and deep neural networks are raising the performance limits at the expense of explainability and computational requirements. This development highlights the need to balance between accuracy, interpretability and scalability in fake news detection. The findings of this review have guided the choice of methods in the given study, namely, the combination of sentiment-based features and explainable machine learning models to handle the peculiarities of headline-based classification.

III. Dataset Description and Constraints

A. Dataset Description

The proposed research is based on the Fake and Real News Dataset, obtained on Kaggle, consisting of $44,898$ news articles written in the English language. The dataset gained much adoption in the research of fake news detection thanks to clear labels on its instances and equal balance between fake and real news categories. It includes news items, gathered on a diversified media base, which provides a representative overview of political and general news discourse from various perspectives.

The dataset employed in this research is publicly available on Kaggle at the following URL: https://www.kaggle.com/datasets/clmentbisaillon/fake-and-real-news-dataset. The five main characteristics of the dataset are title, text, subject, date, and label. The label column classes each entry as either fake or real, of which $23,481$ are fake and $21,417$ are real, approximating to 52 percent fake and 48 percent real. This small class imbalance should not pose much of a problem to binary classification, but could be important when evaluating the model.

B. Initial Exploratory Data Analysis (EDA)

An initial EDA was performed in order to gain more insight into the structure and distributional characteristics of the data. Regarding the headline length, calculated based on the title field, the mean headline length is about 80 characters, and the median one is 73 characters. The minimum length of a headline is 8 characters, and the maximum is 286 characters. The interquartile range (IQR) is located within the boundaries of $63-91$ characters, which means that the majority of the headlines are quite short, which is compliant with the traditional journalistic approach.

The subject attribute gives categorical data on the theme of each article. The most widely presented topics are: politicsNews (11,272 articles), worldnews (10,145 articles), news (9,050 articles), and politics (6,841 articles). More specialized and less numerous but still within the scope are such categories as "left-news" (4,459), "Government News" (1,570), "US_News" (783), and "Middle-east" (778). The diversity of the subjects contributes to the suitability of the dataset to a

broad variety of domain-specific classification tasks and its generalization.

Preliminary inspection of the dataset suggests the potential presence of anomalies such as duplicate entries and extremely short headlines. These issues will be systematically addressed during the preprocessing phase to enhance data quality and ensure the robustness of subsequent classification models.

C. Constraints

Although the dataset is quite comprehensive, it has various limitations. First, it is monolingual and limited to English, thus omitting news content of non-English origins and limiting its use in multilingual settings. Second, the data is unimodal and contains only text; there are no images, videos or other multimedia objects, which are frequently incorporated in the news articles nowadays. Third, it does not contain any interactional and contextual metadata (user comments, shares and likes) which other research has demonstrated can affect the spread and believability of fake news.

Also the sources of collection of the dataset could induce some biases. The news sources present either fake or real news, which are not evenly spread out and could present a certain ideological inclination, which might induce a bias in the model generalization. The other difficulty lies in the nature of the headlines themselves. Both fake and real news are likewise loaded with sensationalism and clickbait-style wording, which can cause models to pick up on shallow stylistic features instead of semantic interpretation and subsequently over-estimate performance scores without actually solving the content-based deception detection problem.

D. Summary of Dataset Description and Constraints

To conclude, the Fake and Real News Dataset is a large and labelled dataset that is suitable for practical tasks in the field of misinformation identification based on natural language processing and machine learning. Nevertheless, its structural peculiarities, class distribution, and representational biases should be carefully considered to guarantee the resilient and interpretable model creation.

IV. Project Approach

A. Design Approach

The overall pipeline for this project is illustrated in Fig. 1. The project began with merging and labeling two publicly available datasets: one containing fake news and the other real news. After concatenation, all rows with missing values were removed to ensure data integrity. The target variable was encoded into binary labels (1 for real, 0 for fake).

Following data cleaning, the project employed TF-IDF vectorization to transform raw textual headlines into numerical feature representations. This transformation captures term relevance across documents and helps mitigate the impact of frequent but uninformative words.

The dataset was then split into features and labels and passed through a 5-fold stratified cross-validation setup to maintain class balance. The following three machine learning classifiers were trained and evaluated:

- Logistic Regression
- Support Vector Machine (SVM) with a linear kernel
- Random Forest Classifier

Each model was assessed using the F1-score metric averaged over the cross-validation folds. The workflow ensures generalizability and robustness in performance estimates.

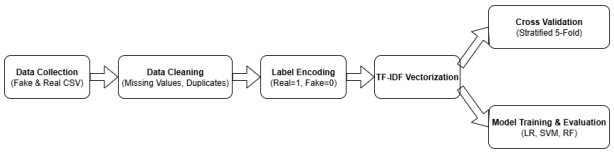

Fig. 1. End-to-End Pipeline for TF-IDF and Classical ML Models

B. Implications of Research

The given study is likely to produce some important findings that have practical as well as theoretical value. First, the research seeks to determine the degree of contribution of sentiment features, i.e., emotional polarity and subjectivity, in enhancing detection performance by training headline classification models. The identification of this relationship will be useful in explaining whether sentiment is a predictive cue or it is just a reinforcement of already existing lexical tendencies in fake news.

Second, model interpretability analysis, especially via feature importance or SHAP values, will aid in identifying those individual linguistic features, like the strength of sentiment or exaggeration in language, that are highly linked with deceptive text. These findings could improve the explanations and the credibility of classification systems.

In practical terms, the findings of this research can be used as a stepping stone towards creating lightweight and scalable extensions, executable as browser extensions or content filtering systems that can indicate potentially misleading headlines on the fly. The fact that the detection at the headline level can be carried out with small and explainable models provides a practical benefit in situations when either computational power or user focus is at a premium.

Last but not least, the research adds to the academic discussion of fake news detection by investigating the relationship between sentiment analysis and model explainability. Its contribution to the current body of literature is empirical confirmation of the claim that the addition of affective cues to traditional NLP features can improve model performance and model interpretability, thus simultaneously overcoming two historical challenges in the literature.

V. Results

In this section, we present the classification results obtained from three machine learning models: Logistic Regression, Support Vector Machine (SVM), and Random Forest. The models were trained on TF-IDF vectorized headline text and evaluated using 5-fold cross-validation. Evaluation metrics include the F1-score and confusion matrices.

A. F1-Score Comparison

Figure 2 illustrates the average F1-score achieved by each classifier. The Support Vector Machine (SVM) model achieved the highest F1-score of 0.948, followed closely by Logistic Regression with an F1-score of 0.942. The Random Forest classifier also performed comparably, achieving an F1-score of 0.942.

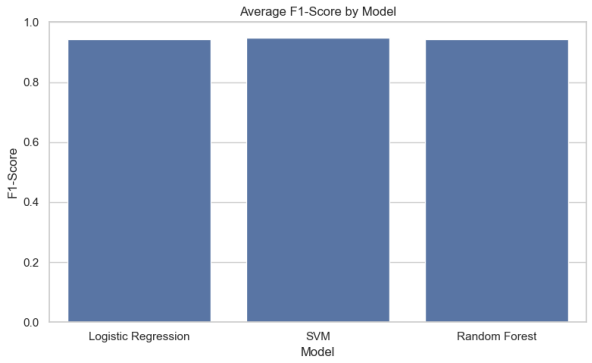

Fig. 2. Average F1-Score by Model

B. Confusion Matrix Analysis

The confusion matrices in Figures 3, 4, and 5 provide deeper insights into model performance by displaying the number of true positives, true negatives, false positives, and false negatives for the Logistic Regression and Random Forest models.

For Logistic Regression (Figure 3), the model correctly classified 21,990 fake news items and 20,402 real news items, with false positive and false negative counts being 1,491 and 1,015 respectively.

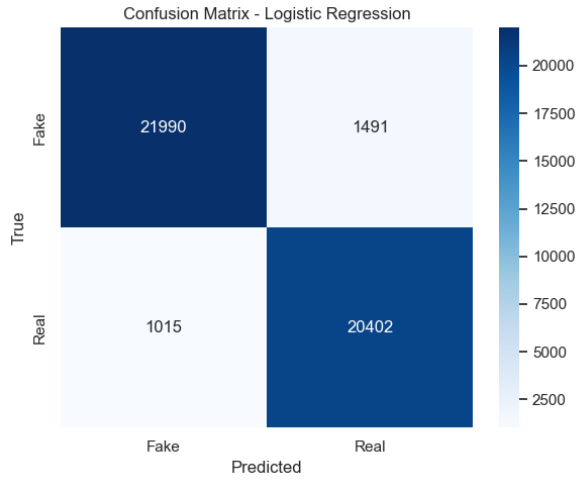

Fig. 3. Confusion Matrix - Logistic Regression

Similarly, the Random Forest model (Figure 4) correctly classified 21,991 fake news and 20,394 real news articles. The false positive and false negative counts were 1,490 and 1,023 respectively, indicating a nearly identical performance to Logistic Regression.

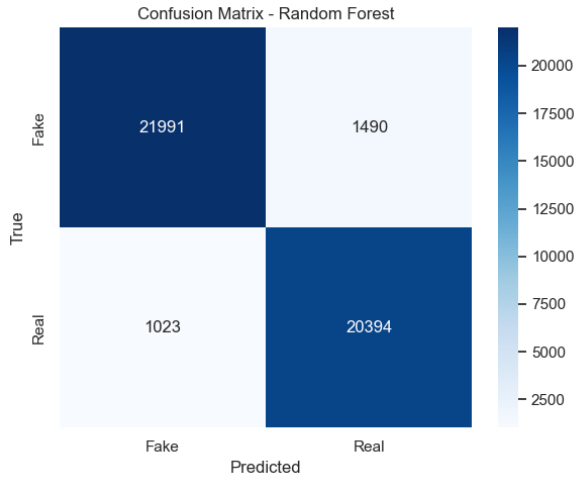

Fig. 4. Confusion Matrix - Random Forest

And, the SVM model (Figure 5) correctly classified 22,208 fake news and 20,444 real news articles. The false positive and false negative counts were 1,273 and 973 respectively, indicating a nearly identical performance to Logistic Regression and Random Forest.

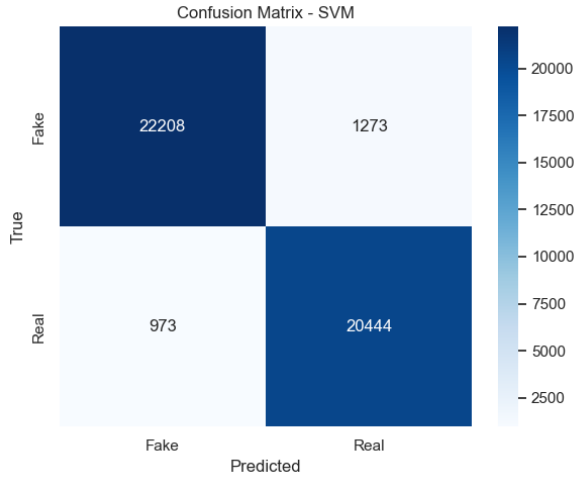

Fig. 5. Confusion Matrix - SVM

C. Summary of Results

The experimental findings support the feasibility of using TF-IDF vectorized news headlines for fake news detection. All three models demonstrated strong performance, with F1-scores exceeding 0.94. Among them, SVM showed the best overall results. The confusion matrices further validate the models' effectiveness, exhibiting high accuracy and balanced error distribution.

VI. Conclusion and Future Work

A. Conclusion

The project in question examined how effective the methods based on natural language processing, namely TF-IDF

vectorization, could be combined with the classical models of machine learning and be applied to the task of classifying the news headlines as fake or real. After conducting a comparative analysis of the three popular classifiers which included Logistic Regression, Support Vector Machine (SVM) and Random Forest, we found that the above achieved top-ranking F1-scores in all the models which were consistent, where SVM scored the best at 0.948.

The findings show that in the case of fake news detection not only with fairly low-level description technique, such as TF-IDF, but also specifically in the task at a textual level of headings, high accuracy can be attained. Their predictive power of Logistic Regression and Random Forest was further confirmed by the confusion matrices that indicated low numbers of false positive and false negative rates.

Notably, the research questions are also supported in the findings. NLP-based features managed to successfully identify the headlines that were fake demonstrating that some lexical patterns and word distributions correlate with the fake news. Although the sentiment polarity has not been immensely involved in the modeling process of this phase, its viability was tested during EDA and it will also remain a significant aspect in future versions.

B. Future Work

A number of improvements can be imagined in order to add dimensions and width to this project:

- Deep Learning Architecture: Next work might address sequential neural network architectures (e.g. LSTMs or transformer-based models (e.g., BERT)) that are better designed to capture contextual flairs in text.
- Explainability: The application of explainable AI methods, like SHAP or LIME would allow to discover which words or other linguistic characteristics produce the strongest impact on classification.
- Sentiment and Semantic Analysis: The polarity of a sentiment in the sentences and the semantic content could be studied more closely, which could provide additional levels of interpretability especially in terms of extracting persuasive or emotionally stimulating language that is prevalent in fake headlines.
- Generalized and Larger Dataset Incorporation: The current Kaggle dataset is rather limited (only English-language headlines and only official Kaggle sources). Incorporation of more sources, in the form of other languages, or even platforms that grew out of the misinformation phenomenon, would help generalizability and robustness.
- Deployment: Since this is a real-life application, a browser-extension or a go-light programming interface to give instant credibility ratings of headlines could be created.

To summarize, the available findings have allowed constructing a robust baseline of light and efficient fake news detectors based on the text of the headline, and there are several exciting potential research directions.

REFERENCES

[1] Syed Ishfaq Manzoor, Jimmy Singla, Nikita, "Fake News Detection Using Machine Learning approaches: A systematic Review," *2019 3rd International Conference on Trends in Electronics and Informatics (ICOEI),* Tirunelveli, India, 23-25 April 2019.

[2] Mykhailo Granik, Volodymyr Mesyura, "Fake news detection using naive Bayes classifier," *2017 IEEE First Ukraine Conference on Electrical and Computer Engineering (UKRCON),* Kyiv, UKraine, 29 May-2 June 2017.

[3] Zhenzhong Li, Wenqian Shang, Menghan Yan, "News text classification model based on topic model," *2016 IEEE/ACIS 15th International Conference on Computer and Information Science (ICIS),* Okayama, Japan, 26-29 June 2016.

[4] Shivani S Nikam, Rupali Dalvi, " Machine Learning Algorithm based model for classification of fake news on Twitter," *2020 Fourth International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC),* Palladam, India, 7-9 Oct. 2020.

[5] Ankit Kesarwani, Sudakar Singh Chauhan, Anil Ramachandran Nair, " Fake News Detection on Social Media using K-Nearest Neighbor Classifier," *2020 International Conference on Advances in Computing and Communication Engineering (ICACCE),* Las Vegas, NV, USA, 22-24 June 2020.

[6] N. Smitha, R. Bharath, " Performance Comparison of Machine Learning Classifiers for Fake News Detection," *2020 Second International Conference on Inventive Research in Computing Applications (ICIRCA),* Coimbatore, India, 15-17 July 2020.

[7] Sherry Girgis, Eslam Amer, Mahmoud Gadallah, " Deep Learning Algorithms for Detecting Fake News in Online Text," *2018 13th International Conference on Computer Engineering and Systems (ICCES),* Cairo, Egypt, 18-19 Dec. 2018.

[8] Shenhao Zhang, Yihui Wang, Chengxiang Tan, " Research on Text Classification for Identifying Fake News," *2018 International Conference on Security, Pattern Analysis, and Cybernetics (SPAC),* Jinan, China, 14-17 Dec. 2018.

[9] Karishnu Poddar, Geraldine Bessie Amali D., K.S. Umadevi, " Comparison of Various Machine Learning Models for Accurate Detection of Fake News," *2019 Innovations in Power and Advanced Computing Technologies (i-PACT),* Vellore, India, 22-23 March 2019.

[10] Chaitra K Hiramath, G. C Deshpande, " Fake News Detection Using Deep Learning Techniques," *2019 1st International Conference on Advances in Information Technology (ICAIT),* Chikmagalur, India, 25-27 July 2019.

[11] Muhammad Umer, Zainab Imtiaz, Saleem Ullah, Arif Mehmood, Gyu Sang Choi, Byung-Won On, "Fake News Stance Detection Using Deep Learning Architecture (CNN-LSTM)," *IEEE Access (Volume: 8),* 26 August 2020.