



City Research Online

City St George's, University of London

Citation: Hussain, S., Khattak, A. S., Abbasi, R. A., Russell-Rose, T. & Chinthalapati, V. L. R. (2024). Identification of Targets in Disinformation News Articles Using Supervised Machine Learning. Paper presented at the 20th International Conference, ADMA 2024, 3-5 Dec 2024, Sydney, NSW, Australia. doi: 10.1007/978-981-96-0847-8_15

This is the accepted version of the paper.

This version of the publication may differ from the final published version. To cite this item please consult the publisher's version.

Permanent repository link: <https://openaccess.city.ac.uk/id/eprint/34317/>

https://doi.org/10.1007/978-981-96-0847-8_15

Copyright and Reuse: Copyright and Moral Rights remain with the author(s) and/or copyright holders. Copies of full items can be used for personal research or study, educational, or not-for-profit purposes without prior permission or charge, unless otherwise indicated, provided that the authors, title and full bibliographic details are credited, a hyperlink and/or URL is given for the original metadata page and the content is not changed in any way. For full details of reuse please refer to [City Research Online policy](#).

Identification of Targets in Disinformation News Articles using Supervised Machine Learning^{*}

Sadam Hussain¹ Akmal Saeed Khattak² Rabeeh Ayaz Abbasi³ Tony Russell-Rose⁴
and Venkata L. Raju Chinthalapati⁵

¹ Goldsmiths, University of London, UK shuss007@gold.ac.uk

² Quaid-i-Azam University, Islamabad, Pakistan akhattak@qau.edu.pk

³ Quaid-i-Azam University, Islamabad, Pakistan rabbasi@qau.edu.pk

⁴ Goldsmiths, University of London, UK t.russell-rose@gold.ac.uk

⁵ Goldsmiths, University of London, UK v.chinthalapati@gold.ac.uk

Abstract. Fake news or disinformation spreads widely among various communities worldwide due to the advancement in technology involving social media platforms such as Facebook, Twitter, and Instagram in our daily lives. Disinformation news is designed to mislead and deceive the public against some entities, particularly countries, the public, religion, etc. This news is frequently disseminated by individuals, organizations, covert agencies, or nations to target particular governments and organizations to damage their international standing. In the past, academics concentrated on issues related to classification problems, identifying fake news, and detecting fake profiles. In the field mentioned above, locating hidden targets is a popular topic of investigation. In the proposed work, we have used the EU Disinfo Lab dataset to identify the targets within the disinformation news articles. The targets in disinformation news are identified using content features, unigram, bigram, unigram with bigram, and unigram with trigram. The proposed model is trained using supervised machine learning techniques such as the Linear Support Vector Classifier (LSVC) and Logistic Regression (LR), as well as three ensemble methods: Random Forest (RF), Passive Aggressive (PA), and extreme Gradient Boosting Classifier (XGB). For nine classes, the LSVC performed better on all four N-grams, including unigram, bigram, unigram with bigram, and unigram with trigram. This classifier also performed better for three classes except for unigram with bigram and unigram with trigram features; for these features, it was the second highest after LR. The targets were correctly identified using contents features by unigram with bigram and unigram with trigram, with a higher accuracy of 77% for each.

Keywords: Russia-Ukraine war, COVID-19, Machine Learning, Target Identification, Disinformation

^{*} Supported by organization x.

1 Introduction

Fake news means disinformation, alternative facts, misinformation, media manipulation, and propaganda. It is a form of news consisting of deliberate disinformation spread via traditional news media such as print, broadcast, and online social media. Fake news is defined as purposely and verifiably misleading information that has the potential to mislead readers [1]. The definition has two main characteristics: authenticity and intent. First, fake news includes false information that can be verified. Secondly, fake news is created with dishonest intentions to mislead people. Disinformation has negative consequences, including reputational harm, financial loss, and even life danger [2]. The faster the spread of false news, the sooner the terrible influence occurs. Fake news usually targets a specific country regarding its secret service, product, religion, cities, organizations, celebrities, and politicians to gain financial or political advantage and give the country a bad reputation [3]. Fake news is an unconventional strategy for targeting anything (entity). Fig. 1 depicts the various stages of fake news, including how it originates, spreads, and is consumed in society to target a specific entity. EU Disinfo lab⁶ is a new independent non-governmental organization dedicated



Fig. 1. The image consists of different stages: how fake news is designed using digital devices, how it spreads, and how it is consumed in societies with the help of news organizations

to researching and combating sophisticated disinformation campaigns aimed at the EU, its member states, core institutions, and core values.

The main purpose of the lab [4] was to monitor disinformation campaigns against European countries due to Russian interference in the US and European elections to counter disinformation [5]. According to the EU Disinfo lab, which debunked in their monitoring process, about 20 countries had been targeted in approximately 6500 disinformation cases since 2015 [5]. Marcoux et al., [6] scraped the data from their lab website⁷ for topic modeling using keywords.

⁶ <https://www.disinfo.eu/>

⁷ <https://euvsdisinfo.eu/disinformation-cases/>

The *geopoliticia.ru* news organization, which discussed the US ideology and NGOs, published an article about African societies on 21.09.2021 and claimed that the US wanted to finish African societies by imposing their ideology, represented by globalism, with the help of NGOs.

News circulated by *geopoliticia.ru* on May 20, 2021, related to Serbia's claims that US, UK, and EU forces have had total control of Serbian security and defense forces since 2000, and it has become their colony. They claimed that the West brainwashed the terrorists in Serbia to use it for their benefit. More than 98,000 Ukrainians crossed the Romanian border in December 2022, when false information began circulating. As time passed, false information began to spread on social media, increasing fear, anxiety, and uncertainty among people everywhere.

COVID-19 vaccines have been targeted by fake news, with a report attributed to well-known virologist and Nobel laureate Luc Montagnier claiming that all vaccinated people will die in two years [7].

Research Questions:

The following are the research questions that have been addressed in this research study:

- **RQ1** How can we create a dataset for automatically classifying targets from disinformation news articles?
- **RQ2** What is the performance of various classifiers and ensemble approaches for identifying different target countries in disinformation news articles?
 - Which classifier and ensemble approach most effectively identify the targets?
- **RQ3** What content-based features are beneficial to identify targets in disinformation news articles?

Contributions:

The following are the major contributions of this research study:

- The initial dataset came from Kaggle, but it wasn't enough to train a research model. therefore, new news articles were scraped to gather the data for training the model, labels were manually designed for each news article, and the dataset's structure was further enhanced to make it more suitable for conducting experimental research work.
- Most past research work in the literature has been conducted on short news stories, i.e., most researchers did not consider the whole news article. However, our research used large news texts and complete news stories. Abbreviations are transformed to their extended form for keyword feature preservation.

- The proposed model was evaluated for unseen data after using hyperparameter tuning with ⁸ functions for all the machine learning classifiers and ensemble methods with N-grams features.
- The proposed framework work was trained and evaluated in two multiclass classification experiments.

2 Related Work

Researchers are interested in working in this field because of the propagation of disinformation by users. The majority of researchers are focused on detecting this news. This news is disseminated on social media by a variety of accounts. They defined many ways, such as crowdsourcing, expert fact-checking, machine learning, natural language processing techniques, hybrid techniques, deep learning, social networking, neural networks, etc. This section will discuss detecting fake news with machine learning classifiers and ensemble methods.

2.1 Classification

Labeled data is used in supervised machine learning to classify the input with its predicted output classes. Datasets that contain feature sets and output labels are called labeled datasets. The dataset labels are binary or multi-class, which could be predicted through text features and known labels. The output class may be in three categories, i.e., continuous, categorical, or binary [8]. The vectors feature predicts the output and assigns it to the associated classes. Researchers commonly use textual features to distinguish between fake and authentic news in news data.

Logistic regression is the earliest method used for classification problems and was introduced by David Cox in 1958 [9]. The main purpose was to predict the model based on its feature matrix and the categorical variable as the output based on independent variables. Logistic regression classifier is used for textual data; for instance, in Twitter, the classifier analyzed the users' reviews about a particular product and identified whether the users' reviews about the product are positive or negative [10].

The Random Forest algorithm is used for text classification and was proposed in 1995 by T. Kam Ho [11]. It developed a random decision tree, which was improved by L. Breiman in 1997. The algorithm performs well in training a model but is comparatively slow in testing data. Random forest algorithm used by Selvi et al., [12] multi-label text classification with Rocchio algorithm, which takes vectors as input and gives the appropriate category of the text document.

Linear SVC(Support vector classifier) The Support Vector Classifier is a supervised machine learning algorithm designed for classification problems. It performs three different tasks: training, optimization, and prediction. It uses a line among the classes to separate each class from the other based on the vector

⁸ GridsearchCV

features set. Using quadratic programming ensures that the data are classified accurately for future predictions. The last step is to decide and classify the used data to predict the future output [13].

Xgboost stands for Extreme Gradient Boosting, an efficient and scalable gradient boosting framework. It should be used in many machine-learning problems like classification, regression, and ranking. Xgboost algorithm inputs different data formats, like dense matrix, sparse matrix, Data file, and GB.matrix, and its classes [14]. It can be used in a deep learning approach called the N-channel convolutional network, which uses embedding and lexicon-based feature approaches. Similarly, the Xgboost was used for detecting emotions on Twitter by comparing two languages, English and Arabic [15].

A passive-aggressive classifier was proposed by Koby Crammer in 2006. The main purpose of the classifier was to solve binary classification problems. Therefore, it has been upgraded to solve the multiclass problems [16], which has used TF-IDF counts and a shared set of learned weights. It is mostly used by the researcher for massive textual data contents, in which it performs better than SVM [17]. Therefore, the authors used the algorithm to detect fake news and outperform the naive Bayes algorithm but have produced a similar result as the SVM algorithm. For instance, classify the document text [18]. The research work of [19] has shown that this classifier performed well in fake news detection. These methods have been used to identify the hidden targets correctly inside disinformation news.

2.2 Feature Extraction

Generating outcomes from cleansed data mostly involves extracting features to anticipate the outcome. Machine learning produces a result based on features that predict the label. For extracting features from text data, several methodologies and strategies are employed. The literature we studied mostly used the three types of features discussed by [20]: contents-dependent, context-dependent, and domain-dependent. Supervised learning algorithms such as SVM, logistic regression, linear regression, and deep learning algorithms such as LSTM [21] and Recurrent Neural Network (RNN) used these features. Because textual data generates features from raw data to filter for classifiers, most content-dependent features are used in Natural Language Processing (NLP). According to the literature, no research has been done to identify the targets inside disinformation articles. Researchers focus on machine and deep learning content features when working with textual data. Different strategies are used in NLP to extract features from raw input and then execute operations to set them up for categorization. Word2vec, Bag Of Words, TF-IDF, FastText, and GLOVE are content features mostly used in machine learning.

3 Proposed Methodology

In this section, the proposed approach has been discussed. The workflow of the proposed approach is presented in Fig. 2.

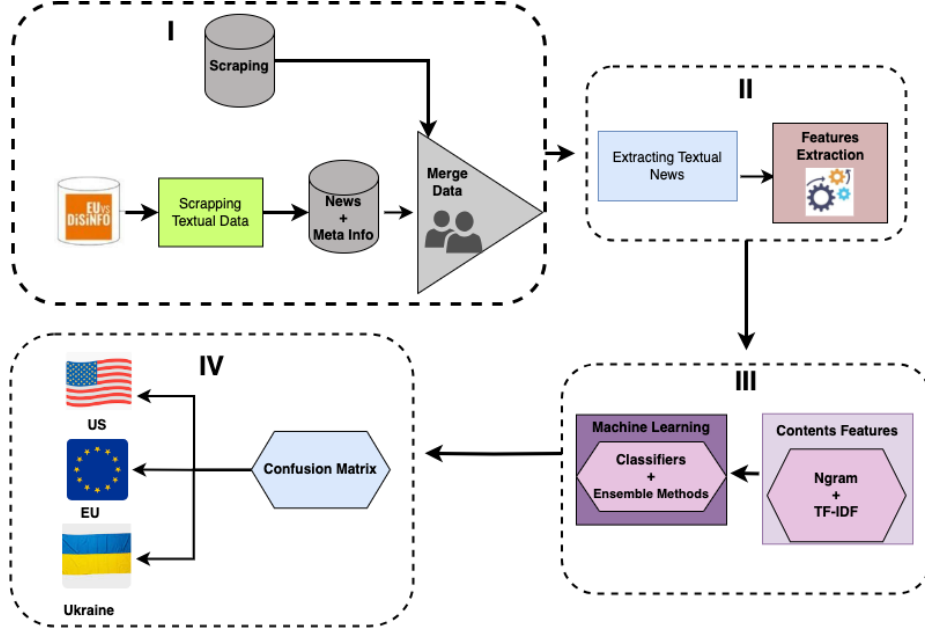


Fig. 2. Workflow of Proposed Methodology

I The source of our dataset, used for experimental work, is Kaggle⁹. This is created in one of the projects in the EU Disinfo lab having 7000 disinformation news articles/cases¹⁰. The lab described the inspiration for the cited dataset as “Which countries are common targets of disinformation? What are frequent topics and recurring narratives? Can we use this data to detect new disinformation campaigns?” but our focus is on the dimension of targeted countries.

The first version of the dataset combined two types of files: a text file containing news, which was scraped on March 8, 2020, and a CSV file containing metadata about the news. The news articles vary among sixteen (16) different languages, but we have only extracted the English-language articles, which were insufficient to train our proposed model. The second version of the data was updated on September 18, 2020, at Kaggle¹¹. This “third version” is the output of merging both datasets and removing duplicate data based on the news title.

RQ1.: How can we build a dataset for automatically classifying targets from disinformation news articles? To address this question, we will first focus on data scraping for the dataset, which will be used in the experimental setup of our proposed model. For the available dataset versions, the third version only contained the URLs of news articles. Therefore, these URLs were used as links

⁹ <https://www.kaggle.com/>

¹⁰ <https://www.kaggle.com/corrieaar/disinformation-articles>

¹¹ <https://www.kaggle.com/datasets/imuhammad/euvdisinfo-disinformation-database>

to identify and fetch the URI of the articles, which were further utilized for retrieving the entire news article's contents through a "Python Package-Beautiful Soup" from their respective sources. These URI were used to scrape the text of news articles, and both URI and scraped texts were stored in the dataset. This updated dataset is named as "Fourth Version of Dataset". There were other discrepancies in the dataset, like the unique number/index of articles not in the same format; for instance, in the index format, one part was consistently named "Media-object" instead of "News-articles". Therefore, all these index/unique article names were converted into the same format, like "News-articles". The provided URLs are used to identify the URI of the missing news articles to deal with the missing data in the dataset. The index of every missing news article is updated according to the URI information.

II In the preprocessing step, all of the abbreviations in this document are changed into non-shortened variants in one of the preprocessing processes to exclude the possibility of missing any crucial keyword. In most of the textual dataset preprocessing processes, remove the stop words from the text during most preprocessing procedures to minimise the dataset's dimension. For instance, words like "UK" and "US" have occurred very frequently in news articles; if the normal pattern of stop word removal has applied, therefore, it has the chance of removing the word "US" because it looks and reads like "us". The Python kit of NLTK (Natural Language Tool Kit) has been used to preprocess the dataset. i) All special characters are removed. ii) All the single characters are removed from text data. iii) The single characters occurring at every sentence's start are removed. iv) Multiple spaces are replaced with single spaces. v) All regular expressions are removed. vi) The entire text of the corpus is converted into lowercase text. vii) The last step was the limitization technique applied to the dataset.

Although few feature extraction techniques exist, we have utilised and applied the TF-IDF, which is most suitable for large volumes of textual data. The TF-IDF technique is applied to our dataset, and tokens and their frequency are generated accordingly. Token frequencies are used to classify the text articles' output target labels. The parameters min-df, which ignores the words that appear less than a specific number; max-df, which ignores the terms that appear most than a specific number; and N-gram, which range the sequence of the term pairwise, are used to calculate various values needed to reduce the dimensionality of the dataset. In our proposed model, the value for min-df is 0.05, meaning the token frequency is less than five percent. The value of max-df is 0.9, meaning a token frequency of more than 90 percent is ignored. Tokenize the data and remove the stopword using the TF-IDF technique. Similarly, N-grams is used with four variants. The results are that tokens of four N-grams are produced, and frequency is assigned to each N-gram-generated token. We have used TF-IDF because it performs better in NLP than other feature extraction techniques like counter vectorizer and DNN approaches [22]. The TF-IDF is suitable for both research area ML and DNN approaches. Jing et al.,[23] used this method in text mining for feature extraction, where it performs exceptionally well. Without los-

ing any text data, it generates effective output. For documents having relevant words, the researcher used this method for different domain documents and, after assigning the TF, DF, and TF-IDF, produced the relevant keywords related to those three documents of the different domains [24]. We have used TF-IDF because it performs better in natural language processing than other feature extraction techniques like counter vectorizer and deep neural network approaches [22]. This algorithm suits both research area machine learning and deep neural networks approaches.

III Similarly, N-grams are used with four variants: unigram, bigram, uni with bigram, and uni with trigram. In the results, tokens of four N-grams are produced, and frequency is assigned to each N-gram-generated token. These N-grams tokens are assigned as input to each classifier and ensemble method to classify the targets.

The researchers used various feature-extraction techniques to detect fake news more accurately. TF-IDF feature extraction techniques usually perform better with content features.

$$\text{tf}(t, d) = \frac{f_{t,d}}{\sum_{t' \in d} f_{t',d}} \quad (1)$$

The first term in Eq. 1 represents the term frequency, meaning the number of times the term appears in the document divided by the Total number of terms in the documents.

$$\text{idf}(t, D) = \log \frac{N}{|\{d \in D : t \in d\}|} \quad (2)$$

Eq. 2 shows the inverse document frequency, and the numerator (N) shows the total number of documents by the number of documents that contain that specific term and, after that, takes the logarithms of the quotient.

$$\text{tfidf}(t, d, D) = \text{tf}(t, d) \cdot \text{idf}(t, D) \quad (3)$$

Eq. 3 shows to multiply TF with IDF, and the result will show the term's relevance. Higher value mean, higher relevance, and lower value mean lower relevance after multiplying the result of both Eq. 1 and Eq. 2. This technique increases the frequency of the words which occur frequently in one document, and the same word occurs in the whole corpus.

The researchers [25] used Passive Aggressive Classifier, Random Forest, and Logistic Regression with TF-IDF techniques to detect fake news correctly. These three techniques perform better with values 99%, 98%, and 97%, respectively. An online learning algorithm called the Passive Aggressive Classifier is appropriate for large-scale classification issues because it can react to streaming data and update its model with little computational overhead. XGBoost is an enhanced iteration of gradient boosting renowned for its rapidity, effectiveness, and exceptional performance in classification and regression tasks, performed with a higher accuracy value of 99% with the same features method.

In classification and regression applications, a random forest is an ensemble of decision trees that work together to increase accuracy and reduce overfitting.

In modeling the probability that a given instance belongs to a specific class, logistic regression is linear and well-suited for binary classification. Linear Support Vector classifier used in proposed work with TF-IDF as it performs better in web-based news [13] with the same feature extraction technique. These classifiers outperformed by comparing twelve models; the highest result was 99.16%.

IV Researchers mostly use confusion matrix in data mining for quantitative analysis to evaluate the model.

Our proposed approach's evaluation measures are precision, recall, and expected results for the confusion matrix attributes True Positive, False Positive, True Negative, and False Negative for nine and three different classes.

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \quad (4)$$

Similarly, we used the standard formula in Eq. 5 for Recall, an attribute of the confusion matrix.

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad (5)$$

As indicated in Eq. 6, the accuracy of all models is defined as the percent of all samples correctly identified by the classifier. Calculating the accuracy using the following formula

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \quad (6)$$

Eq. 7 indicates the harmonic mean of precision and recall, which is a more accurate measure of erroneously classified instances than accuracy. The harmonic mean is used because it penalises extreme values.

$$\text{F1} = 2 * \frac{\text{Precision} * \text{Recall}}{\text{Precision} + \text{Recall}} \quad (7)$$

Our model divided our data into four-word N-gram sequences: i) Unigram, ii) Bigram, iii) Unigram with Bigram, and iv) Unigram with Trigram. We conducted our experiments in two parts, the first of which involved considering the nine classes listed in Table 1 and generating the results. Afterwards, these nine classes were merged into three general classes mentioned in Table 2 and generated the results for three classes.

RQ2. What is the performance of different classifiers and ensemble methods in identifying different target countries in disinformation news articles? Initially, we approached this research question by grouping our findings into four groups. For starters, for our proposed model, we chose unigram. Second, bigram was utilized to generate the results. Then, we used a combination of unigram and bigram to present our findings. Finally, we've integrated the unigram and trigram.

Unigram for 9 Classes and 3 Classes

For the experiment's first phase one with unigram-only features, the LSV generates higher accuracy, precision, recall, and F1-score values with 64%, 66%, 50%, and 53%. LR is the second highest in terms of accuracy, with a value of 62%. RF was poorly performed with the value of 54% after the XGBClassifier. For the second phase of the experiment, LSVC performance is on the top in accuracy with a value of 76%, followed by LR and PA with 75%. RF was the third highest with a value of 75%, five percent more than XGBClassifier.

Bigram for 9 Classes and 3 Classes We have implemented the same methods as we used for unigram. Using LSVC, we achieved the highest accuracy, recall, and F1-score, with values of 59%, 41%, and 44%, respectively. Regarding precision and F1-score, XGB produces the same results as LSVC. For nine classes, the LSVC and LR have given similar results in terms of accuracy, with a value of 72%.

Unigram with Bigram for 9 Classes and 3 Classes

Fig. 3 shows the performance for LSVC outperforms unigram with bigram features for three classes, with an accuracy value of 64%, followed by XGBClassifier, with an accuracy value of 63%. LR is the third highest in accuracy and precision, with values of 62% and 56%, respectively. RF's performance was comparatively lower than all the others, generating an accuracy value of 54% and a precision value of 40%. For recall and F1-score, the XGBClassifier gave the highest value for the proposed model, 62% and 55%. LSVC is the second highest in recall and F1-score values, giving 53% and 49%. LR was the third highest, followed by PA, while RF performance was the same in accuracy and precision with values 36% in recall and 37% in F1-score.

LR and LSVC are the two methods that outperformed the second phase experiment with three classes in accuracy and precision. LR correctly identified the target with an accuracy rate of 77%, followed by LSVC with 76%. Similarly, for precision, LR generates a value of 76%, followed by the second highest method of 74%. PA was the third highest in accuracy and precision performance with values of 71% and 72%. XGBClassifier performed one percent higher than RF with a value of 70% and performed two percent lower in precision, 68%. For recall and F1-score, the LSVC performance was the same with a value of 74%, followed by LR with the same for both 75%. XGBClassifier performed the same for recall, and the F1-score was third highest with a value of 68%, followed by PA with the same value in the F1-score and was one percent lower in the recall. RF performs one percent lower than RF in recall and F1-score.

Unigram with Trigram for 9 Classes and 3 Classes

Fig. 4 The LSVC outperforms in accuracy and precision for unigram-only features, with 64% and 66% values, respectively, for nine classes. LR is the second-highest algorithm, with an accuracy value of 62%. XGBClassifier performed second highest in precision, with a value of 58%, and for accuracy, it accurately classified the targets with a 61% rate. LR and RF perform the same, with 56% each for perception. RF had the lowest accuracy value, 57%, and PA had the lowest value, at 47%. For recall and F1-score, the XGBClassifier outperforms with values 50% and 53%, respectively.

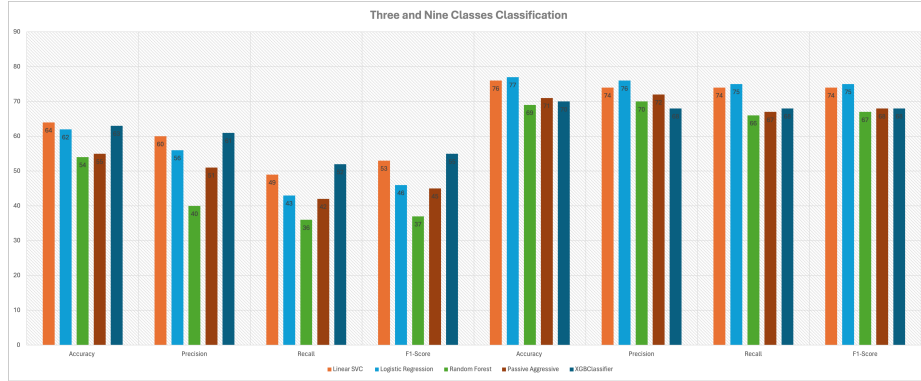


Fig. 3. Results for Unigram with Bigram with Three and Nine Classes

After merging the nine classes into three classes, the LR achieved higher accuracy and precision with values of 77% and 76% , respectively. LSVC is the second highest in accuracy and precision, scoring 76% and 75%. PA gives the lowest accuracy at 66% and 77% in precision. For recall, the LSVC and LR perform the same with the highest value of 75%. For the F1-score, the LR is the top highest value at 76%, followed by LSVS with 75% value. PA performs lower than others in recall and F1-score with a value 65% each.

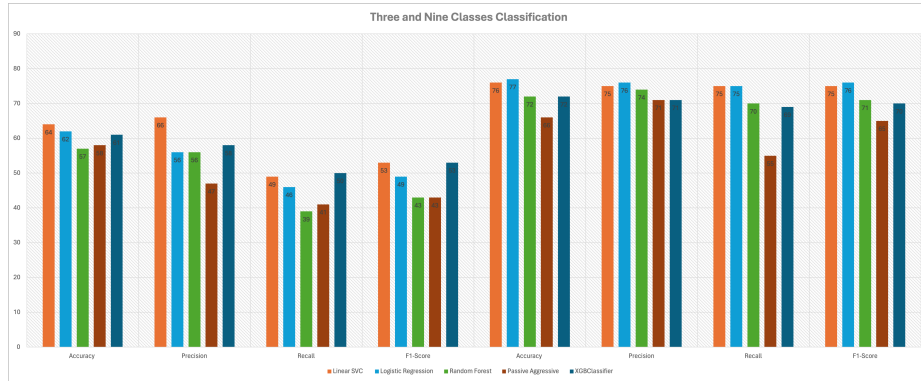


Fig. 4. Results for Unigram with Trigram with Three and Nine Classes

4 Experimental Setup

The number of English language articles collected from the version one dataset was only 311. We used dataset version two to increase the number of disinforma-

tion news articles we scraped. There were 670 URLs in total that we collected before filtration. We completed 553 filtered URI after preprocessing. After scraping, we found 553 more items with this unique URI. Finally, after combining versions one and two of the dataset, we get 864 disinformation news articles, which is sufficient for our model. Experiments were performed in two stages based on the multiple classes shown in Table 1 and Table 2.

Table 1. Classes Details for First Dataset

Classes Names	Count
United States	302
The West	143
European Union	140
Ukraine	123
United Kingdom	50
Netherland	37
Germany	27
Sweden	26
France	16

Table 2. Classes Details after Merging Classes

Classes Name	Merged Classes	Count
United States		302
The West	The West, United Kingdom	193
European Union	European Union, Ukraine, Netherland, Germany, Sweden, France	369

Table 3. Data Reliability and Judges Agreements [26]

Condition	Kappa Values	Data reliability	Agreement
1	0.50 and 0.59	15 to 35%	Weak
2	0.60 and 0.79	35 to 65%	Moderate
3	0.80 and 0.90	64 to 81%	Strong
4	above 0.90	82 to 100%	Perfect

How reliable are our labels, and how accurate is our data? We collaborated with the information technology consultants to find an answer to this question. After reading each article’s text, he double-checked each label, added labels to each article, and changed the record’s status in CSV with 1 and 0 to add one more feature. If the label decision between 2 judges is mismatched, then the

status changes to 1; otherwise, 0 in case of agreement in CSV. The new feature status used for calculating the kappa values, which can be measured by Eq.8 and data reliability present in Table 3 in we have used an authentic tool named Statistical Package for the Social Sciences (SPSS), created by IBM in 2009.

$$K = \frac{\text{Observed agreement} - \text{Chance agreement}}{\text{Maximum agreement} - \text{Chance agreement}} = \frac{\rho_o - P_e}{1 - p_e} \quad (8)$$

Our model uses a stratified split technique to split the dataset into several chunks to train the data for each label. We have used a Python programming language with a machine learning library named Scikit-Learn, Pandas, and Natural Language Toolkit (NLTK) to implement our proposed approach. While scraping the text articles, we used the newspaper API and BeautifulSoup. Grid search API was used to tune and identify the optimum parameters. To separate the data, we divided it into train and test groups, with 20% going to testing and 80 % going to training. Then, we split the training further with 15 percent for validation and 85% for new training, where we performed the Hyperparameter tuning on new training data for classifiers and generated the best parameters for each N-gram individually. The GridSearchCV function has been used in the proposed model in hyperparameter tuning to prevent the model from overfitting. For each label class, the model is trained, and hyperparameters are extracted for each technique.

4.1 Discussion

RQ3. What content-based features are useful in detecting targets in disinformation news articles? This question concerns the performance of the suggested ensemble method in conjunction with N-grams features and various classes, nine and three. We used N-grams features to test the accuracy of each classifier and ensemble approach to address this research question. In both phases of the studies, the N-grams features unigram with bigram and unigram with trigram accurately recognized the target for all classifiers and ensemble methods. LSVC is the sole classifier that performs exceptionally well for nine classes across all features and is second highest for unigram with bigram and unigram with trigram with three classes.

Due to better performance in natural language processing with big text data, we applied the characteristics of the contents TF-IDF. We used the stratified split strategy to train the model with unbalanced classes accurately. We divided the data into train, test, and validation sets. Then, we used training and validation data to fine-tune our model to find the ideal parameter for each classifier and N-grams. Finally, we obtained the findings after passing unseen testing data to each classifier and ensemble technique with their best settings. The results for unigram, bigram only, unigram with bigram, and unigram with trigram were then generated. For all N-grams characteristics, the LSVC and LR classifiers performed better. For 9 Classes with N-grams , LSVC performs superior. With the last two characteristics, unigram with bigram and unigram with trigram, LR

performs better for 3 classes. Unigram with bigram and unigram with trigram are the two features that effectively identified the targets. The proposed work can correctly identify the targets automatically, even with an insufficient amount of text articles. Selecting top content features and hyperparameters can improve the accuracy of the model. We selected and compared the performance of machine learning techniques and top features for the least number of articles.

5 Conclusion

The study aimed to use supervised machine learning to identify targets in large-text articles. We proposed a framework for automatically identifying the hidden targets in fake news. We do not depend on automation alone; therefore, we involve humans in overcoming the chances of errors by detecting the actual target by creating the labels manually. LR and LSVC perform better for all N-game features, comparatively. Unigram with bigram and unigram with trigram are the top content features, with a higher accuracy of 70% each. In future work, more information will be scraped from the EU Disinfo lab, as the website updates the data weekly to increase the benchmark size. Deep learning techniques will be employed to extract the features automatically, which can improve the detection of fake news results.

References

1. Hunt Allcott and Matthew Gentzkow. Social media and fake news in the 2016 election. *Journal of economic perspectives*, 31(2):211–36, 2017.
2. Jeremy Liebowitz. Hate speech, disinformation and political violence in myanmar. *FAKE NEWS*, page 1, 2019.
3. Salome Mikiashvili. Covid-19 and its impact on disinformation and fake news.
4. About us. <https://www.disinfo.eu/about-us/>.
5. Steven Blockmans and Sophia Russack. Deliberative democracy in the eu: Countering populism with participation and debate. 2020.
6. Thomas Marcoux, Esther Mead, and Nitin Agarwal. Studying the dynamics of covid-19 misinformation themes.
7. Alessandro Rovetta and Lucia Castaldo. The influence of mass media on italian web users during covid-19: an infodemiological analysis. 2021.
8. Sotiris B Kotsiantis, I Zaharakis, and P Pintelas. Supervised machine learning: A review of classification techniques. *Emerging artificial intelligence applications in computer engineering*, 160(1):3–24, 2007.
9. David Roxbee Cox and E Joyce Snell. *Analysis of binary data*, volume 32. CRC press, 1989.
10. Anjuman Prabhat and Vikas Khullar. Sentiment classification on big data using naïve bayes and logistic regression. In *2017 International Conference on Computer Communication and Informatics (ICCCI)*, pages 1–5. IEEE, 2017.
11. Tin Kam Ho. Random decision forests. In *Proceedings of 3rd international conference on document analysis and recognition*, volume 1, pages 278–282. IEEE, 1995.

12. S Thamarai Selvi, P Karthikeyan, A Vincent, V Abinaya, G Neeraja, and R Deepika. Text categorization using rocchio algorithm and random forest algorithm. In *2016 Eighth International Conference on Advanced Computing (ICoAC)*, pages 7–12. IEEE, 2017.
13. Md Abdur Rahman, Md Mahbubur Rahman, Md Hafizur Rahman Sumon, Shanta Islam, Md Mahamudul Hasan, Md Shahriar Alam Sakib, Md Moudud Ahmmed, and Hafiz Md Hasan Babu. Web-based machine learning framework for authenticating news: A fake news detection system. 2024.
14. Tianqi Chen, Tong He, Michael Benesty, Vadim Khotilovich, Yuan Tang, Hyunsu Cho, et al. Xgboost: extreme gradient boosting. *R package version 0.4-2*, 1(4), 2015.
15. Mohammed Jabreel and Antonio Moreno. Eitaka at semeval-2018 task 1: An ensemble of n-channels convnet and xgboost regressors for emotion analysis of tweets. *arXiv preprint arXiv:1802.09233*, 2018.
16. Koby Crammer, Ofer Dekel, Joseph Keshet, Shai Shalev-Shwartz, and Yoram Singer. Online passive aggressive algorithms. 2006.
17. Jasmine Shaikh and Rupali Patil. Fake news detection using machine learning. In *2020 IEEE International Symposium on Sustainable Energy, Signal Processing and Cyber Security (iSSSC)*, pages 1–5. IEEE, 2020.
18. Arbana Kadriu, Lejla Abazi, and Hyrije Abazi. Albanian text classification: Bag of words model and word analogies. *Business Systems Research: International journal of the Society for Advancing Innovation and Research in Economy*, 10(1):74–87, 2019.
19. Saloni Gupta and Priyanka Meel. Fake news detection using passive-aggressive classifier. In *Inventive Communication and Computational Technologies*, pages 155–164. Springer, 2021.
20. Mohamed K Elhadad, Kin Fun Li, and Fayez Gebali. Fake news detection on social media: a systematic survey. In *2019 IEEE Pacific Rim Conference on Communications, Computers and Signal Processing (PACRIM)*, pages 1–8. IEEE, 2019.
21. Aishwarya Agrawal, Dhruv Batra, Devi Parikh, and Aniruddha Kembhavi. Don’t just assume; look and answer: Overcoming priors for visual question answering. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 4971–4980, 2018.
22. Ehesas Mia Mahir, Saima Akhter, Mohammad Rezwanul Huq, et al. Detecting fake news using machine learning and deep learning algorithms. In *2019 7th International Conference on Smart Computing & Communications (ICSCC)*, pages 1–5. IEEE, 2019.
23. Li-Ping Jing, Hou-Kuan Huang, and Hong-Bo Shi. Improved feature selection approach tfidf in text mining. In *Proceedings. International Conference on Machine Learning and Cybernetics*, volume 2, pages 944–946. IEEE, 2002.
24. Shahzad Qaiser and Ramsha Ali. Text mining: use of tf-idf to examine the relevance of words to documents. *International Journal of Computer Applications*, 181(1):25–29, 2018.
25. Jumana Jouhar, Anju Pratap, Neharin Tijjo, and Meenakshi Mony. Fake news detection using python and machine learning. *Procedia Computer Science*, 233:763–771, 2024.
26. Ted Byrt et al. How good is that agreement? *Epidemiology*, 7(5):561, 1996.