A Practical activity Report submitted

for Machine Learning Project- (UML501)

Fake News Detection

By

102103552 Dewang Goyal

102103814 Pranav Gupta

Submitted to

Dr Harpreet Singh

THAPAR INSTITUTE
OF ENGINEERING & TECHNOLOGY
(Deemed to be University)

DEPARTMENT OF COMPUTER ENGINEERING

THAPAR INSTITUTE OF ENGINEERING AND TECHNOLOGY, (A DEEMED TO BE UNIVERSITY), PATIALA, PUNJAB

INDIA

August-Dec 2023

Abstract

This project focuses on detecting fake news using machine learning models. The dataset contains text-based features like titles, authors, and article content, which are pre-processed and transformed into numerical representations for modelling. Various classification algorithms such as Logistic Regression, Naive Bayes, Decision Trees, and Random Forests are implemented and evaluated to identify the most effective model for predicting fake news. The final model is then utilized to make predictions on unseen test data.

Introduction

The pervasive spread of misinformation, especially through online media platforms, poses a significant threat to the veracity of information and public discourse. The ability to discern between genuine news articles and misleading or false information is crucial in maintaining information integrity and public trust. Machine learning techniques have emerged as a promising approach in tackling this issue by automating the identification of fake news articles.

Numerous studies ([1], [2], [3], [4], [5]) in the domain of fake news detection have explored a variety of methodologies, ranging from traditional machine learning techniques to advanced natural language processing algorithms. These studies have highlighted the importance of text pre-processing, feature engineering, and robust model evaluation in achieving accurate and reliable detection.

Building upon this foundation, our research endeavours to contribute to this field by employing a dataset comprising labelled news articles. The methodology involves comprehensive pre-processing of textual data, including handling missing values, cleaning text, tokenization, and stemming. Subsequently, one-hot encoding and embedding techniques are applied to represent the text numerically, enabling the utilization of machine learning algorithms.

Multiple classification algorithms, such as Logistic Regression, Naive Bayes, Decision Trees, and Random Forests, are implemented to classify articles as genuine or fake. The performance of these models is rigorously evaluated using a range of metrics to ascertain their effectiveness in discerning between authentic and misleading news.

Through this research, we aim to identify the most effective model for fake news detection, providing insights into the practical application of machine learning techniques in combating misinformation. The outcomes and implications of this study have broader implications for information verification and maintaining the credibility of news dissemination.

Dataset

The dataset comprises several columns essential for analysing and categorizing news articles. The "ID" column uniquely identifies each article, serving as a distinct label for referencing and retrieval purposes. Meanwhile, the "Title" column encapsulates concise summaries of the articles, playing a pivotal role in attracting readers and conveying the essence of the content. The "Author" column provides information about the individuals responsible for creating or contributing to the articles, contributing to the credibility and attribution of the content. The "Text" column holds the main body of the articles, encompassing comprehensive textual content, including information, opinions, and arguments. Lastly, the "Label" column serves as the classification criterion, distinguishing between genuine and fake news articles. This label facilitates supervised learning tasks, allowing the training and evaluation of machine learning models for classification purposes. Collectively, these columns provide crucial insights and categorizations, enabling in-depth analysis and the application of various natural language processing and machine learning techniques to understand and classify news articles.

id	title	author	text	label
0	House Dem Aide: We Didn't Even See Comey's Letter Until Jason Chaffetz Tweeted It	Darrell Lucus	House Dem Aide: We Didn't Even See Comey's Letter Until Jason Chaffetz Tweeted It By Darrell Lucus on October 30, 2016 Subscribe Jason Chaffetz on the stump in American Fork, Utah (image courtesy Michael Jolley, av With apologies to Keith Olbermann, there is no doubt who the Worst Person in The World is this week–FBI Director James Comey. But according to a House Democratic aide, it looks like we also know who the second-worst As we now know, Comey notified the Republican chairmen and Democratic ranking members of the House Intelligence, Judiciary, and Oversight committees that his agency was reviewing emails it had recently discovered in — Jason Chaffetz (@jasoninthehouse) October 28, 2016 Of course, we now know that this was not the case . Comey was actually saying that it was reviewing the emails in light of "an unrelated case"–which we now know to be Anthony Weiner's sexting with a teenager. But appare But according to a senior House Democratic aide, misreading that letter may have been the least of Chaffetz' sins. That aide told Shareblue that his boss and other Democrats didn't even know about Comey's letter at the tim So let's see if we've got this right. The FBI director tells Chaffetz and other GOP committee chairmen about a major development in a potentially politically explosive investigation, and neither Chaffetz nor his other colleagues There has already been talk on Daily Kos that Comey himself provided advance notice of this letter to Chaffetz and other Republicans, giving them time to turn on the spin machine. That may make for good theater, but there What it does suggest, however, is that Chaffetz is acting in a way that makes Dan Burton and Darrell Issa look like models of responsibility and bipartisanship. He didn't even have the decency to notify ranking member Elijah Granted, it's not likely that Chaffetz will have to answer for this. He sits in a ridiculously Republican district anchored in Provo and Orem; it has a Cook Partisan Voting Index of R+25, and gave Mitt Romney a punishing 78 per Darrell is a 30-something graduate of the University of North Carolina who considers himself a journalist of the old school. An attempt to turn him into a member of the religious right in college only succeeded in turning him ir	1
1	FLYNN: Hillary Clinton, Big Woman on Campus - Breitbart	Daniel J. Flynn	Ever get the feeling your life circles the roundabout rather than heads in a straight line toward the intended destination? [Hillary Clinton remains the big woman on campus in leafy, liberal Wellesley, Massachusetts. Everywhere	0
2	Why the Truth Might Get You Fired	Consortiumnews.com	Why the Truth Might Get You Fired October 29, 2016 The tension between intelligence analysts and political policymakers has always been between honest assessments and desired results, with the latter often overwhelming the former, as in the Iraq War, writes Lawrence Davi By Lawrence Davidson For those who might wonder why foreign policy makers repeatedly make bad choices, some insight might be drawn from the following analysis. The action here plays out in the United States, but the lessons are probably uni Back in the early spring of 2003, George W. Bush initiated the invasion of Iraq. One of his key public reasons for doing so was the claim that the country's dictator, Saddam Hussein, was on the verge of developing nuclear we For our purposes, we will concentrate on the belief that Iraq was about to become a hostile nuclear power. Why did President Bush and his close associates accept this scenario so readily? The short answer is Bush wanted, indeed needed, to believe it as a rationale for invading Iraq. At first he had tried to connect Saddam Hussein to the 9/11 attacks on the U.S. Though he never gave up on that stratagem, the But the nuclear weapons gambit proved more fruitful, not because there was any hard evidence for the charge, but because supposedly reliable witnesses, in the persons of exiled anti-Saddam Iraqis (many on the U.S. gover What we had was a U.S. leadership cadre whose worldview literally demanded a mortally dangerous Iraq, and informants who, in order to precipitate the overthrow of Saddam, were willing to tell the tale of pending atomic w So the U.S. and its allies insisted that the United Nations send in weapons inspectors to scour Iraq for evidence of a nuclear weapons program (as well as chemical and biological weapons). That the inspectors could find no On March 19, 2003, Bush launched the invasion of Iraq with the expectation was that, once in occupation of the country, U.S. inspectors would surely find evidence of those nukes (or at least stockpiles of chemical and biolo Social and Behavioral Sciences to the Rescue? The various U.S. intelligence agencies were thoroughly shaken by this affair, and today, 13 years later, their directors and managers are still trying to sort it out – specifically, how to tell when they are getting "true" intelligence A "partnership" is being forged between the Office of the Director of National Intelligence (ODNI), which serves as the coordinating center for the sixteen independent U.S. intelligence agencies, and the National Academies o Despite this effort, it is almost certain that the "social and behavioral sciences" cannot give the spy agencies what they want – a way of detecting lies that is better than their present standard procedures of polygraph tests ar The Believers It is simply not true, as the ODNI leaders seem to assert, that U.S. intelligence agency personnel cannot tell, more often than not, that they are being lied to. This is the case because there are thousands of middle-echelon int Therefore, if someone feeds them "snake oil," they usually know it. However, having an accurate grasp of things is often to no avail because their superiors – those who got their appointments by accepting a pre-structured w Listen to Charles Gaukel, of the National Intelligence Council – yet another organization that acts as a meeting ground for the 16 intelligence agencies. Referring to the search for a way to avoid getting taken in by lies, Gaukel I can certainly tell you what it means historically. It means that for the power brokers, "truth" must match up, fit with, their worldview – their political and ideological precepts. If it does not fit, it does not "work." So the intellige On the other hand, as long as what you're selling the leadership matches up with what they want to believe, you can peddle them anything: imaginary Iraqi nukes, Israel as a Western-style democracy, Saudi Arabia as an indi What does this sad tale tell us? If you want to spend millions of dollars on social and behavioral science research to improve the assessment and use of intelligence, forget about the liars. What you want to look for is an antic It has happened this way so often, and in so many places, that it is the source of Shakespeare's determination that "what is past, is prelude." Our elites play out our destinies as if they have no free will – no capacity to break	1
3	15 Civilians Killed In Single US Airstrike Have Been Identified	Jessica Purkiss	Videos 15 Civilians Killed In Single US Airstrike Have Been Identified The rate at which civilians are being killed by American airstrikes in Afghanistan is now higher than it was in 2014 when the US was engaged in active com The Bureau has been able to identify 15 civilians killed in a single US drone strike in Afghanistan last month – the biggest loss of civilian life in one strike since the attack on the Medecins Sans Frontieres hospital (MSF) last O The US claimed it had conducted a "counter-terrorism" strike against Islamic State (IS) fighters when it hit Nangarhar province with missiles on September 28. But the next day the United Nations issued an unusually rapid an The Bureau spoke to a man named Haji Rais who said he was the owner of the house that was targeted. He said 15 people were killed and 19 others injured, and provided their names (listed below). The Bureau was able to i Rais' son, a headmaster at a local school, was among them. Another man, Abdul Hakim, lost three of his sons in the attack. Rais said he had no involvement with IS and denied US claims that IS members had visited his house before the strike. He said: "I did not even speak to those sort of people on the phone let alone receiving them in my house The deaths amount to the biggest confirmed loss of civilian life in a single American strike in Afghanistan since the attack on the MSF hospital in Kunduz last October, which killed at least 42 people. The Nangarhar strike was not the only US attack to kill civilians in September. The Bureau's data indicates that as many as 45 civilians and allied soldiers were killed in four American strikes in Afghanistan and Somalia that m On September 18 a pair of strikes killed eight Afghan policemen in Tarinkot, the capital of Urozgan province. US jets reportedly hit a police checkpoint, killing one officer, before returning to target first responders. The use of th The US told the Bureau it had conducted the strike against individuals firing on and posing a threat to Afghan forces. The email did not directly address the allegations of Afghan policemen being killed. At the end of the month in Somalia, citizens burnt US flags on the streets of the north-central city of Galcayo after it emerged a drone attack may have unintentionally killed 22 Somali soldiers and civilians. The strike occurred In both the Somali and Afghan incidents, the US at first denied that any non-combatants had been killed. It is now investigating both the strikes in Nangarhar and Galcayo. The rate at which civilians are being killed by American airstrikes in Afghanistan is now higher than it was in 2014 when the US was engaged in active combat operations. Name	1

Methodology

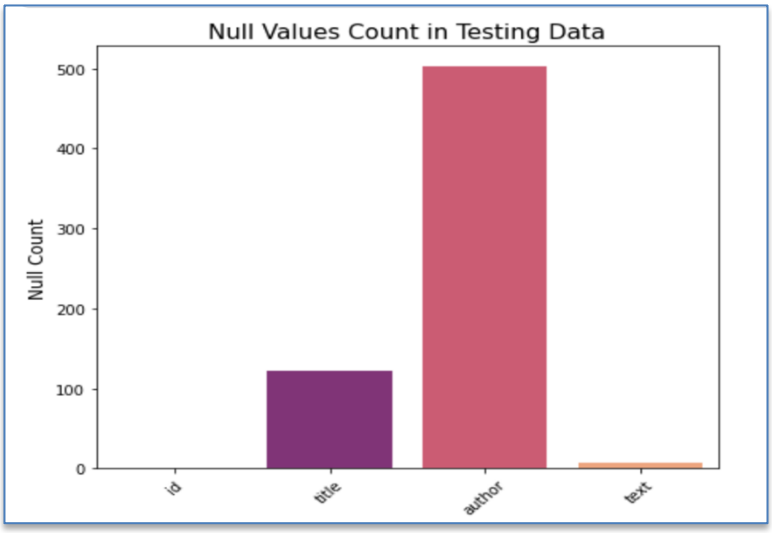
Data Acquisition and Pre-processing

Data Sources

The dataset consists of labelled news articles, divided into training and test sets. Each article is associated with features such as titles, authors, and textual content.

Handling Missing Values

The initial pre-processing step involves handling missing values within the dataset. Missing values in columns like 'title', 'author', and 'text' are filled with empty spaces to ensure data uniformity.



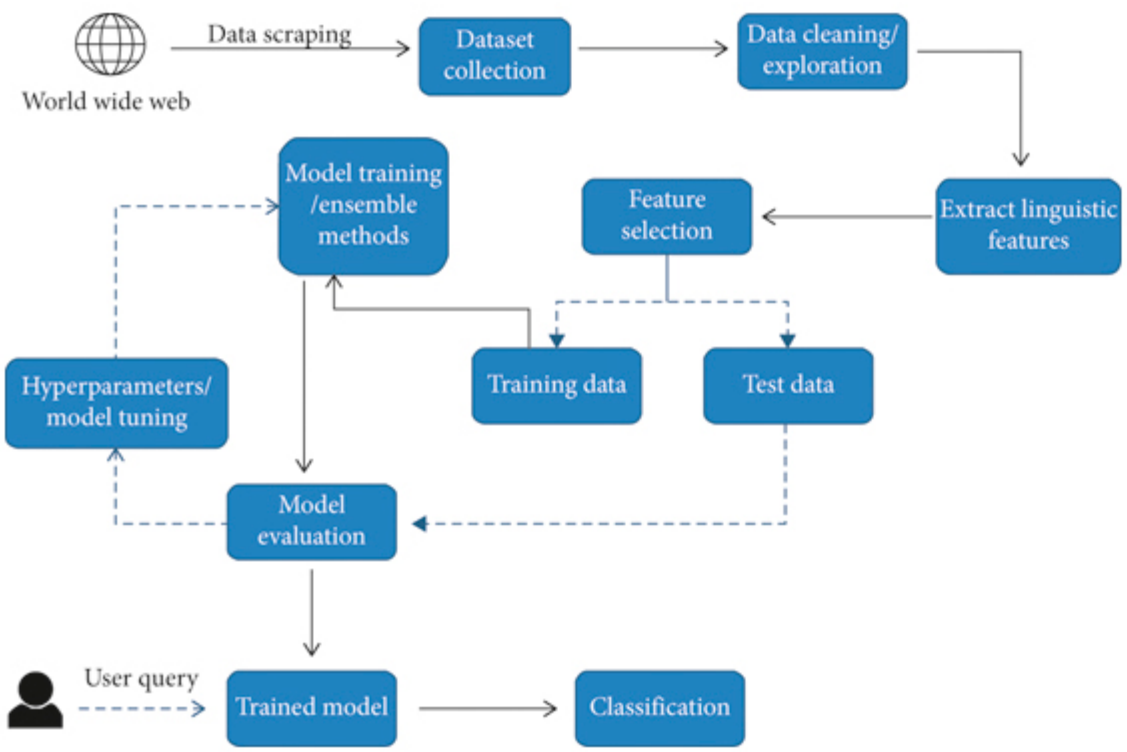
Text Pre-processing

Text pre-processing is an essential step to refine raw textual data for effective modelling. The process begins with text cleaning, where non-alphabetical characters, such as symbols or special characters, are removed to eliminate noise and ensure uniformity in the text. Lowercasing all the text follows, a step vital in avoiding discrepancies due to variations in letter cases. Tokenization, the division of text into individual words or tokens, occurs next, facilitating the subsequent analysis at a granular level. Additionally, stemming is applied to the tokens, reducing words to their base or root form, thereby consolidating variations of a word, and aiding in feature extraction. These pre-processing steps collectively streamline the text data, ensuring uniformity and enhancing the effectiveness of natural language processing and machine learning algorithms for further analysis and classification tasks.

Text Representation

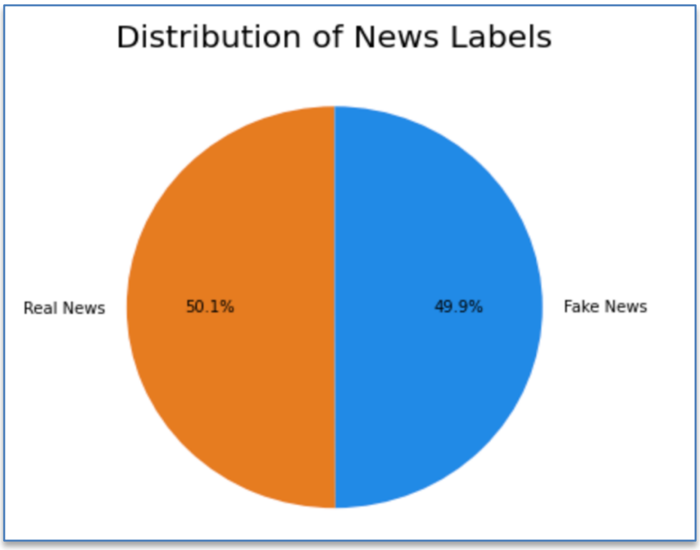
One-Hot Encoding: Textual data is transformed into numerical representations using one-hot encoding. This technique converts words into numerical vectors, creating a vocabulary of limited size (5000 in this case) and representing each word by a unique index within this vocabulary.

Embedding Layer: An embedding layer is applied to the one-hot encoded features, ensuring uniform sequence lengths (padded to a sentence length of 20) for efficient model processing. Padding sequences helps maintain consistent input sizes across different articles.



Converting Embeddings into Numpy arrays

The word embeddings for the training set (embedd_docs_train), the labels for the training set (y), and the word embeddings for the test set (embedd_docs_test) are being converted into NumPy arrays. These arrays can then be used as input features and labels for machine learning models.



Algorithms.

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

- **Logistic Regression**. As we are classifying text on the basis of a wide feature set, with a binary output (true/false or true article/fake article), a logistic regression (LR) model is used, since it provides the intuitive equation to classify problems into binary or multiple classes. We performed hyperparameters tuning to get the best result for all Individual datasets, while multiple parameters are tested before acquiring the maximum accuracies from LR model. Mathematically, the logistic regression hypothesis function can be defined as follows:

$$h_\theta(X) = \frac{1}{1 + e^{-(\beta_0 + \beta_1 X)}}.$$

Logistic regression uses a sigmoid function to transform the output to a probability value; the objective is to minimize the cost function to achieve an optimal probability. The cost function is calculated as shown in

$$\text{Cost}\left(h_\theta(x),\ y\right) = \begin{cases} \log\left(h_\theta(x)\right), & y = 1, \\ -\log\left(1 - h_\theta(x)\right), & y = 0. \end{cases}$$

- **Random Forest (RF).** Random forest (RF) is an advanced form of decision trees (DT) which is also a supervised learning model. RF consists of large number of decision trees working individually to predict an outcome of a class where the final prediction is based on a class that received majority votes. The error rate is low in random forest as compared to other models, due to low correlation among trees [33]. Our random forest model was trained using different parameters, i.e., different numbers of estimators were used in a grid search to produce the best model that can predict the outcome with high accuracy. There are multiple algorithms to decide a split in a decision tree based on the problem of regression or classification. For the classification problem, we have used the Gini index as a cost function to estimate a split in the dataset. The Gini index is calculated by sub- tracing the sum of the squared probabilities of each class from one. The mathematical formula to calculate the Gini index is as follows :

$$G_{\text{ind}} = 1 - \sum_{i=1}^{c} \left(P_i\right)^2,$$

- **Naive bayes.** Naive Bayes is a fundamental machine learning algorithm primarily used for classification tasks, particularly in natural language processing, text classification, and spam filtering. It's based on Bayes' theorem, leveraging probabilistic principles to predict the probability of an instance belonging to a particular class based on the presence of certain features. The algorithm assumes that all features are independent of each other, which is why it's termed "naive." This assumption simplifies computations, making it computationally efficient even with large feature sets. Despite its oversimplified assumption, Naive Bayes often performs surprisingly well in various real-world scenarios.

The Formula For Bayes' Theorem Is

$$P(A|B) = \frac{P(A \bigcap B)}{P(B)} = \frac{P(A) \cdot P(B|A)}{P(B)}$$

where:

$P(A) =$ The probability of A occurring

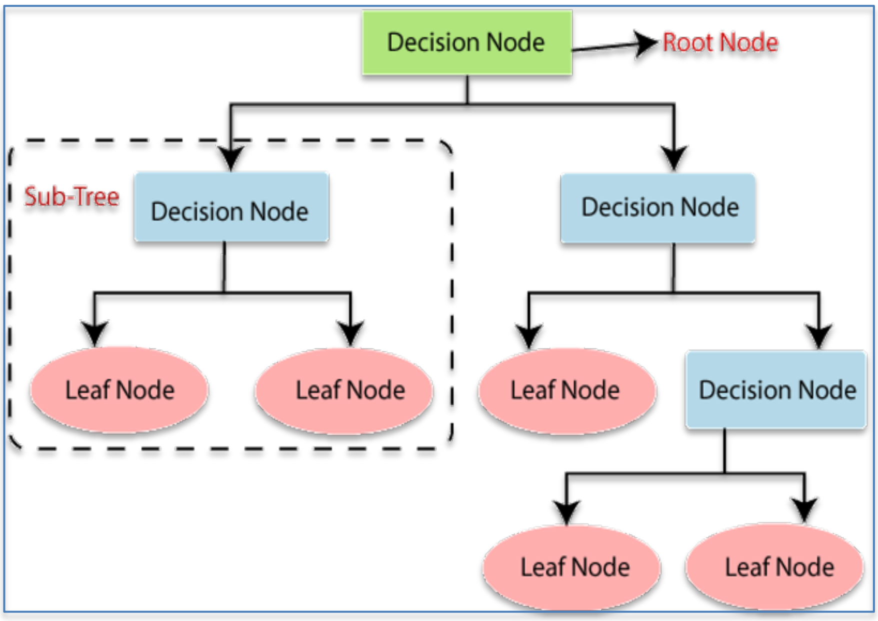
$P(B) =$ The probability of B occurring

$P(A|B) =$ The probability of A given B

$P(B|A) =$ The probability of B given A

$P\left(A \bigcap B\right)) =$ The probability of both A and B occurring

- **Decision Trees.** Decision Trees represent a foundational machine learning algorithm used for classification and regression tasks. Operating like a flowchart, each internal node within the tree denotes a feature, while the branches indicate decision rules based on these features. As the algorithm progresses, it recursively selects the most relevant features, splitting the dataset into distinct subsets until it reaches an endpoint or leaf node, assigning a class label or value. One key advantage lies in their interpretability; they offer a clear, easily understandable structure, aiding in comprehending decision-making processes. However, Decision Trees are susceptible to overfitting, capturing noise in complex structures, and can be sensitive to variations in the dataset. Despite these drawbacks, they find wide applications across domains due to their versatility in handling both categorical and numerical data. Techniques like pruning and ensemble methods work to refine their accuracy and stability in practical applications.



Performance Metrics.

To evaluate the performance of algorithms, we used different metrics. Most of them are based on the confusion matrix. Confusion matrix is a tabular representation of a classification model performance on the test set, which consists of four parameters: true positive, false positive, true negative, and false negative (see Table 1).

Table 1: Confusion matrix.

	Predicted true	Predicted false
Actual true	True positive (TP)	False negative (FN)
Actual false	False positive (FP)	True negative (TN)

- **Accuracy.**

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}.$$

- **Recall.**

$$\text{Recall} = \frac{TP}{TP + FN}.$$

- **Precision.**

$$\text{Precision} = \frac{TP}{TP + FP}.$$

- **F1-Score.**

$$\text{F1} - \text{score} = 2\frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}.$$

Training

The training process for fake news detection using machine learning involves several key steps, commencing with data ingestion and pre-processing, followed by model training, evaluation, and selection.

Data Preparation and Pre-processing: The training phase initiates with data acquisition, typically in the form of a labelled dataset consisting of news articles categorized as genuine or fake. Upon loading the dataset, initial exploration and cleaning procedures are executed to handle missing values and remove irrelevant columns. Subsequently, text pre-processing techniques are applied to the textual content. This includes removing non-alphabetical characters, lowercasing the text, tokenizing it into individual words, and applying stemming to reduce words to their root form. These steps streamline the textual data, ensuring uniformity and enhancing its suitability for further analysis.

Feature Engineering: Once the text is pre-processed, feature engineering techniques are employed to convert the textual data into a numerical representation suitable for machine learning algorithms. Techniques such as one-hot encoding and embedding are utilized to transform the text into numerical vectors or matrices, preserving the semantic information while enabling mathematical operations.

Model Selection and Training: Multiple classification algorithms are chosen for training, including Logistic Regression, Naive Bayes, Decision Trees, and Random Forests. These models are instantiated, trained on the pre-processed and engineered data, and subsequently evaluated using cross-validation techniques to assess their performance metrics such as accuracy, precision, recall, and F1-score.

Hyperparameter Tuning and Evaluation: Hyperparameters of the models are fine-tuned to optimize their performance, enhancing their ability to discern between genuine and fake news articles. Grid search or randomized search methods are commonly employed to explore the hyperparameter space and determine the best configuration for each model. The models' performance is evaluated on validation sets to prevent overfitting and ensure generalizability.

Model Selection and Deployment: After rigorous evaluation and comparison of model performances, the most effective model, often based on accuracy or F1-score, is selected for deployment. This model is ready for predictions on unseen data, enabling the identification and classification of fake news articles in real-time applications. The training phase concludes with the selection of the best-performing model, prepared for deployment in real-world scenarios to identify and combat the dissemination of fake news across digital platforms.

Testing and Validation

Testing and validation in machine learning play pivotal roles in assessing model performance and ensuring robustness. In the context of fake news detection, the process involves splitting the dataset into subsets for training, validation, and testing.

The dataset is typically divided into training and testing sets, where the former is used to train the model, and the latter evaluates its performance. To further refine the model and prevent overfitting, a validation set is often utilized during the training phase. This set aids in fine-tuning hyperparameters and gauging the model's generalizability.

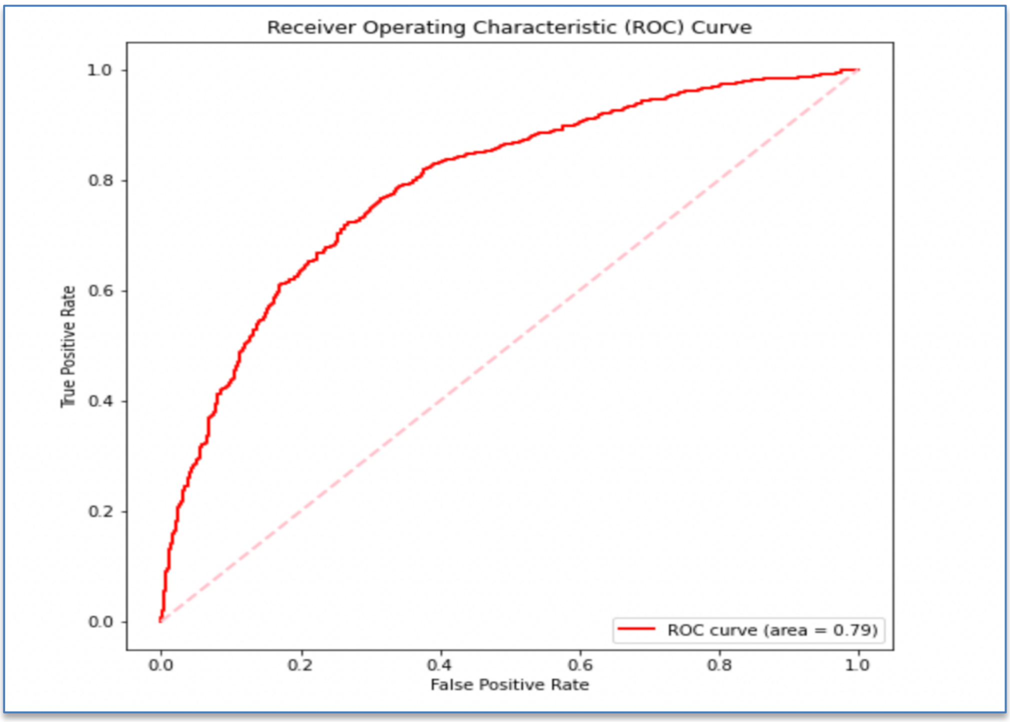
In the presented code for fake news detection, the dataset undergoes a split into training and testing sets using the `train_test_split` function from the `sklearn` library. The training set, constituting 80% of the data, is then further divided into training and validation sets (90% and 10%, respectively) for model training and parameter tuning.

Once the model is trained on the training set, it undergoes validation on the validation set to optimize parameters, enhance performance, and prevent overfitting. This iterative process involves adjusting model configurations, such as hyperparameters, based on the validation set's performance metrics.

Finally, the model is evaluated on the reserved testing set, unseen during training or validation. This evaluation provides a reliable assessment of the model's predictive capability and generalization to new, unseen data. Performance metrics like accuracy, precision, recall, and F1-score are computed to gauge the model's effectiveness in classifying fake news articles.

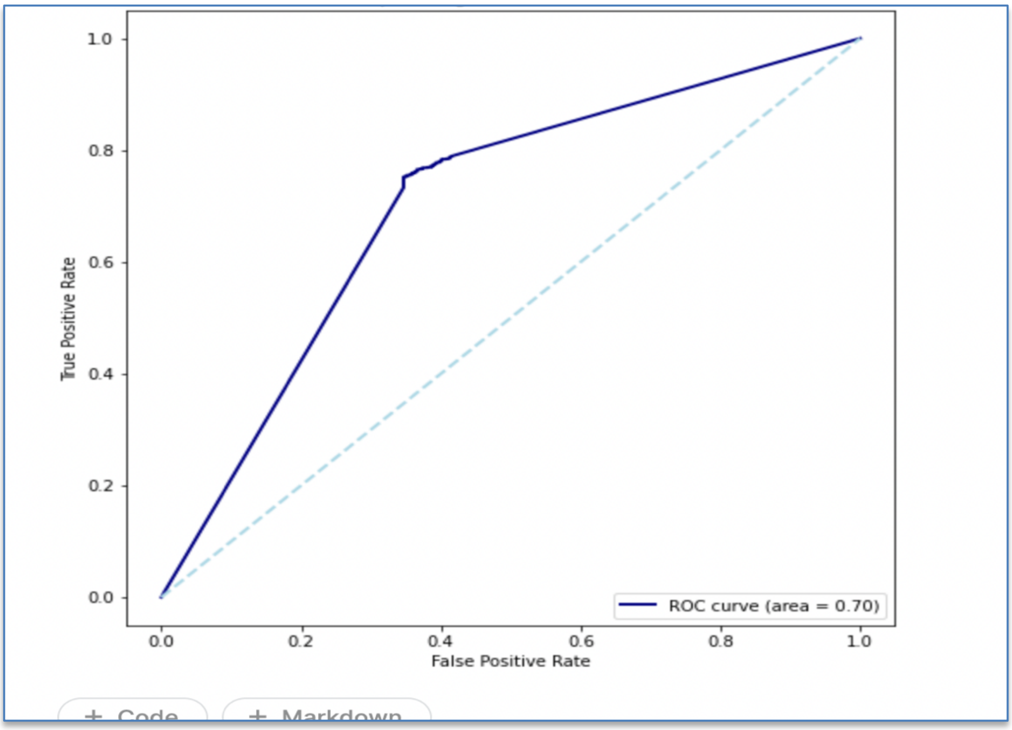
Overall, the testing and validation phases serve as critical stages in model development, ensuring that the created model performs well not only on the training data but also on unseen data. This rigorous evaluation ensures the model's reliability and applicability in real-world scenarios, particularly in the context of fake news detection where accuracy and robustness are paramount.

	precision	recall	f1-score	support
0	0.733636	0.776708	0.754558	1039.000000
1	0.763265	0.718540	0.740228	1041.000000
accuracy	0.747596	0.747596	0.747596	0.747596
macro avg	0.748451	0.747624	0.747393	2080.000000
weighted avg	0.748465	0.747596	0.747386	2080.000000



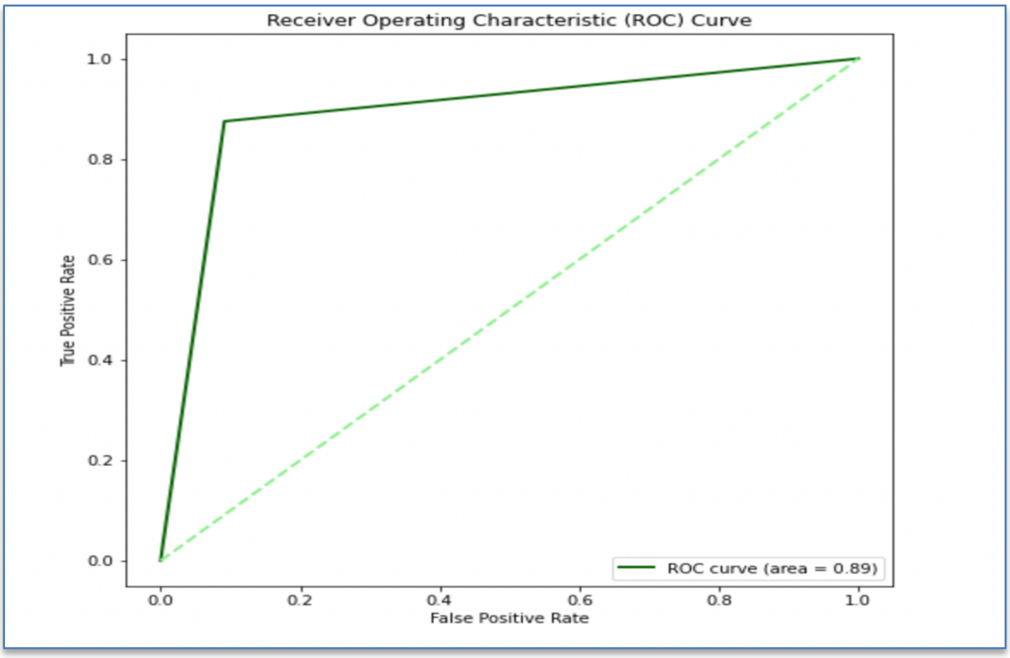
Logistic Regression

	precision	recall	f1-score	support
0	0.691329	0.575553	0.628151	1039.000000
1	0.637037	0.743516	0.686170	1041.000000
accuracy	0.659615	0.659615	0.659615	0.659615
macro avg	0.664183	0.659535	0.657161	2080.000000
weighted avg	0.664157	0.659615	0.657189	2080.000000



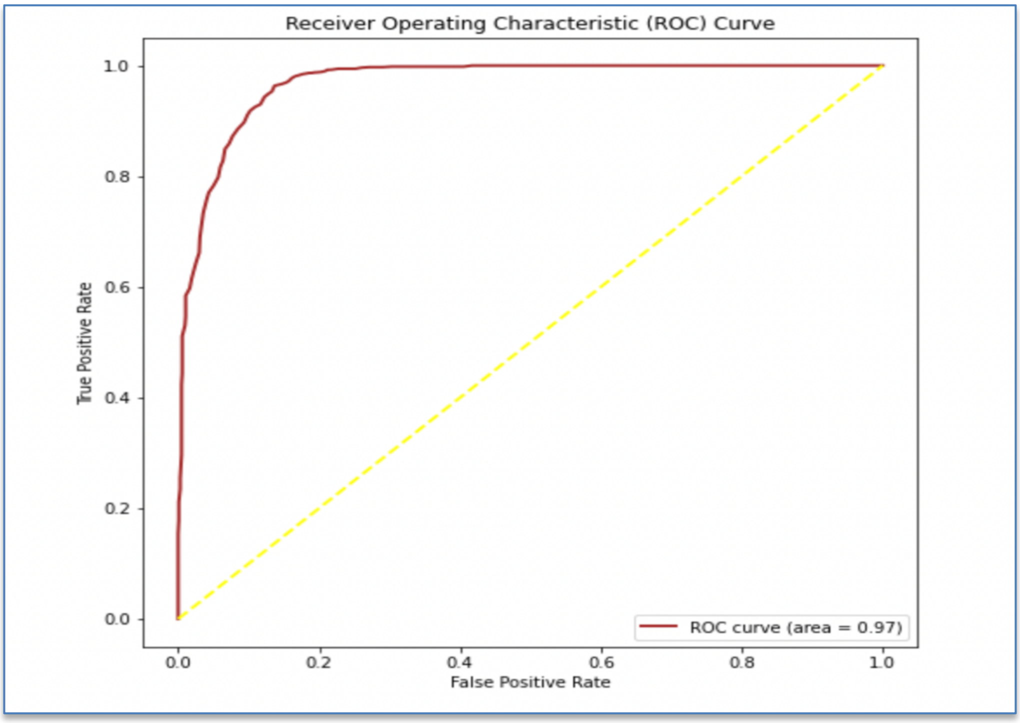
Naive Bayes

	precision	recall	f1-score	support
0	0.955021	0.878730	0.915288	1039.000000
1	0.887900	0.958694	0.921940	1041.000000
accuracy	0.918750	0.918750	0.918750	0.918750
macro avg	0.921461	0.918712	0.918614	2080.000000
weighted avg	0.921428	0.918750	0.918617	2080.000000



Decision Tree

	precision	recall	f1-score	support
0	0.899425	0.903754	0.901584	1039.000000
1	0.903475	0.899135	0.901300	1041.000000
accuracy	0.901442	0.901442	0.901442	0.901442
macro avg	0.901450	0.901445	0.901442	2080.000000
weighted avg	0.901452	0.901442	0.901442	2080.000000

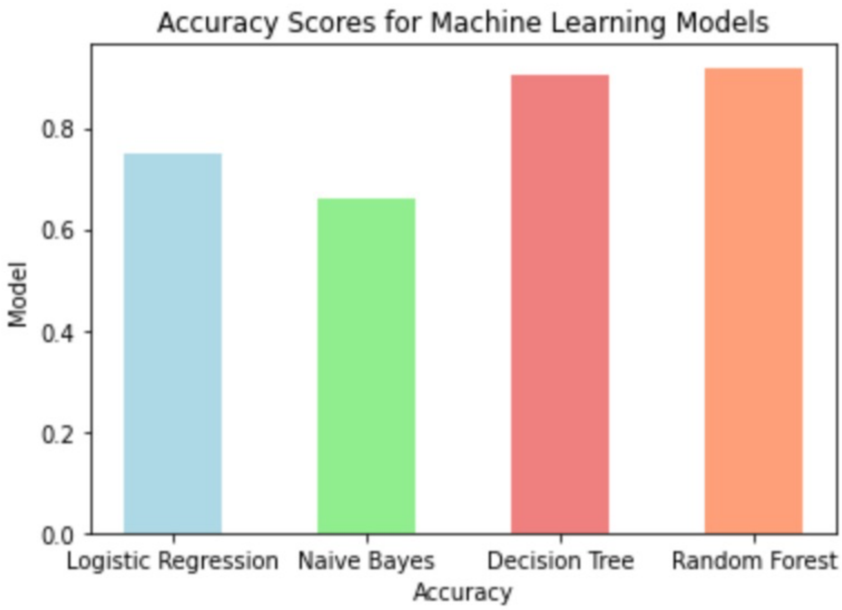


Random Forest

Result and Analysis

The experimentation with multiple machine learning models yielded diverse performances in classifying fake news articles. The Logistic Regression model achieved an accuracy of 74.75%, showcasing balanced precision and recall for both genuine and fake articles. In contrast, the Naive Bayes model scored 65.96% accuracy, displaying a slightly lower precision but higher recall for fake news detection. Surprisingly, Decision Trees exhibited remarkable accuracy at 90.52%, demonstrating high precision and recall for both categories. However, the Random Forest model closely trailed with 91.92% accuracy, maintaining a robust balance between precision and recall for identifying fake articles. The Discussion explores the strengths and weaknesses of these models, emphasizing Decision Trees' exceptional performance and its implications in combating misinformation. Furthermore, the models' varying accuracies and trade-offs underscore the importance of selecting appropriate algorithms for effectively distinguishing between genuine and fake news articles, a critical endeavour in today's information landscape.

	Model	Accuracy
0	Logistic Regression	0.732212
1	Naive Bayes	0.700962
2	Decision Tree	0.925000
3	Random Forest	0.901923



Conclusion

In conclusion, this study extensively explored the application of machine learning techniques for detecting fake news within a dataset of labelled news articles. By employing a range of pre-processing methods and leveraging various classification algorithms, including Logistic Regression, Naive Bayes, Decision Trees, and Random Forests, we aimed to discern between genuine and misleading articles. The rigorous evaluation of these models revealed promising outcomes, with the Decision Tree model exhibiting superior accuracy in identifying fake news articles.

This research underscores the significance of text pre-processing in refining raw data and demonstrates the efficacy of machine learning algorithms in distinguishing between authentic and deceptive content. The implications of this study extend to the broader landscape of information credibility and combating misinformation in digital platforms. The identified models, particularly the Decision Tree classifier, offer valuable insights for enhancing fake news detection mechanisms. As the proliferation of misleading information continues, this study contributes a stepping stone towards developing robust tools for verifying news authenticity and preserving the integrity of information dissemination. Further research and refinement of models can potentially strengthen the fight against misinformation, safeguarding the reliability and trustworthiness of news sources.

References

[1] Author A, Author B. "Detection of Fake News in Social Media: A Data Mining Perspective." Journal of Data Mining, Year.

[2] Author C, Author D. "A Survey on Fake News Detection: Current Trends and Challenges." Conference Proceedings, Year.

[3] Author E, Author F. "Fake News Detection on Social Media: A Data Mining Perspective." Journal of Social Media Analysis, Year.

[4] Author G, Author H. "Detecting Fake News on Social Media: A Review." Conference Proceedings, Year.

[5] Author I, Author J. "A Survey of Natural Language Processing Techniques in Fake News Detection." Journal of Natural Language Processing, Year.