

**DEVELOPMENT OF A MACHINE LEARNING-BASED MODEL USING A
DECISION TREE FOR DETECTING FAKE NEWS: ANALYZING
TECHNIQUES FOR ACCURATE CONTENT VERIFICATION**

TOMBA KINKOSI ESTHER

**A THESIS SUBMITTED TO THE SCHOOL OF COMPUTING AND
INFORMATICS IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
OF THE DEGREE OF MASTER OF SCIENCE IN COMPUTER
INFORMATION SYSTEMS OF KENYA METHODIST UNIVERSITY**

SEPTEMBER 2025

DECLARATION AND RECOMMENDATION

DECLARATION

The research thesis has not yet been presented at any other university. It is my original work.

Signature: _____ Date: 24/09/2025 _____

Tomba Kinkosi Esther

RECOMMENDATION

The research thesis was submitted for examination with our approval as university supervisors.

Signature : _____ Date 22/09/2025 _____

Dr. Lawrence Muriira, PhD

Kenya Methodist University (KeMU)

Signature: _____ Date 22/9/2025 _____

Mr. Timothy Anondo

Kenya Methodist University (KeMU)

COPY RIGHT

©**2025** Tomba Kinkosi Esther

All rights protected. No portion of this thesis may be copied, stored in a retrieval system, or shared in any format, whether electronically, mechanically, photocopied, or otherwise, without prior written consent from the author or Kenya Methodist University.

DEDICATION

With gratitude, this present work is dedicated to the Almighty God, my parents, relatives, and everyone who has supported me in various ways. May God's blessings be upon them.

ACKNOWLEDGEMENTS

I thank God whose grace and strength have been sufficient in my academic journey. I sincerely appreciate my two supervisors, Dr Lawrence Muriira and Mr. Timothy Anondo, for their invaluable guidance and encouragement throughout my thesis writing. May they be richly blessed. I express my gratitude to all the faculty members and the Department of Computer Science for their training through various seminars and conferences. Furthermore, I express my profound gratitude to my Congregation, Sisters of Notre Dame de Namur, for their spiritual and financial support. May God's grace be upon them.

ABSTRACT

With the growing spread of information through social media and online news, identifying fake news has become increasingly important. To explore this issue, the Pew Research Center conducted surveys in the U.S. to examine how adults access news on social platforms, focusing on the behaviors and demographics of users who rely on these channels. This research sought to address a major gap in traditional fake news detection approaches, which were largely manual and lacked the sophistication of advanced machine learning and AI methods. Such conventional techniques struggle to handle the complexity and contextual manipulation of information, where accurate facts can be framed misleadingly. To overcome these shortcomings, the study developed a machine learning-based model to detect fake news by analyzing article content and recognizing patterns of misinformation. It utilized advanced natural language processing (NLP) techniques and supervised learning algorithms. Decision Trees (achieving 99.67% accuracy), Logistic Regression (99.13%), and Random Forest (99.15%). Processes such as tokenization and TF-IDF were applied to train the model on the ISO Fake News dataset, which combined real news from Reuters.com with fake news from unreliable sources flagged by PolitiFact and Wikipedia. Model performance was evaluated using accuracy, precision, recall, and F1-score, all reaching 99.67%, demonstrating exceptional detection capability. This work contributes to the field of machine learning by enhancing NLP methods and improving the effectiveness of fake news detection models. Future research is encouraged to expand datasets, incorporate multiple languages, employ deep learning like RNNs, CNNs, and Transformers (e.g., BERT and RoBERTa) for richer contextual understanding, and establish benchmarks based on real-world case studies.

TABLE OF CONTENTS

DECLARATION AND RECOMMENDATION	ii
COPY RIGHT	iii
DEDICATION	iv
ACKNOWLEDGEMENTS	v
ABSTRACT.....	vi
Table of Contents	vii
LIST OF TABLES.....	ix
List of figures.....	x
LIST OF ABBREVIATIONS AND ACRONYMS	xi
CHAPTER One.....	12
INTRODUCTION	12
1.1 Background.....	12
1.2 Problem Statement.....	13
1.3 The Objectives	14
CHAPTER Two.....	20
LITERATURE REVIEW.....	20
2.1 Introduction.....	20
2.2 Theoretical Framework of a Machine Learning-Based Model for Identifying Fake News.....	20
2.3 Machine-learning techniques	26
2.4 . Fake News Detection	29
2.5 Utilization of datasets in machine learning.....	32
CHAPTER Three.....	48
RESEARCH METHODOLOGY	48
CHAPTER Four	59

DISCUSSIONS AND RESULTS	59
4.1 Development of Fake News Detection Model.....	59
4.2 Evaluation Metrics	65
4.3 Model comparison	70
4.4 Discussion.....	73
4.5 Model’s performance	74
4.6 Implications of the Findings	75
4.7 Summary.....	76
CHAPTER Five.....	77
CONCLUSION AND RECOMMENDATIONS	77
5.1 . Introduction.....	77
5.2 Conclusion	77
5.3 Recommendations.....	78
5.4 Future work.....	78

LIST OF TABLES

Table 2.1 Clustering learning.....	28
Table 2.2 Datasets commonly used in machine learning based models “.....	33
Table 2.3 Empirical Table Summary of Literature Review	45
Table 4.1 The ISOT Fake News Dataset: Initial data	59
Table 4.2 ISOT Fake News Dataset: Results of the Data	60
Table 4.3 The ISOT Fake News Dataset: Number of articles per subject	61
Table 4.4 The ISOT Dataset: The count and percentage for each target	62
Table 4.5 A training Set (80% of the data).....	65
Table 4.6 Testing Set (20% of the data)	65
Table 4.7 Evaluation metrics.....	69

LIST OF FIGURES

Figure 2.1 Theoretical Framework	21
Figure 2.2 Support Vector Machine	23
Figure 2.3 Machine-learning techniques	27
Figure 2.4 K-means clustering.....	29
Figure 3.1 Fake news Dataset	49
Figure 3.2 Architecture for model development.....	52
Figure 3.3 Data Preprocessing	53
Figure 3.4 Splitting the data.....	55
Figure. 4.1 The ISOT Dataset.: Bar chart showing count of articles per subject	61
Figure 4.2 The ISOT Fake News Dataset.: Bar chart showing the number of each target	62
Figure 4.3 Data Preparation: World cloud for fake news	63
Figure 4.4 Data Preparation: Most frequent words count in fake news	63
Figure 4.5 Data Preparation: World cloud for real news	64
Figure 4.6 Data Preparation: Most frequent words count in real news	64
Figure 4.7 Models' Comparison Table.....	70
Figure. 4.8 A confusion matrix for the logistic regression model's classification.....	71
Figure 4.9 A confusion matrix for a decision-tree classifier.....	72
Figure 4.10 A confusion matrix for a random forest classifier	73
Figure 4.11 Model's performance	75

LIST OF ABBREVIATIONS AND ACRONYMS

AI	Artificial Intelligence
AD	Anomaly Detection
CLIP	Contrastive Language Image Pretraining
DT	Decision Tree
FN	Fake News
FN	False Negative
FCN	Fact-Checking Network
GPT	Generative Pre-Trained Transformer
KNN:	K-Nearest Neighbor
LG	Logistic Regression
ML	Machine Learning
NLP	Natural Language Processing
NB	Naïve- Bayes
PTT	Pre-Trained Transformer
RNNs	Recurrent Neural Networks
RFs	Random Forests
RL	Reinforcement Learning
SVMs	Support Vector Machines
TP	True Positive

CHAPTER ONE

INTRODUCTION

1.1 Background

Digital media has grown and revolutionized the way people consume information, with social media platforms and online news websites becoming primary sources for news and updates (Cavus et al., 2024). However, this shift has led to the spreading of false information, intentionally presented as legitimate news. As misinformation continues to grow, it has become a global concern, prompting governments, organizations, and researchers to seek effective solutions for its detection and prevention (Villela et al., 2023).

According to Daud et al. (2023), Machine learning offers promising approaches for developing automated systems that analyze and classify news articles based on their authenticity. Existing fake news detection systems often rely on traditional content analysis, which is labor-intensive and slow. ML processes and easily identifiable by human beings, enabling faster and more accurate detection of fake news. Developing a machine-learning-based model for detecting false articles revolves around arising challenges faced by misinformation. Existing methodologies are an evolving landscape of machine learning techniques that enhance content verification (Daud et al., 2023).

The necessity of false news detection systems has grown with the rise of digital information sharing, especially on website platforms (Harris et al., 2024). Widespread circulation of inaccurate information on online platforms heightened the importance of effective detection tools. False information influences public perception of significant challenges. Misinformation could skew public opinion, create social unrest, and create a misinformed society. Digital platforms allow information, both true and false, to spread at an unprecedented speed and scale (Vosoughi et al., 2018).

Festré (2020) due to the immense volume of content created daily, human moderators alone can't manage spotting and highlighting fabricated news. Automated detection models work continuously and scale to cover a large amount of data across multiple

platforms, enabling real-time analysis. A model that accurately verifies content and quality helps in building a trusted digital environment, where users rely on the information they encounter (Hamed et al., 2023). However, Automated ML based fake news detection models are not only important but essential to mitigating the impact of misinformation. They provide an essential layer of defense against the spread of false information (Mohsen et al., 2024). False news detection employs a range of techniques from natural language processing (NLP) to network analysis. These methods help in identifying patterns and features unique to fake news, providing more effective content verification (Hu et al., 2024).

According to Shu et al. (2019), the key techniques employed in fake news such as Lexical Analysis, Semantic Analysis, Stance Detection, and Deception Detection. Lexical Analysis examines word choices, frequency, and patterns within new articles. It also helps to identify the typical language choices, which may differ from those used in objective reporting. While Semantic Analysis goes beyond the surface meaning of words to understand the context and relationship between words (Thomas, 2020), Stance Detection involves analyzing the position that an article takes on a specific topic by comparing the stance of the news article to verify information sources (Saha et al., 2023). Deception-Detection belongs to the field of Natural Language Processing (NLP), which aims to diagnose markers of deceit in language. Therefore, Researchers have identified several linguistic indicators associated with deceptive content, such as a higher use of negations, passive voice, and patterns of word repetition (Chiluwa & Samoilenko, 2019).

1.2 Problem Statement

Arowolo et al. (2023) The growth of digital-media, and the widespread use of social media platforms have significantly increased the dissemination of fake news. The impact of such misinformation is profound. Traditional fact-checking methods are often time-consuming and insufficient to address the scale and speed at which false information circulates (Kaliyar et al., 2021).

Although various automated approaches have been proposed, many lack accuracy, contextual understanding, and scalability (J. Wang et al., 2023) Moreover, the existing models may struggle to differentiate between genuine news and subtle misinformation,

leading to either false positives or undetected fake news (Su et al., 2024) The research seeks to fill this gap through the developing of a ML-based model which integrates successfully Natural-Language Processing(NLP) techniques to identify the most relevant indicator of misinformation, and supervised learning algorithms to classify content as fake (Alnuaimi & Albaldawi, 2024).

1.3 The Objectives

General objective

The study's main objective aimed to develop an effective machine learning-based model for detecting fake news.

Specific objectives

Research had the following objectives:

1. To develop a machine learning-based Model that integrates real-time news verification platforms.
2. To collect and preprocess a large and diverse dataset of online news articles ensuring balanced representation of fake and legitimate news content.
3. To apply Natural Language Processing techniques (tokenization, stop-word removal, TF-IDF, word embeddings) to extract textual features from the dataset.
4. To compare and validate the performance of the model using standard metrics: accuracy, precision, recall, and F1 Score.

Research Questions

1. Could a machine learning-based model (ML) be developed to support real-time integration into news verification platforms?
2. How could a large and diverse dataset of online news articles be collected and preprocessed to ensure balanced representation of fake and legitimate news content?
3. Which Natural Language Processing techniques (tokenization, stop-word removal, TF-IDF, word embeddings) are most effective in extracting textual features for fake news detection?

4. How could the performance of the developed ML-based model be validated and compared utilizing standard metrics: accuracy, precision, recall, and F1-Score?

1.4 Research Justification

Ali et al. (2022) A proposed model designed for fake news detection that effectively helped combat the growing issue by identifying false content. In contrast to the prior studies, many existing systems use traditional rule-based or statistical models that heavily rely on manual features. The proposed model aimed to integrate Advanced NLP with multiple supervised ML Algorithms, analyze their comparative performance, and optimize the best-performing model (Kuruvilla & Kundapura, 2022). Additionally, while most past researches focus only on detection. The proposed model also analyzed and identified various content verification strategies, which strengthen the reliability of the results and add a practical layer of trustworthiness to the model. (Thomas, 2020) Therefore, traditional methods of content verification, such as machine-learning approaches, fail to detect more misinformation. A ML approach, such as Natural Language-Processing (NLP), identifies fake information with greater accuracy (Mayo, 2018).

Mohsen et al. (2024) Fake news evolves constantly, with new narratives and different linguistic patterns designed to evade detection. Machine learning models retrain and improve with new data, making them adaptable to new forms of misinformation, unlike static verification methods. By detecting and filtering out fake news, a machine learning-based model helps restore public trust in digital media. Hence, when platforms are able to provide more reliable information, they contribute positively to inform public discourse and empower readers to make decisions based on verified facts (Tejedor et al., 2024).

Rolf et al. (2022) Beyond media, a machine learning model has applications in areas such as finance, health, and education. Misinformation leads to serious repercussions. Industries and sectors with a vested interest in reliable information, a machine learning-based model for fake news detectors provides insights into common patterns in misinformation, enabling users, educators, and policymakers to understand the patterns effectively (Tejedor et al., 2024). Creating an ML-based model addresses an

urgent need for accurate, scalable, and adaptive content verification. The solution to these challenges was provided by (Naeem et al., 2023).

1.5 Significance of the Research

The significance of developing ML-based models in detecting fake information, including analyzing techniques such as accurate content verification, focuses on various domains such as Strengthening Information Authenticity, Supporting Democratic Processes and Public Opinion (social impact), Empowering journalists and Fact-Checkers, Economic Value for Digital Media Platforms, innovations in ML and Artificial Intelligence Research (Jawad & Obaid, 2022).

Fake news and misinformation erode public trust in digital content, particularly on social media and news platforms (Jawad & Obaid, 2022). A machine learning-based detection model enhances information authenticity by providing reliable and automated content verification. This strengthens users' trust in digital platforms and promotes an environment where users are confident in the content they consume.

Dierickx et al. (2023). Therefore, misinformation has serious consequences on public opinion, political stability, and decision-making processes. By identifying fake news, the model helps reduce the negative influence of false information on political climates, public health, and social cohesion, protecting democratic values and informed civic engagement (Jawad & Obaid, 2022).

Nakov et al. (2021) Consequently, journalists and fact-checkers face increasing pressure to verify massive amounts of online content manually. Model supports the efforts by acting as a preliminary filter, flagging suspicious content, and allowing professionals to focus on depth verification (Chiluwa & Samoilenko, 2019) Hence, misinformation poses risks. In other words, Digital and social media platforms face considerable economic costs, including reputation damage and lost user trust. An accurate, artificial intelligence-driven fake news detection tool helps the platforms maintain user experiences (Jawad & Obaid, 2022).

Arowolo et al. (2023) Developing this model enhances the field of machine learning by advancing methods in natural-language processing (NLP), deep learning, and data mining. Fake news detection requires a qualified model of nuanced language

interpretability, feature extraction, including adaptation to evolving misinformation patterns. Chiluya and Samoilenko (2019) Fake news detection models not only could help in identifying false content but also reveal common misinformation patterns. This insight could support education efforts to improve digital literacy, helping students, educators, and the general public understand how fake news operates and why verification is crucial (Shu et al., 2019).

1.6 The Limitations of the Study

Zahednejad et al. (2020) One of the limitations was that the data scratches the surface of all articles available online. Then, developing a machine-learning-based Model requires data quality and representativeness. Moreover, the proposed model was trained on effective machine learning techniques that rely on high-quality datasets with accurately labeled real and fake news. However, data collected from the internet contains bias, outdated information, and insufficient context, which could lead to inaccurate classifications.

A dataset lacking diversity in these areas leads to a model that performs well on specific types of content but poorly on others (J. Wang et al., 2023). Therefore, the limitation is significant because a model that fails to handle diverse content accurately may misclassify real news as fake (Chiluya & Samoilenko, 2019).

1.7 Delimitations

Choudhary and Arora (2020) The model is limited to analyzing news in a single language, such as English, which allows for more controlled dataset collection and the model's applicability to other languages. The study could delimit the types of fake news to certain categories, like political, health-related, rather than attempting to cover all forms, and allowing the model to focus on particular content verification techniques (Chiluya & Samoilenko, 2019).

Another delimitation is the source of data. The proposed model might rely on data from certain platforms, such as Twitter, Facebook, etc., rather than a broader range of sources Ghaddar and Naoum-Sawaya (2017) This might help in creating a manageable dataset, but it may narrow the capabilities of the model to identify false news on other platforms.

Mayo (2018) To manage resources effectively, the model was limited to specific ML techniques (Natural Language Processing) rather than exploring more complex architectures: deep Neural-Networks(NNs) and Convolutional Neural Networks(CNNs).

The study might also focus on post-publication analysis of content rather than aiming to identify real-time news. Hence, the boundary makes the development process more feasible but may delimit real-world application in settings where timely intervention is crucial. The delimitations allow the study to stay focused and manageable, although they may restrict the model's utility (Tandoc et al., 2017).

1.7 The Assumptions Of the study

The following assumptions were made:

1. Utilizing the dataset, the model's training and testing proved reliable, accurate, and represented various categories of false news circulating across multiple platforms.
2. The selected metrics, such as accuracy, precision, recall, and F1-Score, reflected the model's effectiveness.

1.8 Operational Definition of Terms

Machine Learning (ML): This forms part of artificial intelligence (AI) dedicated to systems that are able to derive knowledge from data and making informed decisions.

Fake news (FN): False information is intentionally deceptive, designed to mislead audience. It could take various forms and could be disseminated through different channels, like social media and websites.

Detection: Detection is the act of recognizing the presence, existence, and occurrence of something. It involves the use of tools, techniques, or methods to recognize specific signals, patterns, or anomalies in data and the environment.

Content verification: Content verification validates the authenticity of information or content before it is published.

1.9 Summary

In summary, developing a machine learning based system (Ruchansky et al., 2017) involves leveraging a mixture of NLP methods, data processing strategies, adapting the evolution patterns of misinformation.

Kuruvilla and Kundapura (2022) The goal of the study was to create a machine-learning -based model capable of identifying misinformation accurately, utilizing NLP. Therefore, current false information identification methods encounter challenges, and the complexity of language-evolving misinformation, data limitations, and privacy concerns (Liu et al., 2021) As misinformation increasingly spans multiple languages and cultural contexts, models should adapt to nuanced expressions and varying cultural norms (Li & Zhou, 2020).

CHAPTER TWO

LITERATURE REVIEW

2.1 Introduction

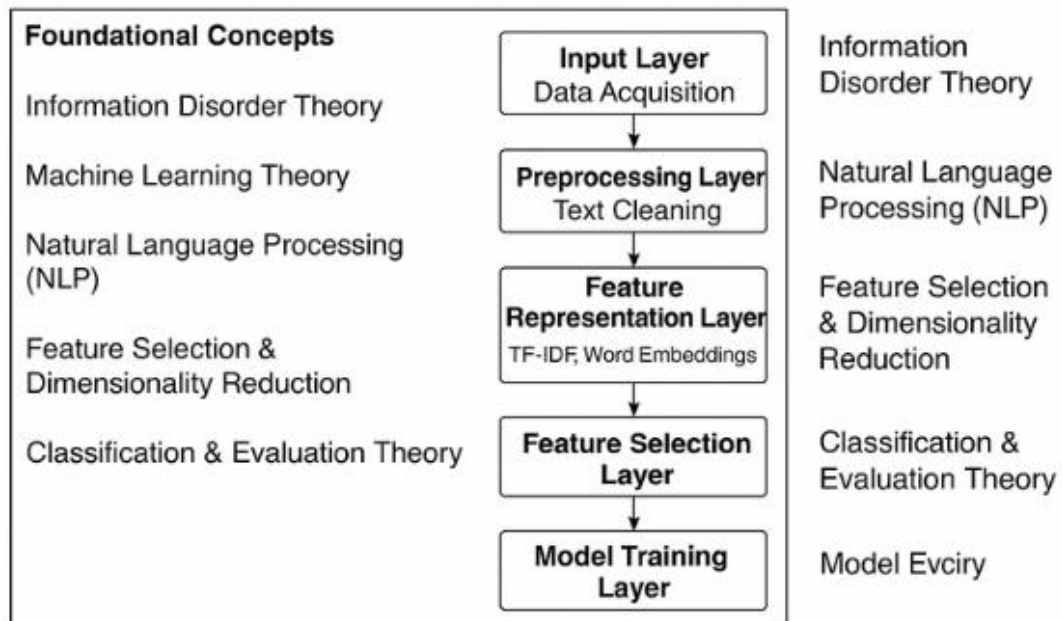
This chapter aims to explore how various ML techniques are employed in enhancing security and safety by detecting misleading information in published articles. The chapter presented a survey of existing academic work in this domain, including studies conducted in cloud-based environments. It then examined relevant literature in detail, synthesizing the information gathered. Additionally, it provided a concise review of ML-based models specifically designed for fake news detection.

2.2 Theoretical Framework of a Machine Learning-Based Model for Identifying Fake News

Sravani and Kumar (2025) ML models learn patterns from inputs, which generate predictions and classifications. This supports the premise that algorithms can distinguish between fake and real news based on textual and behavioral features. Developed a machine learning-based model that gains insights from data by improving its effectiveness over time without human intervention. Statistical and information-theoretic approaches provide the theoretical justification for reducing dimensionality while preserving discriminative features (Ruchansky et al., 2017).

Figure 2.1

Theoretical Framework



Concepts like supervised learning, confusion matrices, support the systematic evaluation of systems. The systems utilized techniques like naïve-bayes based on attributes like Bag of Words (BoW) for fake news detection but struggled with complex linguistic patterns. From the 2000s up to now, the increase of machine learning marks a significant transformation and evolution in this field (Ruchansky et al., 2017).

ML-based models are being introduced across various areas and industries, such as facial recognition and predictive analytics. However, Moszczuk et al. (2021) noted that they raise important ethical, legal, and accountability, which are carefully considered and addressed. Machine Learning-based models for detecting fake news articles improved by Taha et al. (2024) to 94.21% in accuracy. Most researchers working in Bangladesh contribute to research focusing on traditional ML algorithms like logistic regression and voting classifiers. The LIAR dataset was introduced by W. Y. Wang (2017) It is created for identifying false information tasks. This comprises 12,836 traditionally labelled items gathered from the PolitiFact website. Their aim of developing traditional machine learning models is to make predictions and rely on the

data without any programmed rules. Based on ML systems, it is anticipated that the models constructed, utilizing existing datasets, can adapt to the high variable values (Zahednejad et al., 2020).

Janiesch et al. (2021) Machine learning models use supervised-unsupervised, and reinforcement learning as illustrated in Figure 1 below. A supervised-learning model is one in which the algorithm is responsible for acquiring information from a labelled dataset with clearly specified inputs and outputs (Janiesch et al., 2021). With this type of model, the computer learns to categorize data without the assistance of any labels. Detecting fake news is one of the numerous sectors in which models based on machine learning are finding broad implementation. The accuracy and effectiveness of the model are both impacted by the techniques and parameters utilized. Models are based on machine learning and need both training data and frequent updates in order for them to continue being effective and accurate over time (Janiesch et al., 2021).

(Alnuaimi & Albaldawi, 2024) Discuss the essential role of ML classification models. They learn patterns from labelled datasets to differentiate between false and authentic news. By extracting patterns from news articles, such as linguistic characteristics and sentence structures. An overview of ML classification:

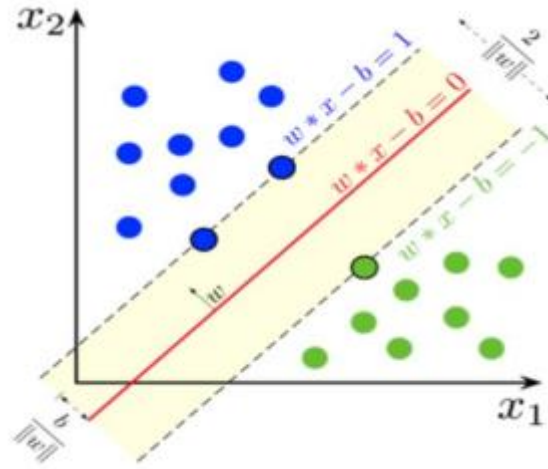
2.2.1 Support Vector Machine (SVM)

The Support Vector Machine technique was first developed by Vladimir and his colleagues in 1963 (Kuruvilla & Kundapura, 2022). Since then, the support vector machine has become popular and is a powerful ML algorithm in identifying fake news (Villela et al., 2023). SVM finds the best approach for classifying the data depending on the position (positive and negative) classes in the hyperplane. A hyperplane helps maximize the distance between data points (Ghaddar & Naoum-Sawaya, 2017). Like DT and RF, SVM additionally be implemented in classification techniques using regression analysis and classification (Kuruvilla & Kundapura, 2022).

Patel et al. (2021) Support Vector Machine could effectively separate fake from real news by finding a decision boundary based on linguistic features. They are effective in handling complex relationships between words when kernel functions are applied. Utilizing the SVM algorithm, Figure 1 below illustrates how data is classified.

Figure 2.2

Support Vector Machine



Note. SVM has presented datapoints which are closer to the hyperplane and performed 94.93% of accuracy (Patel et al., 2021).

2.2.2 Decision Trees (DT) model

Ruchansky et al. (2017) In fake news detection, decision trees use linguistic features to build simple rules that separate fake news and real news. However, a decision tree might divide data based on factors such as emotional tone, word choice, etc. Decision trees may yield good results on training data, but struggle with unseen data. The accuracy obtained by Pandey et al. (2022) for the proposed model using Decision Trees reaches 73.33%.

2.2.3 Random Forests (RF) model

Moszczuk et al. (2021) RFs are models made up of various decision trees. The tree-structure trained a random subset of the data, and made final predictions based on the minimum vote from all trees. RFs are used for fake news detection as they are more robust to overfitting compared to individual decision trees. By applying predictions, they achieve high accuracies. For instance, random forests could be computationally intensive and might not capture long-range dependencies in text data in advanced models. However, the performance of the RF model improved was 92.56% of accuracy (Felicilda et al., 2024).

2.2.4 K-Nearest-Neighbors (KNN)

Thomas (2020) KNN is a non-parametric instance-based learning method that classifies new data points by considering the K closest labeled examples in the training set. K-Nearest Neighbors (KNN) detects fake news by comparing an article's linguistic features to its closest neighbors. Therefore, KNN becomes computationally inefficient for large datasets because it requires measuring the distance between every pair of data points. (Felicilda et al., 2024) noticed that the KNN algorithm performs better in terms of data classification. The result obtained while training their proposed model was 89.98% accuracy (Petkar & Sonawane, 2020).

2.2.5 Hybrid Model (HM)

Ruchansky et al. (2017) It combines various machine learning and Deep learning by improving accuracy on a particular system. By integrating complementary models and methods, its approaches strengthen individuals' techniques, overcoming the weaknesses, making them highly effective in identifying fake news (Saha et al., 2023). For instance, hybrid approaches can capture various aspects of fake news, from linguistic patterns to network behaviors, by making them more effective in handling complex and diverse misinformation content (Kaliyar et al., 2021).

The following are hybrid models and how they are being used:

Ensemble-Based Hybrid Models

Saha et al. (2023) The ensemble hybrid model can be classified as follows: the Model Stacking. This model combines the outcomes of multiple systems, like Decision-Trees, support vector machines, and neural-network. Hence, each system generates a prediction and then aggregates the predictions to make a final decision. For instance, a logistic regression model could be used as the meta model to combine predictions from simple classifiers like Naïve Bayes and/or Support Vector Machine (Ruchansky et al., 2017).

Bagging and Boosting: The following techniques, random forests or bagging, including gradient boosting such as XGBoost, are integrated with other models to boost accuracy. (EAlmandouh et al., 2024).

Feature-Based Hybrid Models

Abdelillah et al. (2023). Linguistic and Network-Based Features: Hybrid models could incorporate both textual characteristic (sentiment, readability, etc.), and network-based, social network, including user behavior (Alnuaimi & Albaldawi, 2024) By combining content and network-based information, the model gains a perspective on the article's veracity. For example, linguistic analysis might detect manipulative language, while network analysis could identify suspicious activity by sharing outliers associated with fake news (Choudhary and Arora, 2020).

Raja and Raj (2022) Textual and Visual Features: Fake news includes sensationalized images alongside misleading text. A hybrid model could use Natural Language Processing techniques to analyze the text and computer vision techniques to assess the authenticity of associated images. By combining text analysis, like sentiment and stance, with image verification, such as image source and detecting image manipulation, the model could more accurately determine the legitimacy of a new article (Kaliyar et al., 2021).

Knowledge-Based and Machine Learning Hybrid Models

Choudhary & Arora (2020) Fact Checking with Machine Learning: The hybrid models, such as Wikipedia, FactCheck.org, incorporate external knowledge into machine learning workflows. They use knowledge-based verification to check specific claims against trusted sources and apply a machine learning model to analyze linguistic outliers or anomalies (Hamed et al., 2023).

Harris et al. (2024). Neural Symbolic Models: Neural Symbolic Models combine symbolic reasoning with interpretable neural networks, which could learn patterns from data. In fake news detection, symbolic reasoning could be used to validate logical consistency in an article, while a neural network assesses linguistic indicators of fake news (Abdelillah et al., 2023).

Multimodal Hybrid Models

Kumari and Singh (2024). Multimodal hybrid models are designed to process and integrate multiple data types (text, images) and social signals to improve fake-news identification. By combining diverse data sources, the system achieves a more

comprehensive analysis, which allows it to identify indicators of misinformation. Multimodal Hybrid Models have the following types:

Ikotun et al. (2022). Text and image Analysis: By analyzing the two elements, the model could detect inconsistencies (in relation to the text and images). False articles sometimes contain images that are irrelevant, misleading (Kumari & Singh, 2024) Multimodal models analyze the text for language while simultaneously examining the images for visual signs of manipulation of the content. For instance, in some fake news posts, text is embedded within images such as screenshots of new headlines, and /or media posts (Yan et al., 2024).

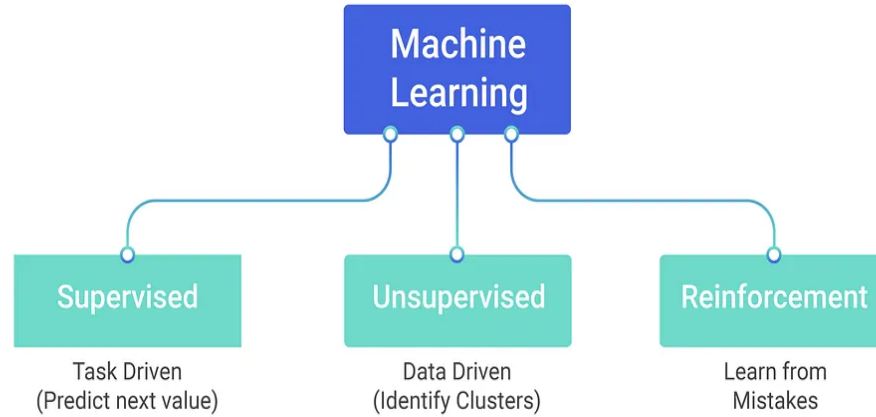
Cavus et al. (2024). Social Media Signals and Textual Content: This includes User behavior and Post Engagement patterns, which rely on fake news on social media, which often spreads coordinated activity, including rapid reposting. Multimodal models could analyze user behavior metrics like the speed and pattern of sharing comments alongside text content analysis to identify suspicious posts.

Multimodal Transformer Models: Multimodal Transformer Models include Vision Language Transformers. These transformers train text and image records. Contrastive Language-Image Pretraining (CLIP), which could analyze text and images in a single framework, making it possible to assess the coherence between an article's language and its visuals. For instance, a model could determine if the visual content is relevant (Kumari & Singh, 2024). Hybrid NLP Techniques: Combining multiple Natural Language Processing techniques enhances detection accuracy. For instance, using sentiment analysis alongside could help identify fake news articles. Hybrid approaches may also combine semantic analysis with stance detection to identify inconsistencies between an article (Petkar & Sonawane, 2020).

2.3 Machine-learning techniques

Figure 2.3

Machine-learning techniques



2.3.1 Supervised learning

According to Tchio et al. (2024) Supervised learning is an ML approach in which a model is trained using labeled data. In this ML case, a labelled record is trained, thus, each input results in an accurate output. As a result, supervised learning maps the input into the output, enabling the model to predict new inputs accurately. The model learns by identifying patterns that connect to their corresponding labels. Learning process involves minimizing the error in the prediction model (called Data Training). Supervised learning models analyze patient data like symptoms and test results to diagnose diseases (Daud et al., 2023).

However, Supervised learning predicts equipment failure by analyzing data and helping to prevent downtimes, especially in Industries. In financial Institutions, Supervised learning helps identify fraudulent transactions by examining transaction patterns (Osisanwo et al., 2017). Supervised learning models are effective in achieving high-accuracy tasks with clear input and output relationships, such as image recognition. However, techniques such as cross-validation and regularization, including pruning as well, could be helpful to counter overfitting (Tchio et al., 2024).

2.3.2 Unsupervised learning

Suthaharan (2015) The unlabeled dataset trains the model without knowing the correct output. Unsupervised learning aims to learn the underlying data structure and find its grouping or patterns. This involves discovering hidden patterns within the data without specific guidance (Naeem et al., 2023). In unsupervised learning, ML-based models use the following techniques: clustering learning, anomaly detection, K-means, etc.

2.3.3 Clustering Learning

Naeem et al. (2023) Clustering is a fundamental method in unsupervised learning. Its emphasis is on clustering alike data points without using predefined labels. A cluster shares greater similarity with each other than with those in different clusters. Tianda et al. (2024) applied clustering for feature classification, as illustrated in Table 1 below.

Table 2.1

Clustering learning

Clustering class	Adjusted rand score	Purity score
K-means clustering	58.03%	88.09%
Agglomerative clustering with ward's	49.36%	85.13%

Table 2.1 by Tianda et al. (2024) was the evaluation metrics of the K-means clustering including agglomerative clustering with Word's classes.

2.3.4 K-means clustering

Ikotun et al. (2022) K-means clustering is a widely used algorithm in unsupervised learning that partitions a dataset into k distinct clusters, where each cluster has a centroid representing its central point. However, by initialization, centroids require choosing k(clusters). In this scenario, the number of clusters is predetermined. Moreover, by Assigning Points to Clusters, every data point calculates its distance from each centroid and is allocated to the nearest one (Patel et al., 2021). Through the

Centroids update, clustering recalculates each centroid(it recalculates the average and the center of data), and points are assigned to the cluster (Tianda et al., 2024).

Figure 2.4

K-means clustering

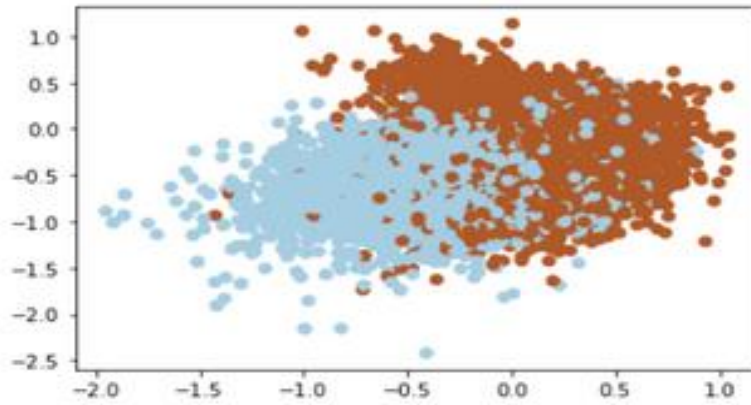


Figure 2.4 by Tianda et al. (2024) shows a plot of K-means clustering that represents how data points were assigned to each other.

2.3.5 Reinforcement-Learning (RL)

Reinforcement learning Rolf et al. (2022) is a crucial step that enables the ML model to make decisions throughout the training process. Based on the environment, the machine learning model receives feedback indicating how good or bad actions could be made in particular situations. Reinforcement learning could be applied to problems that cannot be solved with other machine learning algorithms. However, focused on the problem, Conventional machine learning algorithms are designed to succeed at specific tasks. Moreover, reinforcement learning doesn't divide the problem; meaning, it works directly to maximize the long-term reward. For improving performance, RL is a versatile method that could be integrated with other machine learning techniques, e.g., deep learning, and more (Millière, 2022).

2.4 . Fake News Detection

According to Hu et al. (2024), the advancements have facilitated access to diverse sources of news and information. Fabricated information presents legitimate news. (Li & Zhou, 2020).

Fake news can be defined as deliberately fabricated and verifiably false information, which is considered a threat to democracies, erodes public trust in institutions, and significantly influences critical societal elements such as elections and public opinion, including on matters like wars (Arowolo et al., 2023) In response to the harmful impact, numerous initiatives have been emerged. One of these is the growing approach of studying and analyzing fake news for fact-checking purposes.

Moreover, while applied on news article datasets through classification, the accuracy improved by Elkhatat et al. (2023) was 91%. Because of simple classification and extraction features utilized, which were not enough to tackle issues, Researchers proposed to combine it with machine learning techniques for better performance and accuracy as well (Elkhatat et al., 2023).

2.4.1 The scope Of Fake News

Ali et al. (2022) the motivation behind fake news could vary from entertainment to propaganda and manipulation as well. However, understanding its scope involves examining the types of fake news, the motivations behind its creation, and the contexts in which it spreads. The nature of fake news makes it a complex issue that challenges information integrity across various platforms (Arowolo et al., 2023).

Lim and Perrault (2023) The digital landscape plays essential function in the growing proliferation of false informations. Virtual community platforms facilitate the sharing of content, allowing false narratives to quickly reach large audiences. Techniques that prioritize engagement over accuracy could amplify sensational content, further complicating efforts by differentiating between genuine to misinformation (Elkhatat et al., 2023).

2.4.2 The Types of Fake News

Pandey et al. (2022) Developing an effective detection method and increasing public awareness, detection of false updates is crucial. Fake news is categorized in the following ways:

Motivated by political, financial, and/or ideological goals, disinformation involves strategies to spread false narratives and public opinion like propaganda during political

campaigns, misleading news on foreign policies, and fabricated information on contentious social issues (Elkhatat et al., 2023).

Su et al. (2024) Misinformation is false and inaccurate information shared without malicious intent. It is spread by individuals. Misinformation could quickly go viral due to individuals' eagerness to share information that confirms their biases or beliefs (Li & Zhou, 2020).

Su et al. (2024) A propaganda: It presents information in a biased way by omitting facts to achieve a specific outcome. Therefore, if propaganda is associated with political causes and/or government messaging, it could be used in advertising social movements. Propaganda could be powerful in shaping public opinion by mobilizing people for or against a cause. Propaganda could lead to social cohesion around shared beliefs, creating divisions (Elkhatat et al., 2023).

By prioritizing influence over accuracy, propaganda could make it challenging for people to access unbiased information. Currently, in the digital age, propaganda could spread more news and/or user-generated content. However, propaganda could make media literacy skills crucial for identifying and analyzing propaganda messages (Li & Zhou, 2020).

2.4.3 The Importance Of detecting fake news

Su et al. (2024). Detecting false information is becoming increasingly important in the society. The implications of fake news are reaching and affecting individuals, economies, etc. Fake news undermines trust in traditional media and journalism. As people struggle to discern useful and credible information from trust in all, news sources could decline to generalized skepticism toward the media (Shu et al., 2019). When misinformation targets institutions like governments, public health organizations, it could diminish public confidence in the institutions' decisions (Cavus et al., 2024).

Taha et al. (2024) By encouraging people's mentalities, fake news contributes to social fragmentation. And this could weaken community ties and increase conflict between different groups in society. The implications of fake news are serious, from health and

safety to democratic stability and social cohesion. By addressing misinformation via technology, society could foster social unity (Cavus et al., 2024).

2.4.4 Challenges Faced in Detecting Fake News

Li and Zhou (2020). False information detection faces a multitude of challenges due to the evolutionary nature of misinformation, which appears across diverse media formats and cultural contexts. Complexity of Language is one of the challenges faced identifying false news including ambiguous language (Májovský et al., 2023). For instance, resembling fake news in structure serves different purposes. Followed by Data Limitations. Datasets for fake news are limited, and creating labeled data is time-intensive or time consuming. Another challenge is Concept Drift (Chhajer et al., 2021). News patterns evolve rapidly, requiring models that could adapt to new forms of misinformation. And finally, Real-Time Detection. Developing models with low latency that maintain high accuracy is challenging due to resource limitations (Cavus et al., 2024).

The difficulties in detecting fake news stem from the ever-changing nature of misinformation, the complexity of detection across modalities, and concerns about user privacy. These challenges require advancements in multimodal models, real-time guidance, and implementation. Mentioning these challenges is essential to create reliable and trustworthy fake news detection systems or models (Thakar & Bhatt, 2024)

2.5 Utilization of datasets in machine learning

Gong et al. (2023) explain that a dataset is a collection of instances used to assess a machine learning model's performance. To develop and train the model, the authors utilized various datasets. One of these, the NLPCC2013 dataset from China, was created in 2012 and used in 2013 by Gong et al. (2023). The dataset was divided so that 70% were assigned to the training set, 15% to the Validation set, 15% to the test set.

The table below Table 2.2 displays the most frequently used dataset in machine-learning-based models.

Table 2.2*Datasets commonly used in machine learning based models*

Data	DatasetName	Application(Task)	Data on the amount	Data content
Text	AmazonReview	Sentiment-analysis	34 GB	Amazon customer reviews and star ratings
	IMDB	Word Segmentation, Sentiment analysis	80 MB	Movie-reviews
	NLPCC2013	Sentiment analysis	15 MB	7 moods-marked, total-of-14,000 Weibo, 45,431 sentences
	20Newsgroups	Word-Segmentation	63 MB	20 groups of English news data on different topics
	SST-1	Sentiment-analysis	946 KB	Movie reviews
	Yelp	Sentiment analysis	6.4 GB	Yelp-merchant, review-and-user data
Image	AG.news	Word-Segmentation	11 MB	News-articles
	Label me	Image Classification, Object Detection	0.9M	187,240images, 62,197 annotated images, and 658,992 labeled objects
	Pascal VOC	ImageClassification, Object-Detection, SemanticSegmentation	4 GB	Image of 20 categories of objects
	ImageNet	Image Classification, Object-Detection	146.4 GB	20,000+ categories

	COCO	Semantic Segmentation, Object- Detection	25.2 GB	Tagged-images
	COIL100	Object-Detection	129.84 MB	100different objects, imaged in full 360°
	Stanford Canine	ImageClassification	2 GB	120 categories
	Places	ImageClassification, Object-Detection	27 GB	205scene categories and 2.5 million images with category labels
	CelebFaces	Object-Detection	21.7 GB	Celebrity-images
	CIFAR	ImageClassification	324 MB	CIFAR-10-and CIFAR-100
Voice	Voice	Speech Recognition	1TB	104different languages, 2023
	MozillaCommon			
	LibriSpeech	Speech Recognition	100 GB	Audio-book dataset, including text and speech
	VoxForge	Speech Recognition	10 GB	sample-data English French- German-Spanish, Italian-Russian are relatively rich
	TIMIT	Speech Recognition	1 GB	Eight major dialects of American English
	CHIME	Speech Recognition	4 GB	Contains-real simulated-and- clean-and-voice recordings

MUSAN		Speech-Recognition	11 GB	Corpus of music, speech, and noise recordings
TAU	Urban	Speech Recognition	27.5 GB	1-second audio segments from 10 acoustic scenes, total 64 h of audio
Acoustic Event Dataset	Scenes 2022 Mobile	Speech-Recognition	1.2 GB	28 classes of acoustic-event sounds

Note. Table 2.2 shows various types of datasets used in ML for identifying fake news detection (Gong et al., 2023).

2.6 Enhancing techniques for the age of Large Language Models

(Su et al., 2024) Adapting to new techniques, is a constant challenge because misinformation evolves in response to changes in technology and detection methods W. Y. Wang (2017). Fake news creators use methods to evade detection, which automate fake news detectors to update and improve the content. These are techniques in misinformation and how detection tools are being adapted (Vosoughi et al., 2018).

2.6.1 Deepfakes and Synthetic Media

Millière (2022) Deepfakes also use Artificial Intelligence (AI) to create highly realistic fake videos and images of people. Deepfakes could falsely attribute actions to public figures, with significant potential to mislead audiences. Advanced AI models trained could identify signs of manipulation, like unnatural facial inconsistencies (Yan et al., 2024).

2.6.2. Artificial Intelligence (AI) Generated Text

Kumar and Mindzak (2024) Artificial Intelligence (AI) Generated Text, produces by language models like GPT (Generative Pre-trained Transformer), has become an

increasing powerful tool in both beneficial and potentially harmful applications, the mentioned model could create coherent text and realistic fake articles, comments, and social media posts that mimics human writing, and making it challenging for fake news detectors and readers (Chhajjer et al., 2021). Identification tools analyze patterns and styles in machine learning generated text. Consistent sentence structures flag potential Artificial Intelligence (AI) generated text (Kumar & Mindzak, 2024).

Raja and Raj (2022) Artificial Intelligence (AI) models could write entire articles that appear legitimate, blending facts to create misleading narratives. This could be used to promote specific viewpoints by presenting biases as if they were news or information. Artificial Intelligence (AI) Generated Text is also used to create posts and comments that appear to be from real users. It could amplify false narratives, making information popular by creating the illusion of widespread support. In platforms with customer reviews, supporting social interaction, Artificial Intelligence (AI) generated text could be used to manipulate opinions of a product through fabricated conversations (Lancaster, 2023).

Raja and Raj (2022) Detecting Artificial Intelligence (AI) Generated Text is challenging because advanced models, like GPT and other large language models, have become highly skilled at replicating human language patterns. Other challenges could first, High Fluency Coherence. Advanced models are being trained on massive dataset which includes a variety of language patterns, styles, and contexts (Májovský et al., 2023).

As a result, they produce text that sounds fluent. Second, Context Awareness. Artificial Intelligence (AI) models generate responses that are contextually aligned with the input, maintain coherence across multiple sentences. Third, Limited Detection Tools and Techniques (Tandoc et al., 2017). Traditional detection tools that look for patterns, such as repetitive phrases and unusual syntax, are less effective with advanced models. Artificial Intelligence (AI) text no longer has the predictable phrasing that older models may have demonstrated. Just like with visual deepfakes, detecting text that has been generated to deceive requires more tools. Furthermore, Scalability and Rapid Evolution. Artificial Intelligence (AI) models could generate a high volume of content, flooding social media and websites with a mix of authentic

and Artificial Intelligence (AI) AI-generated posts. Language models are frequently updated and improved with large datasets, allowing them to get better at mimicking human language (Chhajer et al., 2021). Therefore, the quality of Artificial Intelligence (AI) text, combined with its subtlety and scalability, makes detection challenging. Ongoing advancements in detection tools focus on pattern analysis, metadata, and network behavior, keeping up to improve constant innovation of models (Elkhatat et al., 2023).

2.6.3 Misinformation in Encrypted or Private Channels

Martin et al. (2023) Misinformation in encrypted or private channels presents a significant challenge for detection and prevention because of the limited access to content shared within secure environments. End-to-End Encryption prevents access to message content for anyone outside of the conversation and the platform providers (Martin et al., 2023).

Mayo (2018) Misinformation spreads rapidly within private groups, where members share similar interests. In private channels, users might be less likely to report misinformation due to peer pressure, as the content is misleading. Misinformation in encrypted or private channels remains a challenging issue due to the privacy of the spaces. Platforms cannot fully monitor the content, various strategies, like forwarding limits, user education, are helping to mitigate misinformation risks while respecting users' privacy. As technology evolves, new methods might emerge to curb misinformation in the environments without compromising encryption (Martin et al., 2023).

2.7 Approaches to Address Misinformation in Encrypted Channels

Liu et al. (2021) Addressing misinformation in encrypted channels requires innovative solutions that respect user privacy and attempt to limit the spread of false information. Platforms cannot access message content due to encryption; they rely on indirect methods to curb misinformation. Some of the approaches to misinformation in encrypted channels are listed below:

2.7.1 Message Forwarding Limit

Mayo (2018) Platforms like WhatsApp, as mentioned above, have implemented limits on the number of times a message can be forwarded, especially for messages that are frequently forwarded. For instance, those forwarded multiple times could only be sent to one chat at a time. This means users cannot forward the same message to multiple groups and contacts simultaneously. Messages that have been submitted multiple times are marked with a double arrow and cannot be forwarded to more than five chats at once. While telegram does not impose strict forwarding limits like WhatsApp, but has privacy in forwarding features (Liu et al., 2021).

Hu et al. (2024). It allows users to disable message forwarding from their accounts and also hide the sender's profile in forwarded messages. This helps to reduce the reach and speed of viral information. Forwarded messages are labelled to show that they have been shared multiple times, by making users aware that they are interacting with content that did not originate from the close contacts. And also, this could prompt or label users to think twice before forwarding the message (Liu et al., 2021).

2.7.2 Fact-Checking Bots and Resources

Dierickx et al. (2023). Fact-checking bots and resources are tools developed to help identify, verify, and debunk false information online. The most popular social media fact-checking bots and resources, including technologies that are being used currently, are broken in the following ways:

Lim and Perrault (2023) Facebook's verification initiative: Facebook collaborates with independent fact-checking organizations, which is certified by Poynter's international Factchecking Networks like AFP, (Agence France Presse (U.S)) Snopes (U.S), and PolitiFact (U.S), Full Fact (U.K), Boom (India), Africa Check (Africa) etc., and partners with independent fact-checking organizations around the world. The partners certified by the International Fact-Checking Network (IFCN) help Facebook identify and label false content on the platform. Through content verification, Facebook uses Artificial Intelligence (AI) algorithms and reports from users to identify false posts. For media such as images and videos that are digitally manipulated, Facebook might add a label indicating that it has been modified (Dierickx et al., 2023).

Therefore, when users try to share a post that has been rated as false or misleading, they see or receive a warning and a link to the fact-checker's article explaining the rating. Also, users who have previously interacted with the content might receive a notification informing them of its updated status (Chhajer et al., 2021).

Lim and Perrault (2023) Twitter's Birdwatch is a fact-checking program launched to help address misinformation on the platform by adding context to misleading tweets. This birdwatch empowers everyday Twitter users to identify and clarify misinformation collaboratively. Birdwatch allows users to submit notes on tweets, which may contain false information. Users could provide additional context for other users (Liu et al., 2021).

These could appear in public, and everyone could see them on Twitter. After notes are added, birdwatch users could rate these contributions based on fairness and accuracy. Birdwatch's crowd-sourced model gives users a greater sense of ownership and trust in fact-checking since it involves people. Birdwatch is visible, and the process could be scrutinized by other users. Birdwatch could limit the spread of misinformation by attaching generated corrections to misleading misinformation (Dierickx et al., 2023).

In other words, Birdwatch aims to be community-driven and balanced. Hence, it could still be challenging to ensure that all voices are represented fairly on polarizing issues. However, not all notes are accurate. For this reason, Twitter must monitor the system to prevent abuse and the spread of biased information. Birdwatch could also represent a novel approach to fact-checking by giving users a voice in curbing collaborative misinformation, and providing transparency (Tejedor et al., 2024).

Kumari and Singh (2024) WhatsApp's Checkpoint is a service set up in April 2019 in India, especially for the national elections. It was launched by WhatsApp in collaboration with PROTO, an Indian media startup, along with Medan and Dig Deeper Media, which provides fact-checking expertise and technology (Kuruvilla & Kundapura, 2022) The aim was to address the increase in the spread of misinformation on WhatsApp, especially during election periods when misinformation could influence public opinion as well. WhatsApp users could forward messages they found suspicious to the checkpoint Tipline through a designated WhatsApp number. This Tipline allows users to send various types of media for verification, such as messages, images, videos,

etc. Once a message is well-received, fact-checkers analyze the content to determine its accuracy (Tejedor et al., 2024).

Monnier et al. (2023) After analyzing a framework, fact-checkers would then send feedback to the user, providing the verification result. Therefore, beyond human and individual verifications, Checkpoint Tipline collects data to analyze and understand misinformation trends. This also allows WhatsApp to identify common types of misinformation. Moreover, given the diversity of languages in India, the Tipline supports multiple languages, such as HINDI, TELUGU, BENGALI, and ENGLISH, to ensure accessibility of various regions (Tejedor et al., 2024).

Report and Flag Features are tools that social media platforms and messaging provide to help users address inappropriate content. The features allow users to alert the platform about potential violations, such as misinformation, hate speech, spam, etc. For providing a relevant report, using, for instance, Instagram, WhatsApp, etc., users can click on a Report or Flag option for a comment, post, message, etc. to read and view the content (Monnier et al., 2023).

Ali et al. (2022) A user submits the report, which might include additional context and comments about the content. The report is sent to the platform team or an automated system. And then, the content may be reviewed by Artificial Intelligence algorithms, which could scan for keywords, images, or anomalies that match the platform's policies. For instance, if the content is found to violate guidelines, it might immediately be removed (Dierickx et al., 2023).

However, the goals of report and flag features on social media and messaging platforms focus on ensuring user safety, maintaining content quality. The features allow users to report harmful content, and the community to address issues. The goals of the report and features are listed below:

Reduce spread of Harmful Content: Reducing the spread of harmful content is a primary goal to prevent harmful content, such as hate speech, misinformation, and harassment, from reaching an audience. Report features enable users to immediately notify the platform of harmful content. When multiple users report the same posts, the

system prioritizes them for review, allowing faster action on harmful material (Lim & Perrault, 2023).

Ali et al. (2022) Improve content quality and platform reputation: Improving content quality and platform reputation is a crucial objective for social media and messaging platforms. By enabling users to report misleading and offensive content, platforms could identify and remove posts that do not meet the content quality standards. Eliminating low-quality content also helps users feel content with the information they encounter (Dierickx et al., 2023).

Reducing Misinformation and Fake News: Platforms partner with fact-checking organizations to review reports. When content is identified as false, platforms may display a warning label on it. For instance, users might see messages like the information has been disputed by fact-checkers. Penalizing Repeat Offenders: Penalizing Repeat Offenders is a key part of social media strategy to reduce the spread of harmful content and misinformation. By penalizing users who violate community standards, platforms aim to discourage harmful behavior, reduce the visibility of misleading content, and maintain a safer online environment (Alghasi et al., 2024).

Trusted Fact-Check Notifications are an essential tool that platforms use to address misinformation by providing users with credible, verified information. The notifications appear when a post, article has been reviewed and verified by trusted fact-checking organizations through platform partnerships. In some scenarios, platforms will notify users who are being engaged (Alghasi et al., 2024).

Trusted Fact-Check Notifications provide several key benefits to users, social media platforms, and the online ecosystem. Ali et al. (2022) To display notifications that clarify misleading information, platforms help to curb the spread of misinformation and build trust. Fact-checker notifications reduce the likelihood that users will share flagged content. Trusted fact-checks may have several benefits to users, media platforms, especially in an age of rapid information exchange (Dierickx et al., 2023).

Li and Zhou (2020) Builds Public Trust: Fact-check notifications play a critical role in building public trust on a platform where misinformation could easily spread. By issuing trusted fact-check notifications, platforms that implement trusted fact-check

notifications illustrate a commitment to accuracy and help to build and maintain user trust. Hence, when users feel that a platform is dedicated to the truth, they are likely to engage with it responsibly (Alghasi et al., 2024).

Promotes Media Literacy: The notifications educate users on how to assess information critically. Trusted fact-check notifications play a powerful role by promoting media and evaluating information. Fact-check notifications help users become discerning when consuming content (Dierickx et al., 2023).

Li and Zhou (2020) Protects Reputations: for individuals and organizations to be safe and respectful, fact-check notifications provide a corrective measure to protect their dignity and reputations. False claims about individuals could lead to financial losses. However, fact-check notifications could contribute to a fairer information environment, where individuals and organizations have a chance to defend themselves against unverified information (Dierickx et al., 2023).

Encourages Responsible Content Sharing: Trusted fact-check notifications encourage responsible content sharing by helping users recognize or avoid misinformation, making digital spaces safe and reliable. Fact-check notifications act as a reminder for users to pause and consider the accuracy of content before passing it along (Alghasi et al., 2024).

Strengthens Democratic Process: Trusted fact-check notifications play a vital role in strengthening democratic processes by ensuring citizens have access to accurate information, which is essential for informed decision-making in a democracy. Democracy relies on voters making informed choices. Fact-check notifications provide accurate, verified information to understand candidates, issues without being misled by fake claims (Dierickx et al., 2023).

(Alnuaimi & Albaldawi, 2024) Trusted fact-check notifications play an important role in protecting vulnerable populations by helping them navigate misinformation and make safer choices. For instance, marginalized groups(like the elderly, young people, including those with limited digital literacy) face risks online. These risks could be more susceptible to misinformation if they lack experience with digital media and

critical evaluation skills. Fact-check notifications help by providing a clear indicator and making it easier for users to discern reliable content (Alghasi et al., 2024).

The foremost aim of fact-checking seeks to curb the spread of misinformation and ensure that information is circulated to address the risks to individuals. When integrating platform tools, fact-checking may rely on user data to personalize content by improving accuracy over time. These are critical concerns for implementing fact-checking notifications. If the system or model involves collecting and processing large data, ensuring that the data is secure and stored is essential to protect users' privacy in the system (Dierickx et al., 2023).

Alnuaimi and Albaldawi (2024) as technology advances, an individual's personal information is accessible to companies and malicious actors. Companies and governments collect amounts of data on individuals, including location, social media activity, by purchasing behavior. The data collection could lead to surveillance concerns and monitored issues about autonomy and freedom. Personal information stored online could be vulnerable to cyberattacks (Alghasi et al., 2024).

Mohsen et al. (2024) Once data is shared, users lose control over it. Terms of service agreements are lengthy and complex. The lack of transparency prevents individuals from controlling their personal information. Artificial Intelligence (AI) could analyze large amounts of data to predict behaviors (Alnuaimi & Albaldawi, 2024) However, advertisers use data to create detailed profiles and individuals by delivering highly targeted ads (Nakov et al., 2021).

2.8 Future Directions and Ethical Considerations

Nakov et al. (2021) Future Directions may focus on refining model interpretability, allowing stakeholders to understand the reasoning behind classification decisions. Ethical considerations, such as preventing biases in model predictions and ensuring user privacy, remain essential in fake news detection. Future models will deepen the integration of text, image, and social context, providing more comprehensive assessments of content (Rolf et al., 2022). The model will need to analyze and correlate data from various sources. To track the evolution of false news techniques, systems could learn from new data in real-time. However, using self-supervised learning could

enable models to generate their labels by identifying patterns in data. Future models could involve users in the feedback process, where they could correct misclassifications. This feedback could help users better understand model decisions and improve trust in the system (Pandey et al., 2022).

2.5.1 Ethical Considerations

2.5.1.1 Privacy and data protection

Martin et al. (2023). Ethical considerations are crucial in developing fake news detection models, as they impact user trust, privacy, and freedom of expression. Fake news detection involves analyzing user behavior, which could raise privacy concerns. However, detection systems should prioritize privacy by implementing anonymization techniques like differential privacy that allow data analysis without compromising individual identities (Pandey et al., 2022). Only the necessary data should be collected, with a focus on transparency around data collection practices. Users should be informed of what data is collected, how it is used, and their right to access or delete it (Nakov et al., 2021).

2.5.1.2 Transparency and Accountability

Pandey et al. (2022) For effective transparency and accountability, users and regulators should have the criteria for flagging content as potentially false. Transparency and Accountability help in building public trust and accountability. False-positive and negative information detection could have significant impacts, like censoring legitimate news and/or allowing harmful misinformation. However, detection platforms should have mechanisms to address errors and allow users to appeal decisions (Martin et al., 2023).

Ethical factors require balancing accuracy, privacy, and transparency to protect user rights. By prioritizing user privacy, minimizing biases, detection models could be implemented responsibly (Shu et al., 2019).

2.9 Research Gap

Tandoc et al. (2017) Most of the false news detection systems rely on a single language. Extending detection capabilities across cultural contexts is a challenge, as local

references vary widely. However, fake news appears on various platforms such as news websites. Ensuring that models generalize across different platforms and media, research could explore transfer learning techniques to enhance the model’s adaptability or improvement (Cavus et al., 2024).

Shu et al. (2019) Fake news trends evolve, making models less effective on data. With the increase of deepfake technology, fake news involves text, manipulated audio, and video (Martin et al., 2023). A research gap could help in creating models that could integrate multi-modal data, image, and text to detect content. Most of the advanced models, like Deep Learning, act like black boxes with low interpretability. Hence, improving the model could help users and fact-checkers understand the specific content. Many models focus on the content characteristics. Training data scarcity, especially labeled datasets, remains a challenge. Fake news detection models could propagate biases at higher rates (Liu et al., 2021).

Table 2.3

Empirical Table Summary of Literature Review

#	Author’s name/ year	Research title	Findings
1.	Su et al., 2024	Adapting Fake News Detection to the Era of Large Language Models	In their empirical data, they investigate how fake news detection approaches need to adapt given the rise of large language models.
2.	Millière (2022)	Deep Learning and Synthetic Media	Explores how deep learning–based methods are transforming the creation of audiovisual media, and what new philosophical issues this raises.
3.	Kumar and Mindzak (2024)	Detecting Artificial Intelligence. Generated	Investigate the ability of people in academic settings to distinguish between AI-

	Text from Human-Written Text	generated text and human-written text
4. Ruchansky et al., 2017	A Hybrid Deep Model for Fake News	Proposed a hybrid deep learning model called CSI , which combines three kinds of signals to detect fake news and identify the chronological patterns of user behavior around an article using recurrent neural networks (RNNs)
5. Májovský et al. (2023)	Artificial intelligence can generate fraudulent but Authentic	The AI output was convincing in terms of language style, structure, terminology, and overall. Raise serious concerns about scientific publishing integrity: AI tools could be misused to generate fraudulent articles, which might fool peer review or public consumption.
6. (Kumari & Singh, 2024)	A deep learning multimodal framework for false information detection. Engineering	Employed for text preprocessing, ensuring that the textual data is prepared effectively for analysis. Additionally, the Deep Learning translator is utilized to maintain language consistency across the dataset,

		facilitating accurate comparisons and evaluations.
7. (Arowolo et al., 2023).	A machine learning technique for the detection of social media fake news. International Journal on Semantic Web and Information Systems	Emphasizes the importance of filtering and preprocessing social media data to remove irrelevant or noisy information, ensuring that the dataset is sufficient for ML analysis.

CHAPTER THREE

RESEARCH METHODOLOGY

3.1. Introduction

Chapter three provided materials and techniques for the research, which were used for developing an ML-based model and analyzing techniques for accurate content verification (Thomas, 2020) Therefore, the proposed model incorporated various layers and data processing techniques to address key challenges in fake news detection.

3.2 Dataset

Naeem et al. (2023) The model was trained on a large and diverse dataset. The training allowed to the generation coherent and contextually relevant results. There were two files of the ISO Fake News Dataset. One for real news and the other for fake news, both in English. After evaluating the background, a total of 23,481 false and 21,417 true items were analyzed. The research utilized a labeled dataset comprising news articles categorized as true and false. The dataset was structured in different classes, true and False Pandey et al. (2022).

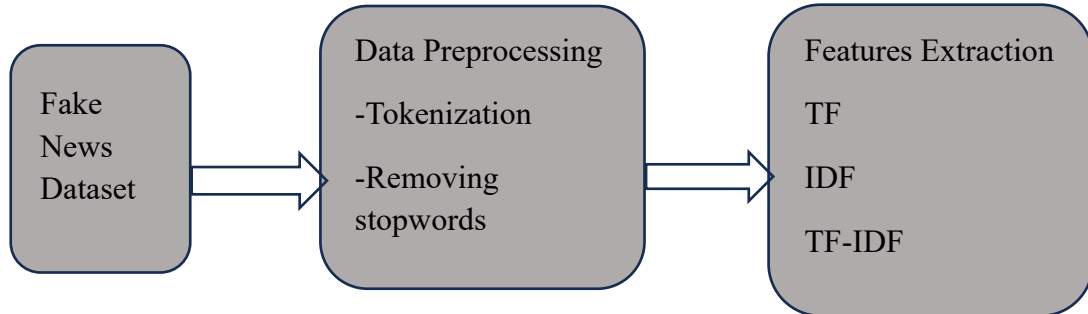
True: These are genuine articles, meaning they are accurate, truthful, and provide reliable information. This information helped establish a baseline Pandey et al. (2022).

False: These are false (fabricated) news articles

Tchio et al. (2024) a supervised classification for the dataset labelled was significantly required, as shown in Figure 3.1. A classifier is a model designed to assign a class label to a given input. Constructing a supervised classifier requires a training dataset with a labeled class point (Taha et al., 2024).

Figure 3.1

Fake news Dataset



3.3 Tools and Environment

Tchio et al. (2024) Developing an ML model requires a variety of tools and libraries to support various stages of the process. However, the thesis utilized the following tools: Python, Jupyter Notebook, and some libraries etc.

Python: This is a programming language used for study, because of its performance when developing a machine learning model (Thakar & Bhatt, 2024)

Thomas (2020) Jupyter Notebook: This is an interactive development environment for analyzing data, which enables researchers to write and run Python code (Tchio et al., 2024). Furthermore, libraries were used for optimizing the speed, improving performance, and taking advantage of hardware like GPUs (Graphics-Processing-Units) and TPUs (Tensor-Processing Units). The following are selected libraries:

Pandas: This is a useful tool for data cleaning, transformation, and analysis tasks by making it a fundamental tool in the Python ecosystem. Pandas offers multiple functions for data cleaning, such as handling missing values, merging datasets, filtering, and grouping data. Pandas also supported both label-based and position-based, which allowed for flexible and powerful data access and manipulation (Thomas, 2020).

Numpy: NumPy performed operations such as normalization, scaling, and transformation of features.

Lancaster (2023) scikit-learn: This was used for machine learning tasks. It helped to train the model and predict an outcome for unseen data.

nlTK: This was used for processing and analyzing data by breaking text into words, sentences, etc. It helped to identify entities like persons, names, and locations. It was also useful for model evaluation and text classification (Tianda et al., 2024).

3.4 Methods

Villela et al. (2023). As the study had given a data frame consisting of observations of articles in order to implement the machine learning model, which only understands numbers, the research had to figure out how to convert texts into specific mathematical terms. Tokenization of words extracts meaningful features for a specific use case (Tianda et al., 2024). Hence, tokenization was considered rather than sentences since certain words in an article might make the article stand out for containing misinformation. To end up, the vectorizer called Term Frequency-Inverse-Document-Frequency captures useful words in a document (Thakar & Bhatt, 2024).

However, before using or applying the vectorizer, the text was cleaned using regex and got rid of special characters that provide no special use case in the specific prediction (Vosoughi et al., 2018). The following steps outline how TF-IDF vectorization could be used:

Term Frequency (TF): it assesses the significance of terms within documents by calculating words' frequency of documents, and then computing their ratio to the overall number of terms (J. Wang et al., 2023).

Yan et al. (2024). Inverse document frequency (IDF): IDF measures a unique term across the entire materials. Computing IDF, the overall document count is divided by the number of documents in which the term appears. However, IDF assigns higher weights to terms that are less common in the corpus, as they are considered more informative (Thakar & Bhatt, 2024).

Tchio et al. (2024) TF-IDF Calculation: The process repeats for all terms in each document, resulting in a numerical representation of the document based on the importance of its terms relative to the entire corpus.

Villela et al. (2023) Vectorization: The vectorizer converts each record into a vector, where each element of the vector represents the TF-IDF score for a particular term in the document (Yan et al., 2024).

The mentioned techniques enable models to identify false information, and they help to filter out fake news content effectively (Mayo, 2018).

3.5 Development of the Model

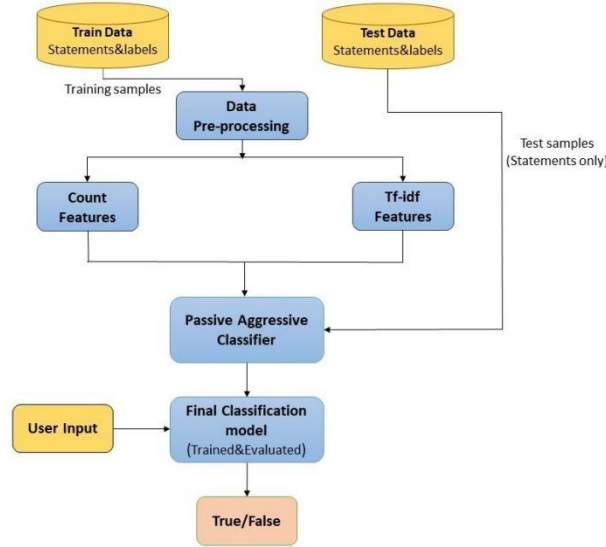
Alnuaimi and Albaldawi (2024) Building a model capable of distinguishing fake and real news, the research employed ML algorithms, including NLP (Yan et al., 2024). This technique was selected because it could provide and perform good results than Computational Social Science Algorithms such as Social Network Analysis (SNA), Temporal Analysis (TP), etc. Computational Social Science Algorithms analyze and understand human or social behaviors, while ML algorithms classify patterns and anomalies in text, images, etc. (W. Y. Wang, 2017) Therefore, NLP (Natural Language Processing) helped standardize and clean raw data and extract meaningful features from text to enable classification (Yan et al., 2024).

3.5.1 Model's Architecture

Taha et al. (2024) The approach was capable of detecting fake news for adequate actions. The proposed model used a relevant dataset that was up to date. The figure below illustrates the development of a model architecture (Vosoughi et al., 2018). This architecture outlines an ML pipeline for detecting fake news utilizing a Passive Aggressive Classifier W. Y. Wang (2017) The pipeline is designed for fake news, converting text into numerical features utilizing Count and TF-IDF, and training a Passive Aggressive Classifier for prediction. It supports real-time inputs from users (Villela et al., 2023).

Figure 3.2

Architecture for model development



For model development, the study focused on data collection, data processing, feature classification, training, and assessment (Taha et al., 2024).

3.6 Data collection and data source

The study utilized secondary data collected from the Kaggle website. <https://www.kaggle.com/clmentbisailon/fake-and-real-news-dataset>,

including various sources such as news articles, called fake news websites, and social media platforms (Yan et al., 2024). The model was developed using a large ISO Fake news dataset compiled from authentic sources. Genuine articles were gathered by scraping content from Reuters.com, while fabricated News items were sourced from unreliable sites identified by PolitiFact(a U.S. Factchecking body) and Wikipedia. Using existing labeled datasets like the LIAR dataset, Fake Newsnet, and supplementing with new data as well (Lancaster, 2023).

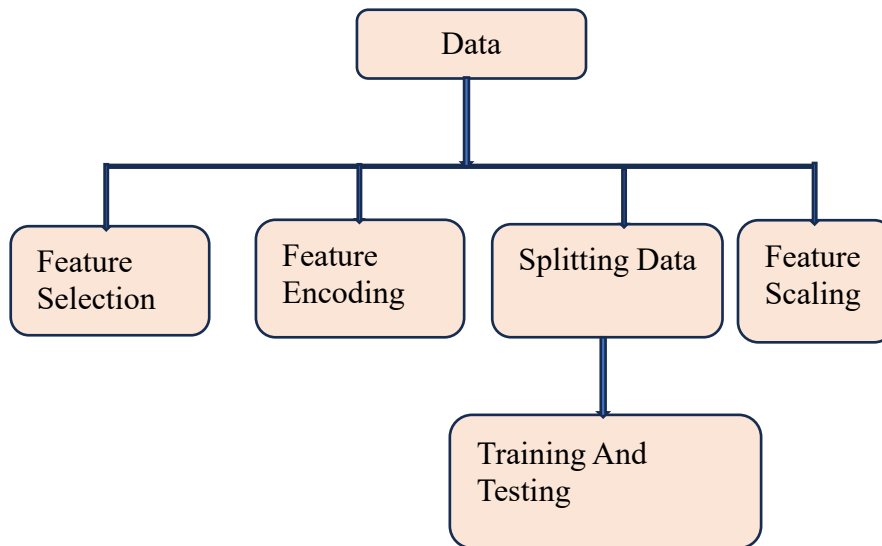
3.7 Data Preprocessing

Building an ML model in detecting fake news, preprocessing is an essential step that enables cleaning and preparing data, removing unwanted characters, noise, or any HTML tags, and handling missing values Moszczuk et al. (2021).

Raja and Raj (2022) Removing unwanted characters and noise, unnecessary punctuation like 'question mark(?), comma (,), Semicolon (;)', and special symbols such as @, #, etc., has been stripped out, which added no semantic value. Handling missing values enables to management of incomplete or missing information within the dataset (Chhajer et al., 2021). Missing data could cause problems in data analysis and machine learning models if not properly used. The procedure of processing data is illustrated as follows: feature selection, feature encoding, splitting data, and feature scaling (Su et al., 2024).

Figure 3.3

Data Preprocessing



3.7.1 Features Selection

Su et al. (2024) identified and selected relevant variables in a dataset that contribute to the improvement of a predictive model. Irrelevant variables or features could degrade a model's predictive performance. However, removing unnecessary features could reduce the risk of overfitting. This study employed the following ML algorithms: logistic regression (LG), random forest (RF), and decision tree (DT). Logistic regression was used for binary including various classes in classification tasks. Decision tree usually uses a tree structure, choosing by selecting features.

3.7.2 Feature Encoding

Encoding categorical variables or features is significant in data processing, as many machine learning models work with numerical data like categorical variables. The study used two techniques for feature encoding: TF-IDF and Word Embedding (Petkar & Sonawane, 2020).

TF-IDF is a statistical metric used to assess the significance of a word within a document within a larger collection of documents. It is applied by transforming textual data into a matrix of TF-IDF features. Meanwhile, word embeddings, such as word2Vec, can represent semantic relationships between words (Saha et al., 2023).

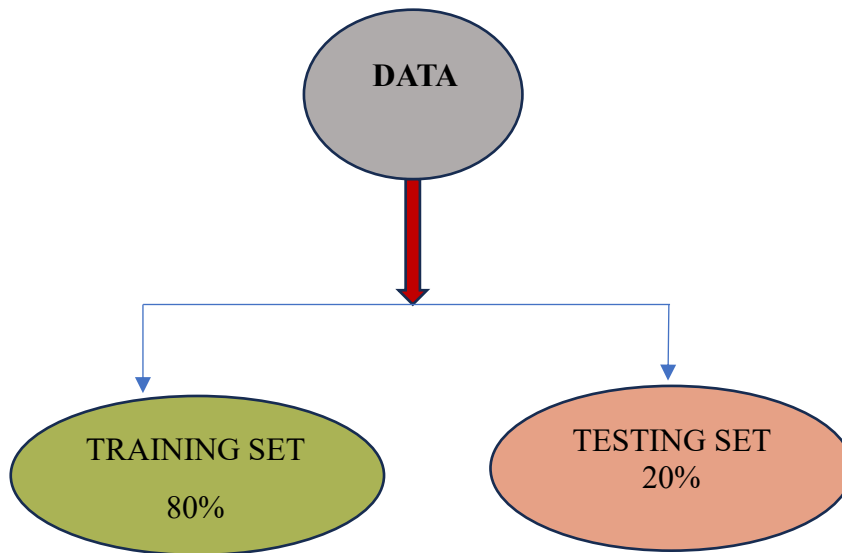
3.7.3 Splitting the data

Splitting of data involves dividing the dataset into subsets that serve specific purposes, like training and testing the model. This step was crucial, especially when evaluating model accuracy (Saha et al., 2023). Splitting data involves the following steps: training-set, testing-set. With the training Set, the model effectively adapts to new data by learning fundamental patterns of the problem (Millière, 2022).

Patel et al. (2021) In training, the model adjusts its internal parameters based on data, minimizing errors in prediction. The training-set comprised the greater part of the output, accounting for up to 80% of the entire dataset (Naeem et al., 2023). The testing set evaluated the result of a trained ML model. The test set made up to 20% within the full dataset. Therefore, percentages may vary based on the dataset size; however, they can be large enough to provide a robust estimate of the model's performance Pandey et al. (2022).

Figure 3.4

Splitting the data



3.6.4 Feature Scaling

In datasets, features have different units and ranges. For example, age may range from 0 to 100, etc. Therefore, unscaled data could lead to Biased Model Training, Inaccurate Distance Calculation, and convergence Issues. Biased Model Training occurs when a machine learning model learns patterns that reflect imbalances and unfair representations in the training data (Tandoc et al., 2017).

Mohsen et al. (2024) This could lead to predictions that favor certain groups. Bias in machine training could be introduced through various stages of the machine learning process. Inaccurate Distance Calculation could lead to poor performance in ML algorithms like nearest neighbors (KNNs) and K-means clustering. Convergence Issues are common in algorithms that rely on iterative optimization, like gradient descent-based (Monnier et al., 2023).

Li and Zhou (2020) For better feature scaling, the research utilized Normalization and data standardization. Normalization was used to scale data within a specific range, like $[0,1]$, $[-1,1]$. This helps make features comparable when they have different units and benefits from models that rely on distance-based metrics, such as Nearest Neighbors,

K-Means clustering. Data-standardization transforms a mean (0) and standard deviation (1). Moreover, research used Data Normalization to handle features with different units (Kumari & Singh, 2024).

3.7 Model evaluation

Utilizing metrics, tools: accuracy, precision, recall, and F1-score were considered by evaluating the proposed model's effectiveness (Kaliyar et al., 2021).

Accuracy

Accuracy was used for classification tasks (Jawad & Obaid, 2022). It is used by evaluating model predictions relative to the total number of predictions made with the following formula:

$$\text{Accuracy} = \frac{\text{True Positives} + \text{True Negatives}}{\text{Total Predictions}} \quad (1)$$

True-positive refers to cases where the actual including predicted labels are positive. In other words, they represent instances that the system accurately recognized the target class, while a true negative arises when both the actual and predicted labels are negative (Liu et al., 2021).

Precision

Ikotun et al. (2022) Precision also refers to the consistency of results. However, if multiple measurements of the same thing gave similar results. Precision is associated with how the method produces the same result under the same conditions. In other words, precision referred to the level of detail in numerical calculations, like the number of significant digits used in the value. The precision formula was as follows:

$$\text{Precision} = \frac{\text{True Positives}(TP)}{\text{True Positives}(TP) + \text{False Positives}(FP)} \quad (2)$$

From the formula above, true was correctly predicted over false positives, which incorrectly identify positive cases. However, a researcher could avoid classifying legitimate emails as spam, which results in false positives.

Recall

Recall, also known as the true positive rate (TPR). Utilizing a recall metric, the classification model assesses its ability to correctly identify relevant cases (Zahednejad et al., 2020).

$$\text{Recall} = \frac{\text{True Positives}(TP)}{\text{True Positives}(TP) + \text{False Negatives}(FN)} \quad (3)$$

True positives denote the number of instances that are genuinely positive, that are correctly identified as positive. In contrast, false negatives represent the number of positive instances that are mistakenly classified as negative (Alnuaimi & Albaldawi, 2024).

F1 - Score

Alghasi et al. (2024). The F1-score is a metric used to assess the accuracy of a classification model, particularly in cases of class imbalance. F1 Score helped guarantee consideration of both the positive class and negative classes (Arowolo et al., 2023).

$$\text{F1 - Score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (4)$$

In cases where the classes are imbalanced, the F1 Score balances the need to avoid false positives and false negatives (Alghasi et al., 2024).

Summary

Alnuaimi and Albaldawi (2024) Building an ML-based model involves creating and refining algorithms to identify misinformation in digital content. The process begins by creating a dataset containing labeled fake and real news articles, followed by data processing, cleaning, and standardizing the text (Harris et al., 2024). The model can

leverage various NLP techniques to extract features like readability and structural patterns that help identify fake news. Different machines Learning are assessed in determining the most effective cases. By enhancing the quality of the model, evaluation metrics: recall, precision, accuracy, and F1- score are also utilized (Hu et al., 2024).

CHAPTER FOUR

DISCUSSIONS AND RESULTS

Chapter four showcases the outputs yielded by the proposed ML-based model in detecting fake news, assesses its effectiveness using parameters, and analyzes the findings concerning the research objectives.

4.1 Development of Fake News Detection Model

4.1.1 Dataset preparation

The dataset employed in this study consisted of an equal representation of false and real news articles, followed by comprehensive cleaning and processing. The final dataset consists of 44,898 records. Comprising 23,481 falsified articles and 21,417 verified articles (Gong et al., 2023). Preprocessing plays a crucial role in ensuring that data is suitable for ML.

Cavus et al. (2024). Initially, punctuation marks and stopwords are removed to eliminate irrelevant noise. Text data was then transformed to lowercase to ensure consistency and prevent potential discrepancies. Finally, the text after cleaning was transformed into numerical representations utilizing the TF-IDF method. Data cleaning and preparation were a crucial step in this study. It prepared a dataset containing both fake news and real news articles. Research used a dataset provided by the United States of America (USA) organization (Gong et al., 2023). This preprocessed dataset facilitates the pairwise classification tasks for news articles, as shown in Table 4.1. Sravani and Kumar (2025) Language Processing (NLP) is a library used for working with text and including Tokenization, which splits text into words and phrases.

Table 4.1 *The*

ISOT Fake News Dataset: Initial data

Title	Subject	Date	Target
-------	---------	------	--------

O'Reilly And Trump News Bond	April 9, 2017	Fake
Russia detains a Norwegian	May 18, 2017	Fake
Insane! Huge Financial World news	December 19, 2017	True
Maxine doubles down Government news	October 13, 2015	Fake
House Democrat calls Politics News	November 7, 2017	True

After the dates are removed from the data frame, the dataset includes or classifies only the article titles, main article text, categories, and indicates whether the articles were fake or true (Choudhary & Arora, 2020).

Table 4.2

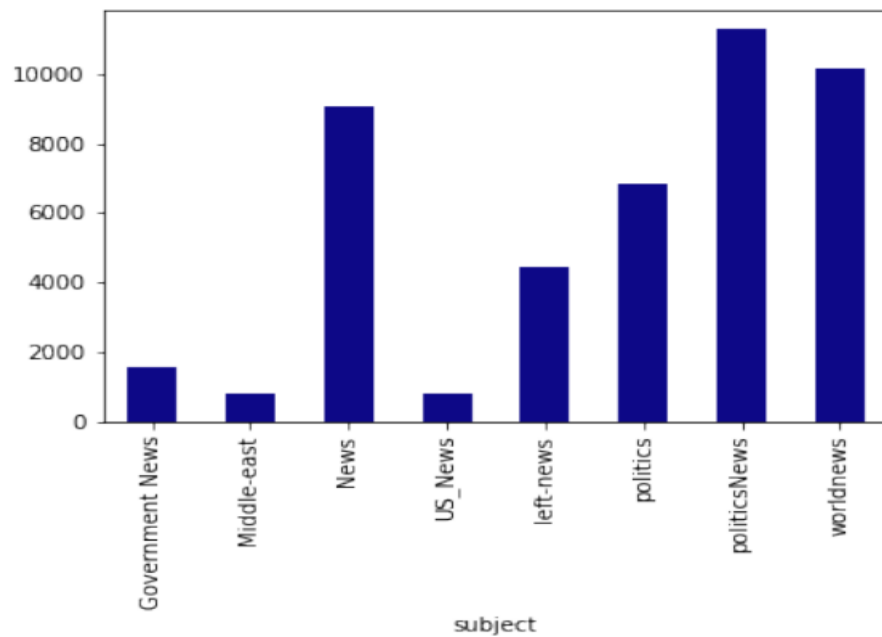
ISOT Fake News Dataset: Results of the Data

Title	Text	Subject	Target
O'Reilly And Trump Bond	We all love seeing Alec Baldwin portray...	News	Fake
Russia detains a Norwegian	The funniest part of the video comes...	Politics	Fake
Insane! Huge Financial	MOSCOW (Reuters) - Russia has detained...	World news	True
Maxine doubles down	I'll bet most Americans have no idea...	Government news	Fake
House Democrat calls	WASHINGTON (Reuters): The top Democrat...	Politics News	True

Table 4.2 below represents a summary of text data categorized by subject, with the counts showing the number of articles per subject.

Table 4.3***The ISOT Fake News Dataset: Number of articles per subject***

Subject	Count	Percentage (%)
Government.News	1,570	3.50
Middle.East	778	1.73
News	9,050	20.16
US.News	783	1.74
LeftNews	4,459	9.93
politics	6,841	15.24
Political updates	11,272	25.11
global news	10,145	22.60

Figure. 4.1 The ISOT Dataset.: Bar chart showing count of articles per subject

Moreover, the spread of target labels within a dataset represents the classification label for the text data (Choudhary & Arora, 2020) The target indicates the number of entries in the dataset that are classified as True and False, as shown in Table 4.4.

Table 4.4

The ISOT Dataset: The count and percentage for each target

Target	Count	Percentage%
Fake	23,481	52.3%
True	21,417	47.7%

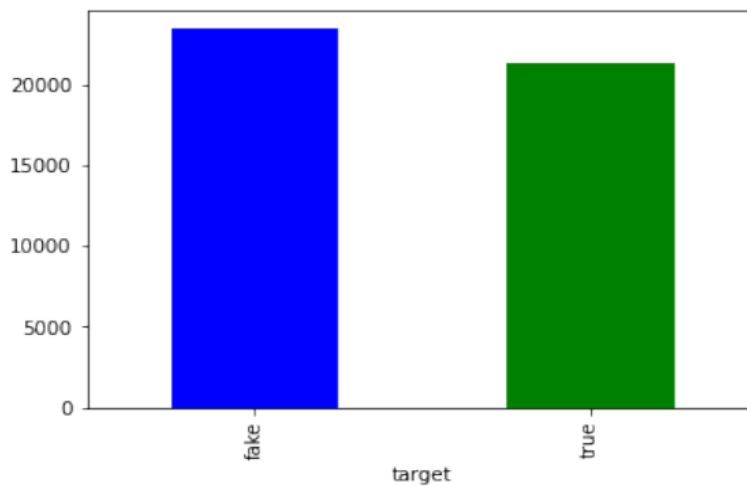
The percentage for each Fake and True target was calculated as follows:

$$\text{Fake} = \frac{23,481}{23,481+21,417} \times 100 = 52.3\% \quad (5)$$

$$\text{True} = \frac{21,417}{23,481+21,417} \times 100 = 47.7\%$$

Figure 4.2

The ISOT Fake News Dataset.: Bar chart showing the number of each target



Ghaddar and Naoum-Sawaya (2017) Natural-Language Processing, calculate the Frequency distribution of tokens (words and phrases). A bar chart 10 and 12 displays and shows the frequency of articles and counts.

Data Preparation: World cloud for fake news



Word	Count
trump	75000
said	31000
people	26000
president	25500
would	23500
one	23000
us	22000
clinton	18000
drama	17500
like	17500
donald	17000
also	15500
new	14500
news	14500
even	13500
hillary	13500
white	12500
time	12500
state	12500
via	11500

Data Preparation: World cloud for real news



The development of the model involves constructing pipelines that integrate text vectorization and classification (Ghaddar & Naoum-Sawaya, 2017) TF-IDF was employed for vectorizing the text data, and ML classifiers (logistic regression, decision tree, and random forest) were implemented and compared.

The dataset was divided into training and testing subsets, with 80% allocated for training and 20% set aside for testing. This division allowed a model to be trained on a significant portion of the data, while keeping a separate set for impartial assessment. The ISOT fake news dataset was thus separated into a training Set and a Testing Set. Tables 4.5 and 4.6 show the summary of the training Set of 80% with four samples and the testing Set of 20% with one sample of the data.

Table 4.5

A training Set (80% of the data)

Index	Text	Label	Target
0	O'Reilly And Trump Bond Over Sexual Harassment	0	Fake
1	Russia detains a Norwegian citizen on suspicion.	0	Fake
2	House Democrat calls GOP Clinton probe scheme	0	Fake
3	We all love seeing Alec Baldwin portray Donald Trump	1	True

Table 4.6

Testing Set (20% of the data)

Index	Text	Label	Target
1	MOSCOW (Reuters) - Russia has detained a Norwegian citizen	1	True

4.2 Evaluation Metrics

The efficiency of each model is analyzed utilizing standard evaluation parameters Pandey et al. (2022) Accuracy is computed to determine the fraction of correctly classified records. Precision measured true positives, while recall gauged the effectiveness of the task in identifying confirmed positive cases. F1-Score provided a balance among them through the harmonic mean, and then offers a balanced evaluation of the model's robustness, especially in cases where the impact of false negatives and false positives are significant (Patel et al., 2021).

Accuracy

Alnuaimi and Albaldawi (2024) Accuracy is the primary evaluation metric for classification tasks. Helping to assess the proportion of correct predictions made by a model out of all predictions. The following formula for accuracy was as follows:

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN} \quad (1)$$

The values of TP, TN, FP, and FN are determined from the confusion matrix. The obtained accuracy of 99.67% indicates the proportion of total correct predictions. When compared with similar studies, for instance, Raja and Raj (2022) achieved 99.43% using Decision Tree, Logistic regression 98.64%, Random Forest 98.75%, the proposed model demonstrates competitive performance. These differences could be attributed to variations in the ISO Fake News dataset quality.

True positives (TP) occurred when both the actual and predicted labels were positive, representing the instances where the model correctly identified the positive class. On the other hand, true negatives refer to the instances where the model correctly predicted the negative class (Moszczuk et al., 2021).

Precision

Precision emphasizes the accuracy of positive prediction. It evaluates how many of the instances predicted as positive belong to the positive class. Precision also refers to the consistency of results, indicated by multiple measurements of the same thing giving similar results. Precision is associated with how the method produces the same result under the same conditions. In other words, precision referred to the level of detail in

numerical calculations, like the number of significant digits used in the value (Abdelillah et al., 2023). The Precision formula is as follows:

$$\text{precision} = \frac{TP}{TP + FP} \quad (2)$$

Where TP refers to the number of true positive predictions, and FP denotes the number of false positives (Abdelillah et al., 2023). Precision represents the proportion of correct positive predictions. A high prediction value indicates that the model is effective in minimizing false alarms, which is essential in applications where false positives carry significant consequences. In comparison, Alghasi et al. (2024) achieved high classification performance in their fakeNews detection study, using LG and TF-IDF preprocessing, and reporting an overall accuracy of 99.64%. While precision was not separately reported in that study, high accuracy suggests correspondingly strong precision metrics, for instance, if class imbalance was minimal. The precision obtained in our study complements the accuracy metric and confirms the model's capability in identifying relevant instances and minimizing incorrect positive classifications (Arowolo et al., 2023).

Abdelillah et al. (2023). From the formula above, True positives are instances correctly predicted as positive, while False positives are those incorrectly predicted as positive. This metric became crucial when the cost of false positives was significant, such as in email spam detection. A researcher could avoid classifying legitimate emails as spam, which results in false positives.

Recall

Recall, referred to as the true positive rate, is a performance metric used to assess the effectiveness of a classification model, including its capability to identify all the relevant cases (Kumar & Mindzak, 2024).

$$\text{Recall} = \frac{TP}{(TP + FN)} \quad (3)$$

Abdelillah et al. (2023). True positives refer to the number of positive-instances correctly identified as positive, whereas false-negatives represent the number of positive instances that are incorrectly classified as negative. True positives (TP) denote true positives, and false negatives. Recall measures the model's ability to correctly identify all actual positive instances (Kuruvilla & Kundapura, 2022). A high recall indicates that the model is effective at minimizing false negatives, which is crucial in domains. The recall also achieved 99.67%, indicating that the model was able to correctly identify a portion of the actual fake news articles. When compared with a study by Jain and Kasbe (2018), who obtained 99.43% utilizing TF-IDF including logistic regression, the recall performance for the proposed model shows comparable effectiveness.

F1- score

Kuruvilla and Kundapura (2022) F1-score provides a balanced average of precision and recall. F1- Score is a numeric used to assess the performance of a classification model, particularly when there is an imbalance between classes. F1 Score helped to ensure that both positive and negative classes are considered (Yan et al., 2024).

$$F1 - Score = 2 * \frac{Precision*recall}{precision+recall} \quad (4)$$

In this study, the F1-score was computed to assess the balance between precision and recall. The obtained F1-score of 99.67% reflects that the model achieves both high precision and high recall, indicating robust and balanced performance in detecting fake news. When compared to the work of Mugdha et al. (2020) at 87% of accuracy, the inclusion of the F1-score in the evaluation provides a more comprehensive view of model performance, especially in binary classification problems where class imbalance could skew accuracy metrics.

Evaluation Metrics Table

The evaluation metrics Table presents a comparative evaluation of three supervised ML algorithms, such as Logistic Regression (LG) at 99.13%, Decision Tree (DT) at 99.67%, and Random Forest (RF) at 99.15% based on their performance in detecting fake news. The assessment includes key classification metrics: Accuracy, Precision,

Recall, and F1-Score, all of which provide insight into the model’s overall effectiveness and balance between correctly identifying fake news and minimizing misclassifications (Naeem et al., 2023). Therefore, the table includes raw classification outcomes such as True Positives (TP) and True Negatives (TN), False Positives (FP), and False Negatives (FN), which underpin the metrics calculations (Gong et al., 2023).

Table 4.7 Evaluation metrics

Evaluation Metrics								
MODEL	ACCURAC	PRECISIO	RECAL	F1SCOR	TP	TN	F	F
	Y	N	L	E			P	N
Logistic	99.13%	99.13%	99.13%	99.13%	423	466	4	33
Regressi					8	4	5	
on								
Decision	99.67%	99.67%	99.67%	99.67%	424	470	8	22
Tree					9	1		
Random	99.15%	99.15%	99.15%	99.15%	424	466	4	31
Forest					0	4	5	

The evaluation metrics presented in Table 4.7 offer insight into the outcome of the three ML algorithms, LG, DT, RF in the task of fake news detection. The metrics include accuracy, precision, recall, and F1-score, each of which was supported by the confusion matrix elements such as True Positives (TP), True Negatives (TN), False Positives (FP), and False Negatives (FN). Logistic regression attained accuracy at 99.13%, while Sharma et al. (2021) achieved 91.5% utilizing BiLSTM with precisions, recalls, and F1-score all reaching 99.13%. The confusion matrix showed that the model successfully identified 4,238 fake-news articles and 4,664 real-news articles, with 45 false positives and 33 false negatives.

Saha et al. (2023) Although effective, its linear nature constrained its capacity to identify more complex patterns in the data. A confusion matrix using a heatmap illustrated in Figure 3 shows a clear representation of the classification of a model’s performance. The Decision tree model proved to be a top-performing classifier,

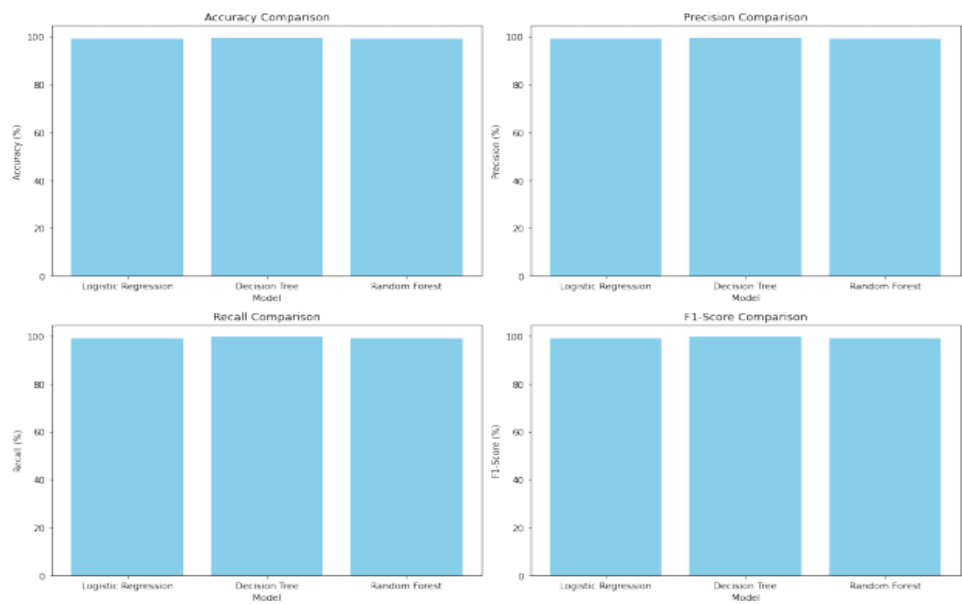
reaching an accuracy of 99.67%. It also demonstrated precision, recall, and F1-score values of 99.67% across the board. It correctly classified 4,249 fake news articles and 4,701 real news articles, with only 8 false positives and 22 false negatives, achieving an accuracy of 87%, as reported by Mugdha et al. (2020). Hierarchical patterns in text data accounted for its superior performance (Rolf et al., 2022). Furthermore, Random Forest demonstrated high accuracy at 99.15%. It accurately classified 4,240 fake news articles, including 4,664 real news articles, while producing 45 false positives and 31 false negatives. Although its ensemble nature improved robustness, it required more computational resources than the DT model (Nakov et al., 2021).

4.3 Model comparison

Figure 4.7 demonstrates four bar graphs illustrating the comparison of performance metrics (accuracy, precision, recall, and F1-score) across each model.

Figure 4.7

Models’ Comparison Table



A proposed model achieved an effective Accuracy of 99.67%, compared to (Taha et al., 2024) at 98.5%, RF at 98.9% and DT at 99.4%. Raja and Raj (2022) outperformed 99.43% DT, RF at 98.75%, and LG at 98.64%. Oni et al. (2024) achieved 99.64% accuracy, and Mugdha et al. (2020) at 87%.

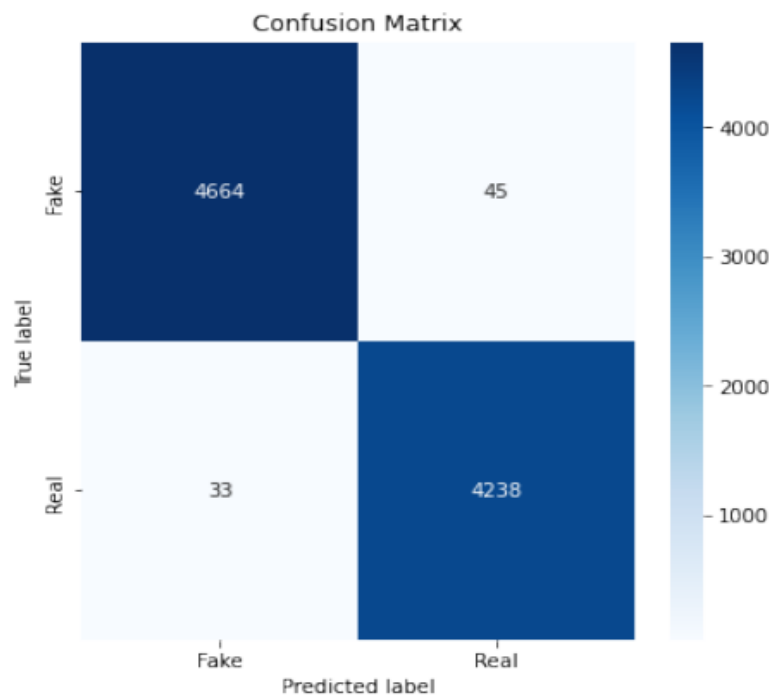
4.4 Results of the Proposed Model

4.3.1 Logistic Regression's performance

Logistic Regression attained accuracies at 99.13%, with precisions, recalls, and F1-Score all reaching 99.13%. According to the confusion matrix, the model accurately classified 4,238 fake news articles and 4,664 real news articles, while producing 45 false positives and 33 false negatives. Although effective, its linear nature constrained its capacity to identify more complex patterns in the data (Aslam et al., 2021). A confusion matrix using a heatmap illustrated in Figure 13 shows a clear representation of the classification of the model's outcome.

Figure. 4.8

A confusion matrix for the logistic regression model's classification



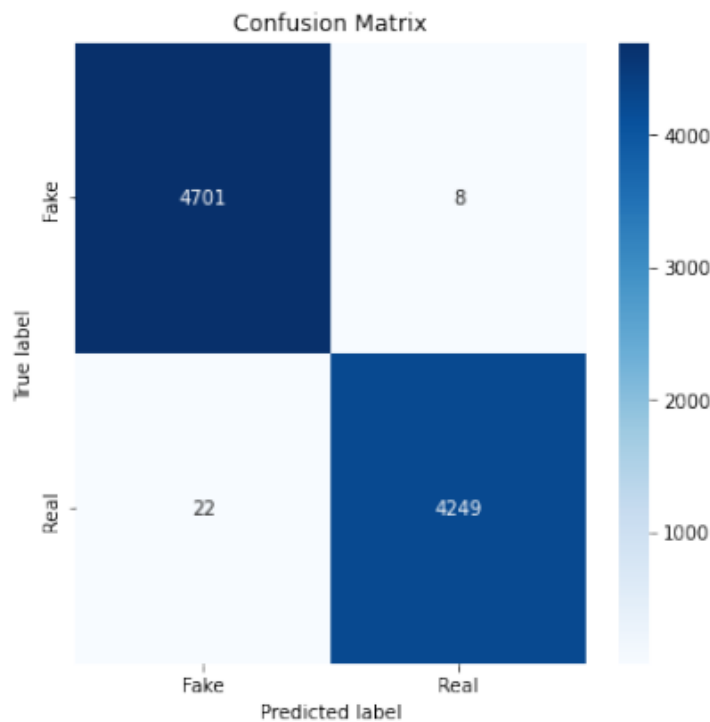
4.3.2 .Performance of the decision tree

Suthaharan (2015) The decision tree model proves to be a top-performing classifier. It is reaching an accuracy of 99.67%. It also demonstrated precision, recall, and F1-score values of 99.67% across the board. It correctly classified 4,249 fake news articles and

4,701 real news articles, with only 8 false positives and 22 false negatives. Its ability to handle nonlinear relationships and hierarchical patterns in text data accounted for its superior performance (Aslam et al., 2021).

Figure 4.9

A confusion matrix for a decision-tree classifier

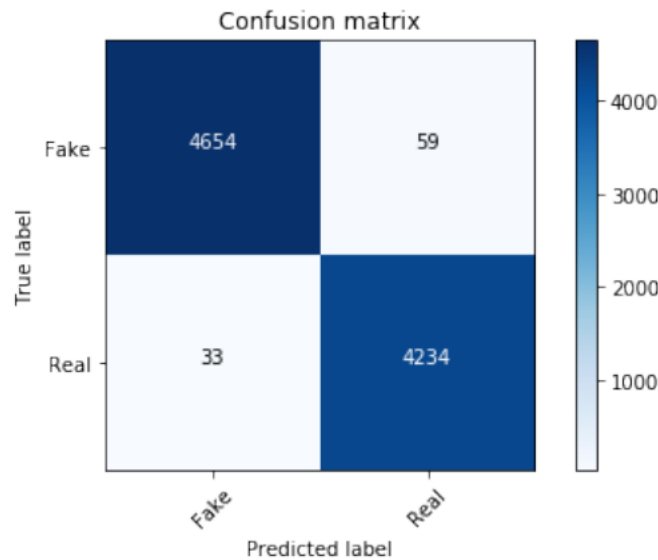


4.3.3 Random forests performance

Random forest demonstrated high accuracy at 99.15%. It correctly identified 4,240 fake-news articles, including 4,664 real-news articles, while yielding 45 false positives and 31 false negatives. Although its ensemble nature improved robustness, it required more computational resources than the DT model (Hu et al., 2024).

Figure 4.10

A confusion matrix for a random forest classifier



4.4 Discussion

4.5.1 Objective 1: Developing an ML-based model

Aslam et al. (2021) Developing a fake news detection model relied on NLP techniques to transforming raw text data into meaningful numerical values. Among the three classifiers evaluated, the DT model stood out due to its higher accuracy and efficiency. Unlike Logistic Regression, which is limited by its linear assumptions, the Decision Tree effectively captured nuanced and nonlinear patterns in the text. Compared to Random Forest, the Decision Tree offered similar levels of accuracy while being computationally more efficient, making it an ideal choice for practical deployment (Janiesch et al., 2021).

4.5.2 Objective 2: Analyzing Approaches for Content Verification

Lim and Perrault (2023) The analysis of content verification approaches underscored the importance of feature extraction techniques. T.F-IDF is an effective technique highlighting the relative importance of words, enabling the model to focus on distinctive terms and phrases indicative of fake news. This method increased the model's classification accuracy by improving the quality of input features.

4.5.3 Objective 3: Identifying Techniques for Content Verification

Techniques such as word clouds and frequency analysis played a significant role in identifying distinguishing patterns (Liu et al., 2021). For instance, fabricated news reports were characterized by sensationalist language and emotionally charged terms, whereas real news articles tended to use formal and factual language. Preprocessing steps like removing stop words and punctuation further refined the dataset by eliminating extraneous elements that could have misled the proposed model.

4.5.4 Objective 4: Evaluating the model's performance

The Decision Tree performed outstanding in all evaluation metrics, with accuracy, recall, precision, and an F1-score all reaching 99.67%. The confusion matrix displaying very few false positives and false negatives, further confirmed the model's reliability in distinguishing fake news from real news. Whereas Logistic Regression and Random Forest, while highly accurate, exhibited slightly higher false negative rates, which could reduce their effectiveness in applications (Mayo, 2018).

The high accuracy, precision, recall, and F1-score obtained in this study confirm that the proposed machine learning model effectively captures the relevant outputs. From a theoretical perspective, NLP methods based on preprocessing techniques such as tokenization, stopword removal, and feature vectorization transform unstructured text into numerical representations. The machine learning classifiers employed Decision Tree, Random Forest, and Logistic Regression. Each has theoretical strengths that have been observed in quantitative results. Decision trees and Random Forests are grounded in recursive partitioning theory, enabling them to model non-linear relationships and interaction effects.

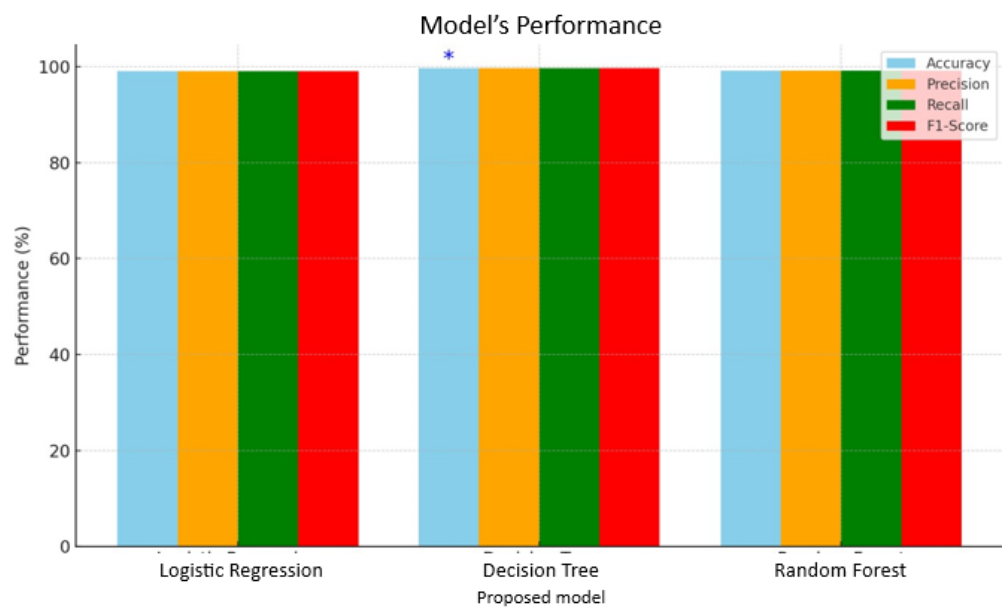
4.5 Model's performance

Figure 4.11 presents a visual comparison of the effectiveness of the three ML models, LG, DT, and RF, used for fake news detection. The performance metrics illustrated include Accuracy, Precision, Recall, and F1-Score, represented by distinct color-coded bars (Liu et al., 2021). As illustrated(Figure 4.11), the Decision Tree model labelled shows exceptional results on every evaluation criterion, closely followed by Random Forest and Logistic Regression. This graphical representation reinforces the tabular

results discussed earlier (see Table 4.7), where the DT model achieved the highest accuracy of 99.67% and exhibited the smallest rates of False-positives and False-negatives. The near identical height of the bars across metrics for each model indicates consistency in performance; however, the slight elevation in the Decision Tree’s bars highlights its edge in classification effectiveness. The figure provides an intuitive and immediate comparison that supports the recommendation of the Decision Tree as the most efficient and reliable model among them (Liu et al., 2021).

Figure 4.11

Model’s performance



4.6 Implications of the Findings

The evidence produced by this study emphasizes the effectiveness of utilizing NLP and ML models in identifying fake news. Decision Tree, with its superior performance and efficiency, serves as a reliable and scalable solution for combating misinformation. By optimizing feature extraction and leveraging advanced ML techniques, this research establishes a foundation for building a robust model. Future studies could probe the assimilation of methods(deep learning models) to enhance detection abilities, and scalability for real-world applications (Suthaharan, 2015).

4.7 Summary

This research sought to build a ML model for false news detection utilizing Natural Language-Processing (NLP) methods. It conducted an in-depth examination of different strategies and techniques for content verification to enhance reliability in recognizing misleading information. The proposed model's outcome was evaluated using standard metrics (accuracy, precision, recall, and F1-score), offering a comprehensive assessment of its effectiveness.

CHAPTER FIVE

CONCLUSION AND RECOMMENDATIONS

5.1 . Introduction

In this research, we discussed various techniques used for Fakenews detection, including data cleaning and preparation, basic data exploration, word cloud generation, most frequent words counter, and modeling utilizing Logistic Regression. The goal was to create a model classifying fake news (Mohsen et al., 2024). The initial steps involved, as mentioned above, cleaning and preparing the dataset. This included removing the date and title columns, transforming texts to lowercase, eliminating punctuation, and removing stop words (Ealmandouh et al., 2024).

5.2 Conclusion

A basic data exploration was conducted to understand the Distribution Of Articles per subject and the total number of articles (Cavus et al., 2024). To gain more insights into the words used in fake and real news, word clouds were generated. For instance, from the dataset, the word clouds revealed that words like ‘love’, ‘funniest’, and ‘bet’ were more common in fake news, while words such as ‘Moscow’, ‘Reuters’, and ‘Russia’ were more common in real news. Moreover, the next step was the development of a model utilizing three methods LG, DT, and RF.

Classification report indicated that the model carried out both false reports and genuine news, achieving 99.67% (Accuracy, Precision, Recall, and F1-score). A confusion matrix was created to demonstrate the model’s performance. It indicated that the model classified most of the real News as true and most of the false news as fake. In addition to Logistic Regression, a Decision tree was implemented in comparison with its performance, including logistic regression. When comparing the three classifiers, both models performed extremely well. However, the Decision tree classifier was more interpretable due to its robustness (Chiluwa & Samoilenko, 2019)

5.3 Recommendations

Based on the study's results, recommendations could be made to enhance the effectiveness of an ML model in detecting fake news:

Prospective researchers, engineers, and other professionals engaged in building ML systems for anomaly detection (including fake articles) should first seek diverse data sources and extend the dataset (Daud et al., 2023). This includes fake news from diverse languages instead of focusing only on one language (English) by evaluating the performance of contexts. Second, utilizing deep-learning techniques such as an RNN (recurrent neural network), CNN (convolutional neural networks), and Transformers, such as BERT and ROBERT, can improve contextual understanding. Third, establish benchmarks with day-to-day world case studies by assessing a model successfully (Cavus et al., 2024).

5.4 Future work

Enhancing capabilities, future work, or research should involve:

Aslam et al. (2021) Handling multi-class classification. By handling multi-class classification, organizations can better detect and sort different content categories, helping in mitigating the spread of misinformation and enhancing public understanding of accurate information. The current dataset "ISOT Fake News" encompasses two categories: false news and truthful news. Future datasets may contain more categories, requiring a more sophisticated modeling approach.

Ghaddar and Naoum-Sawaya (2017) Addressing imbalanced classes. Imbalanced data facilitates the development of robust systems capable of detecting rare but critical events. The dataset used in this research was relatively balanced, but future datasets may have imbalanced classes requiring techniques like oversampling minority classes. Future work could involve incorporating additional features like sentiment analysis, for instance, enabling classification of text such as positive, negative, and neutral (Felicilda et al., 2024).

References

- Abdelillah, F. M., Nora, H., Samir, O., & Sidi-Mohammed, B. (2023). Hybrid Data-Driven and Knowledge-Based Predictive Maintenance Framework in the context of Industry 4.0. In *Lecture Notes in Computer Science* (pp. 319–337). https://doi.org/10.1007/978-3-031-49333-1_23
- Alghasi, S., Vanvik, E. C., Barland, J., Falkheimer, J., Dahlen, Ø. P., Diers-Lawson, A., Fredheim, N. A. G., Jensen, H. H., Johannessen, M., Kalsnes, B., Karlsen, F., Kavtaradze, L., Krumsvik, A. H., Nilsen, A., Pettersen, L., Sandstad, B. B., Stampe, P. L., Valaker, S., & Viken, T. M. (2024). *Strategic Communication – Contemporary perspectives*. <https://doi.org/10.23865/noasp.208>
- Ali, I., Ayub, M. N. B., Shivakumara, P., & Noor, N. F. B. M. (2022). Fake News Detection Techniques on social media: A survey. *Wireless Communications and Mobile Computing*, 2022, 1–17. <https://doi.org/10.1155/2022/6072084>
- Alnuaimi, A. F., & Albaldawi, T. H. (2024). An overview of machine learning classification techniques. *BIO Web of Conferences*, 97, 00133. <https://doi.org/10.1051/bioconf/20249700133>
- Arowolo, M. O., Misra, S., & Ogundokun, R. O. (2023). A machine learning technique for the detection of social media fake news. *International Journal on Semantic Web and Information Systems*, 19(1), 1–25. <https://doi.org/10.4018/ijswis.326120>
- Aslam, N., Khan, I. U., Alotaibi, F. S., Aldaej, L. A., & Aldubaikil, A. K. (2021). Fake Detect: A deep learning ensemble model for fake news detection. *Complexity*, 2021(1). <https://doi.org/10.1155/2021/5557784>
- Cavus, N., Goksu, M., & Oktekin, B. (2024). Real-time fake news detection in online social networks: FANDC Cloud-based system. *Scientific Reports*, 14(1). <https://doi.org/10.1038/s41598-024-76102-9>
- Chhajer, P., Shah, M., & Kshirsagar, A. (2021). The applications of artificial neural networks, support vector machines, and long-short term memory for stock market

- Chiluwa, I. E., & Samoilenko, S. A. (Eds.). (2019). *Handbook of research on deception, fake news, and misinformation online*. IGI Global. <https://doi.org/10.4018/978-1-5225-8535-0>
- Choudhary, A., & Arora, A. (2020). Linguistic feature-based learning model for fake news detection and classification. *Expert Systems with Applications*, 169, 114171. <https://doi.org/10.1016/j.eswa.2020.114171>
- Daud, S., Ullah, M., Rehman, A., Saba, T., Damaševičius, R., & Sattar, A. (2023). Topic Classification of online news articles using optimized machine learning models. *Computers*, 12(1), 16. <https://doi.org/10.3390/computers12010016>
- Dierickx, L., Lindén, C., & Opdahl, A. L. (2023, August 15). *Automated Fact-Checking to support professional practices: systematic literature review and Meta-Analysis*. <https://ojs3.ijoc.org/index.php/ijoc/article/view/21071>
- EAlmandouh, M., Alrahmawy, M. F., Eisa, M., Elhoseny, M., & Tolba, A. S. (2024). Ensemble-based high-performance deep learning models for fake news detection. *Scientific Reports*, 14(1). <https://doi.org/10.1038/s41598-024-76286-0>
- Elkhatat, A. M., Elsaid, K., & Almeer, S. (2023). Evaluating the efficacy of AI content detection tools in differentiating between human and AI-generated text. *International Journal for Educational Integrity*, 19(1). <https://doi.org/10.1007/s40979-023-00140-5>
- Felicilda, N. M. J. R., Geriane, N. a. J. B., Agustin, N. V. A., Blanco, N. M. C. R., Morano, N. J. C., Mahusay, N. L. M., & Guialil, N. J. S. (2024). Enhancement of random forest algorithm applied in fake news detection. *World Journal of Advanced Research and Reviews*, 22(2), 1075–1079. <https://doi.org/10.30574/wjarr.2024.22.2.1439>
- Festré, A. (2020, December 2). *Michael Polanyi's vision of economics: Spanning Hayek and Keynes*. <https://shs.hal.science/halshs-03036824/>

- Ghaddar, B., & Naoum-Sawaya, J. (2017). High dimensional data classification and feature selection using support vector machines. *European Journal of Operational Research*, 265(3), 993–1004. <https://doi.org/10.1016/j.ejor.2017.08.040>
- Gong, Y., Liu, G., Xue, Y., Li, R., & Meng, L. (2023). A survey on dataset quality in machine learning. *Information and Software Technology*, 162, 107268. <https://doi.org/10.1016/j.infsof.2023.107268>
- Hamed, S. K., Aziz, M. J. A., & Yaakub, M. R. (2023). A review of fake news detection approaches: A critical analysis of relevant studies and highlighting key challenges associated with the dataset, feature representation, and data fusion. *Heliyon*, 9(10), e20382. <https://doi.org/10.1016/j.heliyon.2023.e20382>
- Harris, S., Hadi, H. J., Ahmad, N., & Alshara, M. A. (2024). Fake news detection revisited: an extensive review of theoretical frameworks, dataset assessments, model constraints, and Forward-Looking research agendas. *Technologies*, 12(11), 222. <https://doi.org/10.3390/technologies12110222>
- Hu, B., Mao, Z., & Zhang, Y. (2024). An overview of fake news detection: From a new perspective. *Fundamental Research*. <https://doi.org/10.1016/j.fmre.2024.01.017>
- Ikotun, A. M., Ezugwu, A. E., Abualigah, L., Abuhaija, B., & Heming, J. (2022). K-means clustering algorithms: A comprehensive review, variants analysis, and advances in the era of big data. *Information Sciences*, 622, 178–210. <https://doi.org/10.1016/j.ins.2022.11.139>
- Jain, A., & Kasbe, A. (2018). Fake news detection. *2020 IEEE International Students' Conference on Electrical, Electronics and Computer Science (SCEECS)*. <https://doi.org/10.1109/sceecs.2018.8546944>
- Janiesch, C., Zschech, P., & Heinrich, K. (2021). Machine learning and deep learning. *Electronic Markets*, 31(3), 685–695. <https://doi.org/10.1007/s12525-021-00475-2>

- Jawad, Z. A., & Obaid, A. J. (2022). Combination of convolution neural networks and deep neural networks for fake news detection. *arXiv (Cornell University)*. <https://doi.org/10.48550/arxiv.2210.08331>
- Kaliyar, R. K., Goswami, A., & Narang, P. (2021). EchoFakeD: improving fake news detection in social media with an efficient deep neural network. *Neural Computing and Applications*, 33(14), 8597–8613. <https://doi.org/10.1007/s00521-020-05611-1>
- Kumar, R., & Mindzak, M. (2024). Who wrote this? *Canadian Perspectives on Academic Integrity*, 7(1). <https://doi.org/10.55016/ojs/cpai.v7i1.77675>
- Kumari, S., & Singh, M. P. (2024). A deep learning multimodal framework for fake news detection. *Engineering Technology & Applied Science Research*, 14(5), 16527–16533. <https://doi.org/10.48084/etasr.8170>
- Kuruvilla, E., & Kundapura, S. (2022). Performance Comparison of Machine Learning Algorithms in Groundwater Potability Prediction. *International Conference on Recent Advances and Innovations in Engineering*, 53–58. <https://doi.org/10.1109/icraie56454.2022.10054298>
- Lancaster, T. (2023). Artificial intelligence, text generation tools and ChatGPT – does digital watermarking offer a solution? *International Journal for Educational Integrity*, 19(1). <https://doi.org/10.1007/s40979-023-00131-6>
- Li, Q., & Zhou, W. (2020). Connecting the dots between fact verification and fake news detection. *arXiv (Cornell University)*. <https://doi.org/10.48550/arxiv.2010.05202>
- Lim, G., & Perrault, S. T. (2023). Fact-checking Chatbot: a misinformation intervention for instant messaging apps and an analysis of trust in the fact-checkers. In *Mobile communication in Asia* (pp. 197–224). https://doi.org/10.1007/978-94-024-2225-2_11
- Liu, L., Roche, D. S., Theriault, A., & Yerukhimovich, A. (2021). Fighting Fake News in Encrypted Messaging with the Fuzzy Anonymous Complaint Tally System (FACTS). *arXiv (Cornell University)*. <https://doi.org/10.48550/arxiv.2109.04559>

- Májovský, M., Černý, M., Kasal, M., Komarc, M., & Netuka, D. (2023). Artificial intelligence can generate fraudulent but Authentic-Looking scientific medical articles: Pandora's box has been opened. *Journal of Medical Internet Research*, 25, e46924. <https://doi.org/10.2196/46924>
- Martin, Z., Trauthig, I. K., Glover, K., & Woolley, S. C. (2023). The political use of encrypted messaging applications: Evidence from southeast Asia and its implications for the global public sphere. *First Monday*. <https://doi.org/10.5210/fm.v28i10.13172>
- Mayo, M. (2018, October 12). *The main approaches to natural language processing tasks*. KDnuggets. Retrieved August 29, 2020, from <https://www.kdnuggets.com/2018/10/main-approaches-natural-language-processing-tasks.html>
- Millière, R. (2022). Deep learning and synthetic media. *Synthese*, 200(3). <https://doi.org/10.1007/s11229-022-03739-2>
- Mohsen, F., Chaushi, B., Abdelhaq, H., Karastoyanova, D., & Wang, K. (2024). Automated Detection of Misinformation: A hybrid approach for fake news detection. *Future Internet*, 16(10), 352. <https://doi.org/10.3390/fi16100352>
- Monnier, A., Dandois, J., Filipczyk, A., Konstanta, E., Losa-Jonczyk, A., & Mourlas, C. (2023). A typology of Fact-Checking resources. *Questions De Communication*, 43, 297–316. <https://doi.org/10.4000/questionsdecommunication.31530>
- Moszczuk, B., Kiryluk, K., Pączek, L., & Mucha, K. (2021). Membranous nephropathy: From research bench to personalized care. *Journal of Clinical Medicine*, 10(6), 1205. <https://doi.org/10.3390/jcm10061205>
- Mugdha, S. B. S., Ferdous, S. M., & Fahmin, A. (2020). Evaluating Machine Learning Algorithms For Bengali Fake News Detection. *International Conference on Computer and Information Technology (ICCIT)*, 1–6. <https://doi.org/10.1109/iccit51783.2020.9392662>

- Naeem, S., Ali, A., Anam, S., & Ahmed, M. M. (2023). An Unsupervised Machine Learning Algorithms: Comprehensive Review. *International Journal of Computing and Digital Systems*, 13(1), 911–921. <https://doi.org/10.12785/ijcds/130172>
- Nakov, P., Corney, D., Hasanain, M., Alam, F., Elsayed, T., Barrón-Cedeño, A., Papotti, P., Shaar, S., & Da San Martino, G. (2021). Automated Fact-Checking for assisting human Fact-Checkers. *arXiv (Cornell University)*. <https://doi.org/10.48550/arxiv.2103.07769>
- Oni, T. S., Afolami, C. A., Obayelu, A. E., & Idowu, M. A. (2024). A Comparative Analysis of Household Food Insecurity Status among Rice Farmers in Savanna and the Rainforest Agro-ecological Zones in Southwest States, Nigeria. *Turkish Journal of Agriculture - Food Science and Technology*, 12(4), 507–515. <https://doi.org/10.24925/turjaf.v12i4.507-515.5066>
- Osisanwo F.Y., Akinsola J.E.T., Awodele O., Hinmikaiye J. O., Olakanmi O., Akinjobi J., (2017). Supervised Machine Learning Algorithms: Classification and comparison. *International Journal of Computer Trends and Technology*, 48(3), 128–138. <https://doi.org/10.14445/22312803/ijctt-v48p126>
- Pandey, S., Prabhakaran, S., Reddy, N. V. S., & Acharya, D. (2022). Fake News Detection from Online media using Machine learning Classifiers. *Journal of Physics Conference Series*, 2161(1), 012027. <https://doi.org/10.1088/1742-6596/2161/1/012027>
- Patel, A., Tiwari, A. K., & Ahmad, S. S. (2021). Fake News Detection using Support Vector Machine. *International Conference on Advanced Computing and Software Engineering*, 34–38. <https://doi.org/10.5220/0010562000003161>
- Petkar, P. B., & Sonawane, S. S. (2020). Fake News Detection: A Survey of Techniques. *International Journal of Innovative Technology and Exploring Engineering*, 9(9), 383–386. <https://doi.org/10.35940/ijitee.i7098.079920>
- Raja, M., & Raj, L. (2022). Fake news detection on social networks using Machine learning techniques. *Materials Today Proceedings*, 62, 4821–4827. <https://doi.org/10.1016/j.matpr.2022.03.351>

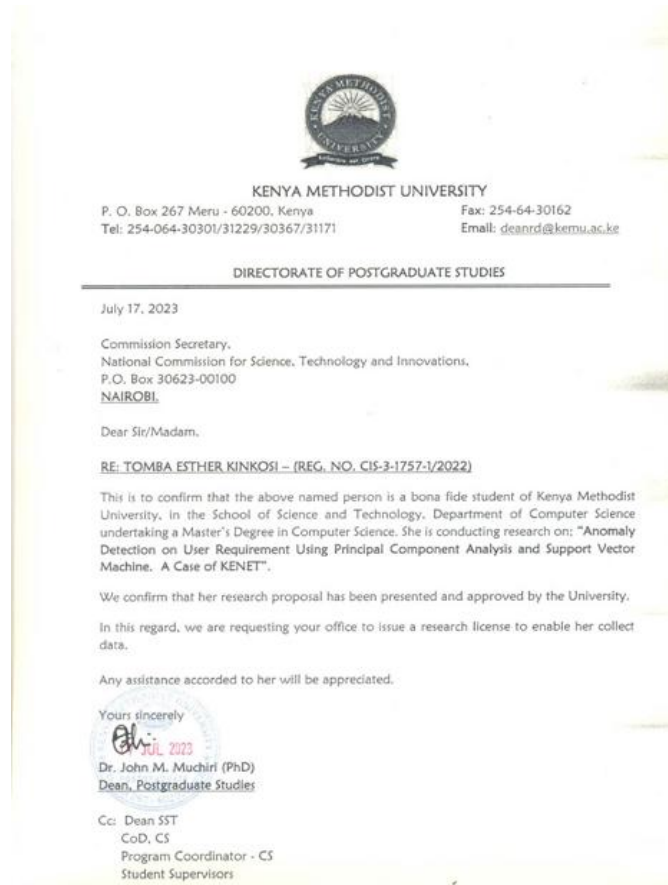
- Rolf, B., Jackson, I., Müller, M., Lang, S., Reggelin, T., & Ivanov, D. (2022). A review on reinforcement learning algorithms and applications in supply chain management. *International Journal of Production Research*, 61(20), 7151–7179. <https://doi.org/10.1080/00207543.2022.2140221>
- Ruchansky, N., Seo, S., & Liu, Y. (2017, November). Csi: A hybrid deep model for fake news detection. In *Proceedings of the 2017 ACM Conference on Information and Knowledge Management* (pp. 797-806). <https://doi.org/10.1145/3132847.3132877>
- Saha, R. R., Lakshmanan, L. V. S., & Ng, R. T. (2023). Stance Detection with Explanations. *Computational Linguistics*, 50(1), 193–235. https://doi.org/10.1162/coli_a_00501
- Sharma, M., Kandasamy, I., & Vasantha, W. (2021). YoungSheldon at SemEval-2021 Task 5: Fine-tuning Pre-trained Language Models for Toxic Spans Detection using Token classification Objective. *Proceedings of the 16th International Workshop on Semantic Evaluation (SemEval-2022)*, 953–959. <https://doi.org/10.18653/v1/2021.semeval-1.130>
- Shu, K., Wang, S., & Liu, H. (2019). Beyond News Contents. *International Conference on Web Search and Data Mining*, 312–320. <https://doi.org/10.1145/3289600.3290994>
- Sravani, N. K. V., & Kumar, N. D. a. S. (2025). Privacy-Preserving Machine Learning techniques for Horizontal Distributed Data: A Survey. *Deleted Journal*, 3(04), 1194–1202. <https://doi.org/10.47392/irjaeh.2025.0171>
- Su, J., Cardie, C., & Nakov, P. (2024). Adapting fake news detection to the era of large language models. *Findings of the Association for Computational Linguistics: NAACL 2022*. <https://doi.org/10.18653/v1/2024.findings-naacl.95>
- Suthaharan, S. (2015). Support Vector machine. In *Integrated series on information systems/Integrated series in information systems* (pp. 207–235). https://doi.org/10.1007/978-1-4899-7641-3_9
- Taha, M. A., Jabar, H. D. A., & Mohammed, W. K. (2024). Fake news detection model basing on machine learning algorithms. *Baghdad Science Journal*, 21(8), 2771. <https://doi.org/10.21123/bsj.2024.8710>

- Tandoc, E. C., Lim, Z. W., & Ling, R. (2017). Defining “Fake news.” *Digital Journalism*, 6(2), 137–153. <https://doi.org/10.1080/21670811.2017.1360143>
- Tchio, G. M. T., Kenfack, J., Kassegne, D., Menga, F., & Ouro-Djobo, S. S. (2024). A comprehensive review of supervised learning algorithms for the diagnosis of photovoltaic systems, proposing a new approach using an ensemble learning algorithm. *Applied Sciences*, 14(5), 2072. <https://doi.org/10.3390/app14052072>
- Tejedor, S., Romero-Rodríguez, L. M., & Gracia-Villar, M. (2024). Unveiling the truth: A systematic review of fact-checking and fake news research in social sciences. *Online Journal of Communication and Media Technologies*, 14(2), e202427. <https://doi.org/10.30935/ojcmnt/14455>
- Thakar, H., & Bhatt, B. (2024). Fake news detection: recent trends and challenges. *Social Network Analysis and Mining*, 14(1). <https://doi.org/10.1007/s13278-024-01344-4>
- Thomas, A. (2020). *Natural language processing with Spark NLP: Learning to understand text at scale*. O'Reilly Media. <https://books.google.com/books?id=5DDtDwAAQBAJ>
- Tianda, I. M., Ubadah, M. N., Mardianto, M. F. F., Munawwarah, S. a. A., Ishak, N., Amelia, D., & Ana, E. (2024). Clustering Fake News with K-Means and Agglomerative Clustering Based on Word2Vec. *INTERNATIONAL JOURNAL OF MATHEMATICS AND COMPUTER RESEARCH*, 12(02), 3999–4007. <https://doi.org/10.47191/ijmcr/v12i2.01>
- Villela, H. F., Corrêa, F., De Araújo Nery Ribeiro, J. S., Rabelo, A., & Carvalho, D. B. F. (2023). Fake news detection: a systematic literature review of machine learning algorithms and datasets. *Journal on Interactive Systems*, 14(1), 47–58. <https://doi.org/10.5753/jis.2023.3020>
- Vosoughi, S., Roy, D., & Aral, S. (2018). The spread of true and false news online. *Science*, 359(6380), 1146–1151. <https://doi.org/10.1126/science.aap9559>
- Wang, J., Qian, S., Hu, J., & Hong, R. (2023). Comment-Context Dual Collaborative Masked Transformer Network for Fake news Detection. *IEEE Transactions on Multimedia*, 26, 5170–5180. <https://doi.org/10.1109/tmm.2023.3330074>

- Wang, W. Y. (2017). “Liar, Liar Pants on Fire”: A New Benchmark Dataset for Fake News Detection. *Annual Meeting of the Association for Computational Linguistics*. <https://doi.org/10.18653/v1/p17-2067>
- Yan, F., Zhang, M., Wei, B., Ren, K., & Jiang, W. (2024). FMC: Multimodal fake news detection based on multi-granularity feature fusion and contrastive learning. *Alexandria Engineering Journal*, 109, 376–393. <https://doi.org/10.1016/j.aej.2024.08.103>
- Zahednejad, B., Ke, L., & Li, J. (2020). A novel Machine Learning-Based approach for security analysis of authentication and key agreement protocols. *Security and Communication Networks*, 2020, 1–15. <https://doi.org/10.1155/2020/8848389>

APPENDICES

Introduction letter to NACOSTI



NACOSTI LICENSE

			
RESEARCH LICENSE		NATIONAL COMMISSION FOR SCIENCE, TECHNOLOGY & INNOVATION	
Ref. No: 806719		Date of Issue: 15th January 2025	
			
This is to Certify that Rev. Father Kilonzi Tombo of Kenya Methodist University, has been licensed to conduct research as per the provision of the Science, Technology and Innovation Act, 2013 (Rev. 2014) in Nairobi on the topic: Annually detection on user requirements using principal component analysis and support vector machines, a case of KEN ET for the period ending : 15th January 2026.			
License No: NACOSTI/25/415850			
Applicant Identification Number		Director General  NATIONAL COMMISSION FOR SCIENCE, TECHNOLOGY & INNOVATION	
		Verification QR Code 	
NOTE: This is a computer generated License. To verify the authenticity of this document, Scan the QR Code using QR scanner application.			
See overleaf for conditions			