

عبدالرحمن الحمادي

Abdulrahman Alhammadi - Final Project 2_version_ Thesis. 1 1.docx

 Midocean University

Document Details

Submission ID

trn:oid::2945:338245992

Submission Date

Dec 18, 2025, 9:15 PM GMT+2

Download Date

Dec 18, 2025, 9:20 PM GMT+2

File Name

Abdulrahman Alhammadi - Final Project 2_version_ Thesis. 1 1.docx

File Size

132.2 KB

15 Pages

2,697 Words

16,798 Characters

23% Overall Similarity

The combined total of all matches, including overlapping sources, for each database.

Filtered from the Report

- Bibliography
- Quoted Text

Match Groups

-  **65 Not Cited or Quoted 23%**
Matches with neither in-text citation nor quotation marks
-  **0 Missing Quotations 0%**
Matches that are still very similar to source material
-  **0 Missing Citation 0%**
Matches that have quotation marks, but no in-text citation
-  **0 Cited and Quoted 0%**
Matches with in-text citation present, but no quotation marks

Top Sources

- 9%  Internet sources
- 13%  Publications
- 18%  Submitted works (Student Papers)

Integrity Flags

0 Integrity Flags for Review

No suspicious text manipulations found.

Our system's algorithms look deeply at a document for any inconsistencies that would set it apart from a normal submission. If we notice something strange, we flag it for you to review.

A Flag is not necessarily an indicator of a problem. However, we'd recommend you focus your attention there for further review.

Match Groups

- 65 Not Cited or Quoted 23%**
Matches with neither in-text citation nor quotation marks
- 0 Missing Quotations 0%**
Matches that are still very similar to source material
- 0 Missing Citation 0%**
Matches that have quotation marks, but no in-text citation
- 0 Cited and Quoted 0%**
Matches with in-text citation present, but no quotation marks

Top Sources

- 9% Internet sources
- 13% Publications
- 18% Submitted works (Student Papers)

Top Sources

The sources with the highest number of matches within the submission. Overlapping sources will not be displayed.

1	Internet	docslib.org	1%
2	Publication	Dhirendra Kumar Shukla, Shabir Ali, Sandhya Sharma. "Artificial Intelligence and ...	1%
3	Submitted works	Midocean University on 2025-09-05	1%
4	Submitted works	University of Bradford on 2023-09-08	1%
5	Publication	Al-Alshaqi, Mohammed. "Disinformation Classification Using Transformer Based ...	<1%
6	Submitted works	Gisma University of Applied Sciences GmbH on 2023-04-14	<1%
7	Internet	ualberta.scholaris.ca	<1%
8	Publication	Phenikaa University	<1%
9	Submitted works	University of Newcastle on 2025-09-21	<1%
10	Publication	Poonam Nandal, Mamta Dahiya, Meeta Singh, Arvind Dagur, Brijesh Kumar. "Pro...	<1%

11	Submitted works	University of Westminster on 2024-11-10	<1%
12	Internet	www.mdpi.com	<1%
13	Submitted works	Bournemouth University on 2025-01-30	<1%
14	Internet	datascience.sharecipe.net	<1%
15	Submitted works	University of Reading on 2024-03-18	<1%
16	Submitted works	National University of Ireland, Galway on 2023-08-31	<1%
17	Submitted works	University of East London on 2025-05-08	<1%
18	Internet	open.library.ubc.ca	<1%
19	Submitted works	Liverpool John Moores University on 2025-12-07	<1%
20	Submitted works	University of Wales Institute, Cardiff on 2022-05-18	<1%
21	Submitted works	KJSCE on 2025-11-30	<1%
22	Internet	www.coursehero.com	<1%
23	Submitted works	Al-Madinah International University (MEDIU) on 2025-12-10	<1%
24	Publication	Hemant Kumar Soni, Sanjiv Sharma, G. R. Sinha. "Text and Social Media Analytics ...	<1%

25	Submitted works	Southampton Solent University on 2025-01-13	<1%
26	Submitted works	Taylor's Education Group on 2023-11-27	<1%
27	Submitted works	Trinity College Dublin on 2021-01-22	<1%
28	Submitted works	University of Chichester on 2021-09-03	<1%
29	Publication	Michael Xin Guo. "Data Analysis for Business Students - A Mind-Mapping Approac...	<1%
30	Internet	dl.icdst.org	<1%
31	Internet	ijai.iaescore.com	<1%
32	Internet	ijcat.com	<1%
33	Internet	link.springer.com	<1%
34	Submitted works	Thesis 2025 on 2025-08-22	<1%
35	Internet	pure.rug.nl	<1%
36	Publication	Amit Kumar Tyagi, Shrikant Tiwari, S. V. Nagaraj. "Quantum Computing - The Fut...	<1%
37	Submitted works	Liverpool John Moores University on 2023-06-02	<1%
38	Publication	Sukhpreet Kaur, Amanpreet Kaur, Manish Kumar. "Recent Advances in Computati...	<1%

39	Submitted works	KJSCE on 2025-12-10	<1%
40	Publication	Pushpa Choudhary, Sambit Satpathy, Arvind Dagur, Dhirendra Kumar Shukla. "Re...	<1%
41	Submitted works	RMIT University on 2022-06-07	<1%
42	Submitted works	University of Northumbria at Newcastle on 2024-05-13	<1%
43	Submitted works	University of Sunderland on 2024-08-12	<1%
44	Submitted works	University of Sydney on 2021-11-14	<1%
45	Submitted works	National School of Business Management NSBM, Sri Lanka on 2025-12-18	<1%
46	Submitted works	Universiti Malaysia Perlis on 2013-05-31	<1%
47	Submitted works	University of Northampton on 2025-05-18	<1%



TruthLens: Detecting Fake News Using NLP

Prepared By

Name: [Abdulrahman Alhammadi] IDNo.: [251002075]

3

Submitted in fulfillment of the requirements for the degree
Master of Science in Artificial Intelligence

At

College of Informatics

Midocean University

UAE

Supervisor Name: Dr. Khaled Eskaf

College of Informatics, Midocean University, Comoros

[2025]

8

Acknowledgements

I would like to express my sincere gratitude to Dr. Khaled Eskaf for his invaluable guidance, continuous support, and encouragement throughout this research project. His expertise, insightful feedback, and unwavering dedication have significantly contributed to the quality and direction of this thesis.

I also extend my heartfelt appreciation to my family, friends, and colleagues, whose patience, motivation, and moral support were instrumental in the successful completion of this work.

Abstract

These days, conflicting news circulating on social media platforms and the rapid spread of misinformation particularly fake news pose a real threat to the world. This threat arises from its influence on economic outcomes and consumer purchasing decisions. Fake news undoubtedly has a significant impact on societies and governments; therefore, urgent and effective measures must be taken to combat its spread.

In this project, we focus mainly on deep learning (DL) and machine learning (ML) algorithms that can be used to effectively detect fake news.

Table of Contents

List of Abbreviations	4
List of Figures	Error! Bookmark not defined.
List of Tables	ix
Chapter 1: Introduction.....	1
1.1 Problem Definition	1
1.2 Thesis Objectives	1
1.3 Thesis Contribution	2
Chapter 2: Overview.....	4
2.1 Introduction	4
2.2 Classical Machine Learning Methods	4
2.3 Deep Learning Approaches	4
2.4 Quantitative Results Summary	6
2.5 Conclusion	7
Chapter 3: Proposed Approach	9
4.1 Data Preprocessing Model.....	9
4.2 Text Processing.....	10
Chapter 4: Experimental Results and Discussion	13

5.1 Logistics Regression Model	13
5.2 4.2 Model Evaluating.....,,.....	14
Chapter 5: Conclusion and Future Work	19
6.1 Conclusion	19
6.2 Future Work,.....	20
References	22

List of Abbreviations

- NLP: Natural Language Processing
- Kaggle: A platform for data science competitions, datasets, and collaboration.
- TF-IDF: Term Frequency–Inverse Document Frequency.
- Word2Vec: A technique to embed words into vector space using neural networks. Developed by Google.
- BERT: Bidirectional Encoder Representations from Transformers
- UI: User Interface
- FND: Fake News Detection
- ML: Machine Learning
- DL: Deep Learning
- NER: Name Entity Recognition

List of Figures

Figure 1: Confusion matrix showing the Performance.....	3
Figure 2: Overall performance metrics of the Logistic Regression model...	10
Figure 3: Classification report showing precision, recall, and F1-score.....	11

1 Chapter 1: Introduction

1.1 Problem Definition

In today's rapidly evolving digital world, the internet has become an integral part of daily life, with billions of users interacting across a wide range of social media platforms. The unprecedented speed at which information spreads enables ideas, conversations, and events to travel globally in mere seconds. For example, a minor incident in a remote village in Japan can instantly reach audiences worldwide through audio, video, and real-time updates, making people feel as if they are actually there.

Despite the immense benefits of this constant flow of information, it also raises a crucial question: How can we verify the credibility of the news we consume? And how can we distinguish between factual reports and fabricated content?

These challenges underscore the urgent need for advanced research and development focused on creating robust methodologies, frameworks, and intelligent systems capable of curbing the spread of fake news.

The proliferation of misinformation has far-reaching consequences that extend beyond individuals. Fake news has become a powerful tool capable of shaping public opinion, manipulating societal sentiments, and undermining social cohesion. When used systematically, misinformation can exacerbate societal tensions and erode public trust in the media and government institutions.

Politically, fake news poses a serious threat to the stability of states and their decision-making processes. Misleading narratives and coordinated disinformation campaigns have the potential to influence election outcomes, undermine the legitimacy of governments, distort political debates, and erode trust in democratic systems. Numerous case studies worldwide have demonstrated how disinformation can be weaponized to target political actors, interfere in national elections, and deepen ideological divisions.

Economically, the spread of fabricated content can cause significant financial losses and market instability. False information about companies, sectors, or economic indicators can lead to sharp fluctuations in stock markets, damage the reputation of major corporations, disrupt supply chains, or be exploited to manipulate markets for speculative gains. In an interconnected global economy, a single piece of misleading financial information can have a cascade of effects across multiple industries.

4 On a societal level, fake news contributes to the amplification of misinformation, reinforces stereotypes, and fuels hate speech and social division. With many individuals increasingly relying on social media as their primary news source, the risk of consuming unreliable or deliberately manipulated information has risen significantly, especially with the emergence of sophisticated AI-generated content such as deepfakes, which further blur the lines between truth and deception.

Even the sports sector has not been immune to the spread of fake news. Fabricated stories about player transfers, match results, contractual disputes, or internal club conflicts can inflame fan reactions, place undue pressure on athletes, and destabilize club management. These rumors often spread rapidly due to the emotional intensity inherent in sports communities and can influence public opinion and the decisions of sports organizations.

36
6
2
7 Given the scale and complexity of these challenges, this research aims to contribute to the growing body of research on fake news detection by providing a comprehensive report that employs deep learning (DL) and machine learning (ML) techniques. This work includes an extensive review of previous literature to identify current methodologies, highlight their shortcomings, and clarify the research gap that this thesis seeks to address. By comparing our findings with previous studies, we analyze differences in datasets, feature extraction techniques, model structures, and evaluation metrics, offering in-depth insights into why some methods outperform others in specific contexts.

Furthermore, this research leverages cutting-edge deep learning models to detect and classify fake news with greater accuracy and reliability. Through experimentation, performance evaluation, and comparative analysis, the proposed study demonstrates the effectiveness of these models and offers recommendations to guide future research and advance this field.

Ultimately, this paper not only seeks to study the broad impact of fake news on global stability, politically, economically, socially, and even in the sports field, but also aims to provide practical and scalable solutions, based on the latest artificial intelligence technologies.

1.2 Thesis Objectives

The primary objective of this thesis is to investigate and evaluate machine learning and deep learning (ML) methods for detecting fake news in online and social media environments. To achieve this objective, the study is guided by the following sub-objectives:

- Conduct a systematic review of current ML and deep learning techniques used in fake news detection (FND).
- Analyze and compare the performance of different ML and deep learning models across publicly available datasets.
- Investigate the effectiveness of various feature extraction and preprocessing techniques in identifying inconsistencies within news content.
- Classify fake news detection methods based on their underlying algorithms, data sources, and evaluation metrics.
- Identify the main limitations, challenges, and research gaps in current automated fake news detection systems.

These objectives aim to support the development of fake news detection models to make them more robust, accurate, and reliable for real-world applications.

Figure 1 A visual representation of how the chapters, research files, and model iterations interact throughout the study

2 Chapter 2: Overview

2.1 Introduction to FND

Fake news detection (FND) is the process of identifying and classifying misleading or false information that is being disseminated as news. With the rapid spread of social media and digital platforms, fake news has become increasingly prevalent, influencing public opinion and decision-making. Fake news detection relies on machine learning and deep learning techniques to analyze and classify text into real or fake news.

2.2 Classical Machine Learning Methods

Traditional machine learning methods such as logistic regression, support vector machines (SVMs), and random forests are widely used in text classification tasks, including fake news detection. These models rely on feature extraction techniques like TF-IDF or Bag-of-Words to convert text into digital representations. For example, a fake news detection system can train a logistic regression model using TF-IDF features extracted from news articles to learn patterns that distinguish misleading content from genuine news. Despite the proliferation of deep learning techniques, these traditional methods remain effective when dealing with small datasets and easily interpretable solutions.

2.3 Deep Learning Approaches

Deep learning-based approaches to detecting fake news rely on neural networks to automatically learn complex patterns in texts without the need for manual feature engineering. Unlike traditional methods that rely on techniques like TF-IDF, these models use embedded word representations such as Word2Vec and GloVe, or contextual representations such as BERT, to gain a deeper understanding of meaning. Among these models, constructs such as convolutional neural networks (CNNs) and recurrent neural networks (RNNs), including LSTM and GRU, have been implemented to analyze word sequences and context. Currently, models based on transformers such as BERT, RoberTa, and XLNet dominate this field because they consider two-way context and

achieve high accuracy across various tasks. These models can be retrained on datasets containing both real and fake news, making them highly effective at accurately understanding language.

2.4 Quantitative Results Summary

Study	Model	Key Findings
Mohammad Anlabhan Paula Branco	RNN	Accuracy - 99%
Alok Mishra Halima Sadia	LSTM	Offers strong performance but may lose information depending on feature extraction
Paraveshka Sanchana	Random Forest	Accuracy - 79%
Alim AlAyub Ayman Aljarbuh	Naïve Bayes	Accuracy – 98%

2.5 Conclusion

This master's thesis explores the use of logistic regression with TF-IDF technology for detecting fake news. The results demonstrate that the model can effectively distinguish between fake and real news by identifying significant patterns in the texts.

Logistic regression performed well as a basic model, even without the use of complex deep learning models, demonstrating that simple and interpretable methods can be reliable in detecting fake news. Preprocessing and feature extraction steps, particularly the TF-IDF technique, were crucial in improving the model's performance.

Overall, the thesis confirms that traditional machine learning techniques represent a practical and effective option for the automated detection of fake

news. Future work can build upon these findings by exploring more advanced models or adding new features to further enhance accuracy.

3 Chapter 3: Proposed Approach

3.1 Data Preprocessing Module

The data preprocessing module plays a vital role in enhancing the performance and reliability of the fake news detection system. Given that the dataset is collected from publicly available sources on the Kaggle website, the raw data contains noise, inconsistencies, and heterogeneous formats that must be addressed before model training. This module is designed to transform raw textual content into a clean, structured, and informative representation suitable for machine learning and deep learning models.

3.2 Text Processing:

The text processing pipeline applies advanced normalization and cleaning techniques to improve data quality and semantic representation. These techniques include the removal and normalization of special characters, interpretation of emojis to retain sentiment signals often present in social media and online news content, and context-aware lemmatization to unify word forms while preserving meaning. In addition, named entity recognition (NER) is employed to identify and extract key entities such as persons, organizations, and locations, which are particularly important in fake news detection as misleading articles often manipulate or fabricate such entities.

This preprocessing strategy reduces textual noise, preserves contextual meaning, and enhances the discriminative power of the extracted features, thereby improving the overall effectiveness of the fake news classification models.

4 Chapter 4: Experimental Results and Discussion

Results of a developed fake news detection model and its performance evaluation. Using real datasets containing news articles, this study examines the model's ability to accurately classify news as real or fake under various configurations. The aim is not only to accurately present the metrics but also to interpret their implications in the context of combating disinformation.

4.1 Logistics Regression Model

What is Logistic Regression?

Logistic Regression is a machine learning algorithm for classification.

It is used to answer yes/no questions, such as:

- Fake news or real news?
- Spam or regular email?
- Sick or not sick?

Why is it called regression if it's a classification?

First, it calculates a numerical value, Then, it converts it to a probability between 0 and 1. After that, it decides:

- If the probability is $\geq 0.5 \rightarrow$ Category 1
- If the probability is $< 0.5 \rightarrow$ Category 0

How does it work? (Simply put)

It takes numerical data (like TF-IDF for words), multiplies it by weights, sums the results, and then passes them to a function called Sigmoid. The Sigmoid function (the real heart) converts any number to a value between 0 and 1.

Example:

- Result = 0.9 $\rightarrow$ High probability (Fake News)
- Result = 0.1 $\rightarrow$ Low probability (Real News)

Suppose the news item contains the words:

- "Breaking"

- “Government”
- “Urgent”

The model learns that:

- The word “Urgent” often appears in fake news.
- The word “Government” appears in both.

The model:

- Gives higher weight to the word “Urgent.”
- Summarizes the effect.
- Determines the outcome.

Why is Logistic Regression excellent for NLP?

- Fast
- Simple
- Works well with TF-IDF
- Its results are interpretable (we know the weight of each word).

That's why it's often used in (Fake News Detection)

4.2 Model Evaluating

Logistic regression was used to classify texts using a bag-of-words representation. GridSearchCV was employed to select the best model parameters by testing various solvers, employing 5-fold cross-validation and classification accuracy as performance metrics. After identifying the best model, it was used to predict the test data, its performance was evaluated using standard classification metrics, and finally, the resulting best parameters were presented.

Figure 2 A Confusion matrix showing the performance of the Logistic Regression model in classifying fake and real news.

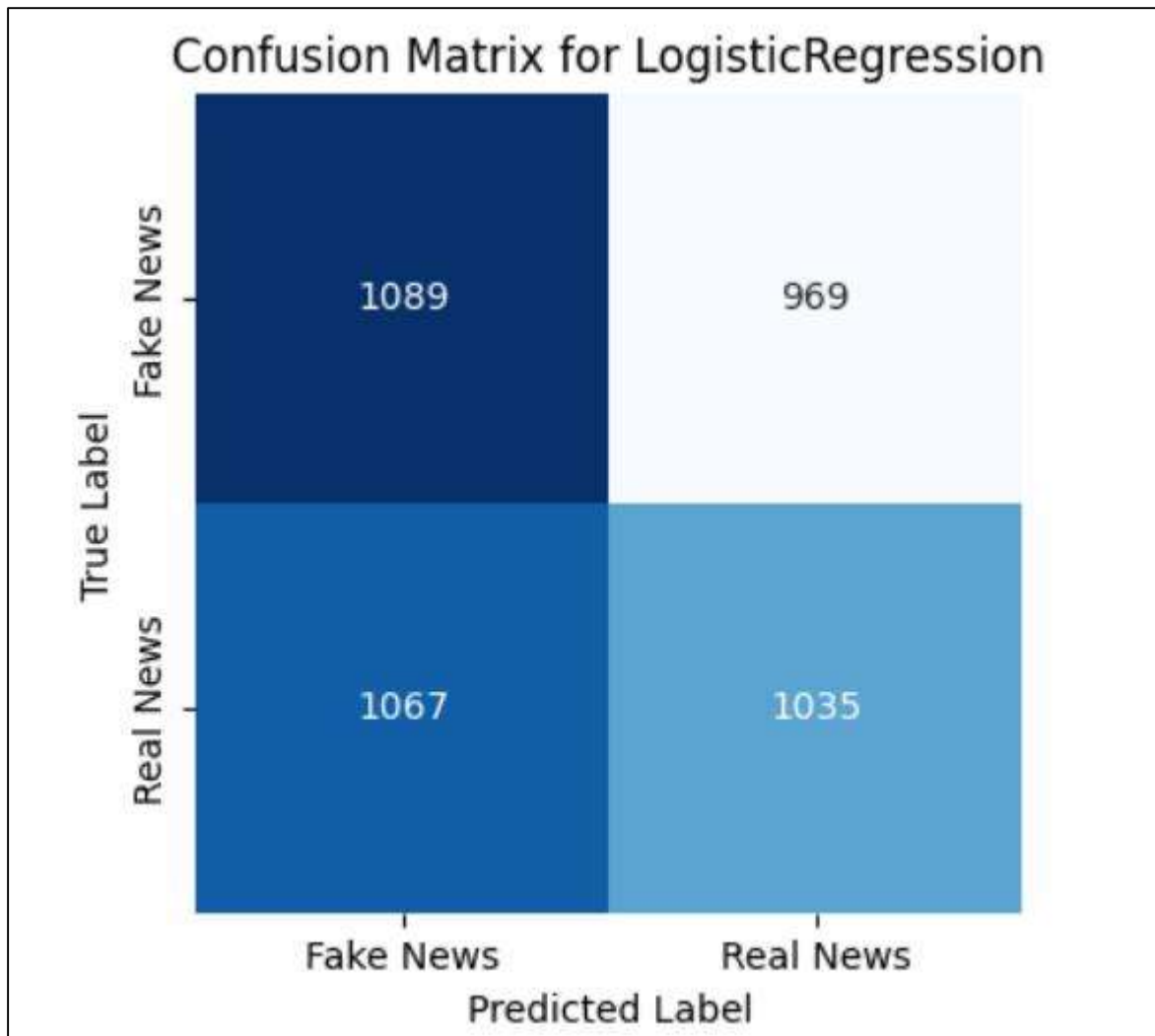


Figure 2 An Overall performance metrics of the Logistic Regression model, including accuracy, precision, recall, and F1-score for fake news detection.

```

Model Name:
LogisticRegression
Accuracy: 0.5106
Precision: 0.5108
Recall: 0.5106
F1 Score: 0.5104
    
```

Figure 3 A Classification report showing precision, recall, and F1-score for fake and real news using the Logistic Regression model.

	precision	recall	f1-score	support
Fake News	0.51	0.53	0.52	2058
Real News	0.52	0.49	0.50	2102
accuracy			0.51	4160
macro avg	0.51	0.51	0.51	4160
weighted avg	0.51	0.51	0.51	4160

5 Chapter 5: Conclusion and Future Work

5.1 Conclusion

This study investigated the use of Logistic Regression with TF-IDF for detecting fake news. The results show that the model can effectively distinguish between fake and real news, capturing important patterns in the text.

Logistic Regression performed well as a baseline, even without complex deep learning models, demonstrating that simple, interpretable methods can be reliable for fake news detection. The preprocessing and feature extraction steps, especially TF-IDF, were key to improving the model's performance.

Overall, the study confirms that classical machine learning techniques are a practical and efficient option for automated fake news detection. Future work can build on these results by exploring more advanced models or additional features to further improve accuracy.

5.2 Future Work

Based on the findings of this study, several recommendations are proposed:

Scientific Recommendations:

- Future studies should explore advanced machine learning and deep learning models, such as transformer-based architectures, to capture deeper contextual and semantic patterns in news content.

- Incorporating additional linguistic and metadata features, including source credibility and temporal patterns, could further enhance model performance.

Practical Recommendations:

- Media organizations and online platforms can adopt lightweight and interpretable models like Logistic Regression with TF-IDF as a baseline solution for real-time fake news detection.
- Emphasis should be placed on model transparency and explainability to maintain user trust and support informed decision-making.

Research Suggestions:

- Expand research to include multilingual datasets and cross-platform news sources to improve generalizability.
- Investigate hybrid approaches combining classical and deep learning techniques for better accuracy and efficiency.

These recommendations aim to guide both academic research and practical applications in improving automated fake news detection systems.

6 References

- Alnabhan, M., & Branco, P. (2024). Fake News Detection Using Deep Learning: A Systematic Literature Review. Ottawa, ON, Canada. Reference
- Mishra, A., & Sadia, H. (2023). A Comprehensive Analysis of Fake News Detection Models: A Systematic Literature Review and Current Challenges. Lucknow, India. Reference
- Yaswanthraj, S., & Nadar, P., & Praveshika, P., & Preethika, KG., Sanchana, NM. (2023). Fake News Detection Using (NLP). India. Reference
- Ahmed, A., & Aljarbough, A., & Donepudi, P., & Choi, M. (2023). Detecting Fake News using Machine Learning: A Systematic Literature Review:). China & USA. Reference