accuracy than the K-NN classification technique. The deep learning technique achieved higher accuracy than the previous two algorithms, as well as from previous studies.

1.6 Thesis layout

In addition to this introductory chapter, this thesis contains the following chapters:

Chapter Two: "Theoretical Background": In this chapter, four main parts will be introduced. The first part talks about natural language processing and text representation (news). Part two is about machine learning and its fundamentals and algorithms and part three is about deep learning, fundamentals, strategies, classes, organization, and optimization. The fourth part focuses on the various evaluation procedures to see how accurately fake news is classified.

Chapter Three: "Designing Machine Learning and Deep Learning Models": There are two parts to this chapter. In the first part, the datasets for fake news are shown, along with what they contain. Also, the second part talks about how the fake news classification systems (machine learning and deep learning classification methods) are set up and how they work.

Chapter Four: "The results and Discussion": This chapter shows the attained results of the implementation of machine and deep learning models and compares them with similar work.

Chapter Five: "Conclusions and Future works": Many conclusions drawn from the implementation of classification models will be presented, as well as some suggestions for future work.

Decision Trees are five different types of machine learning that were employed. The GloVe and BERT weddings used CNN and LSTM but weren't the only ones. The accuracy, F1 score, accuracy, and memory of all the models and marriages used were compared. The LSTM constructed using GloVe vaccines performed the best, according to the data. The misclassified samples were also examined in the study by contrasting them with the rating individuals would have given them.

- Jehada and Yousif [15], 2022, investigated fake news by employing the multi-layer perception algorithm for classification and the term inverted frequency document (TF- IDF) for feature extraction. Three-layer categorization has two steps: feed-forward and propagation-back (an input layer, one hidden layer, and an output layer). When this approach was used for classification, accuracy, precision, and recall reached 95.47%, 96.2, and 95.47 respectively.

1.4 Problem statement

The problem with the thesis is how accurately and to what extent one can distinguish fake news articles using natural language processing and classification algorithms. This thesis is meant to deal with the following issues:

How does this news affect people's opinions and ideas and change their views? What steps can be taken to provide a solution? The core task of detecting fake news involves identifying the language (set of words or sentences) which is used to deceive the readers.

1.5 Aim of the Thesis

The aim of the research is to identify patterns in the text to distinguish fake articles from real news. Different textual features were extracted from the articles (trained and taught learning models) for optimal accuracy. Two common machine learning techniques (NB and K-NN) and a deep learning technique (CNN-1D) were used in the presented scheme, and the NB classification technique achieved higher

Fake or Real News Dataset, and ultimately the Fake News Detection Dataset. The initial dataset was the ISOT fake news dataset. On all test datasets, the Bidirectional LSTM architecture did better than ResNet and CNN. The achieved accuracy percentages were respectively 99.24 percent, 98.99 percent, and 98.24 percent.

- Nasir et al. [13], 2021, integrated convolutional and recurrent neural networks into a new type of deep learning model for identifying fake news. The FA-KES 5 dataset was based on two fake news data sets. It had 804 news stories about the Syrian war (ISO and FA-KES). The number of real articles is 426, while the number of fake articles is 376. This makes for a well-balanced set of articles (53% are real and 47% are fake). On the other hand, the 45,000 news stories in the ISOT 6 dataset were almost equally true and false. In the suggested model, local information was gathered using CNN, while long-term associations were discovered using LSTM. The input vectors are first processed by a one-dimensional CNN layer, which extracts the local features that are at the next level. The CNN layer's output is sent to the next layer, which is called the Recurrent Neural Network and is made up of LSTM units/cells (RNN). CNN tried out the RNN layer, which figures out if a news story is true or not based on how long-term relationships between local parts of the story help figure that out. Experiments with two real-world fake news datasets (100 percent accuracy on ISOT dataset, which has 45,000 articles, and 60 percent accuracy on FA-KES dataset, which has 804 articles) show that hybrid baseline approaches work much better than non-hybrid baseline methods. Last, the amount of data used to train deep learning systems can help them.
- Kishwar and Zafar [14], 2022, produced the first comprehensive dataset for identifying false news in Pakistani news by combining several legitimate news APIs. The created dataset has also been put to the test using a variety of cutting-edge AI methods. Naive Bayes, KNN, Logistic Regression, SVM, and

SOT datasets, respectively. The studies' findings revealed that the ISOT dataset's test set had 2.1% higher accuracy while the LIAR datasets were 7.9% higher.

- Khan et al. [9], 2020, compared the effectiveness of the SVM, LR, DT, Naive Bayes, and k-NN machine learning algorithms on three different datasets in a benchmark study. Accurate results (71% to 90%) were obtained using language models that were pre-trained using k-NN and Naive Bayes.
- Jehad, and Yousif [10], 2020**,** used two different machine learning algorithms to find fake news: random forest and decision tree (J48). In this system, There are 20,761 samples in the dataset as a whole, and 4,345 samples in the testing sample. The first step in preprocessing is cleaning the data by removing extra special characters, numbers, English letters, and white spaces. The last step is getting rid of stop words. After that, the most common way to get features is used (TF-IDF) before the two suggested classification algorithms are used. The findings indicate that the decision tree model's accuracy is highest at 89.11%, while the best accuracy for the random forest model is 84.97%.
- Aslam et al. [11], 2021, suggested determining the veracity of news using an ensemble-based deep learning algorithm. The dataset was assembled in a way that required the usage of two deep learning models. Bi-LSTM-GRU-dense, a deep learning model, was applied to the text attribute "statement." The remaining attributes were employed with the dense deep learning model. The proposed study had an accuracy of 0.898, a recall of 0.916, a precision of 0.913, and an F-score of 0.914 based on experimental data that only considers the statement attribute.
- Sastrawan et al. [12], 2021, presented analysis of a deep learning methodology using four different datasets with a variety of architectures, including CNN, Bidirectional LSTM (Long Short- Term Memory), and ResNet (Residual Network). The False News Dataset came first, then the Fake News Dataset, the

attaching "weights" to certain nerves. Artificial neural networks can learn automatically from outside instructions or grow independently using data [5].

1.3 Related Works

Several studies have been reported on the development of fake news detection. Several deep learning models have been used for rating optimization problems in fake news. Most of them are listed below:

- Bauskar et al. [6], 2019, A special machine-learning model that can identify "fake news" based on both the content and social characteristics of news has been developed using NLP techniques. The proposed model performed well on a sample data set, with an average accuracy of 90.62 percent and an F1 score of 90.33 percent.
- Ahmed et al. [7], 2019, proposed an algorithm to find fake news using n-gram analysis and machine learning. Also, they compared two different feature extraction methods and six machine classification methods: Support Vector Machine (SVM), Linear Support Vector Machine (LSVM), k-Nearest Neighbor (k-NN), Stochastic Gradient Descent, Linear Regression (LR), and Decision Tree (DT). The best results come from using Term Frequency-Inverted Document Frequency (TF-IDF) to pull out features and LSVM as a way to sort things. They find that LSVM classification works at 92% accuracy no matter how many feature values are used. As n-gram (Tri-gram, Four-gram) goes up, the accuracy of the algorithm goes down.
- Goldani et al. [8], 2020, Provided a Convolutional Neural Network (CNN) with margin loss and several embedding techniques. In this analysis, the widely utilized datasets Information Security and Object Technology (ISOT) and LIAR were employed. CNN was utilized to uncover false news while business earnings were declining. The non-static word embedding was employed for the ISOT dataset when the method described above was applied to the LIAR and

1.2 AI based Fake News Detection

Computer science and information technology's field of artificial intelligence (AI) studies how computers can think, learn, and improve themselves as humans. It is said that AI is what lets computers act like humans do. Also, AI is not a separate computer science field; it is closely related to many other fields. There are a lot of active efforts to use artificial intelligence in different areas of information technology to help solve problems in those areas. Systems like automatic translation are already utilized in the area of Natural Language Processing (NLP) [3]. People can talk to computers and share information when more research is done, leading to new ways of using computers. In the field of expert systems, computers can do many of the professional jobs that people do now, such as making diagnoses, figuring out where minerals are found, guessing the structure of compounds, figuring out how much damage an accident caused, etc. In a lot of ways, it was the first step forward. Without artificial intelligence theory, A computer wouldn't be able to figure out what a picture taken by a TV camera is or turn a person's voice into sentences. Both jobs are very tough. Text recognition, robotics, and other fields need imaging and speech recognition. Neural Network is not mathematical logic. It has only been around for a short time. Instead, it uses a network of many basic processors to function similarly to the human brain [4].

AI technologies like machine learning, deep learning, and NLP tools are very important to many studies that aim to make systems that automatically spot fake news. But it is very hard to spot fake news because you need models to summarize it and compare it to the real news to tell if it is fake or not. Also, it's hard to compare the original news to the suggested news because news is very opinionated and subjective. Artificial neural networks come in a variety of forms, including those with one- or two-way feedback loops, complex feedback loops with numerous inputs, and tiers. Overall, how the functions are managed and linked will be determined by the algorithms of these systems. Most systems allow you to alter how they operate by

CHAPTER ONE
INTRODUCTION

1.1 Introduction

Fake news is a form of propaganda or "yellow journalism" in which false information or hoaxes are spread on purpose through print and broadcast media or social media on the internet. Fake news is made up and spread with intending to make money or political gain. Often, fake news is spread through news headlines that are shocking, false, or both to get people's attention. Misleading and purposely misleading fake news is different from satire or parody which is meant to make people laugh rather than trick them [1].

Fake news often uses catchy headlines or completely made-up news stories to get more people to read it online, get them involved online, and earn money from clickstream. In the second case, it's like "click bait" headlines on the internet, which try to get people to click on them, and it depends on the advertising money it brings in, no matter how true the news is. Fake news also makes it harder for journalists to cover important news stories and hurts the credibility of real news coverage [2].

In the report from the European Commission, "fake news" is defined as "all kinds of false, inaccurate, or misleading information that is made, shared, and promoted to hurt the public on purpose or for profit." Fake news that gets around on the Internet is a big problem today. More and more people use digital platforms to get news and share information. Most of the time, they are unable to distinguish between true and false news. Helping people spot fake news is essential to keep bad things from affecting public opinion and people's decisions [3].

CHAPTER ONE

Abbreviation	Full words
MLP	Multi-Layer Perceptron
NB	Naïve Bayes
NLP	Natural Language Processing
NLTK	Natural Language Toolkit
NN	Neural Network
OOV	out-of-vocabulary
Relu	Rectified linear unit.
RNN	Recurrent Neural Network
SLP	Single Layer Perceptron
SSC	Semi-Supervised Classification
SSL	Semi-supervised Single Link
SVM	Support Vector Machine
TF-IDF	Term Frequency –Inverse Term Frequency
TN	True Negative
TP	True Positive
TV	Television
US	United States

List of Abbreviations

Abbreviation	Full words
Adam	Adaptive moment estimation
AI	Artificial Intelligence
ANN	Artificial Neural Network
Bi-LSTM	Bidirectional-Long Short-Term Memory
BoW	Bag of Words
CNN	Convolution Neural Network
DL	Deep Learning
DNN	Deep Neural Network
FBI	Federal Bureau of Investigation
FN	False Negative
FND	fake news detection
FNs	Fake News
FP	False Positive
GAN	Generative Adversarial Networks
GloVe	Global Vectors of Word Representations
GRU	Gated Recurrent Unit
ISOT	Information Security and Object Technology
k-NN	k-Nearest Neighbor
LR	Linear Regression
LSTM	Long Short-Term Memory
LSVM	Linear Support Vector Machine
ML	Machine Learning

List of Tables

List of Algorithms

LIST OF FIGURES

Subject Page No.

Chapter Three

Subject Page No.

Chapter Four

Subject Page No.

Chapter Five

Subject Page No.

Chapter One

Subject Page No.

Chapter Two

Abstract

The Spread of fake news can be defined as one of the social phenomena which might be pervasive at social levels through social media and between individuals. Fake news discussed on social media causes deception and misleading individuals. Such issues aim to change the views of individuals and society. For instance, information is rapidly spreading in social and news media with no accuracy, which might badly impact society and individuals. Thus, it is essential to have a detection mechanism that can predict fake news adequately fast.

A detection model is presented to classify fake news with the effect of Term Frequency –Inverse Term Frequency (TF-IDF) on the dataset. In addition and using two machine learning methods Naïve Bayes and K-Nearest Neighbor (NB, KNN) and one deep learning method Convolution Neural Network one dimension (CNN-1D). and used two different sets of data set. For the first data set, the model achieved the maximum accuracy using two machine learning algorithms (NB and KNN) which are (94% and 87%), and the maximum accuracy for the first data set was (100%) using the deep learning algorithm (CNN-1D). For the second data set, the model achieved the maximum accuracy using two machine learning algorithms (NB and KNN) which are (93% and 97%), and the maximum accuracy for the second data set was (100%) using the deep learning algorithm (CNN-1D). The results achieved are better than the related inline works, so this algorithm enhances the classification accuracy.

Acknowledgment

First, praise God, Lord of all creation, for all the blessings that have been of help in bringing this work to its end.

I thank my supervisor, Dr. Jumana Walid Saleh, for her sincere guidance, valuable instructions, and constructive comments that made completing this work possible.I would also like to express my gratitude and thanks to all the faculty members who were credited with their knowledge that I benefited from. Special thanks to all my friends for their help. Special thanks to the chairman and discussion committee members for their kindness in discussing my thesis and correcting it in a manner that makes it scientifically beneficial. My last words go to my family.

I would like to thank my father, mother, brother and sisters for their constant love and support. And thanks, and gratitude to my wife, the mother of my children, who was the true supporter and the unknown soldier, who gave and endured throughout the study period the trouble of supporting me and completing her duties regarding the family and children as best as possible. May God grant them all the best, for they helped me achieve my goals.

Finally, I do not forget, and I will not forget my brothers who joined the Forgiving, the Merciful.

Many thanks to everyone.

Khalid Abbood

2023

Dedication

To my father and mother ...

To my Family…

To my Wife…

To my children...

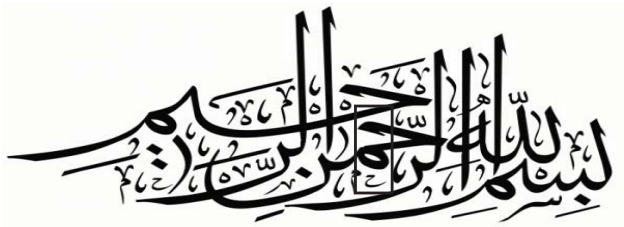
To my supervisor…

To my Friends…

To the soul of my dear Brothers…

To the soul of Captain Ahmed Radi…

With best regards…



﴿قُلْ لَوْ كَانَ الْبَحْرُ مِدَادًا لِكَلِمَاتِ رَبِّي

لَنَفِدَ الْبَحْرُ قَبْلَ أَنْ تَنْفَدَ كَلِمَاتُ رَبِّي

وَلَوْ جِئْنَا بِمِثْلِهِ مَدَدًا﴾

صدق الله العظيم

سورة الكهف

الآية:109

Republic of Iraq

Ministry of Higher Education and scientific Research

University of Diyala

College of Science

Department of Computer Science

Fake News Classification Using Machine Learning Algorithms

A Thesis

Submitted to the Department of Computer Science/ College of Science/ University of Diyala

In Partial Fulfillment of the Requirements for the Degree of Master's in computer science

By

Khalid Abbood Kamel

Supervised By

Asst. Prof. Dr. Jumana Waleed Salih

2023A.D. 1444 A.H.