

## **ACKNOWLEDGEMENT**

I am pleased to acknowledge my sincere thanks to **Board of Management of SATHYABAMA** for their kind encouragement in doing this project and for completing it successfully. I am grateful to them.

I convey my thanks to **Dr. T. Sasikala M.E., Ph.D, Dean, School of Computing, Dr. S. Vigneshwari, M.E., Ph.D. and Dr. L. Lakshmanan, M.E., Ph.D., Heads of the Department of Computer Science and Engineering** for providing me necessary support and details at the right time during the progressive reviews.

I would like to express my sincere and deep sense of gratitude to my Project Guide **Dr. Kalaiarasi G M.E.,Ph.D.**, for her valuable guidance, suggestions and constant encouragement paved way for the successful completion of my project work.

I wish to express my thanks to all Teaching and Non-teaching staff members of the **Department of Computer Science and Engineering** who were helpful in many ways for the completion of the project.

## **ABSTRACT**

Fake news has been in our society for a long time but with the introduction of social media, internet and mobile phones the spread of fake news has severely increased. Social media sites are being used to effectively spread misinformation and hoaxes around the world which not only causes people to change their thinking but also manipulates their opinions and decisions. In today's world it has become nearly impossible to detect if the given news is fake or real. With the arrival of the novel coronavirus-19 pandemic the propagation of fake news is now more than ever. In this time there is a need for something which can classify if a given news is real or not. In this article we aim to develop a model which, using some algorithms, determines if the given news is fake or not.

# TABLE OF CONTENTS

<b>ABSTRACT</b>	<b>5</b>
<b>TABLE OF CONTENTS</b>	<b>6</b>
<b>LIST OF FIGURES</b>	<b>7</b>
<b>LIST OF TABLES</b>	<b>7</b>
<b>ABBREVIATIONS</b>	<b>8</b>
<b>KEYWORDS</b>	<b>10</b>
<b>1. Introduction</b>	<b>11</b>
1.1. Introduction	
1.2. Aim and Objective	
1.3. Motivation	
1.4. Software Requirements Specification	
<b>2. Literature Review</b>	<b>14</b>
2.1. Different approaches used	
2.2. Comparison of Existing methods with merits and demerits	
<b>3. Methodology</b>	<b>16</b>
3.1. Models Used	
3.2. Architecture diagram	
<b>4. Implementation</b>	<b>19</b>
4.1. Data collection and analysis	
4.2. Data Preprocessing	
4.3. Tokenization	
4.4. Feature Extraction	
4.5. Splitting the dataset	
4.6. Model Training and verification	
<b>5. Libraries Used</b>	<b>22</b>
5.1. Numpy	
5.2. Pandas	
5.3. Sklearn	
<b>6. Coding and Testing</b>	<b>25</b>
<b>7. Conclusion and Enchantment</b>	<b>28</b>
<b>8. References</b>	

## **LIST OF FIGURES**

- Architecture Diagram 18

## **LIST OF TABLES**

- Literature Review 14
- Accuracy Table 28

## **ABBREVIATIONS**

**LR** - Logistic Regression

**ML**- Machine Learning

**NLP**- Natural Language processing

**np** - Numpy

**pd** - Pandas

# 1.INTRODUCTION

## 1.1 INTRODUCTION:

As we spend more time connecting online through social media platforms, more people are seeking out and consuming news through social media outlets rather than traditional news organisations. [1] The reasons for this shift in consuming habits are built into the nature of those social media platforms: (i) consuming news on social media is typically more timely and less expensive than traditional journalism, such as newspapers or television; and (ii) sharing, discussing, and discussing the news with friends or other readers is easier on social media. For example, in 2016, 62 percent of adults in the United States received news on social media, compared to only 49 percent in 2012 [1].

It was also discovered that, as a key news source, social media currently exceeds television. Despite the advantages of social media, the quality of social media stories is lower than that of traditional news organisations. Large volumes of fake news, i.e. news pieces with purposely incorrect material, are created online for a variety of reasons, including financial and political gain, because it is inexpensive to supply news online and significantly faster and easier to distribute through social media. Over 1 million tweets have been linked to the false news "Pizzagate" scandal, according to estimates "by the end of the presidential election's first round Given the popularity of this new phenomena, "fake news" is a term that has to be defined "In 2016, the Macquarie dictionary named it the word of the year [2].

The widespread dissemination of fake news has the potential to harm both individuals and society. For starters, fake news has the potential to disrupt the news ecosystem's authenticity equilibrium; for example, during the 2016 presidential election in the United States, the most popular false news was even more widely shared on Facebook than the most widely accepted genuine mainstream news. Second, fake news is designed to induce customers to accept biased or incorrect information. Propagandists use fake news to disseminate

political ideas or influence. For example, according to some reports, Russia has developed fake accounts and social bots to spread misleading tales.

Third, false news alters how people interpret and respond to actual news. For example, some fake news was generated solely to instil distrust and confusion in people, making it difficult for them to tell what is true from what is not.

## **1.2 AIM AND OBJECTIVES**

The main objective behind the development and upgradation of existing projects are the following smart approaches:

- Be Aware of such articles while forwarding to others
- Reveal True stories
- Prevent from false crisis events
- Be Informative

## **1.3 MOTIVATION :**

Machine learning (ML) is a sort of artificial intelligence (AI) that allows software applications to improve their prediction accuracy without being expressly designed to do so.

In order to forecast new output values, machine learning algorithms use historical data as input.

The widespread dissemination of fake news has the potential to harm both individuals and society. For starters, false news has the potential to disrupt the news ecosystem's authenticity balance.

Understanding the truth of new information and its message can have a positive impact on society when used in conjunction with news detection.

## **1.4 SOFTWARE REQUIREMENTS SPECIFICATION:**

- Anaconda Navigator

- Jupyter
- Google Collab
- Windows 10
- Mouse,Keyboard, Laptop



## 2. LITERATURE REVIEW

### 2.1 DIFFERENT APPROACES USED:

Paper Id	Title	Source	Task
1	Fake news detection using discourse segment structure analysis	2020 10th International Conference on Cloud Computing, Data Science & Engineering	to propose and implement a method for automated deception detection
2	Fake News Detection Using Deep Learning Techniques	2021 3rd International Congress on Human-Computer Interaction	Classification techniques like LR, RF, SVM NB and DNN for feature selection and extraction utilized
3	Fake News Detection by Decision Tree	IEEE Southeastcon	Our model can identify the reality of the newest news on the website

4	Fake News Detection Using Deep Learning Techniques	International Conference on Advances in Information Technology	proposed system LR, RF, SVM NB and DNN classification techniques are utilized that will help to detect fake news
5	Fake News Detection on Social Media using K-Nearest Neighbor Classifier	International Conference on Advances in Computing and Communication Engineering (ICACCE)	a specific frame to predict fake news on social media.
6	A Deep Transfer Learning Approach for Fake News Detection	2020 International Joint Conference on Neural Networks (IJCNN)	fake information detection through stance detection

## **2.2 COMPARISON OF EXISTING METHODS WITH MERITS AND DEMERITS:**

### **Limitations of linguistic analysis-based methods**

- Although this method is frequently regarded as superior to cue-based methods, it still fails to extract and fully exploit the content's rich semantic and grammatical information.
- E.g.: The N-gram approach is simple, however it cannot model more complicated contextual dependencies of the text.
- Syntactic features used alone are also less powerful than word-based n-grams and a superficial combination of the two would not be effective in capturing the complex interdependence.

### **Limitations of cue and feature-based methods**

- Various language cues necessitate the creation of a new cue set for each new situation, making it challenging to generalise cues and feature engineering methods across themes and domains.

## 3. METHODOLOGY

### 3.1 MODELS USED:

- **Logistic Regression:**

The logistic function, also called the sigmoid function, was developed by statisticians to describe properties of population growth in ecology, rising quickly and maxing out at the carrying capacity of the environment. It's an S-shaped curve that can take any real-valued number and map it into a value between 0 and 1, but never exactly at those limits.  $\text{sigmoid}(Z) = 1 / (1 + e^{-z})$  Hypothesis  $\Rightarrow Z = WX + B$   $h_{\Theta}(x) = \text{sigmoid}(Z)$

- **Decision Tree Classification :**

Decision Tree is a supervised learning technique that may be used to solve both classification and regression problems, however it is most commonly employed to solve classification issues. Internal nodes represent dataset attributes, branches represent decision rules, and each leaf node provides the conclusion in this tree-structured classifier.

- **Gradient Boosting Classifier :**

Gradient Boosting is a popular boosting algorithm. In gradient boosting, each predictor corrects its predecessor's error. In contrast to Adaboost, the weights of the training instances are not tweaked; instead, each predictor is trained using the residual errors of its predecessor as labels.