# ABSTRACT

Fake News has become one of the major problem in the existing society. Fake News has high potential to change opinions, facts and can be the most dangerous weapon in influencing society.

The proposed project uses NLP techniques for detecting the 'fake news', that is, misleading news stories which come from the non-reputable sources. By building a model based on a K-Means clustering algorithm, the fake news can be detected . The data science community has responded by taking actions against the problem. It is impossible to determine a news as real or fake accurately. So the proposed project uses the datasets that are trained using count vectorizer method for the detection of fake news and its accuracy will be tested using machine learning algorithms

# CONTENTS

# CHAPTER 1

# INTRODUCTION

## 1.1 MACHINE LEARNING AND NLP:

### 1.1.1 MACHINE LEARNING

Machine learning (ML) is the scientific study of algorithms and statistical models that computer systems use to perform a specific task without using explicit instructions, relying on patterns and inference instead. It is seen as a subset of artificial intelligence. Machine learning algorithms build a mathematical model based on sample data, known as "training data", in order to make predictions or decisions without being explicitly programmed to perform the task. Machine learning is closely related to computational statistics, which focuses on making predictions using computers. The study of mathematical optimization delivers methods, theory and application domains to the field of machine learning. "A computer program is said to learn from experience E with respect to some class of tasks T and performance measure P if its performance at tasks in T, as measured by P, improves with experience E." This is Alan Turing's definition of machine learning.

Deep learning is a class of machine learning algorithms that utilizes a hierarchical level of artificial neural networks to carry out the process of machine learning. The artificial neural networks are built like the human brain, with neuron nodes connected together like a web. While traditional programs build analysis with data in a linear way, the hierarchical function of deep learning systems enables machines to process data with a nonlinear approach.

The word "deep" in "deep learning" refers to the number of layers through which the data is transformed. More precisely, deep learning systems have a substantial credit assignment path (CAP) depth. The CAP is the chain of transformations from input to output. CAPs describe potentially causal connections between input and output.

For a feedforward neural network, the depth of the CAPs is that of the network and is the number of hidden layers plus one (as the output layer is also parameterized). For recurrent

neural networks, in which a signal may propagate through a layer more than once, the CAP depth is potentially unlimited.

Deep learning architectures such as deep neural networks, deep belief networks, recurrent neural networks and convolutional neural networks have been applied to fields including computer vision, speech recognition, natural language processing, audio recognition, social network filtering, machine translation, bioinformatics, drug design, medical image analysis, material inspection and board game programs, where they have produced results comparable to and in some cases superior to human experts.

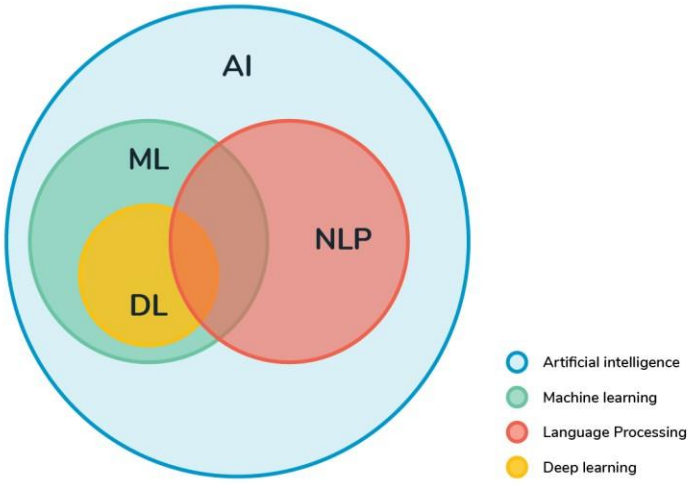


**Fig. 1** : Graphical representation of relationship between various fields in artificial intelligence (source: devopedia.org)

### 1.1.2 NATURAL LANGUAGE PROCESSING

NLP is an area of computer science and artificial intelligence concerned with the interactions between computers and human (natural) languages, in particular how to program computers to fruitfully process large amounts of natural language data.

Integration gives the meaning based on all the sentences given before it.Eg. Consider the sentence “Water is flowing on the bank of the river” But bank has two meanings One Financial Institute and Two River of the bank here System has to consider the second meaning.

#### 1.1.2.1.5 PRAGMATIC ANALYSIS

During this, what was said is re-interpreted on what it actually meant. It involves deriving those aspects of language which require real world knowledge.

## 1.2 MOTIVATION OF WORK

The rise of fake news during the 2016 U.S. Presidential Election highlighted not only the dangers of the effects of fake news but also the challenges presented when attempting to separate fake news from real news. Fake news may be a relatively new term but it is not necessarily a new phenomenon. Fake news has technically been around at least since the appearance and popularity of one-sided, partisan newspapers in the 19th century. However, advances in technology and the spread of news through different types of media have increased the spread of fake news today. As such, the effects of fake news have increased exponentially in the recent past and something must be done to prevent this from continuing in the future.

I have identified the three most prevalent motivations for writing fake news and chosen only one as the target for this project as a means to narrow the search in a meaningful way. The first motivation for writing fake news, which dates back to the 19th century one-sided party newspapers, is to influence public opinion. The second, which requires more recent advances in technology, is the use of fake headlines as clickbait to raise money. As such, this paper will focus primarily on fake news as defined by politifact.com, “fabricated content that intentionally masquerades as news coverage of actual events.” This definition excludes satire, which is intended to be humorous 8 and not deceptive to readers. Most satirical articles come from sources. Satire can already be classified, by machine learning techniques Therefore, our goal is to move beyond these achievements and use machine learning to classify, at least as well as humans, more difficult discrepancies between real and fake news.

The dangerous effects of fake news, as previously defined, are made clear by events in which a man attacked a pizzeria due to a widespread fake news article. This story along with analysis provide evidence that humans are not very good at detecting fake news, possibly not better than chance . As such, the question remains whether or not machines can do a better job.

There are two methods by which machines could attempt to solve the fake news problem better than humans. The first is that machines are better at detecting and keeping track of statistics than humans, for example it is easier for a machine to detect that the majority of verbs used are "suggests" and "implies" versus, "states" and "proves." Additionally, machines may be more efficient in surveying a knowledge base to find all relevant articles and answering based on those many different sources. Either of these methods could prove useful in detecting fake news, but we decided to focus on how a machine can solve the fake news problem using supervised learning that extracts features of the language and content only within the source in question, without utilizing any fact checker or knowledge base. For many fake news detection techniques, a "fake" article published by a trustworthy author through a trustworthy source would not be caught. This approach would combat those "false negative" classifications of fake news. In essence, the task would be equivalent to what a human faces when reading a hard copy of a newspaper article, without internet access or outside knowledge of the subject (versus reading something online where he can simply look up relevant sources). The machine, like the human in the coffee shop, will have only access to the words in the article and must use strategies that do not rely on blacklists of authors and sources. The current project involves utilizing machine learning and natural language processing techniques to create a model that can expose documents that are, with 9 high probability, fake news articles. Many of the current automated approaches to this problem are centered around a "blacklist" of authors and sources that are known producers of fake news. But, what about when the author is unknown or when fake news is published through a generally reliable source? In these cases it is necessary to rely simply on the content of the news article to make a decision on whether or not it is fake. By collecting examples of both real and fake news and training a model, it should be possible to classify fake news articles with a certain degree of accuracy. The goal of this project is to find the effectiveness and limitations of language-based techniques for detection of fake news through the use of machine learning algorithms including but not limited to convolutional neural

networks and recurrent neural networks. The outcome of this project should determine how much can be achieved in this task by analyzing patterns contained in the text and blind to outside information about the world.

## 1.3 PROBLEM STATEMENT

News consumption is a double-edged sword. On the one hand, its low cost, easy access, and rapid dissemination of information lead people to seek out and consume news.It enables the wide spread of “fake news”, i.e., low quality news with intentionally false information. The extensive spread of fake news has the potential for extremely negative impacts on individuals and society. Therefore, fake news detection has recently become an emerging research that is attracting tremendous attention.First, fake news is intentionally written to mislead readers to believe false information, which makes it difficult and nontrivial to detect based on news content.

To develop a FAKE NEWS DETECTION system using natural language processing and its accuracy will be tested using machine learning algorithms. The algorithm must be able to detect fake news in a given scenario.

# CHAPTER 2

# LITERATURE SURVEY

## 2.1 INTRODUCTION

In the world of rapidly increasing technology ,information sharing has become an easy task. There is no doubt that internet has made our lives easier and access to lots of information. This is an evolution in human history, but at the same time it unfocusses the line between true media and maliciously forged media. Today anyone can publish content – credible or not – that can be consumed by the world wide web. Sadly, fake news accumulates a great deal of attention over the internet, especially on social media. People get deceived and don't think twice before circulating such mis-informative pieces to the world. This kind of news vanishes but not without doing the harm it intended to cause. The social media sites like Facebook, Twitter, Whatsapp play a major role in supplying these false news. Many scientists believe that counterfeited news issue may be addressed by means of machine learning and artificial intelligence.

Various models are used to provide an accuracy range of 60-75%. Which comprises of Naive Bayes classifier,Linguistic features based, Bounded decision tree model, SVM etc. The parameters that are taken in consideration do not yield high accuracy. The motive of this project is to increase the accuracy of detecting fake news more than the present results that are available. By fabricating this new model which will judge the counterfeit news articles on the basis of certain criteria like spelling mistake, jumbled sentences, punctuation errors , words used .

## 2.2 REVIEW OF LITERATURE

There are two categories of important researches in automatic classification of real and fake news up to now: