# Text, logo Description automatically generated

# Classifying Misinformation using Deep Learning

## Project Report

Logo

Description automatically generated

## Vamshee Deepak Katta

## MSBA

MIS 64061

## Introduction

Misinformation is a type of propaganda wherein false or inaccurate information is provided which is deliberately intended to deceive. Even if later retracted, misinformation can continue to influence actions and memory. The role of social media has made information readily available to us at any time, and it connects vast groups of people along with their information at one time. Advances in technology has impacted the way we communicate information, and the way misinformation is spread. People often perceive whatever they read/heard as True, and it is affecting the world on a large scale, both politically and financially. To tackle problems that arise due to such misinformation, there is need to ensure that the information that we absorb corresponds to truth. This project gives us a way to tackle such misinformation.

## Summary

As stated above, misinformation, nowadays, can cause tremendous issues if used with negative intentions. It is used for personal/political advantage cashing on public perception to believe in the information read/heard over news or social media without fact checking the information. This leads to misleading, loss of faith and incorrect decisions due to misinformation.

By implementing this project, we will be able to distinguish between true information and false information. This will help us make wise decisions with the knowledge of which information is true and help ensure others have the right information to make decisions of high proportions that impact others.

To conduct this project, a large dataset consisting of over 72,000 news articles with over 35,000 real and 37,000 fake news has been used.

As a contribution to the project, I will be assessing the underlying data in how any kind of news may be classified as either misinformation or true information. This could help us further develop models that could classify the information better and help make faster decisions giving out the reasons to why an information can be classified as misinformation.

## Problem Statement

In current socio-economic situation, where information is abundantly available from innumerous sources which can be either giving out true information or forwarding claims and information which is not fact-checked, we can be overwhelmed with so many clashing claims of information on same subject. This can lead to incorrect decisions due to ambiguity and can further lead to problems arising from such incorrect decisions.

Especially when someone in a position of power, who can influence future of many others, has to make important decisions, misinformation can lead to disasters that might have been averted with accurate and factual information. With developing technology, there is also increase in misinformation in forms such as deep fakes of people which further increase the ambiguity in available information and makes it harder to make wise decisions. We can tackle such misuse of information with technology as well.

## Technique and Technology

For this project, I have used the technique of semi-supervised learning method using BERT model and TF-IDF model.

### BERT

BERT stands for Bidirectional Encoder Representations from Transformers. It consists of several Transformer Encoders stacked together. It uses Transformers to understand the contextual relation between words in a sentence/text. BERT Transformer generally has two mechanisms: An encoder that reads the text input and a decoder that predicts for a given task. The model is trained for a specific task that enables it to understand the patterns of the language. BERT works by randomly masking word tokens and representing each masked word with a vector based on its context. A typical BERT model can be seen as below

Graphical user interface, diagram

Description automatically generated

### TF-IDF

In information retrieval, TF–IDF, short for term frequency–inverse document frequency, is a numerical statistic that is used to reflect how important a word is to a document in a collection. It can be defined as the calculation of how relevant a word is to a text in a document.

The above two models can help us understand on how words are put into context in each of the models giving us insights into how information is read and interpreted to predict the truthfulness of information. The BERT model is presumed to be the best in industry for the purpose of text classification or NLP currently.

## Conclusions

Upon applying BERT model to the available dataset, we were able to classify the input into true information and misinformation. The model is able to classify misinformation with overall accuracy of above 95%. This means that this model can help classify misinformation and provide useful true information that can be used in decision making with high confidence of it being the correct action.

This can help all kinds of media to provide accurate information by dismissing all the misinformation, filtering the abundant information using this model. It can also help the general public in making important personal and political decisions which can lead to general development of the society as a whole.

Though misinformation can be useful in helping make distinction between right and wrong, and be entertaining, it can also lead to very bad situations when used in matters of high importance and influence on the society. Hence, we need to ensure that misinformation is not allowed to spread in situations where it is imperative that correct information is available to make decisions.

## Contributions

To help tackle misinformation and to contribute to the development of the techniques further to solve the problem, I am assessing analyzing the underlying data in the classification of misinformation to better understand how the models classify misinformation on the basis of word association and frequency, in case of TF-IDF and on the basis of contextual relationship between words in case of BERT.

Understanding this information can help improve the models to better classify the input and provide highly accurate predictions on whether an input is misinformation or not.

## Appendix

1. <https://www.analyticsvidhya.com/blog/2022/03/fake-news-classification-using-deep-learning/>
2. <https://ieeexplore.ieee.org/document/9395133>
3. <https://pub.towardsai.net/fake-news-detection-using-bert-model-python-de005c5809ed>