ABSTRACT:

The widespread use of the internet has given rise to media and with it the use of social media platforms for everyday use. Using social media has kept people updated faster nowadays than TV channels and newspapers can keep up with. This is because of how fast data can change and with it the news. This also then gives rise to rumors and fake news.

The purpose of this project is to create an NLP model that can classify fake news as true or false. The use of python libraries to perform tokenization and extraction of text data as well as selection methods will be able to show the best fit features of the model to obtain its highest precision.

PURPOSE:

The purpose of a fake news detection project is to develop a system or algorithm that can accurately distinguish between real and fake news articles or stories. The spread of fake news can have a significant negative impact on individuals, communities, and society as a whole, leading to misinformation, polarization, and even dangerous outcomes. By detecting and filtering out fake news, a detection project can help promote the spread of accurate and trustworthy information and support the goal of building a more informed and responsible society. Furthermore, fake news detection projects can be useful for journalists, fact-checkers, and other professionals who need to verify the accuracy of news and information before publishing or sharing it with others.

PROJECT INTRODUCTION:

Fake news consists of misleading information that needs to be checked. Misleading information can be dangerous as it could cause potential problems between different parties. Difference in statistics and exaggerated service costs of a country can cause unrest. However, their scope is so limited because they depend on human manual detection. In a globe with millions of articles either removed or being published every minute, this cannot be accountable or feasible manually. A solution could be by the development of a system to provide a credible automated index scoring or rating for credibility of different publishers and news context.

This project proposes to create a model that will detect if an article is authentic or fake based on its words, phrases, sources and titles, by applying supervised machine learning algorithms on an annotated (labeled) dataset, that are manually classified and guaranteed. Then, feature selection methods are applied to experiment and choose the best fit features to obtain the highest precision, according to confusion matrix results. We propose to create the model using different classification algorithms. The product model will test the unseen data, the results will be plotted, and accordingly, the product will be a model that detects and classifies fake articles and can be used and integrated with any system for future use.