**COMSATS University   
Park Road, Chak Shahzad, Islamabad Pakistan**

**Semester Project**

**(Final Document)**

**for**

**Fake News Detector**

***By***

**Shariq Ahmed CU/SP18-BCS-151/ISB**

**Umar Khalid CU/SP18-BCS-164/ISB**

**Waleed Butt CU/SP18-BCS-170/ISB**

***Course Instructor*Mr. Usman Yaseen**

Class/Section: BCS-6B

Submission Date: 12/25/2020

*Bachelor of Science in Computer Sc­ience (2018-2022)*

# Introduction

In the following analysis, we will talk about how one can create an NLP to detect whether the news is real or fake. Nowadays, fake news has become a common trend. Even trusted media houses are known to spread fake news and are losing their credibility. So, how can we trust any news to be real or fake?

# Dataset

1) train.csv: A full training dataset with the following attributes:

2) id: unique id for a news article

3) title: the title of a news article

4) author: author of the news article

5) text: the text of the article; could be incomplete

6) label: a label that marks the article as potentially unreliable. Where 1: unreliable and 0: reliable. (Verma, 2019)

# Methodology

## What is Fake News

Before we get into the ways of detecting fake news, we need to know what the fake news are. Fake news encapsulates pieces of news that may be hoaxes and is generally spread through social media and other online media. This is often done to further or impose certain ideas and is often achieved with political agendas. Such news items may contain false and/or exaggerated claims, and may end up being viralized by algorithms, and users may end up in a filter bubble.

## All Steps:

Following are the steps involved in this machine learning model:

* First we loaded the dataset and then we separated features and targeted columns.
* We did some data pre-processing after separating our features and targeted columns.
* In data pre-processing, we checked null values and then we dropped these values from our dataset.
* Then we separated independent features from our dataset. And after that we added some checks to our dataset like news must be in alphabets, news must not contain more than 5000 words.
* We converted our news to one hot representation for better modeling.
* We added some padding to our datasets so that the length of sentences in news will remain same.
* After data pre-processing, we started building our machine learning model.
* We first imported tensorflow and keras as we are using them to build our model.
* We created our model with 40 features.
* The model we used is sequential, along with activation function sigmoid.
* We used binary cross entropy as our loss function and adam as our optimizer.
* After building our model, it’s time to train our model.
* We split our dataset into testing and training. Test size is 0.10 and the random state is 42.
* Then we fit out dataset into the model, with 20 epochs and batch size of 64 in each epoch.
* At the end, we did some predictions on an unseen dataset.
* At first, we did some data pre-processing on our test dataset and made it same as the training dataset i.e only alphabets allowed, then converted to one hot encoding, then did some padding to the sentences etc,
* After test dataset in pre-processed, we did some prediction on our test dataset.

# References

data-flair. (2020). *Advanced Python Project - Detecting Fake News*. Retrieved from data-flair.training: https://data-flair.training/blogs/advanced-python-project-detecting-fake-news

Solutions, P. T. (2019, November 22). *Fake News Detection Using Python*. Retrieved from pantechsolutions.com: https://www.pantechsolutions.net/fake-news-detection-using-machine-learning