

# **Project Summary**

**Group Name:** Bhartiya Miners

**Group Members:**

Aranya Aryaman (170101011), Sayak Dutta (170101061),  
Shubham Kumar (170101064), Shyam Sundar Rav (170101065)

**Teaching Assistant:** Mr. Sujit Kumar

**Project Category:** Fake News Detection

## **Phase - 1**

- 1) Studied & understood the paper “Detecting Cross Modal Inconsistency to defend against Neural Fake News”.
- 2) Implemented the paper using one-directional LSTM & bi-directional LSTM and ran the code on a subset of Good News Dataset provided in the project repository.
- 3) Gave a presentation explaining the code-flow using block diagrams to mark Phase-1 completion.

## **Phase - 2**

- 1) In the first meeting, we presented an idea of implementing three different models, one for detecting political fake news (online sources), one for detecting clickbait & one for detecting fake news articles.
- 2) For Political Fake News, we proposed a model where we would be using two separate models for text and images. As a part of processing text, we performed tokenization, stemming of data, punctuation removal, expanding URLs & doing an analysis over the political tweets. For images, we decided to stick with using a RNN & bi-directional LSTM model to find semantic relations b/w the text and image. Both the parts have been implemented & trained on over 32000 data samples. The data used in this module was generated using Streaming API.

- 3) For Clickbait, we proposed to use a bi-directional LSTM attention model & fine-tune it over the data samples. We couldn't implement it due to lack of time & data.
- 4) For fake news articles, we implemented a matlab based program to do an overall analysis of the fake news articles only. In this analysis, we found the most frequent words, mentions & citations along with finding the general temperament of the tweets. In the analysis, it was found that the fake tweets mostly had a negative sentiment (verified by NSR). The dataset was the standard twitter training dataset available publicly on the internet.
- 5) Besides the above, we tried to contact a few news agencies including the likes of Lallantop to create an Indian News Dataset since most of the publicly available dataset are based on data provided by Western Media. We also considered using models like R/fakeedit but an internet survey proved that Indian Users didn't use Reddit as a frequent medium to discuss things & thus, such a dataset was of no use from an Indian Fake News Detection Perspective.