

---

# Fact or Fake

## A fake news detection classifier

---

Aayush Saxena, Harsh Agrawal, Kanchan Bisht, Sai Kiran Mayee Maddi  
asaxena6(200262043), hagra2(200257388), kbisht(200203292), smaddi(200257327)

### 1 Datasets

1. Kaggle - Getting Real about Fake News
2. Kaggle - Fake News : Build a system to identify unreliable news articles
3. Github - Fake and Real News Dataset

### 2 Idea

In today's world, the outreach of online media has been extremely massive. Spreading false news and fake propaganda has become easier than ever and is being used for political and commercial purposes. The impact of such fraudulent propagation can adversely impact and influence people.

We aim at training a classifier which classifies news as real or fake based upon features like the originating source, text of the article, Facebook metrics etc. We plan to use Natural Language Processing to extract more features from the text of the document to gauge the prevalent patterns in a fake or real document, to enhance the classifier.

### 3 Software

We plan to design a Python model using libraries skikit-learn, NumPy, SciPy, Matplotlib, Pandas.

### 4 Papers

1. The current state of fake news: challenges and opportunities
2. FakeNewsTracker: A Tool for Fake News Collection, Detection, and Visualization
3. Automatic Detection of Fake News

### 5 Teammates and Work Division

We will individually do exploratory data analysis to extract more features from the text. Simultaneously we will split the different aspects of data preprocessing: data cleaning, data imputation, data transformation and others among all the team members. We plan to implement and evaluate different classifiers Decision Trees(Aayush), Naive Bayes(Harsh), SVM(Kanchan), KNN(Kiran) and create an ensemble model.

### 6 Midterm Milestone

We plan to preprocess the data, select relevant features and start training the model for an initial analysis of the problem.