

# FAKE NEWS DETECTION (Team: Trace)

→ References/ Inspirations used in our project:

→ Research Paper Reference:

- ◆ <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8485117/#Sec3title>
- ◆ <https://www.sciencedirect.com/science/article/pii/S1877050918312559>
- ◆ [https://link.springer.com/chapter/10.1007/978-3-030-49186-4\\_34#Fn4](https://link.springer.com/chapter/10.1007/978-3-030-49186-4_34#Fn4)
- ◆ <https://link.springer.com/article/10.1007/s10207-022-00625-3>

## 1) Hong Kong Protests: Using Natural Language Processing for Fake News Detection on Twitter ([reference](#))

### ● Section-3.2 Real News Dataset

- To address the challenge of real news dataset, the real news dataset has to be constructed from the ground up. Tweets from the accounts of trusted news agencies and govt agencies are retrieved, rather than articles.

### ● Section-3.4 Algorithms

- The research paper discusses Four different algorithms used for training and evaluation: Naive Bayes, SVM, C4.5, and Random Forests.
- For SVMs, the Radial Basis Function kernel is used, and parameter optimization is performed using grid search.
- For Naive Bayes and SVMs, all feature values are normalized to a range of 0-1 using a transformation formula.
- Decision trees are left unprocessed due to the difficulty in understanding them.

## 2) Detection of Turkish Fake News in Twitter with Machine Learning Algorithms

(reference)

### • Section- Literature Review:

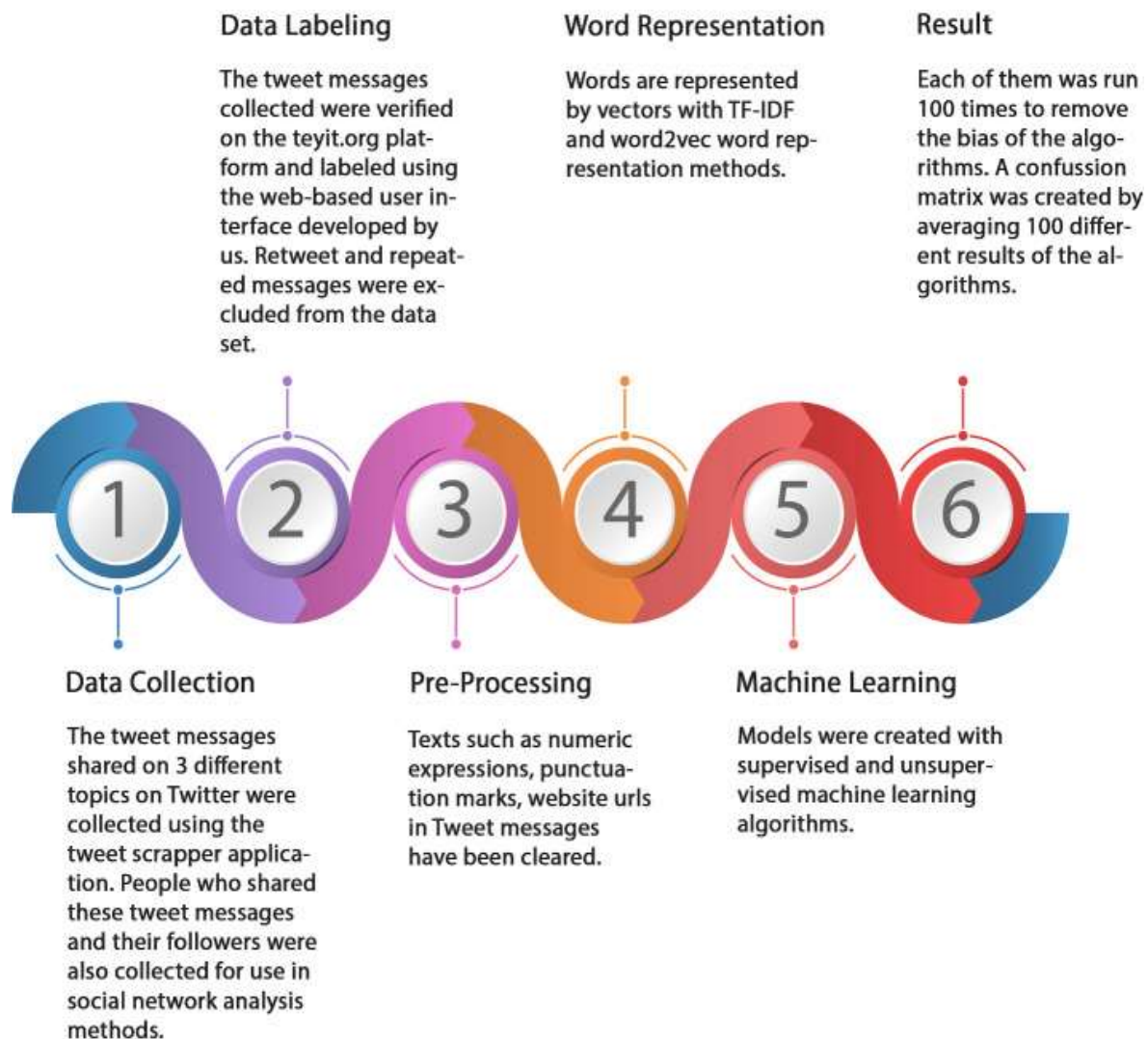
- Studies have been conducted to detect fake news using non-Artificial Neural Network (ANN) based supervised learning algorithms.
- The algorithms used in these studies include Naive Bayes, Linear Regression, K-Nearest Neighbors, Support Vector Machines, Stochastic Gradient Descent, and HBLC.
- Among these algorithms, the Support Vector Machines algorithm is found to be more successful in detecting fake news compared to other algorithms.
- Conclusion: Non-ANN-based supervised learning algorithms, such as Support Vector Machines, can be effective in detecting fake news.

Table 1

Fake news detection studies made by using non-ANN-based supervised learning algorithms.

Refs.	Dataset	ML	Success ML	Performance measure	Best result
[10]	Mobile phone reviews from mobile01.com	LR, SVM	SVM	F1-score	0.61
[74]	Hotel reviews, restaurant review, gay marriage, and gun control, fake and real news articles from kaggle.com	SGD, SVM, KNN, LR, DT	SVM	Accuracy	0.9
[75]	3 large Facebook pages each from the right and from the left and Facebook pages of Politico, CNN and ABC News	NB	NB	Accuracy	0.75
[76]	The authors present a list of Facebook pages divided into two categories: scientific news sources and conspiracy news sources.	LR, HBLC	HBLC	Accuracy	0.99
[11]	Articles on sport, politics, rumor, health and other were collected with web crawler.	SVM, NB	SVM	F1-score	0.79
[77]	2 satirical news sites (The Onion and The Beaverton) and 2 legitimate news sources (The Toronto Star and The New York Times): varying across 4 domains (civics, science, business, and "soft" news)	SVM	SVM	F1-score	0.87
[78]	Collecting legitimate news from mainstream news websites such as	SVM	SVM	F1-score	0.73

• Reference ([Fig.2](#))



### 3) Identifying Fake News and Fake Users on Twitter ([reference](#))

#### ● Section-3 Use Cases:-

#### Use Case 2 – The Second Real Tweet

Tweet text: “Donald Trump uses State of the Union address to say: “There has never been a better time to start living the American dream””.

We have obtained the following result and statistics (see Table 2):

Table 2. The result returned by the NLP services for the second tweet.

Twitter text	Named entities	Topics - confidence	Sentiment score
Donald Trump uses State of the Union address to say: "There has never been a better time to start living the American dream"	Donald Trump	Labor - 0.865	Text - 0 Emoji - -1 Hashtag - -1

```
stats: "Tweet is not fake. Confidence: 50.0. Only one source found.",
stats_code: -50,
hashtag_sentiment: "-1",
similar_tweets: {
  nr_analyzed_tweets: 84,
  tweets: [
    {
      Donald Trump uses State of the Union address to say: "There has never been a better time to start living the Americ...: {
        similiraty_entities: 1,
        user_name: "BBC News (World)",
        user_descripton: "News, features and analysis from the World's newsroom. Breaking news, follow @BBCBreaking. UK news,
        similiraty_text: 0.8285714285714286
      }
    }
  ],
  nr_verified_users: 1
},
```

Fig. 5. The content of the JSON returned by the main service for the second tweet

The tweet is not fake, with a 50% confidence, because a single tweet (posted by a verified user - BBC News) was found with a text similarity over 25% (see Figure 5). The searched tweet has a single entity and the similar tweet contains this entity, it turned out that the trust score is 50 points (10 points for text similarity over 25% and 40 points for identical entities). A single source of trust has been found and for that, the *stats\_code* is -50.

## • Section-2.1- General Architecture of the System

### 2. Proposed solutions

#### 2.1 General Architecture of the System

The general architecture of the system is presented in Figure 1. There is a **Twitter crawler** component, which collects tweets and adds them to our **database**. When we will need tweets from trustworthy sources to compare with our current one, we can retrieve them directly from our database. The **Processing module**: when a user wants to know the credibility of a new tweet, he inputs the link of the tweet in our interface. Our algorithm then uses an **NER** (Named Entity Recognition) **component**, which split the text into its composing parts: it brings out the entities (generally, nouns and their relative importance in the context), the topics, the social tags, the overall tweet sentiment and the hashtag sentiment.

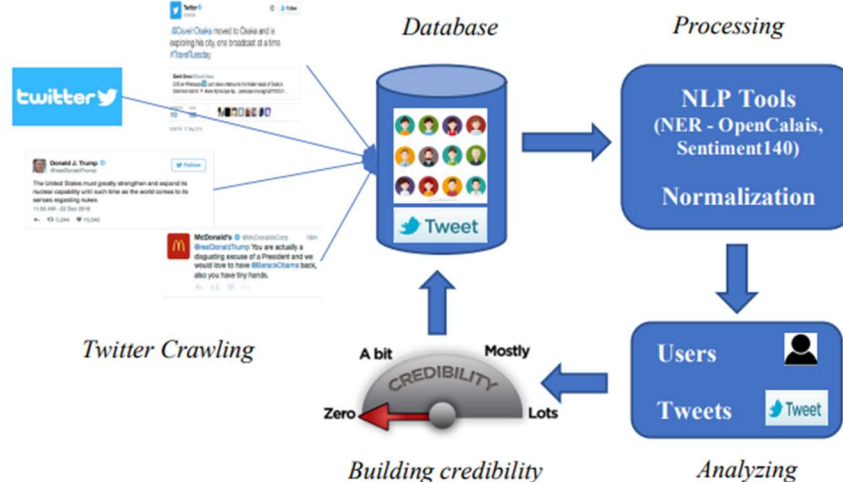
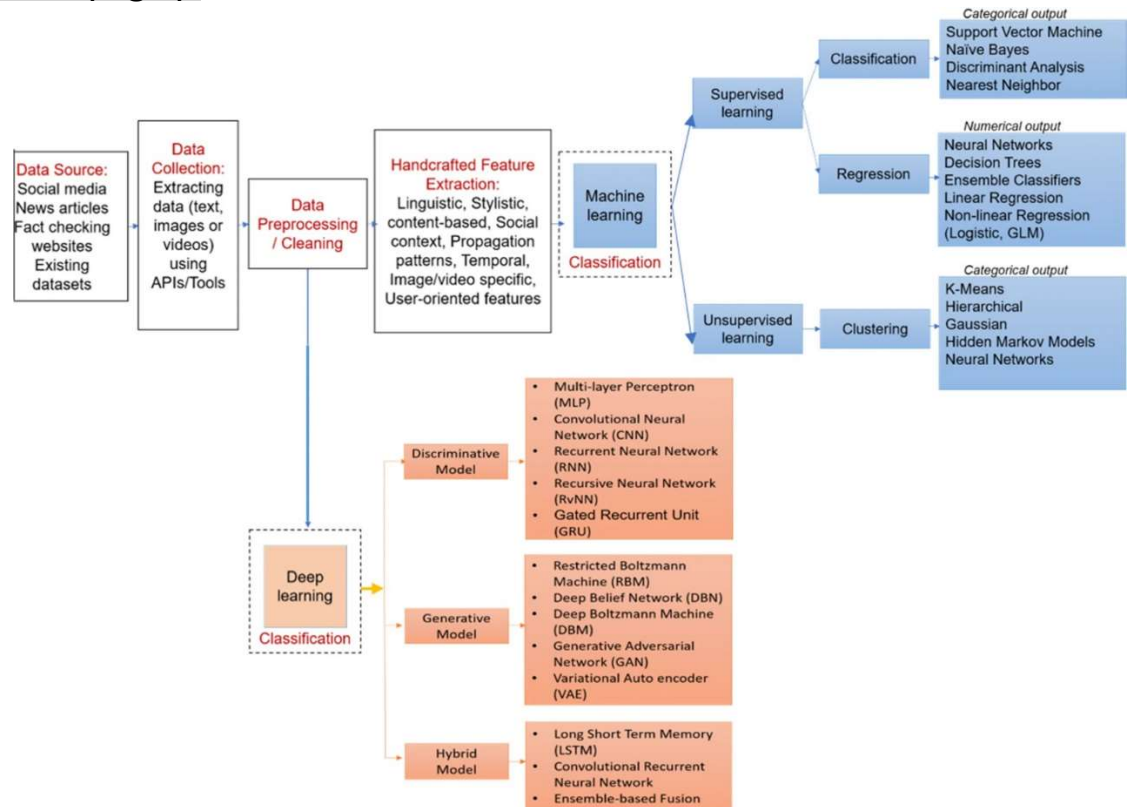


Fig. 1. System architecture

#### 4) A review on fake news detection 3T's: typology, time of detection, taxonomies ([website](#))

- Reference(Fig.2)-



-----END-----