

Covid-19 Fake News Detector

By :

Sumedh Manjrekar

Pinak Jani

Sameeksha Bhatia

Kavya Moolya

K.J Somaiya College Of Engineering, Mumbai

Team ID: #1950

800 million



FAKE NEWS



Idea Implementation

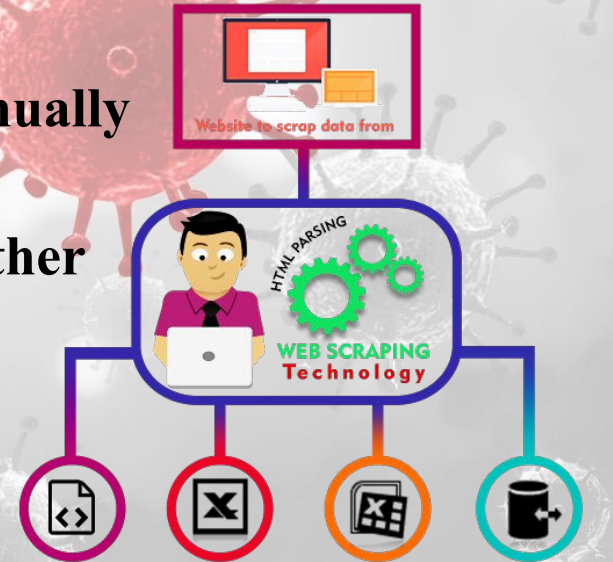
The idea implementation consisted of following parts:

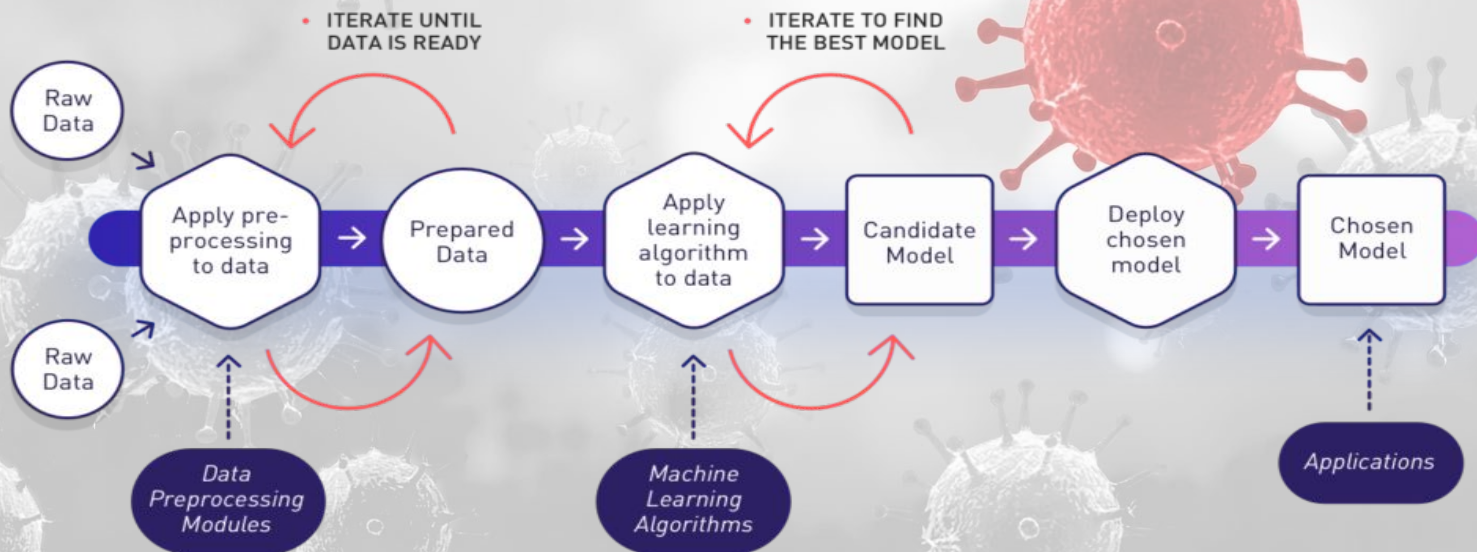
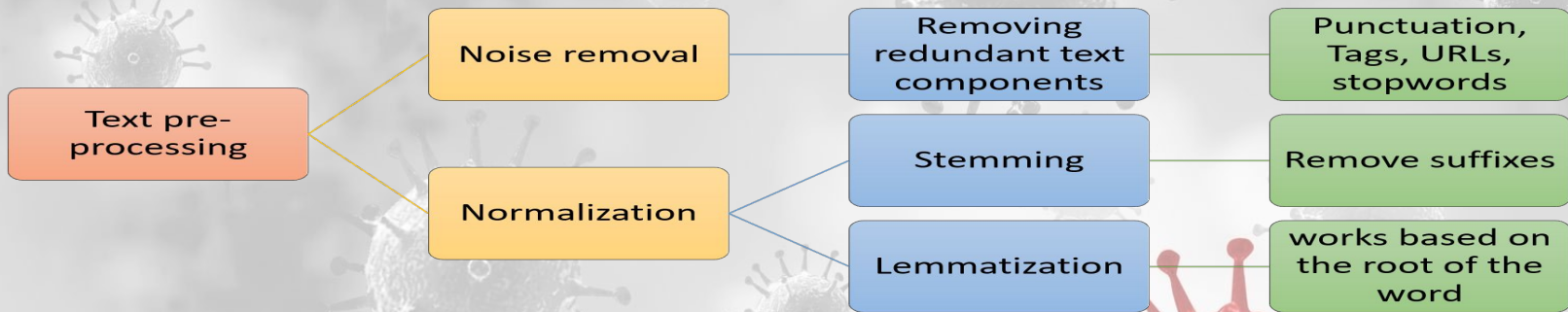
- 1. Web Scrapping and data collection.**
- 2. Natural language Processing on dataset using Python.**
- 3. Running Machine Learning Algorithms.**
- 4. Integrating to Web Application with GUI.**



Web Scrapping

- For Training the ML model the data for fake and genuine news related to covid-19 is necessary.
- This data is achieved by web scrapping the news from websites or other social media platform.
- The data is not only web scrapped but also manually fed from Whatsapp forwards.
- The data scrapped is stored in a csv file for further use.





Natural Language Processing using Python

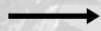
Aa **Lowercasing**



All uppercase characters are lowercased



Tokenization



Split sentences into single words/tokens

.,?!%

Remove punctuation



Remove comma, question marks, etc.



Remove stopwords



Remove most frequent words in a language e.g “this”, “that”, “and”, etc.

1234

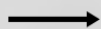
Remove numbers



Remove numeric characters



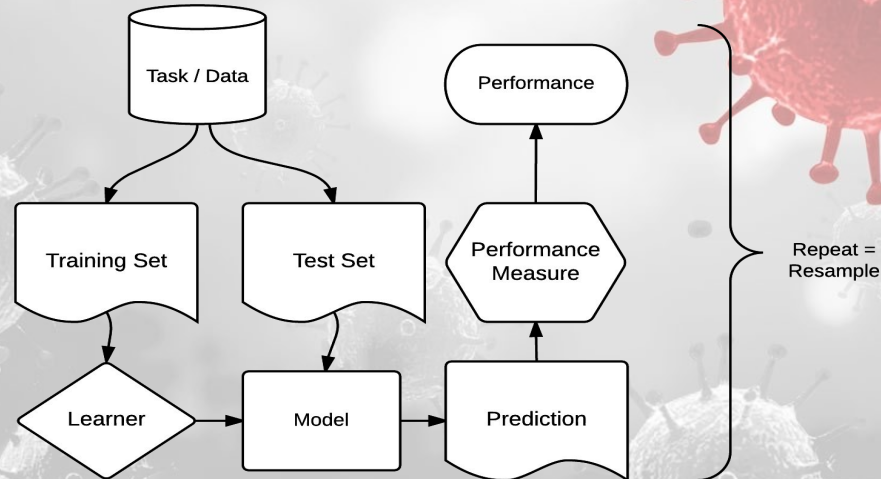
Stemming or lemmatization

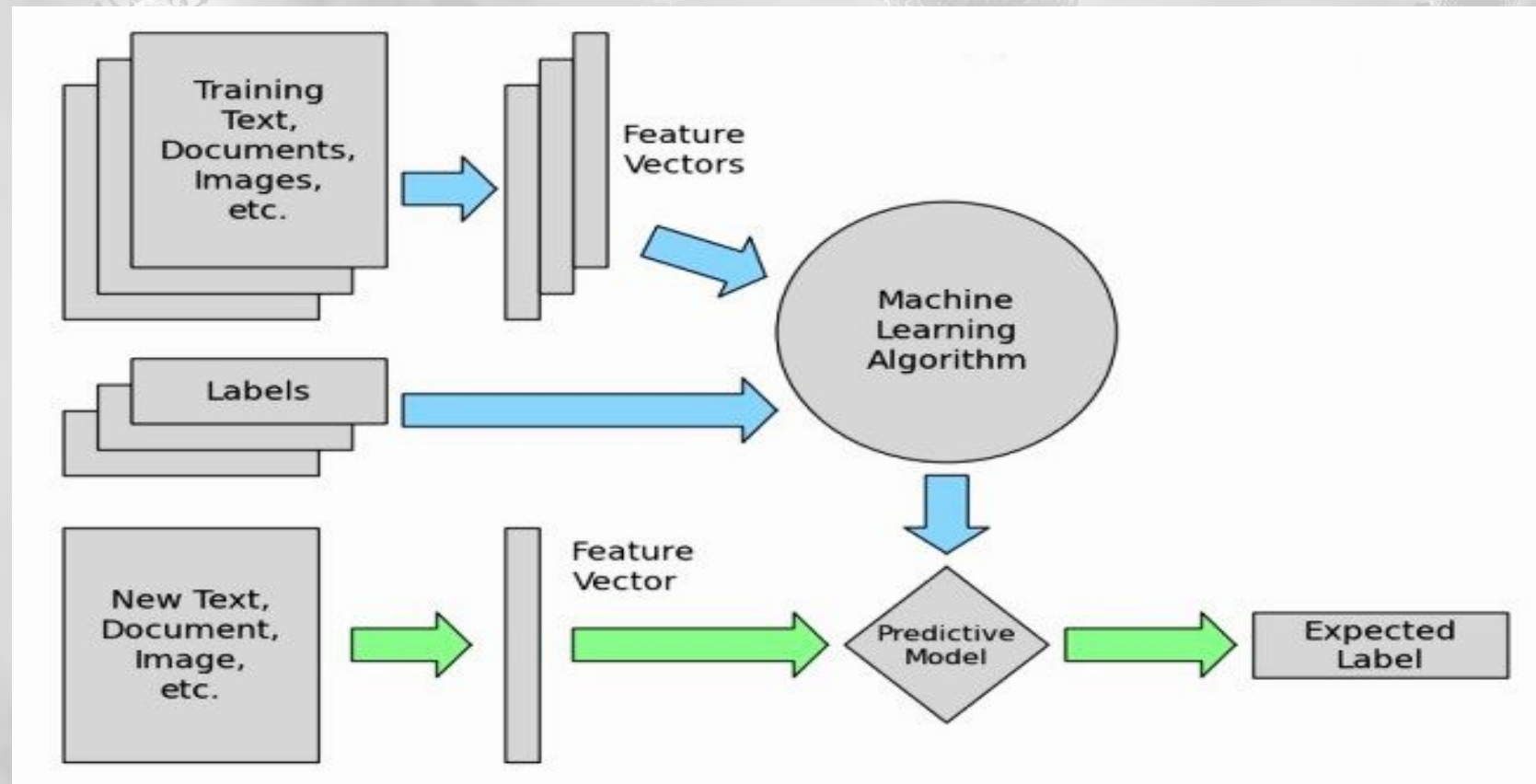


Extracting the root of verbs e.g “running” -> “run”, “eating” -> “eat”

Machine Learning

- After the data collection and implementing NLP over the dataset, we generated a bag of words and their labels.
- This dataset is then trained using ML model to predict the fake news.
- The ML model used is 'Naive Bayes model', by considering its accuracy.
- After training the model, it is then used to predict output for any custom input.





Challenges Faced

- 1) **Collection and Scrapping of data from various sources like websites, Whatsapp forwards and other social media platforms**
- 2) **Verifying the scrapped dataset and manually labelling them 'Fake' or 'Genuine'**
- 3) **Increasing the ML model accuracy from '60%' to '90%'**
- 4) **Integrating the project to a web app with GUI using Django.**

Completion Status

- The project is ready and implementable for web application with a GUI.
- The project is also tested for current fake Whatsapp forwards and provides promising results.
- The current accuracy of the detection is 90% and could be increased further by increasing the size of dataset.
- The current version of the project targets fake text messages and not images, audio or video.
- This prototype is trained only for English language, and not for other regional languages.