T5 Online Data Science BootCamp

# Sentiment Anlaysis

covid-19 tweets

Taif AlWafi
Computer science graduate

# Table of content


1 Introduction

2 Problem statement

3 Dataset

4 EDA

5 Data preprocessing

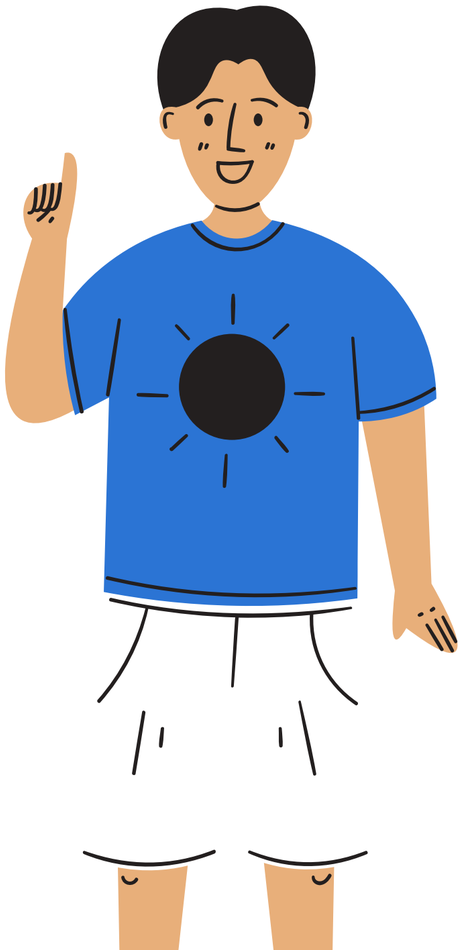
6 Model

# Introduction

In late December 2019 a **coronavirus** was identified in China causing severe respiratory disease including pneumonia. Impacts of the pandemic, have threatened people's mental health. This project aimed to use classification models to predict how a person feels about the coronavirus.

# Problem statment

In this project, I apply a **tweets sentiment** analysis model that helps people detect their sentiment about the pandemic and contacting with a clinic to avoid risk mental health.

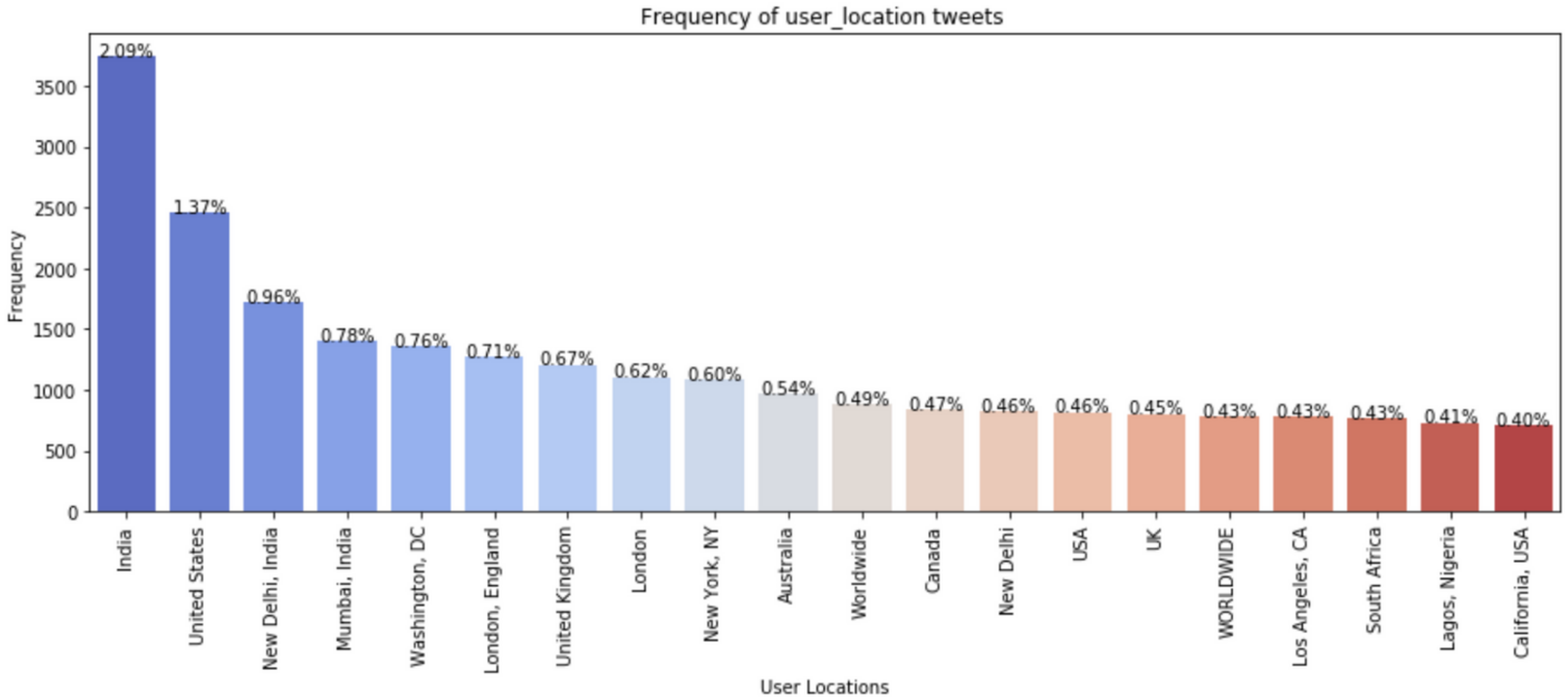


# Dataset

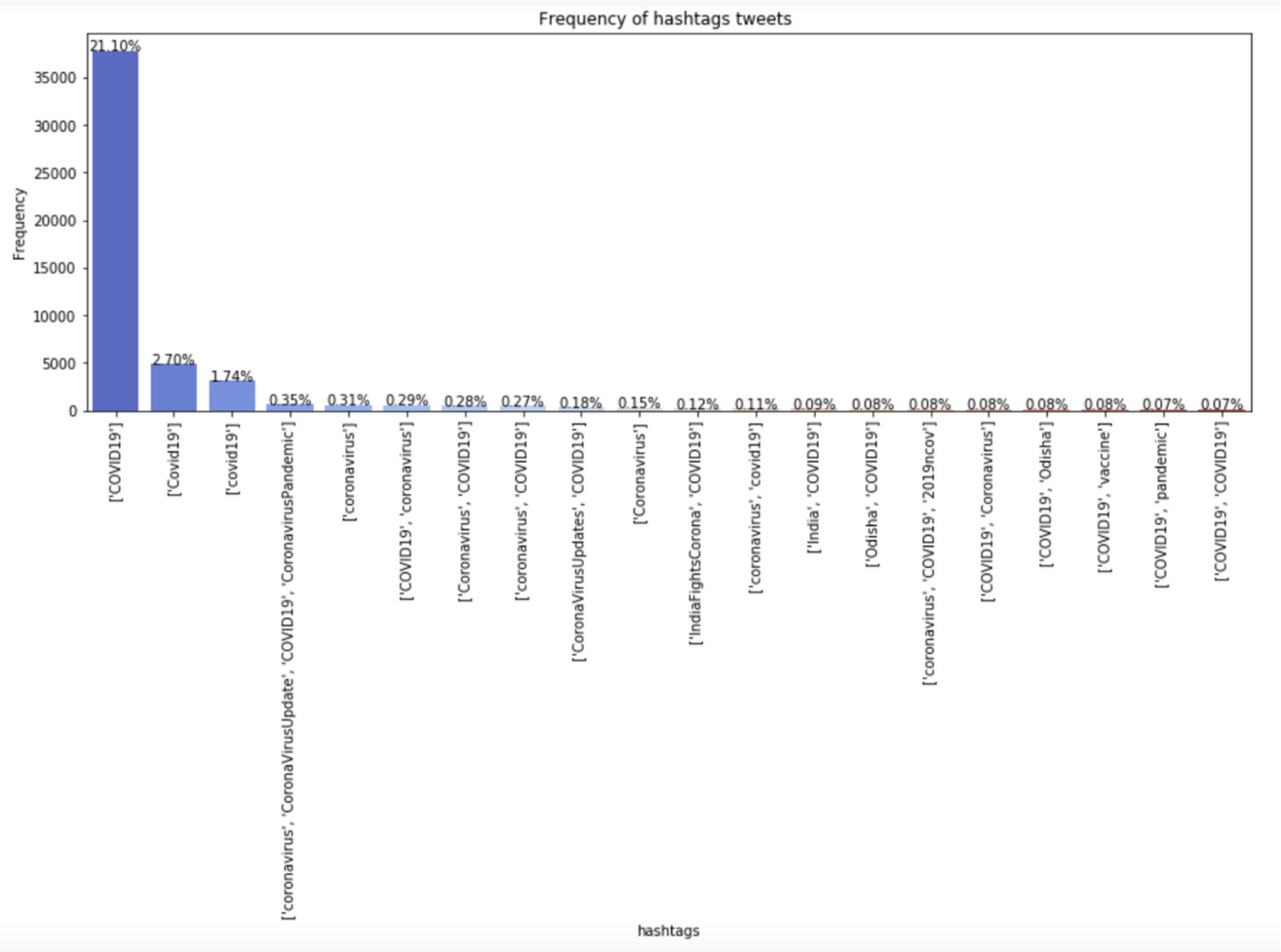
The provided dataset (COVID19 Tweets) consists of 179108 tweets (179108 rows × 13 columns). The dataset can be found at Kaggle.

**'user_name'`, 'user_location', 'user_description', 'user_created', 'user_followers', 'user_friends', 'user_favourites', 'user_verified', 'date', 'text', 'hashtags', 'source', 'is_retweet'**

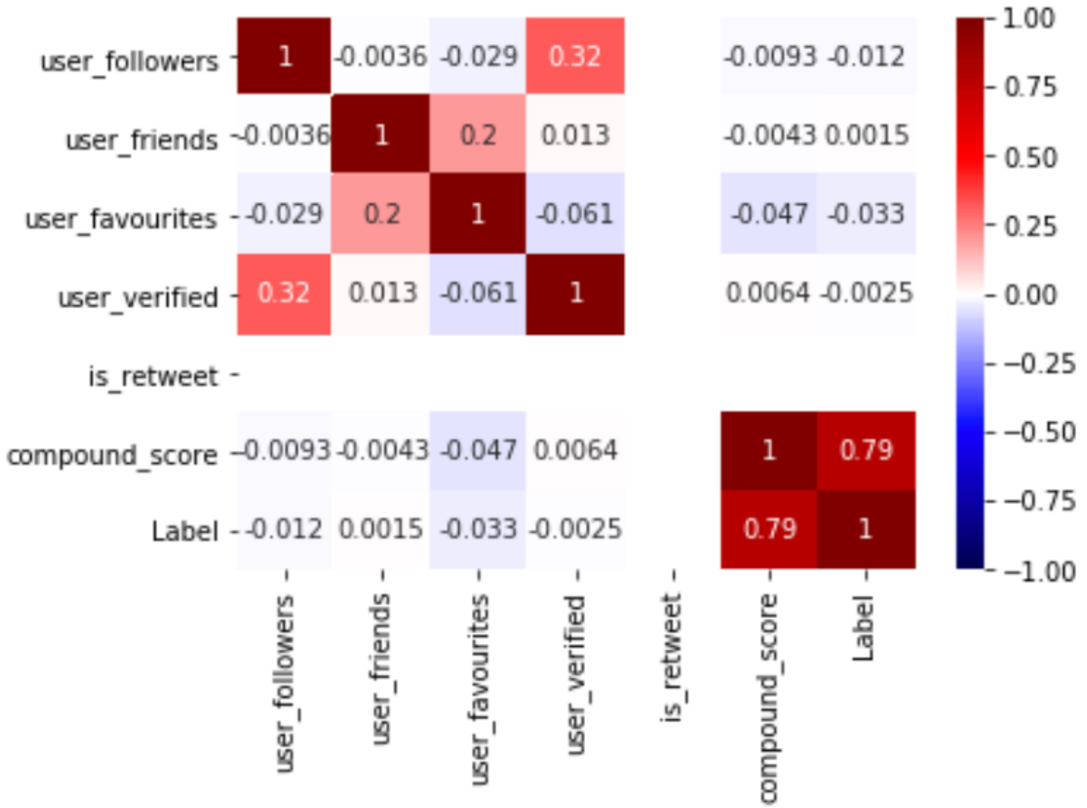
# Exploratory Data Analysis



# Exploratory Data Analysis



# Exploratory Data Analysis



# Data Preprocessing

clean data



labelling data

dividing the data

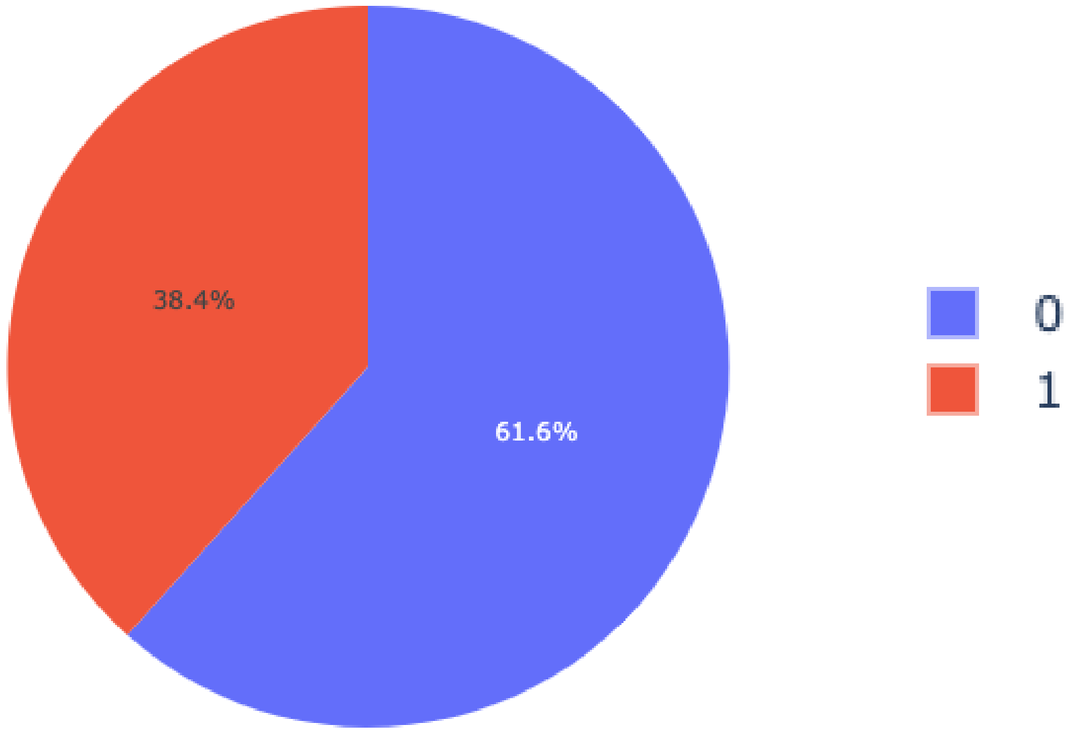
# Data Preprocessing

**Preprocessing the text feature using the following NLP and re techniques :**

- Converting to lowercase
- Remove text in square brackets,
- Remove links,
- Remove punctuation
- Remove words containing numbers
- Removing stopwords
- Lemmatization

# How do you feel about Coronavirus ?

# How do you feel about Coronavirus ?



# Finally

Machine learning model

# Models

1. Naive bayes → Training: 91.18% &Test set: 84.76%
2. Logistic regression → Training: 99.03% &Test set: 94.36%
3. Neural network → Training: 86.68% &Test set: 86.15%
4. Support Vector machine → Training: 96.22% &Test set: 93.96%

# Thanks for listening