# Corona Tweet Classification System

**SYNOPSIS**

OF MAJOR PROJECT
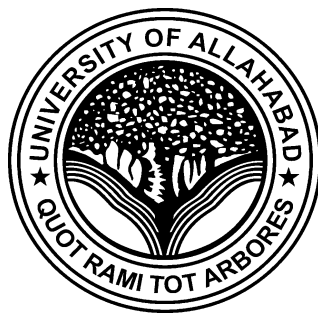
**BACHELOR OF COMPUTER APPLICATIONS**

SUBMITTED BY

Ahmad Faraz Ansari (B.C.A. 6$^{th}$ Sem)
Batch Year – 2019 -22
Enrollment No. - U1946007

**PROJECT GUIDE- Ms. Shreya Agrawal**



**Centre of Computer Education & Training**
**Institute of Professional Studies**
**University of Allahabad, Prayagraj**
**Uttar Pradesh**

## INTRODUCTION

Corona Tweet Classification System is a Machine Learning based Tweet Classification System.

It developed using Python 3 for platform-independency.

It is one of the most accurate and precise Tweet Classification System.

The proposed System is able to classify the Tweets based on the magnitude of emotions like Positivity or Negativity or Neutrality underlying it.

## PROBLEM DEFINITION

Supervised Machine Learning and Classification.

Based on the input (Tweet), the System can classify it in the given classes.

## ALGORITHM

Algorithm used by the System is **Support Vector Machine (SVM).**

## OBJECTIVES

The prime objectives for the Corona Tweet Classification System are as follows:

- To create a System capable of Classifying Tweets.
- Creation of System that has great chance of accuracy and precision.
- Classification is done on the basis of severeness of the underlying emotion.

## OVERALL DESCRIPTION

The proposed Corona Tweet Classification System will take on other Classification System based by tackling the basic underlying problem.

It will take care of all the user resources without requiring any user Interaction.

User is not supposed to get into the details of System's Working or its underlying complexities and its model (abstraction).

## USER CHARACTERISTICS AND ASSUMPTIONS

- User is supposed to have basic knowledge of a computer.
- User has to provide a valid input to gain a valid output.
- User should always try to provide a valid input i.e., Tweet.
- User is required to be patient especially if he/she is using the System on an older hardware.

## USER REQUIREMENTS

- System should be able to handle all logic effectively.
- System should not throw any unchecked exceptions.
- System should be able to determine output i.e., the positivity or negativity of Tweet without any ambiguity.
- System should not crash without providing any concrete result.

# REQUIREMENTS ANALYSIS

The System is designed to work on any machine with newer hardware to harness the maximum power of modern Hardware.

**Hardware Requirements**

- 4GB of RAM or more.
- i3 - 7$^{th}$ Generation Processor or later.
- No External GPU Required.

**Software Requirements**

- Python 3 Interpreter has to be pre-installed.
- PyCharm or Jupyter Notebook pre-installed
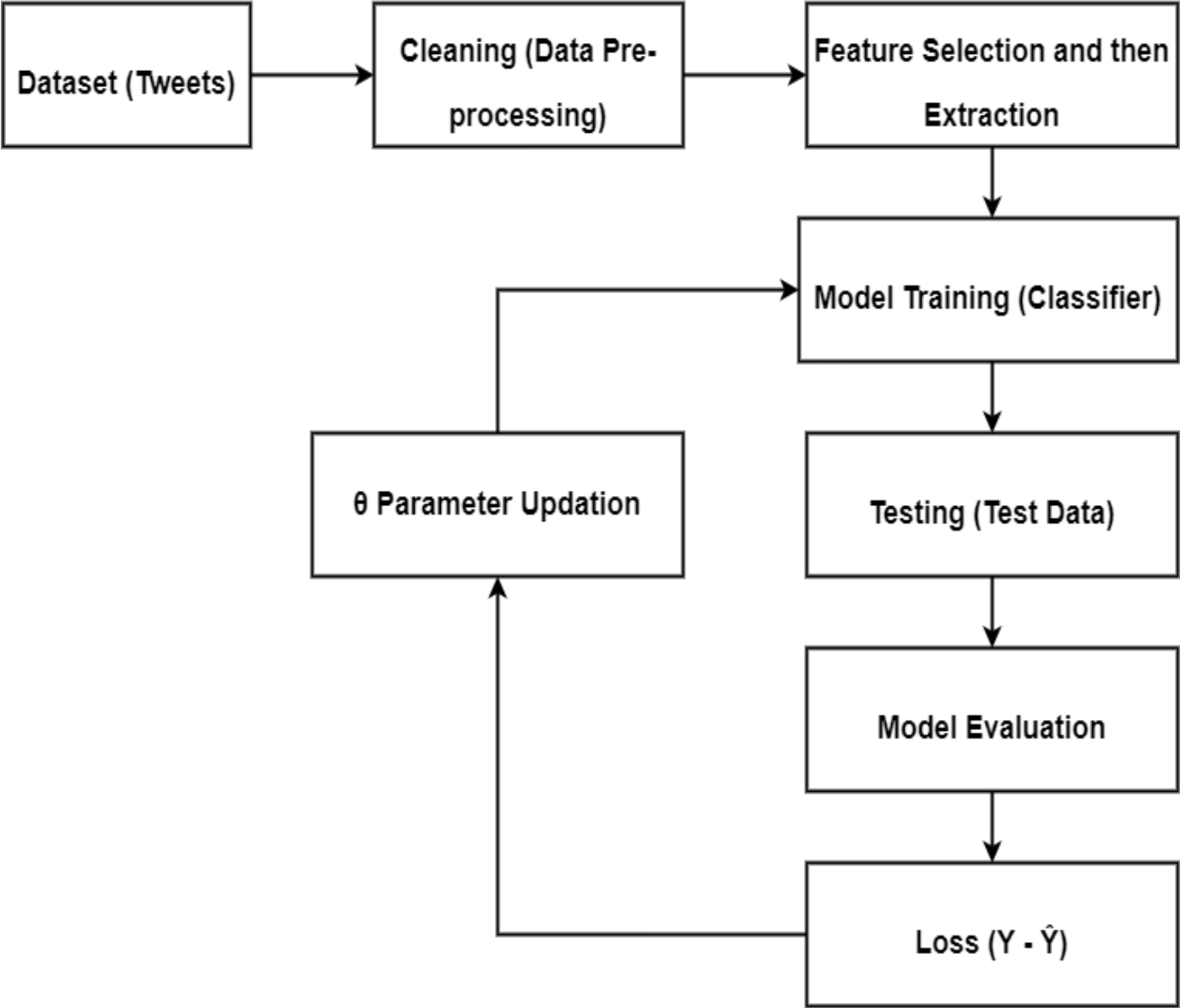- Any 64-bit Desktop OS (Microsoft Windows 10 or later is recommended).

# ATTRIBUTES RELATED TO APPLICATION

Corona Tweet Classification System, has many attributes that makes it outstanding if compared to other Classification System in the same genre.

Following Points provide some of the insights related to the Classification System.

- **Adaptability:** Corona Tweet Classification System, can be easily adapted by new Users without any hassle.
- **Availability:** System can be made available to general User everywhere in the world. It can be accessed by any user irrespective of their location (Some countries excluded).
- **Accuracy:** System can determine the outcome of the Tweet i.e., positivity or negativity with maximum precision and accuracy.
- **Maintainability**- No extra maintenance is required by User.
- **Portability:** Since the System is itself built using Python (An interpreted platform independent language) so that means it can be ported to any desktop platform including macOS, Linux, Windows. It requires no extra overhead to be ported to any other platform.
- **Reusability:** The proposed Classification System can be used any number of times as it does not require to evaluate all the Tweets each time, a new Tweet is provided as input.
- **Cost:** Corona Tweet Classification System is free to use.

# MACHINE LEARNING PIPELINE DIAGRAM

```
┌──────────────────┐      ┌──────────────────┐      ┌──────────────────┐
│ Dataset (Tweets) │ ───> │ Cleaning (Data   │ ───> │ Feature Selection│
│                  │      │ Pre-processing)  │      │ and then         │
│                  │      │                  │      │ Extraction       │
└──────────────────┘      └──────────────────┘      └──────────────────┘
                                                             │
                                                             ▼
                   ┌──────────────────────────┐     ┌──────────────────┐
          ┌──────> │ Model Training (Classifier)│    │                  │
          │        └──────────────────────────┘     │                  │
          │                     │                    │                  │
┌──────────────────┐           ▼                     │                  │
│                  │    ┌──────────────────┐         │                  │
│ θ Parameter      │    │ Testing (Test    │         │                  │
│ Updation         │    │ Data)            │         │                  │
└──────────────────┘    └──────────────────┘         │                  │
          ▲                     │                     │                  │
          │                     ▼                     │                  │
          │            ┌──────────────────┐           │                  │
          │            │ Model Evaluation │           │                  │
          │            └──────────────────┘           │                  │
          │                     │                      │                  │
          │                     ▼                      │                  │
          │            ┌──────────────────┐            │                  │
          └────────────│ Loss (Y - Ŷ)     │            │                  │
                       └──────────────────┘            │                  │
```

## MILESTONE

| S. No. | Project Activity | Estimated Start Date | Estimated End Date |
|--------|------------------|----------------------|--------------------|
| **1.** | Synopsis Submission | 22/03/2022 | 25/03/2022 |
| **2.** | Presentation Submission | 23/03/2022 | 25/03/2022 |
| | | | |

## MEETING WITH THE SUPERVISOR

| Date of Meet | Mode | Comments by the Supervisor | Signature of the Supervisor |
|--------------|------|----------------------------|-----------------------------|
| 10/03/2022 | Offline | | |
| 24/03/2022 | Offline | | |
| | | | |
| | | | |

# BIBLIOGRAPHY AND REFERENCES

- https://app.diagrams.net/
- Kaggle.com/pre-processing-in-NLP
- Statistical Analysis using Machine Learning