PUNE INSTITUTE OF COMPUTER TECHNOLOGY
DHANKAWADI, PUNE – 43.

SUBMISSION OF S.E. PROJECT SYNOPSIS

DEPARTMENT: COMPUTER ENGINEERING   ACADEMIC YEAR: 2021-22

Class: SE   Sem- II
Subject: - Project-Based Learning

Project Group Members:

Roll No.	Name	Guide Name
21101	AADEESH BALI	SHRUTI KUDAGI
21103	ADITI MULAY	
21109	BHAGWAT ANIKET SURYKANT	
21115	BRAMHECHA SAMEER MANOJ	
21122	DEVANSH CHOUDHURY	

Project Title: **Automated and Semi-automated Architecture for Gathering Intelligence Using Different Datasets.**

Domain: Machine Learning, Natural Language Processing (NLP) & Web-Development

Synopsis:

Title of the project: Automated and Semi-automated Architecture for Gathering Intelligence Using Different Datasets.
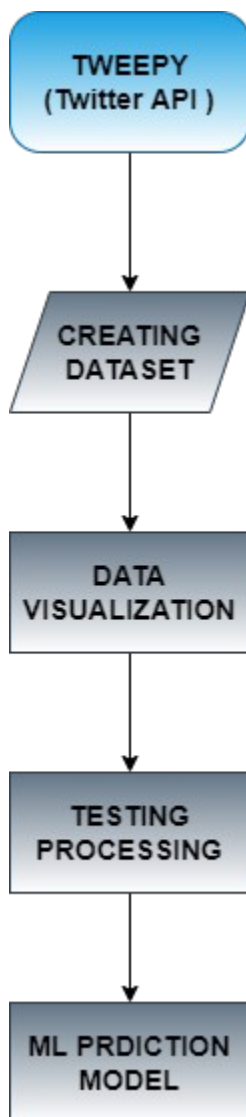
Technical Key Words: -
    i)   Natural Language Processing
    ii)   Database
    iii)   Supervised Learning*
    iv)   Unsupervised Learning*
    v)   API
    vi)   Bootstrap
    vii)   CSS
    viii)   Frontend, Backend
    ix)   Java Script
    x)   PHP
    xi)   Wireframe*
    xii)   Algorithm

Abstract: Our team aims to develop an automated and semi-automated architecture for gathering intelligence using Twitter API. As a specific awareness about mental health and nontoxic positivity rises, we seek to maintain a persisting scale of positivity by filtering out demeaning/deteriorating data from the given data set of tweets on Twitter. The developed model can not only be used for just this cause but can also be trained at various levels to get a specific desired data set filtered as per the user's needs.

Basically, by targeting a social media structure we hope to achieve a level of balance aspired by many. Furthermore, training this model in ways that are customary and suitable to one's needs can also be used for governmental bodies to analyze and identify threats to the societal structure and sentiments of the public.

Problem Definition: With the increase in the data present around us the need to analyze and gather the precious information from it is also on the boom. So, to manipulate the data according to the user's needs, a centralized model can be programmed. One such example of personalized need has been taken into consideration while training this model.

Nowadays, people are becoming increasingly aware of the significance of mental health and mental fitness, but still, some feel a bit of traction while sharing their thoughts with people close to them. To them, it's easier to just share on social media on such a social network as Twitter. So, in this project, we will be training a model to identify the SUICIDAL INTENTION in the given dataset of tweets.

TWEEPY (Twitter API)

↓

CREATING DATASET

↓

DATA VISUALIZATION

↓

TESTING PROCESSING

↓

ML PRDICTION MODEL

Problem Solution: We are working on a Machine Learning project that would solve some real-life issues like suicidal problems, etc. We will collect the relevant data from the TWEEPY (Twitter API), store it in a suitable format including the output in one place so that it becomes very convenient.

After this, all the data will be visualized and will be processed in such a language that is well understood by the machine. The next step will be Testing and processing where we will try and implement some examples to test the accuracy of this model. And then, the entire data will be fed into the model and collected the output result, and stored for future reference.

Software Requirement: -

P:F-PRJ-UG/03/R0

i)   **Python 3.9** – Programming Language.

ii)   **Anaconda** – A distribution of python programming language for scientific computing.

iii)   **PyCharm** – Integrated Development Environment (IDE).

iv)   **Jupyter Notebook** – A software that allows to creation and sharing of documents that integrate live code.

v)   **Visual Studio Code** - A code editor to be used for Web Development

Future Scope: - The model can be further trained to analyze and gather the required information from different datasets having different origins. For example, finding out the target consumers for particular products by different LLC or Government Bodies. The government can use it to track the social media structure of a particular target posing a threat to our national well-being and also can be used to analyze the sentiments of the public over the decision made by the governments.