# **Synopsis - Final Year Project**

# **Twitter Problem Detection**

**Table of Contents**

1. Introduction

2 . Project Concept

2.1 Abstract

2.2 Objective

2.3 Literature Review

2.4 Problem Definition

2.5 Scope

2.6 Technology Stack

2.7 Benefits to the Society

2.8 References

**1. Introduction :**

**PROJECT TOPIC - DETECTING CITIZEN PROBLEMS AND THEIR LOCATIONS USING TWITTER DATA.**

Everyday, over millions of tweets are posted on Twitter by many individuals that receive hundreds of likes, comments and retweets. Apart from tweets that promote products, share news and information, and engage with their audiences, many users also tweet about issues that they face in their localities in their day-to-day lives. Our project focuses on such tweets to detect problems faced by citizens in a city so that the concerned authorities are informed and can take appropriate actions. By using such twitter data, our project aims to provide an interface that displays citizens’ problems along with their locations, in order to provide easy access to the authorities.

**2. Project Concept**

**2.1 Abstract :**

In this project, the problems faced by residents of a city or citizens’ requests are detected by analyzing tweets posted by users concerning the various localities in a city.

* Tweets are extracted from twitter that contain keywords denoting areas, places or localities present in a city. (Thane)
* A Naive Bayes model is trained to detect whether a tweet signifies a problem faced by a citizen or not. Hence, the model classifies a tweet as related or unrelated.
* The extracted tweets are passed through the model and the list of related tweets is obtained.
* The related tweets are analyzed to get information of any city event, citizen's complaint or requests about a problem and stored in the database, along with the location.
* The tweets denoting citizens’ problems and the locations are obtained from the database and displayed on the front-end interface.

**2.2 Objective :**

The objective of this project is to help keep track of problems in cities. Tweets will be extracted based on locations present in the text of these tweets. A website will be created to help keep a record of all the problems using tweets from twitter that are extracted via twitter API.

**2.3 Literature Review :**

Gizem Abalı et al published a paper under IEEE in 2018 based on a Turkish city and uses Turkish language tweets for the study. Aim was to form a smart system, which detects problems of citizens and extracts the problems’ exact locations from tweet texts. Naive Bayes classifier is trained on the tweets and tested on a separate set of tweets [8-10]. The high accuracy which is obtained by the classifier indicates that it is desirable to use this classifier for our study. This project will be based on these research paper findings and is mainly focused on Thane city.

When the past studies using Twitter data were studied, it was seen that many of them are about analyzing traffic tweets [1-3], some of them were done to understand the happiness in a city [4], finding the locations of where a user tweets[6], analyzing disaster tweets [5, 6], and detecting place names that are passed on a tweet text [7].

**2.4 Problem Definition :**

In the past few years, technology has developed a lot and brought about a big change in the lives of people. Many people started to migrate to the cities. The increment in the number of citizens brought along many city problems like pollution, accidents, traffics, etc. Plenty of information is available over the internet, but it is very difficult to get the right information. Thus, an application will be created that will fetch data from one of the largest social media platforms i.e. Twitter. User tweets will be extracted to get information about citizens’ problems and their locations. This application will display these problems with their locations which will be obtained by analyzing the tweets. All this information will be made available to everyone.

**2.5 Scope**

There are many sources present for obtaining information about events, accidents and various problems in citizens’ day to day lives. People post about these problems on different social media platforms. Thus, an application will be created that will detect such citizens' problems with twitter. Tweets will be fetched with the locations for the dataset. Subsequently, this dataset will be used to train and test the machine learning model. As for frontend, a website will display all the tweets that mentions the citizens’ problems along with the location. This website will enlist almost all the problems faced by citizens in the city. Several filters will be added to this website for the authorities to get the information with ease.

**2.6 Technology Stack :**

* Machine learning
  + Train the system using Naive Bayes Classifier (machine learning model) to detect information-containing tweets about the area.

* Google Colab
* Database - MongoDB or SQL
  + MongoDB/SQL is used to store, retrieve and access the data.
* Twitter API
  + In order to collect data, Twitter API is used.
  + The Twitter API provides nearly 1% of tweet data that streams from the selected area.
* Programming language - Python
  + Used for building machine learning model
  + Used for building a code to recognize locations from the tweet data
  + Tweepy - Python package used to access the Twitter API.
  + Pandas - Python library for data manipulation and analysis.
  + Sklearn - Python library used for machine learning and statistical modeling (here for naive Bayes classifier model)
  + Flask - An API of Python used to build web-applications(front - end)

**2.7 Benefits to the Society :**

1. In the past few years, the number of people who live in cities and rural areas has changed and a big increment was observed in the cities' population. This brought along many city problems like pollution, accidents, traffic, etc.

This project can be used to detect these problems along with their locations which can be helpful in order to solve the problems quickly.

1. A smart city can be explained as the area where intelligent functions are used to collect the data and synthesize it to improve the efficiency of services, equity, sustainability and quality of life.

Some of the other features of smart city:

* A city that uses smart computing technologies to create all its infrastructures and services (including health-care, education, transportation).
* A city that watches all its critical transportation, communication, energy and water infrastructures and also major buildings.
* A city which is more effective, liveable, fair and sustainable.

This project can be used to develop a city into a smart city.

**2.8 References :**

[1] N. Wanichayapong, W. Pruthipunyaskul, W. Pattara-atikom, and P. Chaovalit, "Social-based traffic information extraction and classification," in Proc. 2011 International Conference on ITS Telecommunication (ITST), pp. 107-112.

[2] M. Hasby and M. L. Kodra, "Optimal path finding based on traffic

information extraction from Twitter social-based traffic information,"

in Proc. 2013 International Conference on ICT for Smart Society

(ICISS), pp. 1-5.

[4] S. B. Marupudi, "Framework for semantic integration and scalable processing of city traffic events," M.Sc. Thesis, Wright State

University, 2016.

[5] W. Guo, N. Gupta, G. Pogrebna, and S. Jarvis, "Understanding happiness in cities using Twitter: Jobs, children and transport," in Proc. 2016 IEEE International Smart Cities Conference, pp. 1-7.

[6] T. Sakaki, M. Okazaki, and Y. Matsuo, "Earthquake shakes Twitter users: Real-time event detection by social sensors." in Proc. 2010 International Conference on World Wide Web, pp. 851-860.

[7] A. Acar and Y. Muraki, "Twitter for crisis communication: Lessons learned from Japan's tsunami disaster," International Journal of Web Based Communities, vol. 7.3, pp. 392-402, 2011.

[8] M. Pennacchiotti and A. M. Popescu, "A machine learning approach to Twitter user classification," in Proc. 2011 International AAAI Conference on Weblogs and Social Media, pp. 281-288.

[9] A. Z. H. Khan, M. Atique, and V. M. Thakare, "Combining lexiconbased and learning-based methods for Twitter sentiment analysis," International Journal of Electronics, Communication and Soft Computing Science & Engineering, vol. 4(4), pp. 89-91, 2015.

[10] G. Sidorov, and et al., "Empirical study of machine learning based approach for opinion mining in tweets," in Proc. 2012 Mexican International Conference on Artificial Intelligence, Springer Berlin Heidelberg, pp. 1-14.