

STREAMING HIGHWAY TRAFFIC ALERTS USING **TWITTER API**

INT 200 Internship

Final Review – 29th July, 2022

Project Guides / Mentors:

- 1. Dr. Jayanthi. G**
- 2. Prof. Ramya. M**

Team Members

Ayushmaan Das
(E0121037)



SRI RAMACHANDRA

INSTITUTE OF HIGHER EDUCATION AND RESEARCH

(Category - I Deemed to be University) Porur, Chennai

SRI RAMACHANDRA FACULTY OF ENGINEERING AND TECHNOLOGY

INTRODUCTION

- ❖ Twitter can be used as a huge storehouse of data and information of various kinds.
- ❖ This project aims to generate a Traffic Monitoring System from Twitter Stream Analysis.
- ❖ A complete research was done regarding the project and existing cases and models were studied.
- ❖ Tweets are fetched from Twitter API and are *pre-processed* for proper model building.
- ❖ This process also requires *Vectorization* and *Tokenization* for the machine to understand the words.
- ❖ Multiple algorithms has been used that *classified* the tweets to serve the purpose.
- ❖ The ultimate aim was to create a UI for the user to interact. This was achieved using *StreamLit*.
- ❖ The UI exists as a *Web Application* that provides the users complete freedom to grab tweets related to traffic.
- ❖ Efforts have been made to make the app interactive, versatile and user-friendly.

PROBLEM STATEMENT & OBJECTIVES

“Designing of a E-Traffic alert System by mining tweets from Twitter Stream API”

OBJECTIVES

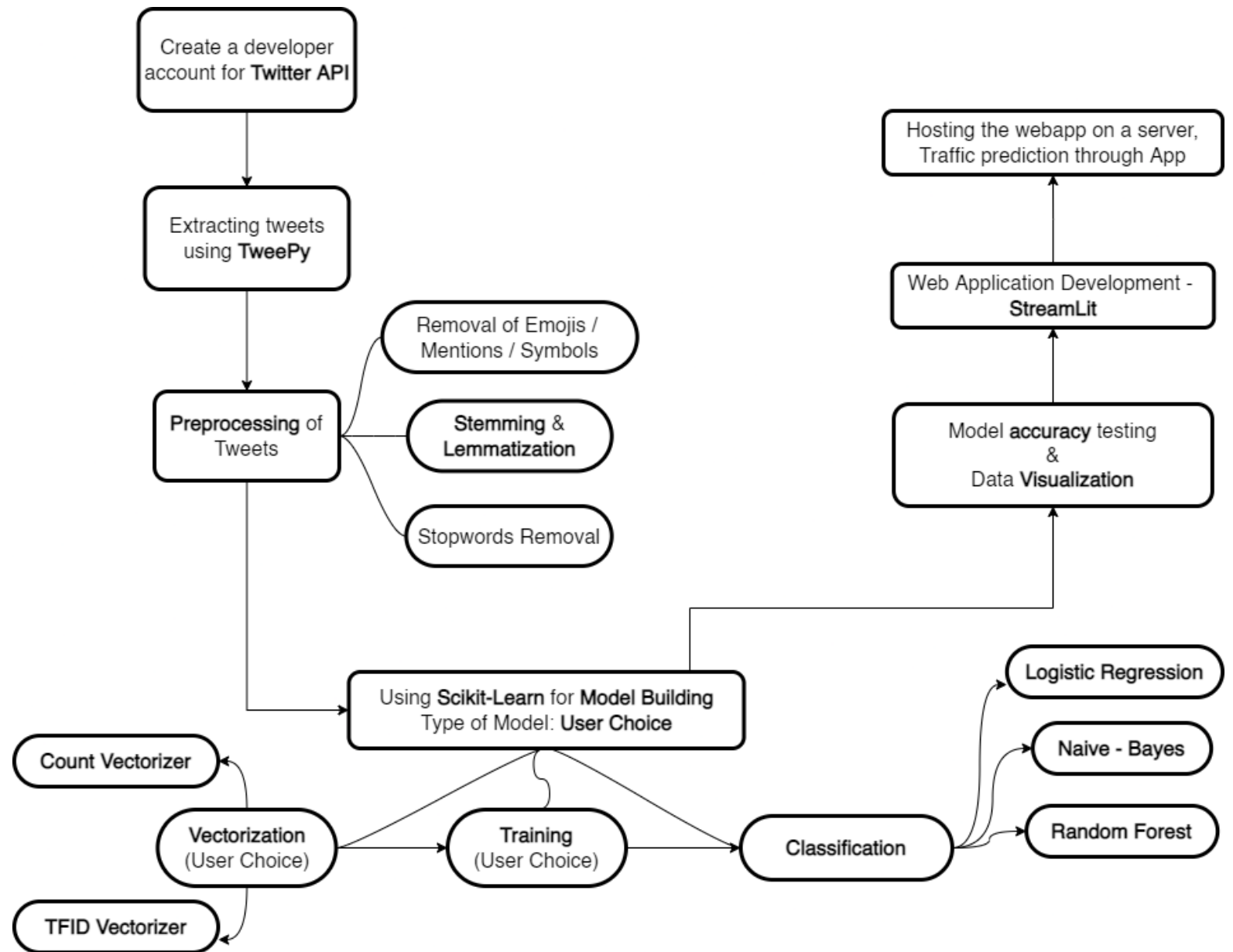
The project has been divided into the following modules for easier approach and implementation:

- **Module 1:** Fetching and storing tweets from Twitter Stream API using *Tweepy*
- **Module 2:** *Pre-Processing* of tweets and *Lemmatization*
- **Module 3:** *NLP – Vectorization*, Model building for *Classification* of Tweets
- **Module 4:** Data *Visualization* and model accuracy testing
- **Module 5:** Designing an UI for user interaction using the *StreamLit* library

RESEARCH & PRODUCT SURVEY

YEAR	AUTHOR	PRODUCT	REMARKS
2016	Faculty CSE, SG Polytechnic, Atigre, IN	Ongoing Detection of Traffic from Twitter Stream Analysis	Benefits: Classification model used works with great accuracy
			Application: Proper classification of traffic related tweets
			Limitation: No data visualisation and no frontend
			Challenges: Developing a proper UI for user interaction
2017	Pouza Rezai	Detection classification and location identification of traffic congestion from Twitter stream analysis	Benefits: Proper data visualisation and accuracy
			Application: Finding traffic and avoiding congested routes
			Limitation: Lack of UI implementation for user to interact
			Challenges: Developing a user-interface and improve the model
2018	Sivagurunathan V	Traffic-Detection-from-Twitter-using-Spark	Benefits: Use of Kafka and ElastiSearch
			Application: Obtaining a Traffic HeatMap
			Limitation: No data visualisation, tweets are old
			Challenges: Using data visualisation and recent Tweets

DATA PIPELINE - Pre-processing tweets and Machine learning workflow for classification



TECHNOLOGIES USED



Python



Tweepy



Pandas



Matplotlib



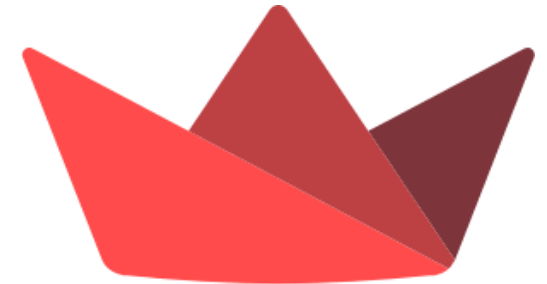
Seaborn



Scikit - Learn



Natural Language Toolkit



StreamLit

WORK DONE

Module 1:

- Carrying out proper research work and product survey.
- Streaming and Storing tweets from Twitter using Tweepy.

Module 2:

- Pre-Processing of Tweets, Labelling followed by proper Lemmatization.
- Natural Language Processing of the tweets.

Module 3:

- Splitting of Dataset and Vectorizing the tweets.
- Defining various models, followed by Training, Prediction and Classification.

Module 4:

- Plotting ROC Curves for every defined models, statistically evaluating the tweets.
- Comparing the accuracy of various models and identifying the most appropriate ones for the process.

Module 5:

- Implementing all the aforesaid concepts within a web-app using StreamLit.
- Creating a proper UI which enables the users to fetch tweets as per their choice.
- Enable the users to select the model they want, making the app more versatile.

	A	B	C	D	E	F
1		username	tweet	location	date	time
2	0	mattgrocoff	@DirtyTesLa @Tesla @elonmusk I just told my Tesla	Ann Arbor, MI, Ear	09-07-2022	01:07:49
3	1	BoojiBoy6	@tentwentsixpm Living in NYC is expensive and finding	Brooklyn, NY	09-07-2022	00:46:43
4	2	CynfullySweetXO	@Slim_Luck_Alex drove thru NYC today ðŸ˜– I do like it	Nowhere	08-07-2022	23:41:17
5	3	jordonaut	@lantzarroyo Iâ€™m particularly scared for my folks in	Seattle, WA	08-07-2022	23:02:33
6	4	En_AmbientMusic	City Rain Traffic Sounds for Sleep and Study ASMR		08-07-2022	22:49:18
7	5	NYC81966570	@JohnnyGoodberry @alefeusch @wealth I think a lot of people's opinions		08-07-2022	22:48:19
8	6	Dzollo_	@NYCMayorsOffice @nycgov @NYPDPC	New York, USA	08-07-2022	22:09:15
9	7	ZalezVickie	@sanjeeva7 @casgroenigen05 @javroar Unless u in NYC. Them mfs smart		08-07-2022	22:03:36
10	8	lolaxlachiva	Okay but nyc traffic ðŸ™„		08-07-2022	21:50:39
11	9	TVariunessKing	NYC needs trams. Replaces these buses that get stuck in	40.7236448,-74.00	08-07-2022	21:50:09
12	10	JMartinezNYC	@TransitNinja205 @Ollie_Cycles Weâ€™re not casting	New York, NY	08-07-2022	21:31:41
13	11	Songbird99	@NYCTSubway @MTA @NYC_DOT There's a crazy whi	Hooverbloob High	08-07-2022	21:23:56
14	12	NYPDnews	If you are planning on spending time in NYC this	New York City	08-07-2022	21:00:09
15	13	BIGKay95	NYC is exactly like the movies when it comes to this traf	Tonawanda, NY	08-07-2022	20:43:55
16	14	ACofRL	Here's this evening's train and traffic update. Have a wc	New York, NY	08-07-2022	20:30:23
17	15	AwkwardCeleb	@j_pineo This kind of thinking is so fkn backwards.... It'	Michigan, USA	08-07-2022	20:27:22
18	16	MickEnglewood	Being a cop in NYC is a pretty good gig these days. You c	Down South Slang	08-07-2022	19:57:30
19	17	NYC_DOT	This Summer, NYC DOT will begin safety improvements	New York City	08-07-2022	19:45:00
20	18	joshkiej	@JarekFA @fresh72_ @Ollie_Cycles @BarmanNYC Ah yes use an example		08-07-2022	19:41:36
21	19	NYPDbot	The NYPD Announces Street Closures and Expected Tra	New York, NY	08-07-2022	19:00:06
22	20	Bill66917830	@BuckSexton I got a ticket once in NYC for riding my BICYCLE against traf		08-07-2022	18:55:33
23	21	Fullblast4Miha	@stevenbodzin @voteshekar @NYC_DOT the ramp traffic uses to enter B		08-07-2022	18:54:41
24	22	UVicECS	Thanks to @amazon Software Engineer & @UVic_	Victoria, British Co	08-07-2022	18:45:03
25	23	king_baffled	@KarenJKwa @ole_cv @brianpmangan New york has		08-07-2022	18:43:46
26	24	scarcebtw	mumbai better than nyc bro wtf is this traffic	h mart	08-07-2022	18:35:06
27	25	FlairScholar	@ana_banza Dallas is always underconstruction. Lol	Virginia, USA	08-07-2022	18:27:42
28	26	nftarabcat	Iâ€™ve been to Cairo , Bangkok, LA and NYC and nothin	Orlando, FL	08-07-2022	17:52:48
29	27	CWCBExpo	Our favorite kind of NYC traffic!	Paramus, NJ	08-07-2022	17:40:02
30	28	DutchKillsCivic	NYPD114Pct: @StevenR10 @NYC_DOT Traffic Rule 4-0	Long Island City, Q	08-07-2022	17:34:08
31	29	jenbharwani	Ohh god!!! So much traffic in NYC !!!!! So annoying!!!!		08-07-2022	17:31:37

FETCHED TWEETS FROM TWITTER API

username	tweet	location	date	time	processed_tweet	predicted_class
mattgrocoff	@DirtyTesLa @Tesla @elonmusk I just told n	Ann Arbor,	2022-07-09	01:07:49	i just told my tesla to take my daugh	0
BooiBov6	@tentwentsixpm Living in NYC is expensive	Brooklyn, N	2022-07-09	00:46:43	living in nyc is expensive and finding a	0
AquilesMp	One thing about NYC it's that not matter the	SomeWhe	2022-07-03	09:36:37	one thing about nyc its that not matte	0
DutchKillsCivic	@NotifyNYC: .@FDNYAlerts Three Alarm Fir	Long Islanc	2022-07-03	08:28:43	three alarm fire th road amp nd stree	1
NotifyNYC	.@FDNYAlerts Three Alarm Fire: 85th Road &	New York (2022-07-03	07:40:18	three alarm fire th road amp nd stree	1
MrAFelix	@diemauerthewall Struck me since living he	Lübben (Sp	2022-07-03	07:02:56	struck me since living here when i wa	0
TotalTrafficNYC	Accident. Two lanes blocked in #NYC:Onthe	New York (2022-07-03	06:25:43	accident two lanes blocked in nycontf	1
cycling_nyc	@StreetwallNY @NYC_DOT @NYSDOT Encc	New York,	2022-07-03	06:08:31	encouraging washington heights resi	0
TotalTrafficNYC	Accident. Two lanes blocked in #NYC:OnThe	New York (2022-07-03	05:40:43	accident two lanes blocked in nycontf	1
1Goodfriend12	@nypost Illegal Motorbikes on NYC Streets	has reached	2022-07-03	04:53:18	illegal motorbikes on nyc streets has	0
drjamima	I'm thinking traffic stops are not going to be	Oregon	2022-07-03	04:12:43	im thinking traffic stops are not going	0
TotalTrafficNYC	Accident. Two lanes blocked in #NYC:OnThe	New York (2022-07-03	03:45:43	accident two lanes blocked in nycontf	1
TotalTrafficNYC	Accident. Left lane blocked in #NYC:OnHenr	New York (2022-07-03	03:25:43	accident left lane blocked in nyconhe	0
CovfeFeAnon	@hjeutysd "Hey, NYC has murders and stabbin	g on the	2022-07-03	02:38:27	hey nyc has murders and stabblings or	0

SAMPLE DATASET AFTER CLASSIFICATION

DIFFERENCE BETWEEN TWEETS BEFORE AND AFTER PRE- PROCESSING

tweet	processed_tweet
@DirtyTesLa @Tesla @elonmusk I just told my Tesla to take my daughter and me from supercharger in New Rochelle to Upper West Side of Manahattan. It went through towns, construction, traffic and down Hudson Pkwy with zero disengagements!! All while NYC drivers tried best to kill us. #FSDBeta https://t.co/IXK4mJrgrh	i just told my tesla to take my daughter and me from supercharger in new rochelle to upper west side of manahattan it went through towns construction traffic and down hudson pkwy with zero disengagements all while nyc drivers tried best to kill us fsdbeta
Living in NYC is expensive and finding an apartment is a mess but I don't find it stressful. What I do find stressful is having to sit in traffic and driver everywhere. I like visiting LA for the food/culture but would never want to live D5	living in nyc is expensive and finding an apartment is a mess but i dont find it stressful what i do find stressful is having to sit in traffic and driver everywhere i like visiting la for the food/culture but would never want to live d5
@Slim_Luck_Alex drove thru NYC today 🤯 I do like it upstate but man oh man, traffic is such a nightmare	drove thru nyc today i do like it upstate but man oh man traffic is such a nightmare
@lantzarroyo I'm particularly scared for my folks in densely populated areas with poor ventilation (cities like NYC) that have high traffic in and out of there. Our public health system isn't doing as much as it COULD be.	im particularly scared for my folks in densely populated areas with poor ventilation cities like nyc that have high traffic in and out of there our public health system isn't doing as much as it could be
City Rain Traffic Sounds for Sleep and Study ASMR Ambience Relaxing ... https://t.co/hEj177VwgU via @YouTube City Rain Traffic Sounds for Sleep and Study #rainambience #rainvideos #rain #city #citytraffic #cityambience #loveny #ilovenyc #nyc	city rain traffic sounds for sleep and study asmr ambience relaxing via city rain traffic sounds for sleep and studyrainambience rainvideos rain city citytraffic cityambience loveny ilovenyc nyc

A GLIMPSE OF THE WEB – APPLICATION

&

DEMO OF THE FEATURES OF THE APP

The StreamLit App has 4 different pages

- ❖ **HOME PAGE**
- ❖ **CUSTOM SEARCH PAGE**
- ❖ **QUICK SEARCH PAGE**
- ❖ **ABOUT THE APP PAGE**

HOME PAGE



TRAFFIC ANALYSIS FROM TWITTER STREAMS

Twitter is one of the most widely used Social Media Platform for users to share their thoughts and ideas. However, Twitter can also be interpreted as a huge storehouse of data, which on proper processing can serve as a huge source of information. This type of Data is known as **Big Data**.

This application streams tweets from **Twitter API** and uses **Machine Learning** to filter and classify the tweets in order to have an idea about the situation of traffic in popular cities.

In order to increase its efficiency, various models have been tried and tested and the best possible one can be seen from the scores and results.

The user can use "**Popular Search**" to quickly obtain traffic related tweets about some common cities based on pre-defined model. They can also have complete control over the model-building in the "**Custom Search**" section.



Visit the [GitHub](#) repository or [Google Drive](#) for source code and references:



Sample Python Code for Tweets Fetching



Sample Python Code for Tweet Preprocessing



Sample Python Code for Model Building



HOME

CUSTOMISED SEARCH

POPULAR CITIES

ABOUT THE APP

Documentation Links:

- [Tweepy](#)
- [Scikit-Learn](#)
- [Pandas](#)
- [Seaborn](#)
- [NLTK](#)
- [StreamLit](#)

Follow / Contact me for collaboration:



USER – DEFINED SEARCH



- HOME
- CUSTOMISED SEARCH**
- POPULAR CITIES
- ABOUT THE APP

CUSTOMISED SEARCH - Search the way you want !!!

Please follow the instructions stated below :

- Enter the **correct spelling** of the city
- After entering the name, do not forget to **press enter**
- Select the Model you want to be used from the dropdown
- Press the submit button to check the results

TYPE IN THE NAME OF THE CITY

Example : Seattle

DATASET SPLITTING SIZE (Training : Testing)



CHOOSE THE VECTORIZER

COUNT VECTORIZER

CHOOSE THE MODEL TO BE USED

LOGISTIC REGRESSION

SEARCH FOR TWEETS

QUICK SEARCH – DEAFULT SETTINGS



-  HOME
-  CUSTOMISED SEARCH
-  POPULAR CITIES
-  ABOUT THE APP

COMMONLY SEARCHED CITIES

United States



LOS ANGELES



NEW YORK



SEATTLE

Canada



ABOUT THE APP PAGE – BASIC DESCRIPTION



- HOME
- CUSTOMISED SEARCH
- POPULAR CITIES
- ABOUT THE APP

ABOUT THE CREATOR

AYUSHMAAN DAS



◦ **STREAM** : B.Tech CSE - Artificial Intelligence and Machine Learning (**AI & ML**)

◦ **INSTITUTE** : Sri Ramachandra Engineering and Technology, Sri Ramachandra Institute of Higher Education and Research (**SRIHER**), Porur, Chennai - 600116

◦ **PERIOD OF WORK** : First Year Internship (INT 200 - Internship 1)

This app/project has been made under the guidance of **PROF. JAYANTHI G** and **DR. RAMYA M** of SRET (SRIHER) - Chennai.

LIBRARIES USED

The following Python Libraries have been used to build this app:

- **Tweepy** - Fetching of Tweets from Twitter API
- **Pandas** - Datframes and datasets
- **Scikit-Learn** - Model Buidling and Classification
- **Matplotlib & Seaborn** - Data visualisation
- **NLTK** - Preprocessing of text
- This application has been designed using **STREAMLIT**

SAMPLE TEST CASE

TYPE IN THE NAME OF THE CITY

los angeles

DATASET SPLITTING SIZE (Training : Testing)

0.20

0.30

0.80

CHOOSE THE VECTORIZER

TFIDF VECTORIZER

CHOOSE THE MODEL TO BE USED

NAIVE - BAYES MODEL

LOGISTIC REGRESSION

NAIVE - BAYES MODEL

RANDOM FOREST CLASSIFIER

SEARCH FOR TWEETS

Testing for the city – Los Angeles
Using TFIDF Vectorizer and Naïve-Bayes
Model

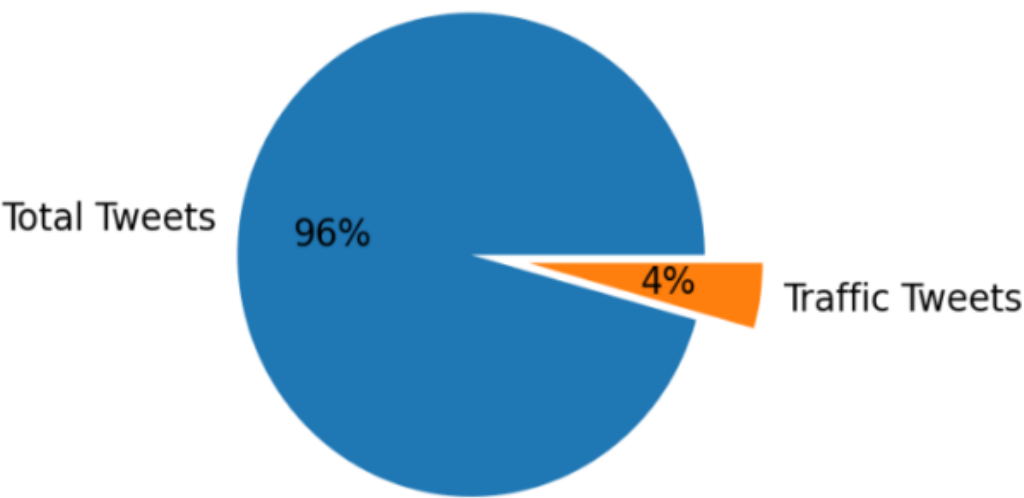
Traffic – related tweets fetched from Twitter
API after classification

	username	tweet	date	time
0	TotalTrafficLA	Brush fire has ONramp blocked in #SouthLA on 110 (I-110 Hbr Fwy) NB at Gage Ave, stopped traffic bac	2022-07-26	78643000000
1	TotalTrafficLA	Accident. Shoulder in #Torrance on 405 NB at Normandie Ave, stopped traffic back to Carson St #LAtra	2022-07-26	57343000000
2	TotalTrafficLA	Crash carpool lane in #Riverside on I-215 NB after Center St (Exit 36), stopped traffic back to Hwy 60/Hv	2022-07-25	83743000000
3	TotalTrafficLA	Brush fire two right lanes blocked in #Pacoima on I-5 SB at Hwy 118/ Paxton St, stopped traffic back to	2022-07-25	75044000000
4	TotalTrafficLA	Accident cleared in #RanchoCucamonga on I-15 NB at Summit Ave, stopped traffic back to I-210 #LAtra	2022-07-25	6945000000
5	TotalTrafficLA	Accident cleared in #SouthLA on 110 (I-110 Hbr Fwy) SB at 51st St, stopped traffic back to Adams Blvd #	2022-07-24	22543000000
6	TotalTrafficLA	Accident. Two middle lanes blocked. in #SouthLA on 110 (I-110 Hbr Fwy) SB at 51st St, stopped traffic t	2022-07-24	21643000000
7	TotalTrafficLA	Accident cleared in #Topanga on Pacific Coast Hwy (N) SB at Coastline Drive, stopped traffic back to La	2022-07-24	14443000000
8	TotalTrafficLA	Accident. Shoulder in #Arleta on I-5 SB at Osborne St, stopped traffic back to Hwy 118/ Paxton St #LAtr	2022-07-22	52243000000

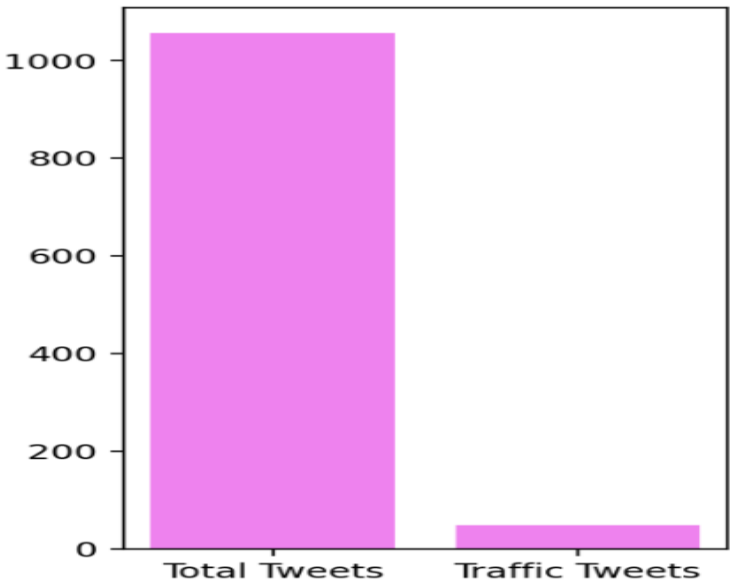
SAMPLE TEST CASE

SHOW ACCURACY SCORES FOR VARIOUS MODELS			
	Logistic Regression	Naive-Bayes	Random Forest
Test Score	0.9510	0.9534	0.9608
Train Score	0.9600	0.9737	0.9968
Accuracy Score	0.9510	0.9534	0.9608

Accuracy scores for various algorithms used

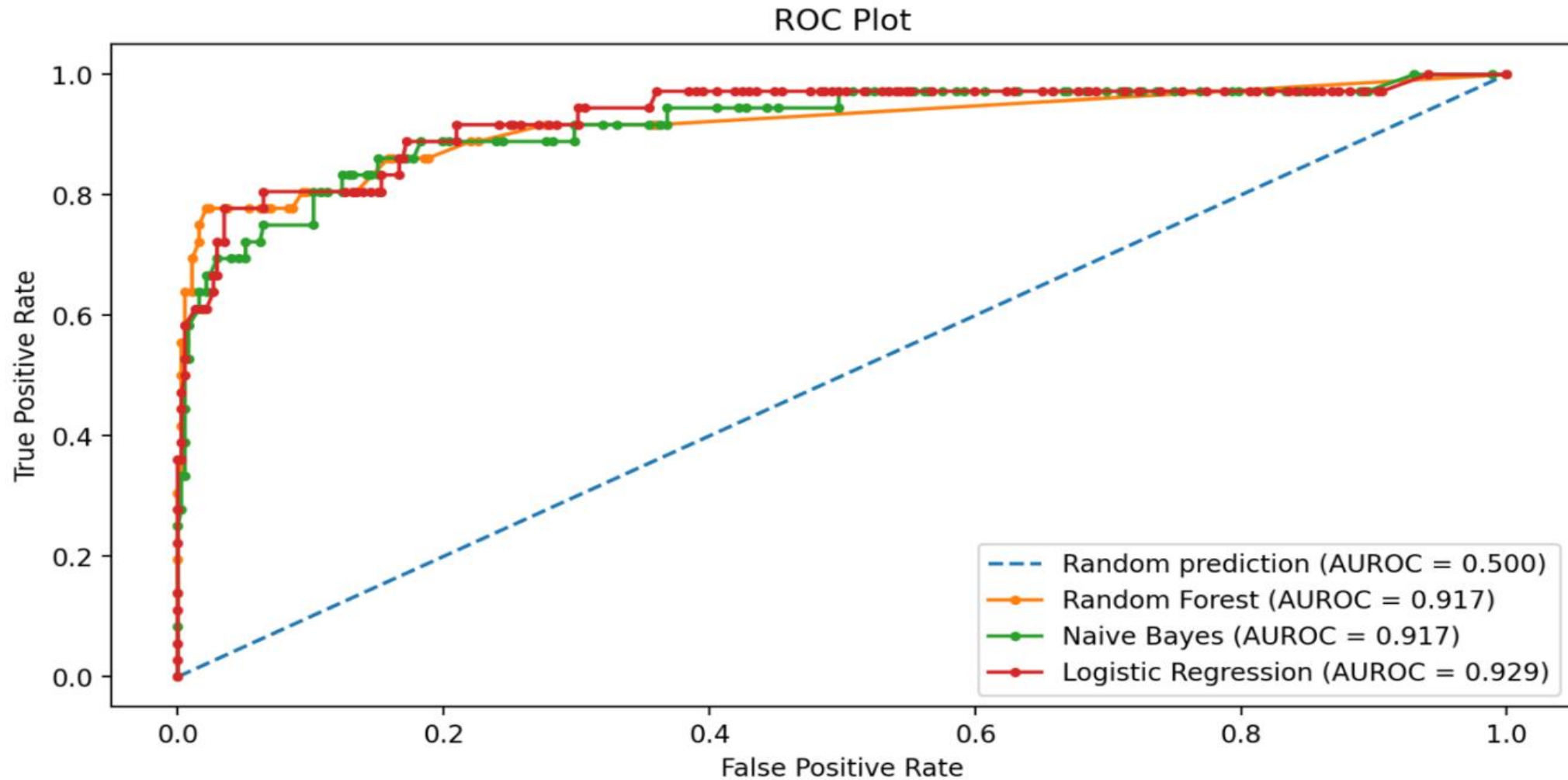


Pie-Chart



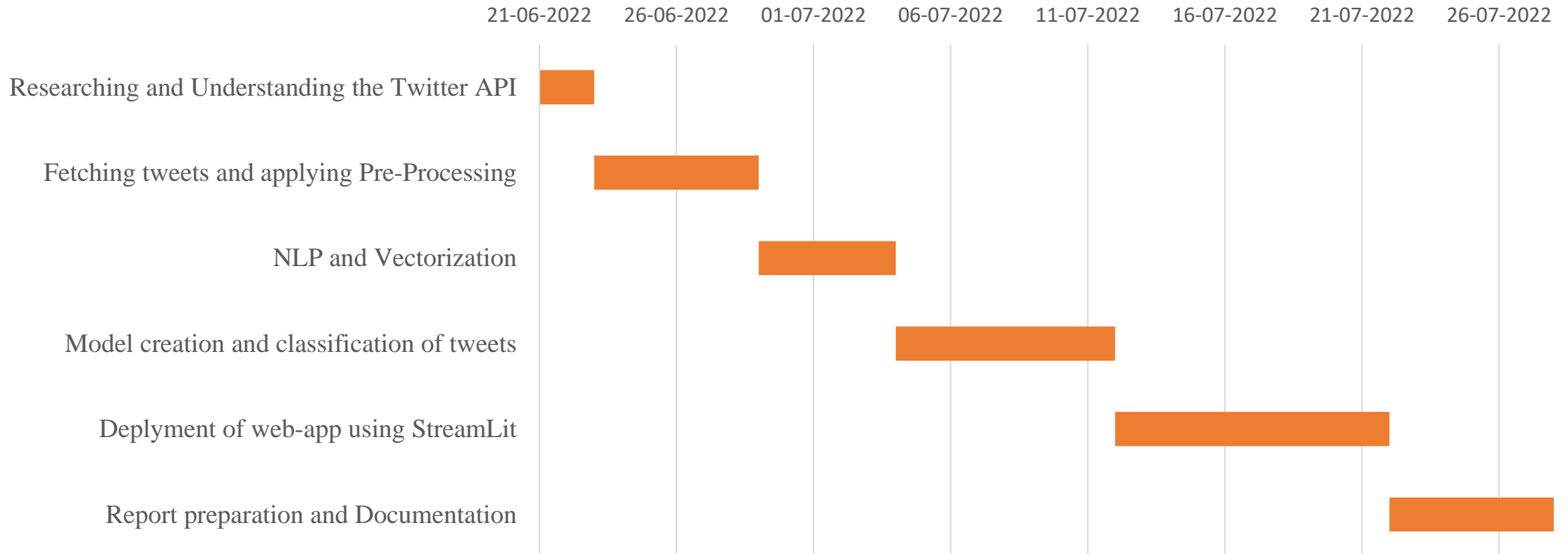
Bar Plot

SAMPLE TEST CASE



ROC Plot for the sample test case

TIMELINE – Gantt Chart



REFERENCES

- ***“Traffic Detection from Twitter using Spark”*** – Sivagurunathan V
(<https://github.com/SivagurunathanV/Traffic-Detection-from-Twitter-using-Spark>)
- ***“From Streaming Data to Twitter Analysis: Using Spark and AWS Kinesis”*** – Zhong Hongsheng
(<https://ieeexplore.ieee.org/document/7057672>)
- ***“ROC Curve”*** – Chanin Nantasenamat
(https://github.com/dataprofessor/code/blob/master/python/ROC_curve.ipynb)
- ***“Detection Traffic Congestion Based on Twitter Data using Machine Learning”*** – Muhammed Taufiq Zulfiqar, Suharto
(<https://www.sciencedirect.com/science/article/pii/S187705091931066X>)
- ***“Detection, Classification and Location Identification of Traffic Congestion from Twitter Stream”*** – Pouza Rezai
(<https://github.com/pouyarz/DETECTION-CLASSIFICATION-LOCATION-IDENTIFICATION-OF-TRAFFIC-CONGESTION-FROM-TWITTER-STREAM-ANALYSIS>)

WEB APP (StreamLit) references:

- ***“Covid 19 - Italy”*** – Tommaso Bonomo (<https://github.com/tommasobonomo/covid19-italy>)
- ***“Playground”*** – Ahmed BESBES (<https://github.com/ahmedbesbes/playground>)

THANK YOU

FOR WATCHING THIS PRESENTATION