“A Country-level Location Classification System for Worldwide Tweets”

DISSERTATION SEMINAR BY:

MALHAR ULHAS AGALE (Reg No. 2016MNS008)

Mtech Final yr (2017 - 2018),
Computer Networks & Information Security,
Dept of Computer Science & Engineering,
SGGSIE&T.

UNDER THE GUIDANCE OF :

PROFESSOR. P.S. NALWADE

Dept of Computer Science & Engineering,
SGGSIE&T

DATE:- 28th JULY, 2018

CONTENTS

- Introduction
- Objectives
- Review of Litreture
- Structure of Tweet
- Retrieval of Tweets
- Proposed System
  - Stages of realizing the system
- Results obtained
- Conclusion and Future Scope
- Bibliography

1. Introduction

Twitter Data & its Usefulness

- Twitter Concentrates more on what users post rather than who is posting it.
- Since Information posted on twitter is widely available and accessible (open for viewing),
- hence popular amongst researchers.

Research focused on:

- How much attention does an event or issue gets?
- Overall public opinion on that particular topic

Applications (Involving Tentativeness):

- User's Senitment Analysis based on public opinions and making predictions.

Real time trend analysis

Early detection of events

Topic detections

Building Recommender systems
(Targeted Advertising)

Examples:

Box office movie performance , Stock market movements , Flu outbreaks , News events (political, sports), Income prediction, Personality predictions, Gender predictions.

Importance of Demographic Details

[43] Concluded: "How much is the representation of the Twitter users from the overall population ?"

Demographic details enables us to focus on one particular set of people or subgroup in order to validate the posted information.

Example: (PEW Research Social media Survey)
Considers User Demographics in the USA to perform analysis on certain Topics and estimate public opinions and draw conclusions.

The Characteristics of Twitter user's are crucial to move forward for more advanced observations, predictions & to draw conclusions without bias, based on some available information.

...Characteristics may change..region to region.

Significance of location as a demographic detail

The **location of tweets** or the location from where the tweet was actually posted, we consider as one of the **most important demographic details**.

Along with the **location of tweets**, determination of **User's home & visiting locations** are also crucial.

Potential Applications:

- ✓ Disaster Management
- ✓ Spreading alerts or awareness about matters concerning Public health.
- ✓ Events detection and personalized advertising
- ✓ Tracking & prediction of Human movements (Urban Planning)
- ✓ Tracking Vehicular Traffic (Traffic engineering)
- ✓ Detect, Track and predict Criminal & Terrorist related activities on social media.
- ✓ Helping to Identify User's involved in "Cyberbullying" .

Classifying tweets based on their posting location and user home location can act as a "precursor" other data mining related tasks like Opinion mining, Sentiment analysis, Topic detection, recommendation tasks and finally performing predictions.

LOCATION INFERENCE SCENARIOS	Location Recognition and Linking
	Predicting Tweet's Posting Location
	User Home Location Inference
	User's Current and Future Location Prediction
	Location Type Categorization
	User Activity Location Inference

Challenges Faced in determining Tweet location

- **Not all Tweets are Geotagged**

  According to [2], an estimated number of tweets containing geotagging information **are less than 1%**

  Reasons: Privacy Concerns [3] & Save battery of Cell Phone [4] .

- **Empty or irrelevant or Ambigous User Location Field**

  According to [5], Some Twitter users also input fake, imaginary or irrelevant in the location field provided by Twitter.

- **Unhelpful and Incomplete Tweet Metadata**

  Not all tweets has **the exact Time zone, place and location coordinates** specified in their respective fields in the tweet metadata, which states that these fields are virtually not much of help in absence of the exact values of interest.

2.Objectives

What this System is all about ?

- The proposed system performs **country level location classification of tweets** posted over Twitter.
- All previous works were aimed at **inferring tweet location** were only restricted to **certain cities** or to any **particular country**.

Enhancing an Existing System:

- A system proposed by [6], performed a similar task of **"Realtime country level location classification of global tweets"**.

- Since the work done by [6], was a <u>**Realtime**</u> location classification task, a very important feature the locations <u>**User friendship network (followers & followees)**</u> were <u>**not**</u> taken in to <u>**consideration**</u>**.**

Reasons:

- **Delay** in the retrieval of user's friends location was significantly higher, intolerable in a Realtime system.

Our Enhancement

- Incorporating all existing features used in [74], along with the User friends & followers location as an extra feature introduced in the existing system.

3.Review of Litreture (Previous works)

Reference	Geographic Scope	Tweet Language	Granularity Level	Location Inference Scenario
[7]	worldwide, 3362 cities	Multilingual	City	Whether tweet contains locations of cities
[8]	worldwide	English	City	Predicting the location of Twitter users and tweets
[9]	St. Louis city Missouri (USA)	English	City	Inferring Tweet origin and categorizing them based on user activity zones
[10]	Korea	Korean	City	Inferring the Twitter user's location
[11]	Dublin, Manchester Boston	English	City	Identifying Twitter users in a city and attaching coordinates to those who are unlabeled by referring the known ones
[12]	USA	English	City	Predicting the home location of Twitter user
[13]	USA	English	City and state	Finding location specific tweets and topic prediction
[14]	worldwide	English	Place of Activity	location type categorization
[15]	London	English	place	Location inference and categorization
[16]	Worldwide	English	Country/City	Location prediction and extraction from tweets

Reference	Geographic Scope	Tweet Language	Granularity Level	Location Inference Scenario
[17]	Brazil (Sao Paulo, Rio de Janeiro, Belo Horizonte)	Portuguese	City	Inferring the Twitter user's location
[18]	UK	English	City	User location inference based on social network
[19]	Japan	English	City	Location estimation
[20]	Worldwide	Multilingual	Country/City	Inferring User home location and origin of tweets
[21]	Worldwide	Multilingual	City/State/Country	Inferring origin of tweets
[22]	USA	English	City	Estimation of Twitter user location
[23]	Worldwide	Multilingual	Country	Estimation of Twitter user location
[24]	USA	English	City /State	Profiling User's home location
[25]	Worldwide	English	Country/State/City	User location identification, future location prediction and categorization
[26]	Worldwide	English	Country/City	Inferring Twitter User's home location
[27]	India	Hindi, English	State/City	Location prediction from tweets and event classification
[28]	Worldwide	Multilingual	Country /City	Predicting Geographic location of tweet creation
[29]	Worldwide	Multilingual	Country/State/City	Inferring user's location
[30]	worldwide,3k cities	English	City	Predicting the location of tweet posted
[6]	worldwide	Multilingual	Country	Predicting Tweet's country of origin

Reference	Combination of Features
[7]	Tweet content, UTC Offset, Tweet creation time, Time zone, User location, User account creation time
[8]	Tweet content
[9]	Geotagging information, Time of tweet posting, GIS data (Google places API)
[10]	Geotagging information, Tweet content
[11]	Tweet content, User social graph
[12]	Tweet content, User social graph, LSBN data
[13]	Tweet content, User friends network, Geotagging information
[14]	Tweet content, Posting time of the tweet, Time zone, User history (check-ins)
[15]	Geotagging information, Time of tweet posting, Foursquare data
[16]	Tweet content
[17]	Tweet content, User friendship networks
[18]	Tweet content, Geotagging information, User Social friendship network
[19]	Geotagging information, Tweet text
[20]	Tweet content, User location field, Web URLs, Time zone, UTC Offset, Geotagging information
[21]	User location, User description, Time zone, Tweet language, Tweet content
[22]	Tweet content

Reference	Combination of Features
[23]	User Social friendship Network, User location, Username, Geotagging information
[24]	Tweet content, User social network
[25]	Tweet content
[26]	[Twitter: Geotagging information, User Location],[Google+: Places Stayed, Education and Employment Location], [Foursquare: User home city and Venue city]
[27]	Tweet content, User history (Geotagging information from Previous Tweets)
[28]	Tweet content, User description, Username, User location, Tweet language, User interface language, Time zone, Tweet Posting Time (UTC)
[29]	Tweet content (Location Specific Words), User Profiles Features (User language, Gender, Age, Number of Followers and Followees)
[30]	Geotagging information, Tweet content, User location, Time Zone, Posting Time of Tweet
[6]	Tweet content, User location, Tweet language, User interface language, Time zone, Offset, User Description, Username

4.Structure of Tweet

5.Structure of Tweet JSON(JavaScript Object Notation)

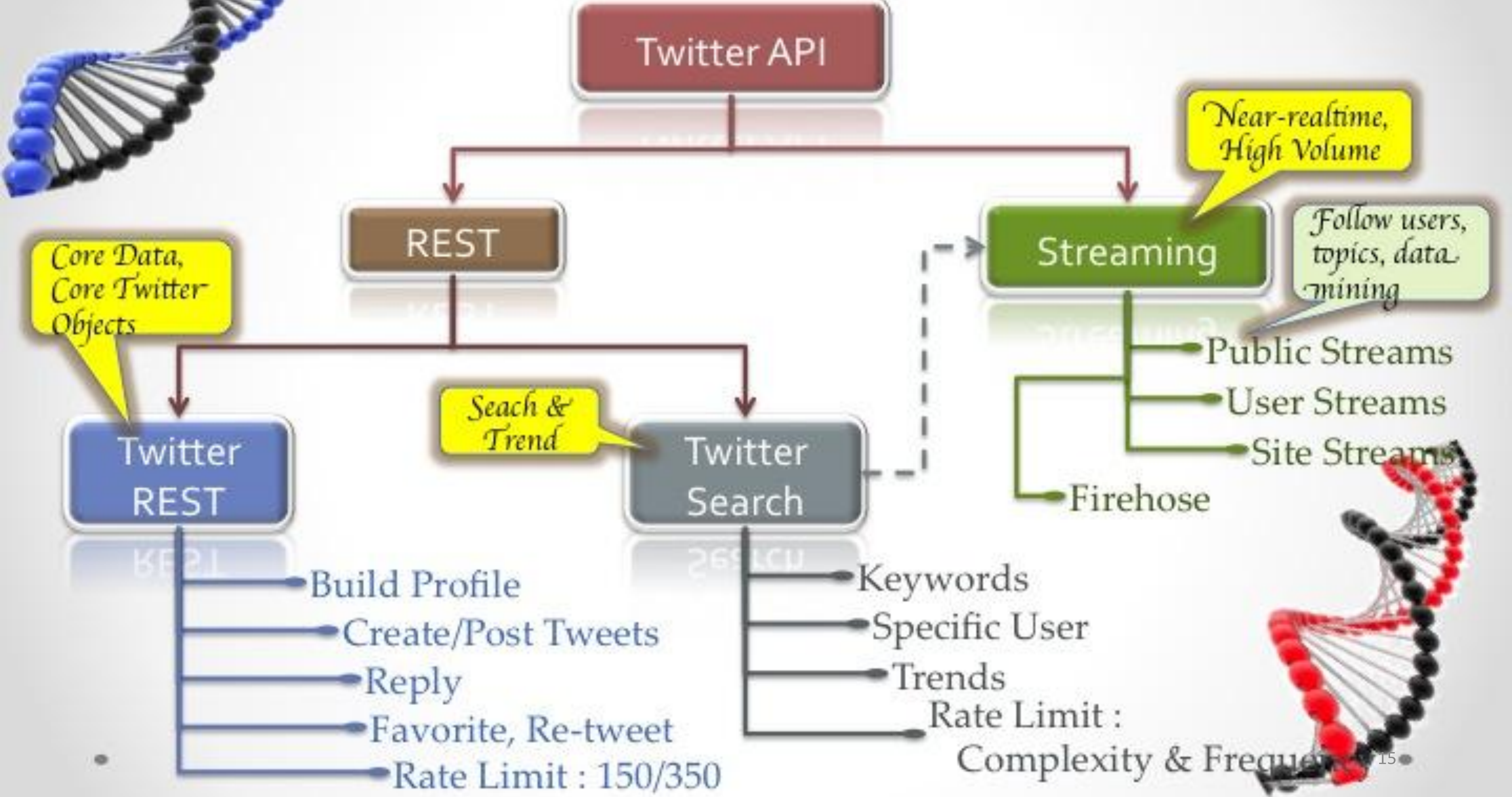
JSON
created_at: Wed Mar 28 08:07:01 +0000 2018
id: 978906290381557800
id_str: 978906290381557766
text: Patience is not simply the ability to wait -
source: Instagram
truncated: false
{} user
{} geo
{} coordinates
{} place
is_quote_status: false
quote_count: 0
reply_count: 0
retweet_count: 0
favorite_count: 0
{} entities
favorited: false
retweeted: false
possibly_sensitive: false
filter_level: low
lang: en
timestamp_ms: 1522224421483

{} user
id: 4786494492
id_str: 4786494492
name: Ruddi Cabalida
screen_name: ruddididodu
location: Dumaguete City, Central Visayas
url: http://www.facebook.com/ruddi.cabalida
translator_type: none
protected: false
verified: false
followers_count: 63
friends_count: 149
listed_count: 0
favourites_count: 685
statuses_count: 407
created_at: Wed Jan 20 02:23:49 +0000 2016
utc_offset: 28800
time_zone: Beijing
geo_enabled: true
lang: en
contributors_enabled: false
is_translator: false

{} geo
type: Point
coordinates: []
[0]: 9.3105722
[1]: 123.3092379
{} coordinates
type: Point
coordinates: []
[0]: 123.3092379
[1]: 9.3105722
{} place
id: 00c0d8fe55323853
url: https://api.twitter.com/1.1/geo/id/00c0d8fe55323853.json
place_type: city
name: Dumaguete City
full_name: Dumaguete City, Central Visayas
country_code: PH
country: Republic of the Philippines
{} bounding_box
type: Polygon
coordinates: []
[0]: []
[0]: []

[1]: 9.263678
[1]: []
[0]: 123.240723
[1]: 9.326806
[2]: []
[0]: 123.314431
[1]: 9.326806
[3]: []
[0]: 123.314431
[1]: 9.263678
attributes: [object Object]
is_quote_status: false
quote_count: 0
reply_count: 0
retweet_count: 0
favorite_count: 0
{} entities
favorited: false
retweeted: false
possibly_sensitive: false
filter_level: low
lang: en
timestamp_ms: 1522224421483

6.Retrival of tweets using Twitter APIs



- **REST API** - To get information, user must specifically request it
  - Well over 50 different REST API "Resources"
- **Streaming API** - Once request is made, provides continuous stream of updates without further input from user (up to 1% full twitter stream)
  - Has "Public", "User", and "Site" streams

The APIs provide 18000 user profiles per minute. The streaming API return about 36000 tweets per 15 minutes and a search returns 18000 tweets per minute. The APIs provide 1% of location tweets posted on twitter.

	Cost	Rate Limit	Learning Curve	Support	Analytic Features	Customization
Social Media Monitoring	High	Depends	Low/Mid	High	High	Low
Twitter Authorized Reseller	High	No	Mid	Mid	Mid	High
Twitter API	Low	Yes	High	Low	Low	High

More on API ratelimits: https://developer.twitter.com/en/docs/basics/rate-limits

4.Proposed System

Problem statement

- Improving the accuracy of country level location classification of the existing system [6], by introducing an extra feature called as user friends & followers along with all the features already used in the existing system.

- Since, user friends and followers feature cannot be used in a Realtime scenario the model which is going to be built will be a non-realtime system.

Objectives and Proposed System

- Build a classification system which will determine a tweets country of origin using **a total of 11 features.**

- Use our training data to train different classifier models and then the testing data as an input to the models to predict a country for the tweet.

- We also determine which classifier does perform best by giving more accurate classification results then others.

Stages in Realizing the Proposed System

Primarily there are 4 stages involved:

1) Creating a Twitter Application

2) Accessing the Twitter streaming API, downloading and saving the streaming tweets to a .json file.

3) Extraction, processing, conditioning & training-testing dataset generation

    a) Features Extraction Phase
    b) Processing & Conditioning Phase
    c) Training & Testing set generation

4) Classification Phase

Application Management

johnnydepp.tweetlocator

Test OAuth

Details | Settings | Keys and Access Tokens | Permissions

Application Settings

Keep the "Consumer Secret" a secret. This key should never be human-readable in your application.

Consumer Key (API Key)	[redacted]
Consumer Secret (API Secret)	[redacted]
Access Level	Read, write, and direct messages (modify app permissions)
Owner	MALHAR_JOHNNY
Owner ID	913730626561495040

Application Actions

Regenerate Consumer Key and Secret | Change App Permissions

Your Access Token

This access token can be used to make API requests on your own account's behalf. Do not share your access token secret with anyone.

Access Token	[redacted]
Access Token Secret	[redacted]
Access Level	Read and write
Owner	MALHAR_JOHNNY
Owner ID	913730626561495040

Token Actions

Regenerate My Access Token and Token Secret | Revoke Token Access

Stages in Realizing the Proposed System

<u>Primarily there are 4 stages involved:</u>

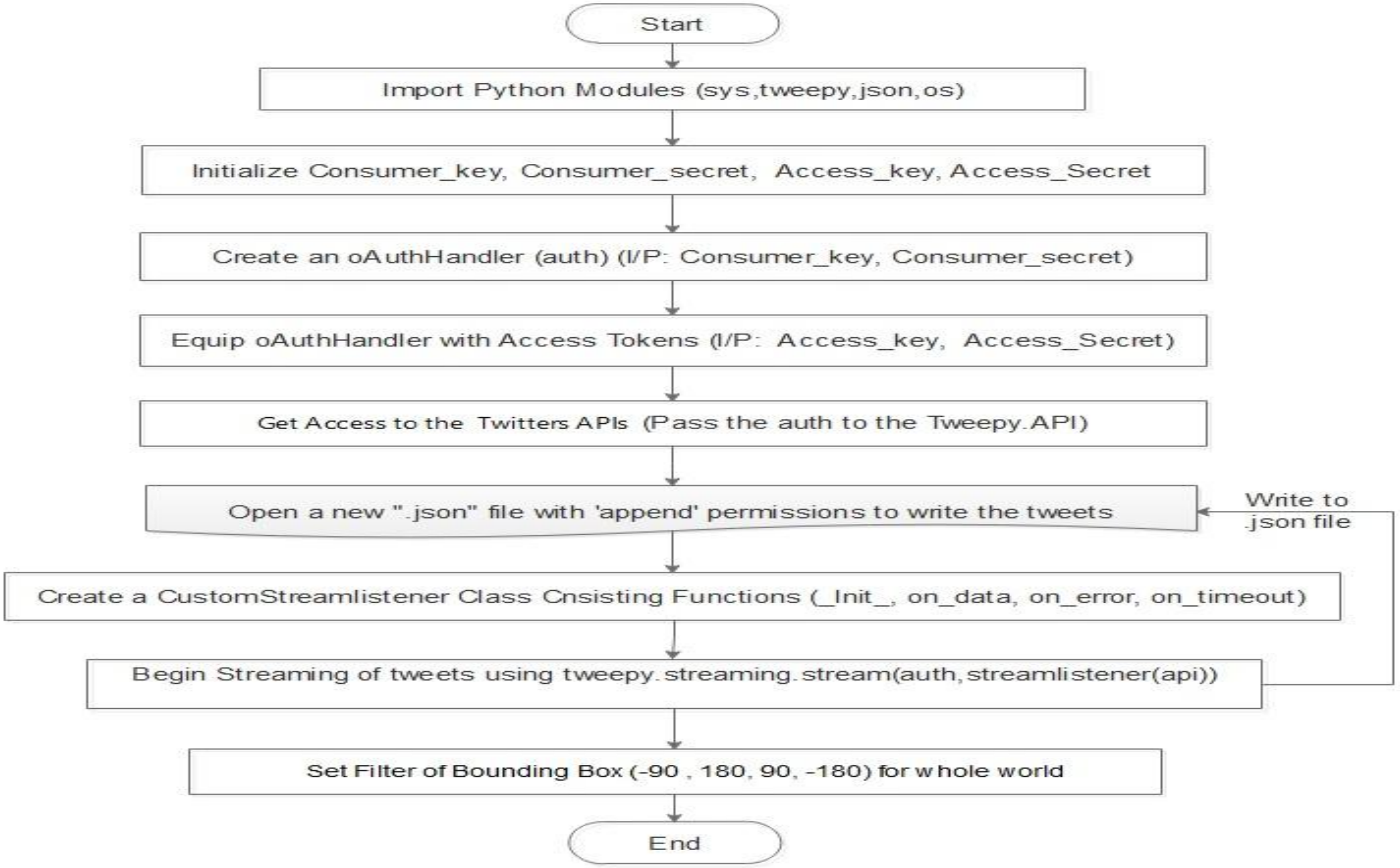
1) **Creating a Twitter Application**

2) Accessing the Twitter streaming API, downloading and saving the streaming tweets to a .json file.

3) Extraction, processing, conditioning & training-testing dataset generation

    a) Features Extraction Phase
    b) Processing & Conditioning Phase
    c) Training & Testing set generation

4) Classification Phase


Start
Import Python Modules (sys,tweepy,json,os)
Initialize Consumer_key, Consumer_secret, Access_key, Access_Secret
Create an oAuthHandler (auth) (I/P: Consumer_key, Consumer_secret)
Equip oAuthHandler with Access Tokens (I/P: Access_key, Access_Secret)
Get Access to the Twitters APIs (Pass the auth to the Tweepy.API)
Open a new ".json" file with 'append' permissions to write the tweets
Write to .json file
Create a CustomStreamlistener Class Cnsisting Functions (_Init_, on_data, on_error, on_timeout)
Begin Streaming of tweets using tweepy.streaming.stream(auth,streamlistener(api))
Set Filter of Bounding Box (-90 , 180, 90, -180) for whole world
End

Stages in Realizing the Proposed System

Primarily there are 4 stages involved:

1) **Creating a Twitter Application**

2) **Accessing the Twitter streaming API, downloading and saving the streaming tweets to a .json file.**

3) Extraction, processing, conditioning & training-testing dataset generation

    a) Features Extraction Phase
    b) Processing & Conditioning Phase
    c) Training & Testing set generation

4) Classification Phase

Stage 3 Extraction, Processing, Conditioning & Training-Testing dataset Generation

(1.) Features Extraction Phase

- Place (Country)
- User Location (Location words)
- Coordinates
- User Location (Organization words)
- User Interface Language
- Tweet content
- Tweet Language
- User name
- User Description
- User Friends Locations
- Time zone

(2.) Pre-Processing & Conditioning Phase

The basic idea behind this stage is as follows:

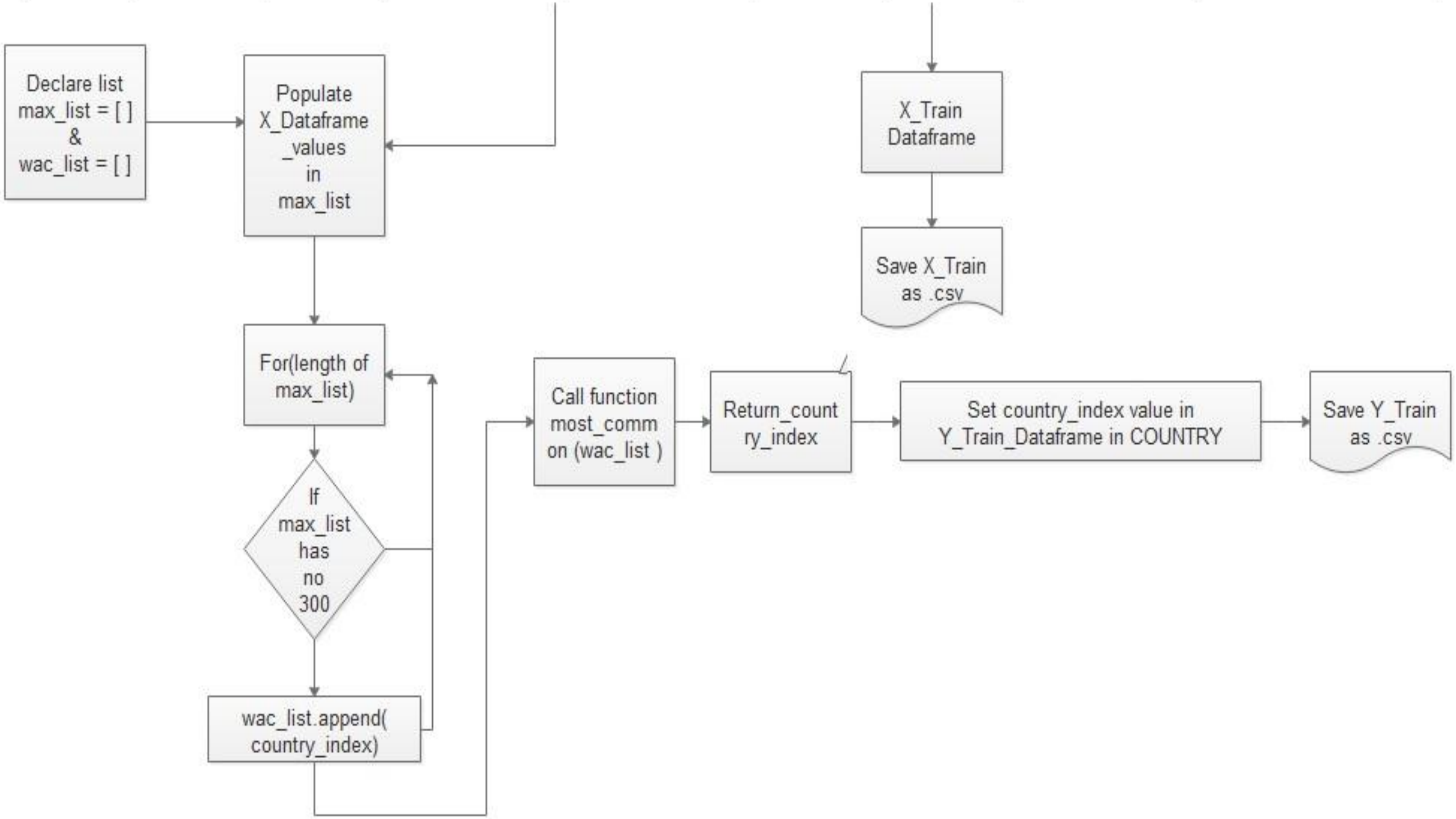
- The preprocessing stage should give us **name of country derived from processing the features**. For some features we have extensively used **Language detectors, Language Translators and Nominatim Gazetteer**.

- In the Conditioning phase, **the country name derived from the user location is converted to a numeric equivalent so that it can be given as an input to the classifiers used for training and prediction** .If a feature doesn't provide us with any country name then in the conditioning phase we set the None value as '300' which denotes 'No country determined'.

- For all **eleven features** we have the pre-processing and conditioning phases applied to them.

(3.) Training & Testing Set Generation

Index	Place	Coordinates	Timezone	Userlocation_loc	Userlocation_org	Username	Textcontent	User description	User_Interface_language	User_Tweet_Langugae	User_Friends_Location
0 to ..	var_plce	var_coord	var_TZ	var_UL_LC	var_UL_OG	var_usrnm	vat_txt	var_desc	var_Intfrc	var_twt	var_usr_friends



Representation of Features(X) and Country_class(Y) in .csv file

place	coordinates	Timezone	LOCFLD_Loc	LOCFLD_Org	USERNAME	Textcont	Descripcont	Intrfc_Lang	Twt_Lang	User_Friends_location	Country_class
300	300	109	109	300	109	300	300	109	109	109	109
300	300	235	300	300	235	300	300	235	235	182	235
300	300	235	232	300	235	300	300	235	235	39	235
300	300	208	208	300	208	300	300	208	208	208	208
300	300	13	13	300	13	13	13	13	13	235	13
300	300	300	300	300	300	300	300	300	300	235	235
300	300	234	300	300	234	234	300	234	234	234	234
300	300	39	147	300	39	300	300	39	39	39	39
300	300	234	181	300	234	300	234	234	234	39	234
300	300	234	234	300	234	234	300	234	234	234	234

Details of Dataset

- A total of **1150 tweets** were preprocessed and conditioned.

- The reason for so less number of instances is the **preprocessing phase consumes lot of time**.

- For the retrieval of user friends, a request is sent to twitter whereby twitter which **calculates the number of friends of users** and returns it back to us but it is a **time consuming process** especially if the **number of friends per user are very high.**

- The **twitter endpoint API service** is **rate limited** hence we have to give input to the API with **some limited tweets(approximately 6 to 7)** everytime and with a slight amount of delay.

- There is a slight unbalancedness in our dataset, to **counter this unbalancedness** we need to have a **large and a refined dataset** where all classes i.e. **countries are well represented**.

Stage 4 Classification Stage

The model performs **multiclass classification** as there are **248 countries** all over the world which will ultimately count **248 classes**.

But due to scarcity of instances of tweets in our dataset, there are **only 83 countries represented** which makes about **83 classes** to be classified.

The main concern for us was **which classifiers do give best results** in terms of **Accuracy, Precision, Recall and F1 measure.**

We have tried a total of 9 classifiers:

- Logistic Regression
- Kneighbors
- Random Forest
- Gaussian Naive Bayes
- Multinomial Naive Bayes
- Decision Tree
- Support Vector Machine (using Linear, RBF and Polynomial Kernels)

Since there is an unbalancedness in our dataset we derived results using averages. There are three averages we have used and they are as follows:

1) **Macro Average:** A macro-average will **compute the metric independently for each class** and **then take the average (hence treating all classes equally)**,i.e. calculate precision values of each class and sum them up. In **macro-averaging, we average the performances of each individual class.**

2) **Micro Average:** A micro-average will **aggregate the contributions of all classes to compute the average metric**, ie in Micro-average method, **we sum up the individual true positives, false positives, and false negatives of the system for different sets and the apply them to get the statistics**.

Macro-average method can be used when you want to know how the system performs overall across the sets of data. You should not come up with any specific decision with this average. On the other hand, **micro-average can be a useful measure when your dataset varies in size.**

In a multi-class classification setup, micro-average is preferable if you suspect there might be class imbalance (i.e you may have many more examples of one class than of other classes).

3) **Weighted Average: Calculate metrics for each label, and find their average, weighted by support (the number of true instances for each label).** This alters 'macro' to account for label imbalance; it can **result in an F-score that is not between precision and recall**.
4) **None:** For None average the scores for each class are returned.

2x2 Confusion Matrix		ACTUAL TRUTH	
		CORRECT	INCORRECT
SYSTEM PREDICTION	SELECTED	TRUE POSITIVE (TP)	FALSE POSITIVE (FP)
	NOT SELECTED	FALSE NEGATIVE (FN)	TRUE NEGATIVE (TN)

- **TP:** A true positive is an outcome where the model correctly predicts the positive class.
- **TN:** A true negative is an outcome where the model correctly predicts the negative class.
- **FP:** A false positive is an outcome where the model incorrectly predicts the positive class.
- **FN:** A false negative is an outcome where the model incorrectly predicts the negative class.

Metrics Used:

1) Accuracy:

Gives us the number of correct predictions to the total number of predictions.

$$\boldsymbol{Accuracy} = \frac{(\boldsymbol{TP}+\boldsymbol{TN})}{(\boldsymbol{TP}+\boldsymbol{FP}+\boldsymbol{TN}+\boldsymbol{FN})} \qquad \textbf{(1)}$$

2) Precision:

Identifies what proportion of positive identifications was actually correct. It states "Of all the samples we classified as true how many are actually true?

$$\boldsymbol{Precision}(\boldsymbol{P}) = \frac{(\boldsymbol{TP})}{(\boldsymbol{TP}+\boldsymbol{FP})} \qquad \textbf{(2)}$$

3) Recall:

It tries to identify what proportion of actual positives was identified correctly?, i.e. Of all the actual true samples how many did we classify as true?

$$\boldsymbol{Recall}(\boldsymbol{R}) = \frac{(\boldsymbol{TP})}{(\boldsymbol{TP}+\boldsymbol{FN})} \qquad \textbf{(3)}$$

4) F1 score:

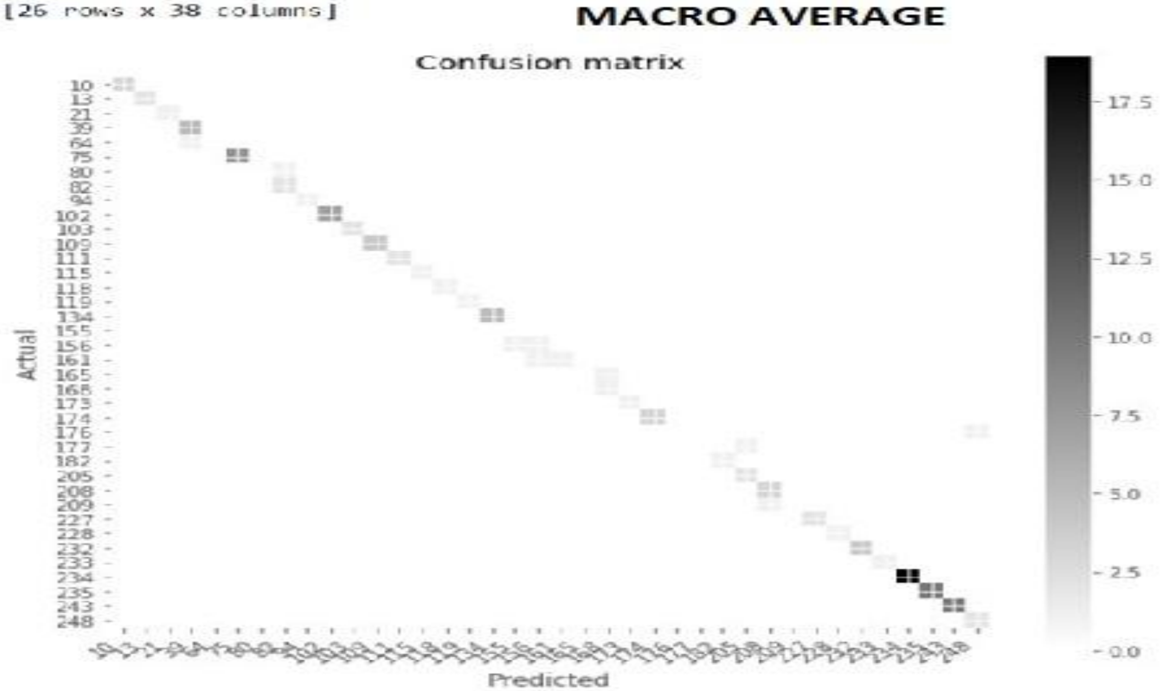
It is the weighted Harmonic mean of Precision and Recall. It returns the minimum value of the average of the precision and Recall. Since **precision and recall are a kind of a tradeoff,** F1score helps us by giving a single value in determining which classifier is better than whom.

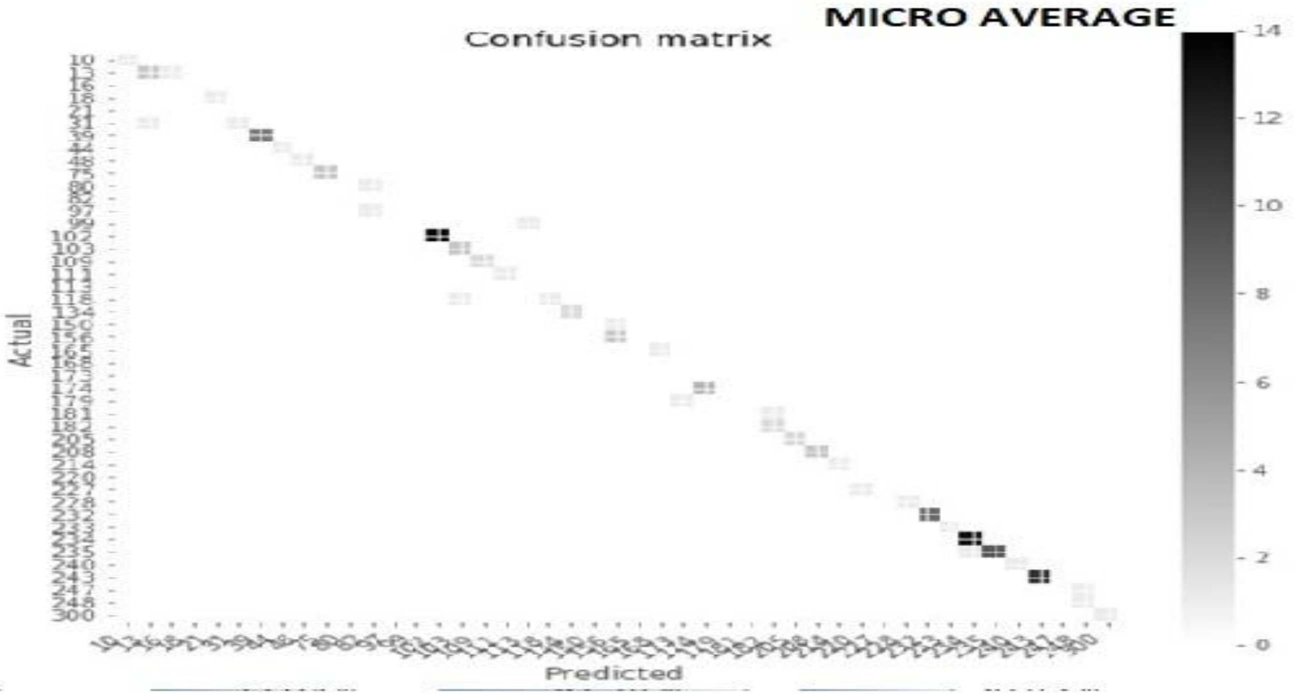
$$\boldsymbol{F1-measure} = \frac{(\boldsymbol{2*PR})}{(\boldsymbol{P}+\boldsymbol{R})} \qquad \textbf{(4)}$$

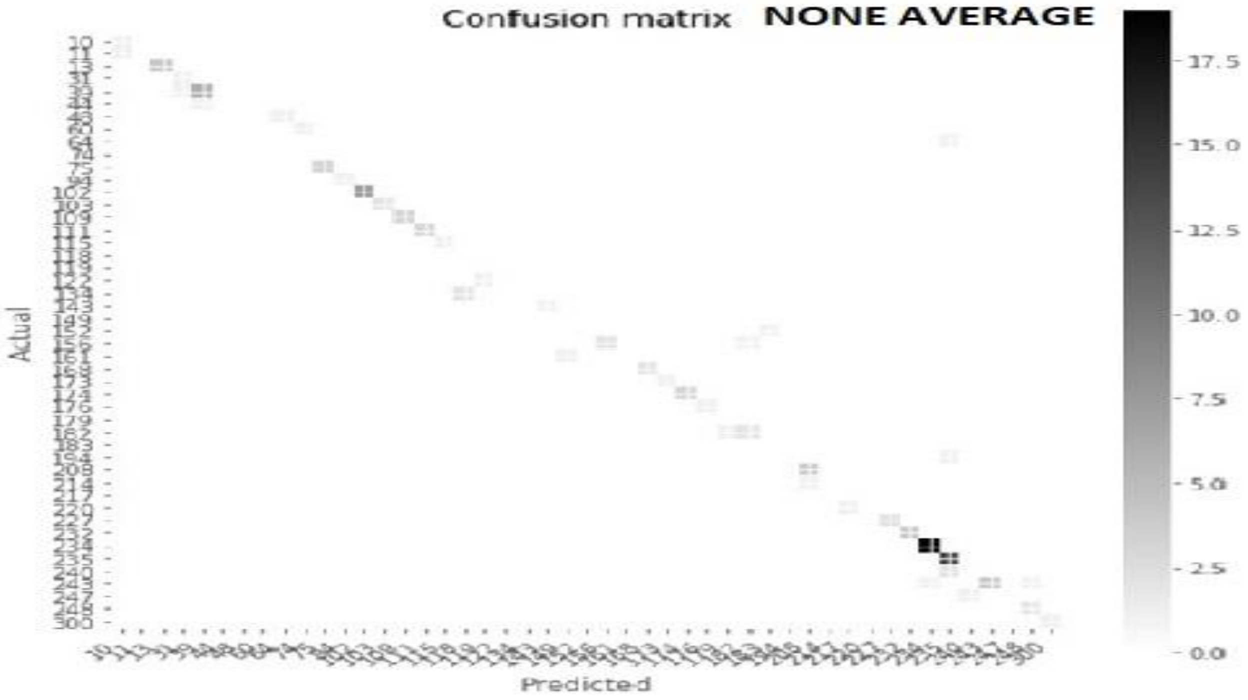
	Multinomial NB	Decision Tree	Linear SVM Kernel	RBF svm Kernel	Polynomial Kernel
Precision					
Macro	4.48%	60.26%	32.24%	56.11%	40.62%
Weighted	11.74%	86.17%	43.04%	74.64%	74.77%
None	10.00%	94.00%	48.00%	66.00%	75.00%
Micro	19.13%	86.95%	43.47%	53.91%	68.69%
Recall					
Macro	6.15%	57.28%	30.48%	47.38%	41.22%
Weighted	11.30%	88.69%	43.47%	60.86%	73.91%
None	13.00%	90.00%	44.00%	57.00%	67.00%
Micro	19.13%	86.95%	43.47%	53.91%	68.69%
F1 Score					
Macro	4.75%	57.92%	28.60%	49.28%	38.39%
Weighted	11.38%	86.81%	40.81%	61.26%	73.53%
None	10.00%	92.00%	42.00%	55.00%	69.00%
Micro	19.13%	86.95%	43.47%	53.91%	68.69%
Accuracy					
Macro	15.65%	86.95%	46.09%	59.13%	65.21%
Weighted	11.30%	88.69%	43.47%	60.86%	73.91%
None	13.04%	84.34%	44.34%	56.52%	66.95%
Micro	19.13%	86.95%	43.47%	53.91%	68.69%

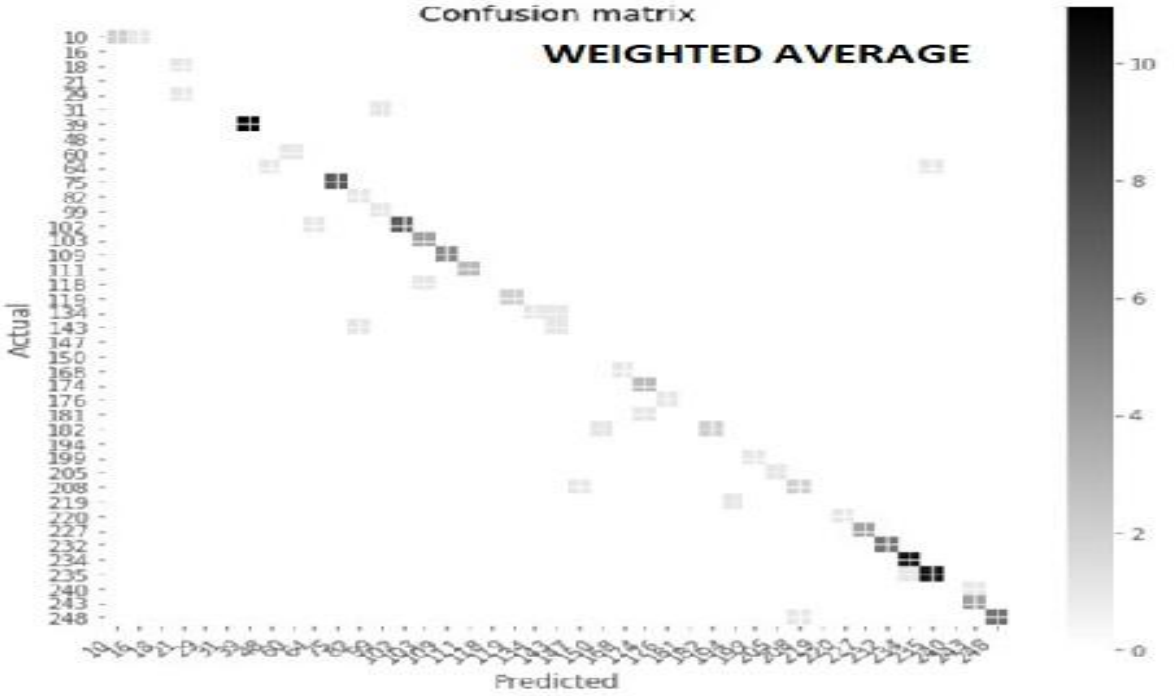
	Logistic Regression	Kneighbors	Random Forest	Gaussian Naive Bayes
Precision				
Macro	11.06%	30.49%	75.22%	26.23%
Weighted	23.43%	57.12%	85.86%	30.79%
None	24.00%	55.00%	90.00%	31.00%
Micro	26.95%	55.65%	81.74%	30.43%
Recall				
Macro	14.06%	33.06%	78.95%	26.15%
Weighted	27.82%	59.13%	85.22%	26.08%
None	24.00%	57.00%	91.00%	27.00%
Micro	26.95%	55.65%	81.74%	30.43%
F1 Score				
Macro	10.87%	30.46%	76.31%	25.37%
Weighted	22.06%	56.11%	84.87%	25.97%
None	21.00%	53.00%	90.00%	28.00%
Micro	26.95%	55.65%	81.74%	30.43%
Accuracy				
Macro	23.47%	60.86%	93.04%	20.86%
Weighted	27.82%	59.13%	85.22%	26.09%
None	24.34%	57.39%	91.30%	26.95%
Micro	26.95%	55.65%	81.74%	30.43%

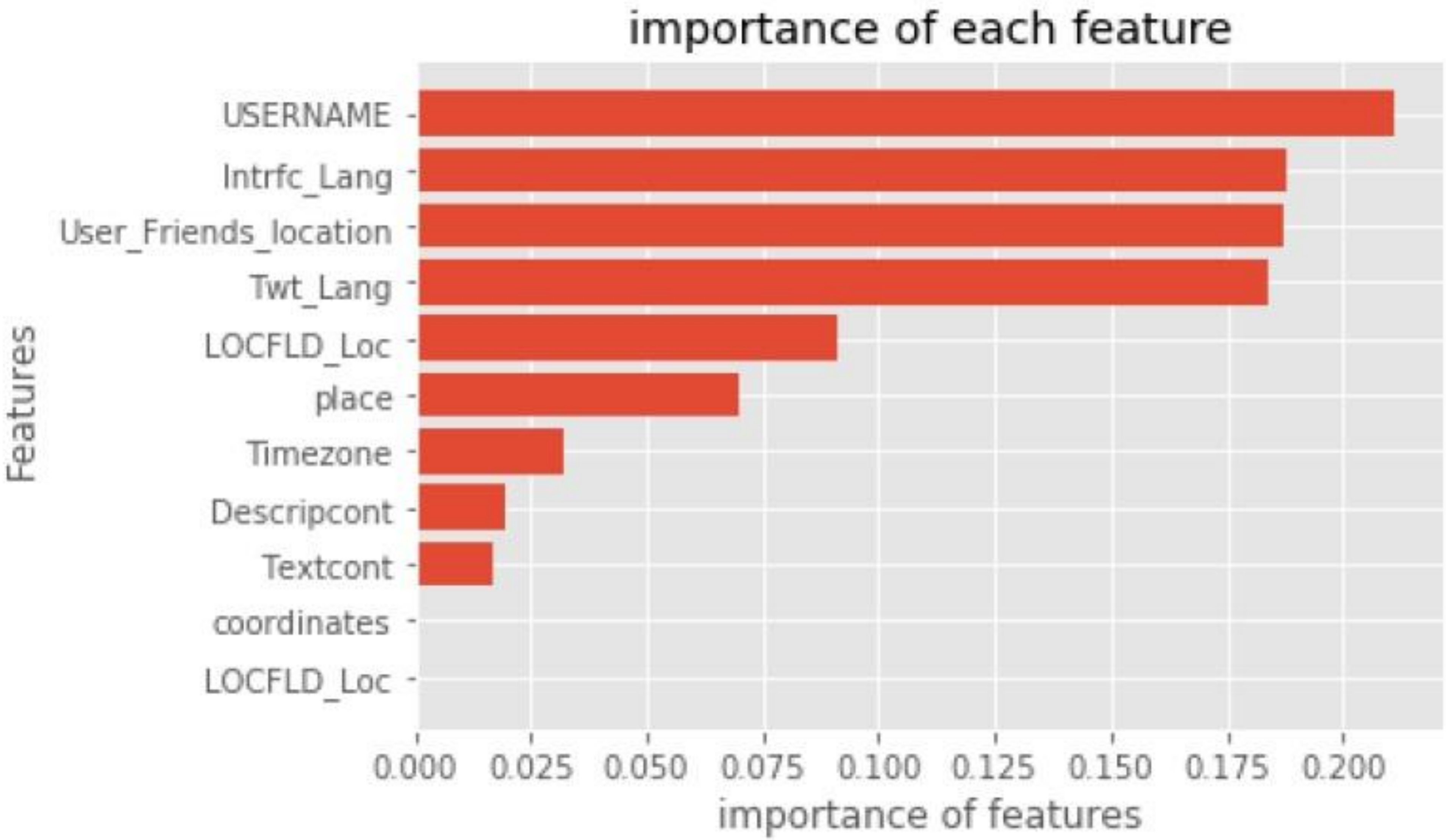
[26 rows x 38 columns]










importance of each feature
USERNAME
Intrfc_Lang
User_Friends_location
Twt_Lang
LOCFLD_Loc
place
Timezone
Descripcont
Textcont
coordinates
LOCFLD_Loc
Features
0.000 0.025 0.050 0.075 0.100 0.125 0.150 0.175 0.200
importance of features

CONCLUSIONS

- The proposed model performs country-level classification of tweets posted on twitter.
- We considered the features from the tweet metadata and performed preprocessing and conditioning of the data.
- Our main contribution in this works is incorporating the user friends location along with the features used in existing work.
- The proposed system does not deal with real-time tweets but certainly performs better then the previous systems.
- This system is able to handle multilingual tweets.
- The generation of the training and testing sets is time consuming process but from the obtained results it is proved that results are better obtained if the data set is larger and a balanced one.

Bibliography

[1]. Mislove, A., et al., Understanding the Demographics of Twitter Users.ICWSM, 2011. 11 (5th): p. 25-25.

[2] Sloan, L., et al., Knowing the tweeters: Deriving sociologically relevant demographics from Twitter.Sociological research online, 2013. 18 (3): p. 1-11.

[3] Li, R., S. Wang, and K.C.-C. Chang, Multiple location profiling for users and relationships from social network and content.Proceedings of the VLDB Endowment, 2012. 5 (11): p. 1603-1614.
Hecht, B., et al.

[4] Energy-accuracy trade-off for continuous mobile device location. in Proceedings of the 8th international conference on Mobile systems, applications, and services.

[5] Tweets from Justin Bieber's heart: the dynamics of the location field in user profiles. in Proceedings of the SIGCHI conference on human factors in computing systems.
Lin, K., et al.

[6] Zubiaga, A., et al., Towards real-time, country-level location classification of worldwide tweets.IEEE Transactions on Knowledge and Data Engineering, 2017. 29 (9): p. 2053-2066.

[7]. Lau, J.H., et al., End-to-end network for twitter geolocation prediction and hashing.arXiv preprint arXiv:1710.04802, 2017.

[8]. Chi, L., et al. Geolocation prediction in Twitter using location indicative words and textual features. in Proceedings of the 2nd Workshop on Noisy User-generated Text (WNUT).

[9]. Huang, Q., G. Cao, and C. Wang. From where do tweets originate?: a GIS approach for user location inference. in Proceedings of the 7th ACM SIGSPATIAL International Workshop on Location-Based Social Networks.

[10]. Ryoo, K. and S. Moon. Inferring twitter user locations with 10 km accuracy. in Proceedings of the 23rd International Conference on World Wide Web.

[11]. Kotzias, D., T. Lappas, and D. Gunopulos. Addressing the Sparsity of Location Information on Twitter. in EDBT/ICDT Workshops.

[12]. Elmongui, H.G., H. Morsy, and R. Mansour. Inference models for Twitter user's home location prediction. in Computer Systems and Applications (AICCSA), 2015 IEEE/ACS 12th International Conference of.

[13]. Rakesh, V., C.K. Reddy, and D. Singh. Location-specific tweet detection and topic summarization in twitter. in Proceedings of the 2013 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining.

[14]. Liu, H., B. Luo, and D. Lee. Location type classification using tweet content. in Machine Learning and Applications (ICMLA), 2012 11th International Conference on.

[15]. Falcone, D., et al. What is this place? Inferring place categories through user patterns identification in geo-tagged tweets. in Mobile Computing, Applications and Services (MobiCASE), 2014 6th International Conference on.

[16]. Hoang, T.B.N. and J. Mothe, Location extraction from tweets.Information Processing & Management, 2018. 54 (2): p. 129-144.

[17]. Rodrigues, E., et al., Exploring multiple evidence to infer users' location in Twitter.Neurocomputing, 2016.171: p. 30-38.

[18]. Rout, D., et al. Where's@ wally?: a classification approach to geolocating users based on their social ties. in Proceedings of the 24th ACM Conference on Hypertext and Social Media.

[19]. Kawano, M. and K. Ueda. Where Are You Talking From?: Estimating the Location of tweets Using Recurrent Neural Networks. in Proceedings of the Second International Conference on IoT in Urban Space.

[20]. Schulz, A., et al. A Multi-Indicator Approach for Geolocalization of Tweets. in ICWSM.

[21]. Priedhorsky, R., A. Culotta, and S.Y. Del Valle. Inferring the origin locations of tweets with quantitative confidence. in Proceedings of the 17th ACM conference on Computer supported cooperative work & social computing.

[22]. Chandra, S., L. Khan, and F.B. Muhaya. Estimating twitter user location using social interactions--a content based approach. in Privacy, Security, Risk and Trust (PASSAT) and 2011 IEEE Third Inernational Conference on Social Computing (SocialCom), 2011 IEEE Third International Conference on.

[23]. Alonso-Lorenzo, J., E. Costa-Montenegro, and M. Fernández-Gavilanes. Language independent big-data system for the prediction of user location on Twitter. in Big Data (Big Data), 2016 IEEE International Conference on.

[24]. Li, R., et al. Towards social user profiling: unified and discriminative influence model for inferring home locations. in Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining.

[25]. Jaiswal, A., W. Peng, and T. Sun. Predicting time-sensitive user locations from social media. in Proceedings of the 2013 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining.

[26]. Pontes, T., et al. Beware of what you share: Inferring home location in social networks. in Data Mining Workshops (ICDMW), 2012 IEEE 12th International Conference on.

[27]. Singh, J.P., et al., Event classification and location prediction from tweets during disasters.Annals of Operations Research, 2017: p. 1-21.

[28]. Huang, B. and K.M. Carley. On Predicting Geolocation of Tweets using Convolutional Neural Networks. in International Conference on Social Computing, Behavioral-Cultural Modeling and Prediction and Behavior Representation in Modeling and Simulation.

[29]. Qian, Y., et al., A Probabilistic Framework for Location Inference from Social Media.arXiv preprint arXiv:1702.07281, 2017.

[30]. Dredze, M., M. Osborne, and P. Kambadur. Geolocation for twitter: Timing matters. in Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies.

THANK YOU!!!