

# Location prediction on Twitter using machine learning Techniques

## Step-1: Data Scraping

we will scrap data from twitter using **Twitter API** with the help of Python's **Tweepy** Library. The twitter data API makes incredibly easy to query tweets by Location , Keywords etc.

To get started ,you will need to do the following things

1.Set up twitter account if you don't have it already

2.Using Your Twitter account, you will need to apply for Developer Access and then create an Application that will generate the API credentials that you will use to access Twitter from Python

3.Import **Tweepy** Library

Once you have done these things , you are ready to begin querying Twitter's API to see what you can learn about tweets.

Once you have your Twitter app set-up , you are ready to access tweets in Python . Begin by importing necessary Python Libraries

To access the Twitter API, you will need 4 things from your twitter App page. These keys are located in your Twitter app settings in the **Keys and Access Token** tab. The keys are consumer\_key, consumer\_secret , access\_token, access\_token

## Step-2: Data Preprocessing and Feature Engineering

As you know twitter data always contains a lot of noise like emojis , keyword. Hence it becomes necessary to clean it. In this session we will clean the text and locations using different techniques for example removing irrelevant words like punctuation, articles and symbols. Also we need to remove url which is at the end of each text.

## Step-3: Model Building

Before proceeding into model building we need to convert our text into numerical values so that our model is able to understand it. For this I would apply two techniques , first one is bag of words model and second one is TF-IDF Techniques , which converts our text into numerical vectors. After converting our text into numerical values then we would split our dataset into train and test set, so that we can test our model on unseen data. Then we would apply machine learning algorithms. In this problem, we would only use GaussianNB ,MultinomialNB, SVC, DecisionTreeClassifier algorithms. Then we would track accuracy for each model

