**SYSTEM ANALYSIS**

**EXISTING SYSTEM:**

In the Existing system to the problem of finding location from social media content. The Social Networks from and motivated by Term frequency (TF) and inverse document frequency (IDF), they arrived Inverse City Frequency (ICF) and Inverse Location Frequency (ILF) respectively. They raked the features by using these frequency values and TF then by TF values. From this they arrived that local words spread in document in few places and have high ICF and ILF values. They approached model for identifying local words indicative or used in certain locations only. They aimed to identify automatically by ranking the local words by their location, and they find their degree of association of location words associated to particular location or cities.

**DISADVANTAGES OF EXISTING SYSTEM:**

* The issue of location prediction related named as geolocation prediction is examined for Wikipedia and web page documents.
* Entity recognition from these formal documents has been researched for years.
* The location prediction problem from twitter depends highly on tweet content.
* **Algorithm**: Term Frequency (TF) and Inverse Document Frequency (IDF)

**PROPOSED SYSTEM:**

Live stream of twitter data is collected as dataset using authentication keys. The aim of proposed system is to predict the user location from twitter content considering user home location, tweet location and tweet content. To handle this we used three machine learning approaches to make prediction easier and finding the best model amongst them. Live tweet stream from twitter for keyword “apple” is collected and stored in Tweettable. Live twitter data can be collected by registering a consumer\_key, consumer\_secret, access\_token, access\_token\_secret for authentication and collecting live stream of tweets. We have collected more than 1000 tweets of particular keywords such as Indian city hashtag names. You can also search tweets based on hashtags.

**ADVANTAGES OF PROPOSED SYSTEM:**

* The information extracted from live includes tweetid, name, screen\_name, tweet\_text, HomeLocation, TweetLocation, MentionedLocation.
* Tweet text is compared with natural language tool kit package available in python to extract data from Cursor to Pandas Dataframe.
* Python programming, with few libraries used are scikit learn, numpy, pandas and geography.

**Algorithm**: Naive Bayes, Support Vector Machine, Decision Tree