# Extraction of Information From A Social Media Platform Like Twitter

**Prerana Gandhi[1], Chandresh Parekha[2], Jaimin Somani[3]**
[1, 2] Department of IT& Telecommunication
[1, 2]Raksha Shakti University,Ahmedabaad
[3]Avlance Global Solutions, Vadodara

***Abstract-****The platform of twitter a social media can be used to take the information using programming languages such as python, this obtained data can be used for various analysis. It also gives the views of privacy of information in certain platforms. The kind of data can also be used to identify the impact it has on a large scale as well as a person's preferences and personality can also be assessed.*

## I. INTRODUCTION

The technology advances have led to a large amount of data being uploaded in the social media platforms on a large scale. One of such platform is twitter, Twitter is one of the extensively used social media platform to express the point of view. Information of users in social media keeps on increasing day by day.

Users Post their view, thoughts, life events on social media and that too without any constraint and unwillingness. Some of the social media allow users to interact with only with their friend's and sharing their post with very easy level of privacy or without any privacy. due to simple and easy privacy policies, and easy availability of some social media, users are drifted from traditional means of communication such as blogs or mailing list to microblogging site such as Twitter, Facebook etc. very large amount of text data in the form of messages on social media make it very attractive medium for data analysis for the researchers.

This has led to countless tools, data services and analytics platforms. However, this easy convenience of social media data for various research may change suggestively due to in demand pressures.

## II.TERMINOLOGY

Natural language processing— It is the application of computational techniques to the analysis and synthesis of natural language and speech. It is a part of computer science, artificial intelligence and linguistics concerned with the interfaces between computers and human languages. Precisely, it is the process of a computer mining meaningful data from natural language input or producing natural language output.

Opinion mining—is sentiment mining, opinion or sentiment extraction. It is a type of natural language processing for tracing the mood of the public about a certain artefact. Opinion mining, which is also known as sentiment analysis, involves building a system to gather and classify opinions about a product.

Scraping— It is collecting online data from social media and other sources in the form of unstructured format. It can also be termed as site scraping, web harvesting and extraction of web data.

Sentiment analysis—It refers to the use of natural language processing, computational linguistics and text analytics to recognize and extract particular information in source materials.

Text analytics—It involves information recovery, lexical analysis to study word frequency distributions, pattern recognition, tagging or note, information extraction, data mining techniques including link and association analysis, picturing and predictive analytics.

API —is a set of subroutine characterizations, protocols, and tools for structure application software. In widespread terms, it is defined as a set methods of communication between various software components. An application programming interface may be used for a web system or operating system, database system, computer hardware or software library. An application programming interface specification may take have many forms, but commonly includes conditions for routines, data structures, object classes, variables or remote calls.

## III. RESEARCH CHALLENGES

Social media scraping and analytics provides an annoying source of academic research challenges for social scientists, computer scientists and funding bodies.

**Challenges Includes :**

Scraping— while social media data is reachable through APIs, due to the profitable value of the data, most of the major sources such as Facebook and Google are making it bit by bit difficult to obtain complete access to their raw data. very few social data sources provide reasonable data assistances to college circles and researchers. many of platforms like news services such as Thomson Reuters and Bloomberg usually charge a premium for access to their data. In difference, Twitter has recently announced the Twitter Data Grants program, where researchers can apply to get access to Twitter's public tweets and historical data in order to get insights from its enormous set of data.

Data cleansing— cleaning of unstructured textual data such as normalizing of text, especially high-frequency streamed real-time data, still presents numerous problems and research challenges. It presents aa challenge to obtain the data in easy form for humans.

Holistic data sources—researchers are gradually bringing together and combining novel data sources such as social media data, real-time market & customer data and geospatial data for analysis.

Data protection—once you have created a data store, the data has to be secured, ownership and IP issues resolved such as storing scraped data is in contradiction of most of the publishers' terms of service and users provided with different levels of access.

Data analytics—refined analysis of social media data for opinion mining such as sentiment analysis still raises uncountable of challenges due to foreign languages, internet slang, spelling errors, short forms and progressing of language.

Analytics dashboards—many platforms of social media like twitter require users to write APIs to access feeds or program analytics models in a programming language, such as Java or python. While sensible for computer scientists, these skills are typically beyond most researchers. For this we need non-programming interfaces are required for giving what might be referred to as deep access to raw data, for example, constructing APIs, integration social media feeds, combining complete sources and emerging analytical models.

Data visualization—visual representation of data for the information that has been obtained needs some representation form with the goal of interactive information and effectually through graphical means. Due to the magnitude of the data involved, visualization is becoming more and more important.

## IV. TYPES OF DATA

Although we focus on social media researchers are continually finding new and innovative sources of data to bring together and analyse. Using numerous data sources is certainly the prospect of analytics.

**Broadly data sub divided into:**

Historic data sets—previously gathered and stored social tweets, data or news. it may also be financial and economic data.

Real-time feeds—live data feeds from various social media, financial exchanges, telecoms services, news services, or from real time applications such as GPS devices and conferences speech.

Raw data—this is a form of unprocessed computer data straight from source that may have errors or is generally unanalysed.

Cleaned data—it is referred as correction or removal of erroneous data caused by irregularities, keying mistakes, missing bits, outliers, etc.

Value-added data—The data that has been cleaned, analysed, identified and improved with knowledge.

## V.CONCLUSION

The easy availability of APIs provided by Twitter, Facebook and News services has led to an explosion of data services and software tools for scraping and sentiment analysis, and social media analytics platforms.

Perhaps, the biggest concern is that companies are increasingly restricting access to their data to monetize their content. It is important that researchers have access to computational environments and especially for big social media data for experimentation.

## REFERENCES

[1] .https://arxiv.org/vc/arxiv/papers/1312/1312.4617v1.pdf : A Survey of Data Mining Techniques for Social Media Analysis

[2] https://www.ripublication.com/acst17/acstv10n5_61.pdf : Advances in Computational Sciences and Technology ISSN 0973-6107 Volume 10, Number 5 (2017) pp. 1337-1354

[3] https://link.springer.com/article/10.1007/s00146-014-0549-4 : Social media analytics: a survey of techniques, tools and platform.