

# [Abstract available social media, users are drifted](https://assignbuster.com/abstract-available-social-media-users-are-drifted/)

Abstract The platform of twitter, a social media platform is a source to takethe information using programminglanguages such as python; this obtained data hasbeen a source for various analysis. It also gives the views of privacy of information in certainplatforms. The kind of data is uses to identify the impact it has on a large scale as well asa person’s preferencesandpersonality may findan assessment.

1.      Introduction The technology advances have led to a large amount of data thatis constantly finds it way in the social media platforms i. e. uploaded on a large scale. One of such platform is twitter, Twitter is one of theextensively used social media platform to express the point of view. Information of users in socialmedia keeps on increasing day by day. UsersPosttheir view, thoughts, life events on social media and that too without anyconstraint and unwillingness.

Some of the social media allow users to interact with onlywith their friend’s and sharing their post with very easy level of privacy or without any privacy. Dueto simple and easy privacy policies, and easy available social media, users are driftedfrom traditional means of communication such as blogs or mailinglistto microbloggingsite such as Twitter, Facebook etc. very large amount of text data in the form ofmessages on social media make it very attractive medium for data analysis forthe researchers. This has led to countless tools, data services andanalytics platforms. However, thiseasy convenience of social media data for various research may change suggestively due toin demand pressures.           2. Terminology Natural language processing— It is the application ofcomputational techniquesto the analysis and synthesisof natural language and speech.

It is a part of computer science, artificialintelligence and linguisticsconcerned 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 naturallanguage output. Opinion mining—is sentiment mining, opinion or sentiment extraction.

It is atype ofnatural language processing for tracing the moodof the public about a certain artefact. Opinionmining, which can be sentiment analysis, involves buildinga system togatherandclassify opinions abouta 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 webdata. Sentimentanalysis—it refers to the use of natural languageprocessing, computationallinguistics and text analytics to recognize and extract particular informationin source materials. Text analytics—It involves information recovery, lexical analysis to study word frequency distributions, pattern recognition, taggingor note, information extraction, data mining techniquesincluding link and associationanalysis, picturing andpredictive analytics. API —is a set ofsubroutine characterizations, protocols, and tools for structureapplication software. In widespread terms, it can be reflected a set method ofcommunicationbetweenvarious software components.

An application-programming interface may use for a web system or operatingsystem, 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.                        3. Research Challenges Social media scraping and analytics provides an annoyingsource of academic researchchallenges forsocial scientists, computer scientists and funding bodies.  Challenges Includes: Scraping— while social mediadata is reachable through APIs, dueto the profitable valueof the data, most of the major sources such as Facebook and Google are making it bit bybit difficult to obtain complete access to their raw data. Veryfewsocial 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 theirdata. In difference, Twitter has recently announced the Twitter Data Grants program, whereresearchers can apply toget access to twitter ispublic tweets and historical data in order to getinsightsfrom its enormous setof data.

Data cleansing— cleaning of unstructured textual data such as normalizing of text, especially high-frequencystreamed real-time data, still presents numerous problems and researchchallenges. It presents a challenge toobtainthedata in easy form forhumans. Holisticdata sources—researchers are gradually bringing togetherand combining noveldata sources such as social media data, real-time market & customer data and geospatial data foranalysis. Data protection—once you have created a data store, the data remains safeguarded, ownership and IP issues resolved such as storingscraped data is in contradiction of most of the publishers’terms of service and usersprovided with differentlevels 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, spellingerrors, short forms and progressing of language. Analytics dashboards—many platforms of social media like twitter require users to write APIs to accessfeeds or program analytics modelsin aprogramming language, suchas Java orpython.

While sensible for computer scientists, these skills are typically beyond most researchers. For this, we need non-programming interfaces are required for givingwhat might refer to as deep access to raw data, for example, constructing APIs, integration social media feeds, combining completesources and emerging analyticalmodels. Data visualization—visual representation ofdata for the information that has obtainedneeds some representationform with the goal of interactive informationand effectually through graphical means. Due to themagnitude of thedata involved, visualization is becoming moreand more important.                         4.

Types of data  Although we focus on social media researchers are continuallyfinding new and innovative sources of data to bring together and analyze. Using numerous data sources is certainly the prospect of analytics.   Broadly, data sub divided as: Historic data sets—previously gatheredandstored 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, telecomsservices, news services, or from real time applications such as GPS devices and conferencesspeech. Raw data—is a form of unprocessed computerdata straight from source that haserrorsor is generally unanalyzed. Cleaned data—referred as correction orremoval of erroneous datacaused byirregularities, keying mistakes, missing bits, outliers, etc.

Value-added data—the datathat has been cleaned, analyzed, identified and improved withknowledge. 5. Conclusion The easy availability of APIs provided by Twitter, Facebook and News services has ledto an explosion of data services and softwaretools for scraping and sentiment analysis, and socialmedia analytics platforms.

Perhaps, the biggest concern is that companies are increasinglyrestricting access to theirdata to monetizetheir content. It is important that researchers haveaccess tocomputationalenvironments and especially for bigsocial media data for experimentation.