

Implementing System for Rating Generation by Analysis of Social Media

Ashish Pancham, Dharmesh Sharma, Vivek Sharma, Harshali Patil



Abstract: the increasing utilization of online content, provides a large amount of data resources. People share their experiences, opinions, as well as day to day activities. This results in a large amount of online data which attracts developers to carry out data mining and data analysis. Thus, there is a necessity of social media screening so as to obtain results which can be used for analysis. This paper presents a system, that facilitates the generation of single rating which is obtained by analysis of social media platforms. This rating is generated by carrying out various process like Data extraction, pre-processing, Data analysis, Rating Generation. "ROCIAL" means "Rating Socially", where each enrolled user will be rated upon his social conduct. The idea behind this initiative came up because of the present requirement for online screening. So, basically what our system would do is connect the user's different social media accounts and analyze them using the criteria which are usually required during the screening process. The main application of our system is during visa allotment process where social media screening is used. So, rather than screening each and every platform individually, our system Rocial can be used to generate a rating which will act as a criterion for visa allotment. Another significant application is during the recruitment process. Various companies check the candidate's social media behavior as a secondary criterion for judging the person. Thus, using Rocial this can be obtained quickly and efficiently.

Keywords: Data extraction, pre-processing, Data analysis, Rating Generation

I. INTRODUCTION

This project presents a system that facilitates the generation of a single rating which is generated with the analysis of the social media activity of a particular individual. The term "ROCIAL" stands for "Rating Socially", where every registered user could link their social media accounts like Twitter, Facebook, etc. and generate their social rating upon their social behavior. The idea behind this project came up due to the current need for social media background check for visa allotment and also for getting jobs in major companies.

Revised Manuscript Received on February 05, 2020.

* Correspondence Author

Ashish Pancham *, B.E. Department of Computer engineering, Thakur College of Engineering and Technology, Mumbai, India. Email: panchamashish@gmail.com

Dharmesh Sharma, B.E. Department of Computer engineering, Thakur College of Engineering and Technology, Mumbai, India. Email: dharmeshsharma649@gmail.com

Vivek Sharma, B.E. Department of Computer engineering, Thakur College of Engineering and Technology, Mumbai, India. Email: vivekradhotra7@gmail.com

Harshali Patil, Associate Professor, Thakur College of Engineering and Technology, Mumbai, India. Email: harshali.patil@thakureducation.org

© The Authors. Published by Blue Eyes Intelligence Engineering and Sciences Publication (BEIESP). This is an [open access](#) article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>)

So, in brief, what our system would basically do is connect the user's different social media accounts and analyze them using the criteria which are usually analyzed during the visa allotment and job recruitment process. Thus, to sum things up, what our framework would essentially do is associate the client's distinctive web-based life accounts and examine them utilizing the criteria which are normally dissected during the visa distribution and occupation enlistment process. Online networking is changing into a popular report as of late, in the fundamental attributable to the very truth that long-range informal communication locales grasp online clients who are allowed to express. Truth be told, these days, any very selling business is by and by submerging inside the new patterns of organizations. Aside from composed overviews, the organizations moreover expand their customer fulfillment investigation through the net, to gather a curiously large amount of information. Not many examinations on investigation have just been given. These investigations are focused to Twitter, for tweet refreshes a couple of explicit subjects, absolutely on brands of item.

The Chinese government has just been working upon a similar application since 2014 and at present (2019) began actualizing it all through the state. By 2020, it should institutionalize the evaluation of residents and organizations and is named as social rating, or 'Social Credit'. By 2018, few limitations had been set on voters, whom state-possessed media outlined in view of the head toward making a national social credit framework. The platform is viewed as a kind of mass police examination that utilizes immense information investigation innovation the govt of ongoing China has furthermore kept up frameworks of paper records on individuals where every citizen beginning off with a score of 1,000. While the govt hasn't made the specific philosophy that needs to ascertain scores open, one's positioning will succumb to each major and minor infractions. Genuine infringement grasps alcoholic driving, defalcation, and misrepresentation. inexhaustible littler infringement that lead to a down score grasp getting a charge out of too a few computer games; spreading "counterfeit news," especially connected with psychological militant assaults, or rejecting military assistance, will bring down one's score. Normally, people are proclaimed "deceptive" for submitting infractions the govt doesn't accept they're truly feeling sorry for. For those announced "conniving," the ability to look for business-class train tickets or to hold up at sure lodgings might be cancelled. At times, the possibility for their adolescents to go to their most well-known secondary school or staff could likewise be moved out, as might work openings.



(The administration urges managers to counsel the boycott before making employing determinations.) Voters World Health Organization act without thought freely, such as strolling their pooches off-rope, will have their canines seized related be expected to require an assessment to ask the pets back.

Albeit "conniving" people are punished for hazardous scores, voters World Health Organization positioned the best inside the new framework will trade out of advantages like business limits or booking structure rooms while not stores.

II. PROBLEM DEFINITION

Social record screening inside the area of smaller scale blogging might be a relatively new investigation subject therefore there's still stores of room for any examination during this space. A decent amount of associated past work has been done on estimation investigation of client surveys, reports, net web journals/articles, and general expression level notion examination. The least difficult outcomes came to in feeling order abuse regulated learning systems like Naive Bayes and Support Vector Machines, be that as it may, the manual marking required for the managed approach is staggeringly overrated. Some work has been done on solo and semi-managed approaches, and there's loads of room for development. There is a need of appropriate and formal correlations between these outcomes landed through various highlights and grouping strategies so as to choose the best highlights and for specific applications, for example, being utilized for visa assignment or having utilized for worker's social action check.

Mockery is one of the principle challenges round-looked in Sentiment Analysis. Twitter conjointly enables the clients to explicit their thoughts and feelings with each other that enables the organizations to get a handle on the overall population assessment on their item or administrations so they will give a timeframe customer help. A scathing comment might be a kind of communicating negative sentiments exploitation positive words. The acidic comment is also once people mean one thing else from what they talk. The acidic comment is utilized not exclusively to make fun anyway conjointly for scrutinizing individuals, sees, thoughts, and so forth in light of that scathing comment is inconceivably bounteous utilized on twitter. Harsh comments are regularly sent from numerous points of view in which kind of an immediate language, discourse, content, and so on. It is regularly reflected exploitation the rating of stars by giving less assortment of stars. There square measure a few applications for police work harsh comment. It's acclimated let the expert handle the expectation of the client and along these lines the situation during which it's aforementioned. The acidic comment is extra overarching in places any place their square measure capital letters, emojis, and outcry marks, and so on. Scathing comment location is one of the extraordinary undertakings in slant investigation. On Amazon and looking sites, it gets a handle on the survey of the product. The customer's inclinations and conclusions are regularly dissected to get a handle available conduct for much better shopper mastery.

1. OBJECTIVES OF PROJECT

- To design a single platform which can be used to generate rating by analysis of various social media platforms rather than an individual screening of various platforms
- To be able to tell the impact of an individual on social media.
- To be able to indirectly refrain users from posting abusive content on social media.
- To be able to minimize the increasing social media crimes and threats.
- System designed should be simple and User-friendly.
- The system should be open source as well as free to use.
- Indirectly restricting individuals from misconducting on social media.
- Achieve better accuracy rates for the generation of rating compared to conventional methods.
- Enhance data for maximum efficiency using pre-processing
- Achieve the output in the least time possible.

2. APPLICATION OF PROJECT

A. Social media screening for recruitment processes.

The utilization of Social Media in Recruitment is an alternate ball game out and out. Also, this is one which today impacts likely practically all occupation searchers and competitors. All the more critically, it is most likely something you as work searcher can control. Associations employ you increasingly more for 'something beyond a vocation' and would not need potential possibility to wear a cover while at work and transform into another person after that.

So what does the utilization of web based life in enrollment comprehensively include?

It includes associations – and that incorporates selection representatives, chiefs, and pioneers – utilizing internet-based life channels to:

1. Understand better the competitors whom they will meeting, meeting and enlisting
2. Attempt to approve different theories shaped during interviews and the enlisting procedure
3. Do clandestine individual verifications in a progressively helpful way

B. Background check during visa allotment

In May 2019, the US has announced that all applicants for US visas will have to submit their social media details under newly adopted rules. Sooner or later other countries will too follow the screening process.

Social media rating can play an important role in terrorist activity and analysis. Social media rating will be a vital tool to screen out terrorists, public safety threats, and other dangerous individuals from gaining immigration benefits and setting foot on foreign soil.

Here one of the major applications of our system ROCIAL kicks in and can be used intensively.



C. Monitor activities for current employees

After the recruitment process, it is necessary to monitor the social behavior of the employees. In recent years we have seen a number of high-profile figures being sacked by their organizations for the comments they made on social media in the past.

Again, rather than scanning every other post of the employee, one can simply give a glance at the rating generated by our system ROCIAL.

If the rating of the individual is lower than the standard set by the organization, than a warning to the employee can be considered.

If the rating keeps on deteriorating, then the organization may consider letting off the employee from their organization.

III. PROPOSED WORK AND METHODOLOGY

On finish of the task, our framework ROCIAL would have the option to produce a solitary rating of each person by examining different online life stages. Presently this rating got can be utilized at various associations just as in different fields. Then again, ROCIAL will facilitate the procedure of web-based life screening for example rather than screening a solitary stage at once, we can utilize the rating acquired by the framework. The information would be investigated, and the previously mentioned results would be given at all measure of time with however much exactness as could reasonably be expected.

This venture upon fulfillment with all out functionalities of approaching every single social medium handles of the enlisted client could investigation everything without exception on the handles and produce one single social rating which could be utilized as an auxiliary criteria for passing judgment on individuals upon social conduct, money related status, fundamental whereabouts of their voyaging areas, and so forth. Be that as it may, connecting each and every dynamic social record of the client must have their information get to consent from those online life designers, which is the thing that ought to be done to acquire full usefulness of this venture.

1. METHODOLOGY

The methodology being obtaining users' information from social media platforms in structured/unstructured types. The extracted information is then cleansed and filtered for our specific needs. {The information| the info |the information} is essentially in-text phrase data frame, then text blob library is employed for applying classification over information

Data Preprocessing

In the information preprocessing stage at first, the info information is taken, and hashtags are distinguished. These hashtags are expelled from the information. This module incorporates Tokenization, Stemming, Noise Removal, and PoS Tagging.

A. Noise Removal

Noise Removal from the information is the way toward expelling the stop words, URLs, accentuation and so on. These are expelled from information so the washed down

information can be effectively handled for the following stage.

B. Tokenization

Tokenization is the way toward parting the info information into little units called tokens. The information is isolated into tokens so as to effortlessly process the information.

C. Feature Selection

Highlight Selection is the way toward removing the necessary highlights from the accessible information dataset. The necessary highlights resemble tweet id, tweet, Date of the tweet and so on.

D. Classification

In the characterization stage, the item surveys gathered are arranged utilizing Naive Bayes Classification and AdaBoost Classification calculations so as to distinguish the mockery in the tweets. For tweets, the arrangement is certain or negative. For the item surveys the characterization method used to distinguish audit thinking about a size of 5. Sentiment Analysis is the way toward distinguishing whether a tweet is certain, negative. Sentiment Analysis is useful for the advertisers to perceive the popular feeling about their organization and items, and furthermore to think about consumer loyalty.

E. Sentiment Analysis

is the process of identifying whether a tweet is positive, negative.

F. TextBlob

is a Python library that is utilized for handling literary information. It gives a basic API to plunging into normal regular language preparing (NLP), the Naive Bayes Classifier Method wherein a notice is being delegated positive or negative.

G. Naive Bayes Classifier Bayesian classifiers

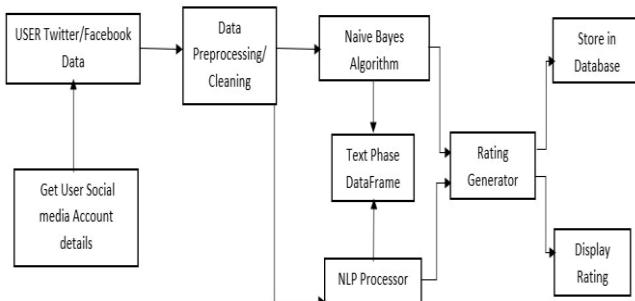
are based around the Bayes rule, technique for viewing restrictive conceivable outcomes that grants you to flip the condition around in a helpful manner. A contingent likelihood is presumably that occasion X will happen, given the proof Y. That is commonly given as $P(X | Y)$. The Bayes rule enables us to decide this likelihood when the sum total of what we have is the likelihood of the contrary outcome and of the two segments exclusively:

$$P(X | Y) = P(X) P(Y | X) / P(Y).$$

The characterized outcomes are put away in rundown and calculation for rating age is applied over rundown information. The rating produced is shown over the screen and furthermore put away in the administrator database.

2. Block Diagram

Implementing System For Rating Generation By Analysis Of Social Media



3. Rating Generation Algorithm:

After Classification of data, the classified results are stored in list and algorithm for rating generation is applied over list data. List contain polarity of data i.e. positive, negative and neutral. Polarity 1 is for positive sentiment, 0 is for neutral sentiment and -1 is for negative sentiment. For rating generation, count the total number of sentiments from the list. Remove the neutral sentiments from the total number of sentiments in the list as they will not be used for rating generation. Rating is generated by considering positive sentiments from the total number of positive and negative sentiments. Final rating is generated out of 5.

Algorithm: Rating Generation

Input: List of Classified Result

Output: Rating

//Initialize variables

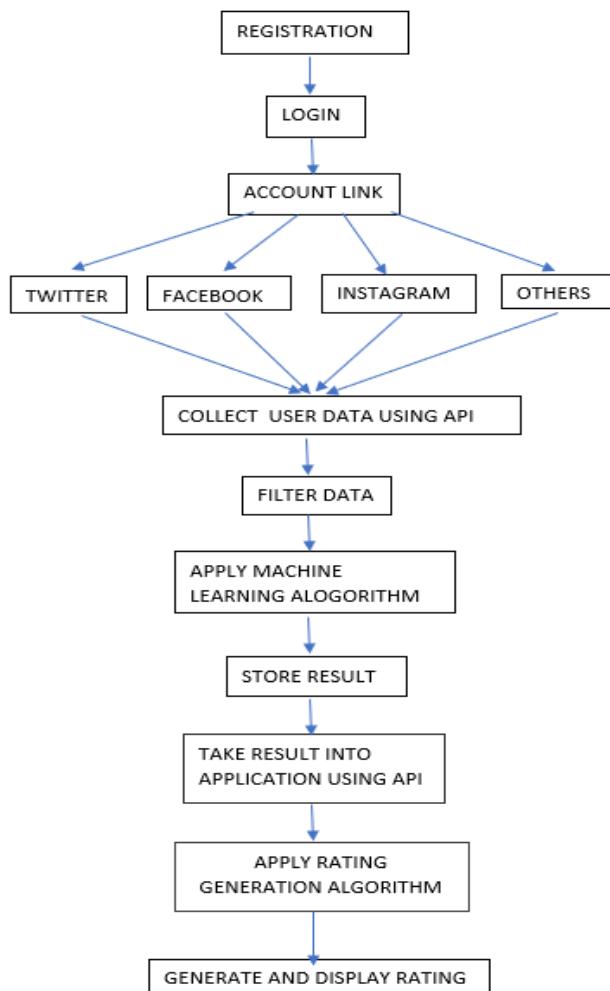
```

1: Initialize count=0      //count=total number of sentiment
                           count
Initialize pcount=0        //pcount=total number of
                           positive sentiment count
Initialize ncount=0        //ncount=total number of
                           negative sentiment count
Initialize zcount=0        //zcount=total number of
                           neutral sentiment count

2: for i in list do
3:   count=count+1          //increment count
4:   if i=1 then
5:     pcount=pcount+1       //increment pcount
6:   else
7:     if i=0 then
8:       zcount=zcount+1     //increment zcount
9:     else
10:      ncount=ncount+1    //increment ncount
11:    end if
12:  end if
13: final=count-zcount      //Removing
                           neutral sentiment count from total number of sentiment
                           count
14: initial rating=(pcount/final)*100 //Calculating
                           percentage of positive sentiment count from the final
                           sentiment count obtained in previous step
15: final rating= (initial rating/100)*5 //converting the
                           initial rating obtained in previous step into a grade up to 5.

```

4. Flowchart



IV. TECHNOLOGY

The proposed project work is under the Intelligent System Design and Development Domain which requires a detailed study of different libraries.

Various Python Libraries used to implement the project:

1. Tweepy –

Tweepy is a Python library for getting to the Twitter API. It is incredible for straightforward mechanization and making twitter bots. Tweepy has numerous highlights. Twitter enables us to mine the information of any client utilizing Twitter API or Tweepy. The information will be tweets extricated from the client. The primary activity is get the customer key and customer secret key. These keys will help the API for validation.

2. TextBlob –

TextBlob gives an API that can perform diverse Natural Language Processing (NLP) errands like Part-of-Speech Tagging, Noun Phrase Extraction, Sentiment Analysis, Classification (Naive Bayes, Decision Tree), Language Translation and Detection, Spelling Correction, and so on. TextBlob is based upon Natural Language Toolkit (NLTK).



TextBlob, which is another incredibly ground-breaking NLP library for Python. TextBlob is based upon NLTK and gives a simple to utilize interface to the NLTK library. TextBlob has numerous highlights, for example, Noun state extraction, Part-of-discourse labeling, Sentiment examination, Classification (Naive Bayes, Decision Tree), Tokenization (parting content into words and sentences), Word and expression frequencies, Parsing, Spelling redress, Add new models or dialects through expansions.

3. NumPy –

NumPy is a universally useful exhibit preparing bundle. It gives an elite multidimensional cluster article, and devices for working with these exhibits. It is the crucial bundle for logical figuring with Python. It contains different highlights including these significant ones:

- An incredible N-dimensional cluster object
 - Sophisticated (telecom) capacities
 - Tools for incorporating C/C++ and FORTRAN code
 - Useful direct variable-based math, Fourier change, and irregular number capacities besides its conspicuous logical uses, NumPy can likewise be utilized as a proficient multi-dimensional compartment of conventional information. Subjective information types can be characterized utilizing Numpy which permits NumPy to consistently and expeditiously coordinate with a wide assortment of databases

4. Pandas –

Pandas is an opensource library that permits to you perform information control in Python. The Pandas library is based over Numpy, which means Pandas needs Numpy to work. Pandas give a simple method to make, control and wrangle the information. Pandas is likewise a rich answer for time arrangement information. Pandas is the most famous python library that is utilized for information examination. It furnishes exceptionally enhanced execution with back-end source code is absolutely written in C or Python. We can break down information in pandas with Series and Data Frames.

Other technologies used are

Graph API –

The Graph API is the essential path for application to peruse and keep in touch with the Facebook social chart. This gives you a path by which you can get information from Facebook stage or you can put information to the Facebook stage. It's a REST based API and used to question information, deal with your advertisements on Facebook, transfer Photos/Videos, post any of your new stories to Facebook automatically. You can utilize this API to get your own Facebook account information, however on the off chance that you need/need to get other client's information then for this you have to take a few authorizations from clients. You have to execute OAuth convention to accomplish this with the goal that anybody can validate and give your authorizations, after this you will get an Access Token (primary concern which you have to use for each demand).

Tkinter –

Tkinter is an inbuilt Python module used to make basic GUI applications. It is the most ordinarily utilized module for GUI applications in the Python. Python offers numerous alternatives for creating GUI (Graphical User Interface). Out of all the GUI techniques, tkinter is most generally utilized strategy. It is a standard Python interface to the Tk GUI toolbox sent with Python. Python with tkinter yields the quickest and most effortless approach to make the GUI applications. Making a GUI utilizing tkinter is a simple undertaking. Bringing in tkinter is same as bringing in some other module in the python code.

V. RESULT ANALYSIS

Our goal was to develop a system which would generate a rating by analysis of social media accounts. The initially obtained results are

s.no	User name	Rating
1	@narendramodi	4.594
2	@realdonaldtrump	3.152

A screenshot of the Python 3.7.2 Shell window. The title bar reads "Python 3.7.2 Shell". The menu bar includes File, Edit, Shell, Debug, Options, Window, and Help. A status bar at the bottom displays the Python version and build information: "Python 3.7.2 (tags/v3.7.2:9a3ffc0492, Dec 23 2018, 22:20:52) [MSC v.1916 32 bit (Intel)] on win32". Below the status bar, a message says "Type "help", "copyright", "credits" or "license()" for more information."

```
Python 3.7.2 Shell
File Edit Shell Debug Options Window Help
Python 3.7.2 (tags/v3.7.2:9a3ffc0492, Dec 23 2018, 22:20:52) [MSC v.1916 32 bit
(Intel)] on win32
Type "help", "copyright", "credits" or "license()" for more information.
>>>
= RESTART: D:\program files\python\tw\sentiment_anaylsis_twitter_data (1).py =
Enter user name ?@realdonaldtrump
User Socail Media Rating : 3.1521739130434785
```

Just using any person's user-ids for various social media accounts, an analysis of the obtained data can be used to generate a rating of the person on a scale of 0-5.

VI. CONCLUSION

The main outcome of this paper is to develop a standard platform where social media accounts of individuals are taken into consideration to generate a rating which is used for social media screening. This platform could be developed using machine learning and data analysis of social media. Usage of this platform is not limited to a smaller level but can also be implemented in order to ease many day to day requirements.

This system upon completion with total functionalities of having access to all social media handles of the registered user could analyze anything and everything and generate one single social media rating which could be used as a secondary criterion for judging people upon social behavior, financial status, basic whereabouts of their travelling locations, etc. So a single platform can be used to facilitate a standard rating with help of screening of various social media platforms. There should no longer be any need for scanning and analyzing different social media accounts on different platforms, or even no need to open various social handles for checking. On the other hand, rather than screening every other social media account individually, the rating generated by our system ROCIAL can be used which will result in effective evaluation in less period of time. The system can be used to monitor as well as examine the social behavior of an individual. The main impact of the system will be that if social media rating is considered as a necessary criterion by the organization then this will indirectly promote the good behavior of people on social media. So, in order to improve the social media rating every individual will control their behavior on social media platforms and will themselves refrain from posting abusive content on social world. This will eliminate social media threats and will also have an impact on ever increasing social media crimes.

REFERENCES

1. Project Report Twitter Emotion Analysis Author: Dr. David Rossiter Year: 2015
2. Twitter Sentiment AnalysisAuthor: Afroze Ibrahim Baqapuri (NUST-BEE-310)
3. Sentiment analysis of Facebook statuses using Naive Bayes Classifier for language learningAuthor: Christos TroussasYear: 2013https://www.researchgate.net/publication/261497806_Sentiment_analysis_of_Facebook_statuses_using_Naive_Bayes_Classifier_for_language_learning
4. How to use Facebook Graph API and extract data using Python.Author: Ravi RanjanYear: 2016<https://towardsdatascience.com/how-to-use-facebook-graph-api-and-extract-data-using-python-1839e19d6999>
5. Chapter 2. Mining Facebook: Analysing Fan Pages, Examining Friendships, and More:<https://www.oreilly.com/library/view/mining-the-social/9781449368180/ch02.html>
6. Author: Justin LittmanYear: 2018 <https://gwu-libraries.github.io/sfm-ui/posts/2018-01-02-facebook>
7. Sentiment analysis on Twitter Author: Mr. Amit Khan (Assistant Professor (IT), RCCIIT)Year: 2018
8. Sentiment analysis of Facebook statuses using Naive Bayes classifier for language learningAuthor: Kurt Junshean Espinosa, Kevin Llaguno, Jaime Caro Department of Computer Science University of the PhilippineYear: 2013
9. Opinion Why India needs to be wary of China-style social credit ratings Author: Nitin Pai Year: 2019<https://www.livemint.com/opinion/columns/opinion-why-india-needs-to-be-wary-of-china-style-social-credit-ratings-1550423726392.html>
10. How China Is Using “Social Credit Scores” to Reward and Punish Its Citizens: Author: Charlie Campbell / Chengdu Year: 2019<https://time.com/collection/davos-2019/5502592/china-social-credit-score/>
11. Spend “frivolously” and be penalized under China’s new social credit system: Author: Nadra Nittle Year: 2018<https://www.vox.com/the-goods/2018/11/2/18057450/china-social-credit-score-spend-frivolously-video-games>
12. TextBlob Sentiment: Calculating Polarity and Subjectivity:Year: 2015 https://planspace.org/20150607-textblob_sentiment/
13. The US just added one more requirement for visa applicants to stress about Author: Jackie Bischof Year: 2019 <https://qz.com/1633061/us-visa-applications-will-now-include-social-media-screening/>
14. Role of Social Media (Facebook, Twitter, LinkedIn) in Hiring & Recruitment Author: Ameya Joshi Year: 2016<http://www.careerizma.com/blog/social-media-in-hiring-recruitment/>
15. How To Get Facebook Data With Python By: Mydatahack Year: 2017 <https://www.mydatahack.com/python-facebook-graph-api-feed-data-ingestion/>
16. Sarcasm Detection for Sentiment Analysis in Indonesian Tweets Author: Yessi Yunitasari1, Aina Musdholifah2, Anny Kartika Sari*3 Year: 2011.

AUTHORS PROFILE



Harshali Patil is associate professor at Thakur College of Engineering. She has done Ph.D. (CSE), M.E (CE), B.E (CSE). The areas of interest are data mining, web mining & data analytics.

harshali.patil@thakureducation.org



Ashish Pancham is pursuing BE in Computer Engineering currently in 8th semester at Thakur College of Engineering. His area of interest is in social media analysis and its applications.

panchamashish@gmail.com



Dharmesh Sharma is pursuing BE in Computer Engineering currently in 8th semester at Thakur College of Engineering. His area of interest is in social media analysis and its applications.

dharmeshsharma649@gmail.com



Vivek Sharma is pursuing BE in Computer Engineering currently in 8th semester at Thakur College of Engineering. His area of interest is in social media analysis and its applications.

vivekradhotra7@gmail.com



Published By:

Blue Eyes Intelligence Engineering
& Sciences Publication

WWW.IJEAT.ORG
Exploring Innovation