

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/382796508>

NLP FOR SOCIAL MEDIA DATA PROCESSING

Article · March 2021

CITATIONS
0

READS
6

1 author:

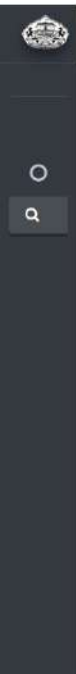


[Prashant Wadkar](#)

Yashaswi Education Society-International Institute of Management Science

9 PUBLICATIONS 0 CITATIONS

SEE PROFILE



UGC-CARE List

UGC-CARE List

You searched for "2229-3620". Total Journals : 1

Search:

Sr.No.	Journal Title	Publisher	ISSN	E-ISSN	UGC-CARE coverage years	Details
1	Shodh Sanchar Bulletin (print only)	Shodh Sanchar Educational and Research Foundation	2229-3620	NA	from September-2019 to April-2021	Discontinued from April 2021

Showing 1 to 1 of 1 entries

Previous 1 Next

GOVT. OF INDIA RNI NO.: UPBIL/2015/62096

UGC Approved Care Listed Journal

ISSN
2229-3620

SIC



SHODH SANCHAR

Bulletin

An International
Multidisciplinary
Quarterly Bilingual
Peer Reviewed
Refereed
Research Journal

Vol. 11

Issue 41

January to March 2021

Editor in Chief

Dr. Vinay Kumar Sharma

D. Litt. - Gold Medalist



sanchar
Educational & Research Foundation

CONTENTS

S. No.	TITLE	NAME OF AUTHORS	PAGE No.
1.	AN ANALYTICAL STUDY OF FLEXI LEAVE POLICY AND ITS IMPACT ON PBO'S EMPLOYEES	Dr. Vijay Dhole	1
2.	STUDY OF IMPACT OF CORONA PANDEMIC AND THE GLOBAL RECESSION ON JOB OPPORTUNITIES FOR INDIAN YOUTH IN KPO SECTOR	Dr. Pallavi Sajanapwar	7
3.	INNOVATION IN OUT OF HOME (OOH) MEDIA ADVERTISING RESEARCH: A LITERATURE REVIEW	Prof. Rupa Rawal Dr. Amod Markale	13
4.	ROLE OF COMPUTER TECHNOLOGY ON PHARMACY IN INDIA	Dr. D. R. Vidhate Ms. D. D. Vidhate	19
5.	UP SKILLING THE SKILLS FOR FUTURE EMPLOYMENT CHALLENGE-2030	Prof. Manoj Sathe Dr. Sudarshan Pawar	24
6.	INVESTORS PREFERENCES TOWARDS GOLD AS AN INVESTMENT OPTION WITH REFERENCE TO PUNE DISTRICT.	Mr. Swapnil Patil Dr. Eknath B. Khedkar	29
7.	A STUDY OF THE IMPACT OF PANDEMIC COVID 19 ON THE HEALTH CARE SECTOR	Dr. Gauri Prabhu	33
8.	LIFELONG MACHINE LEARNING IN SENTIMENT CLASSIFICATION: CONCEPTS AND IMPLEMENTATIONS	Krishna Priya S. Kavita S. Oza	37
9.	A COMPARATIVE STUDY OF URBAN AND RURAL HOUSEHOLDS INVESTMENT AVENUES IN MAHARASHTRA.	Ghodake Shamrao P. Dr. E. B. Khedkar	43
10.	LITERATURE REVIEW ON EMPLOYEE ENGAGEMENT PRACTICES	Dr. Pushpraj Wagh Ms. Pooja Salvekar	52
11.	NEW WORLDVIEW IN SECURITY FOR BANKS: DECEPTION TECHNOLOGY	Ms. Ashwini R. Chavan Dr. R. D. Kumbhar	58
12.	COMPARATIVE STUDY OF ONLINE V/S OFFLINE MODE OF EDUCATION	Mrs. Rupali Kalekar Ms. Aanchal Priya	64
13.	A STUDY OF INBOUND MARKETING AND ITS IMPORTANCE IN INNOVATION MANAGEMENT	Prof. Asmita Abhijeet Gaikwad Dr. Amod Markale	68
14.	POST COVID-19: A TECHNICAL APPROACH FOR HIGHER EDUCATIONAL EXCELLENCE	Ms. Sarika Choudhari Mrs. Arati Patil	73
15.	REDEFINING HR WITH ARTIFICIAL INTELLIGENCE: CHATBOTS THE NEW ASSISTANT!	Sarah Dsouza Saylee Anil Karande Angshupriya Datta	80

16.	CONTRIBUTIONS OF LOGISTICS AND SUPPLY CHAIN IN ENTREPRENEURIAL ECOSYSTEM IN DEVELOPING COUNTRIES WITH SPECIAL REFERENCE TO TANZANIA	Dr. Maige Mwakasege Mwasimba Mr. Alfred Sallwa	87
17.	CRITICAL ANALYSIS OF THE CORONA DISEASE PANDEMIC AND ITS IMPACT ON EDUCATION SYSTEM IN INDIA	Prof. Manoj Ashok Sathe Prof. Yugandhara R. Patil	93
18.	PERCEPTIVE CROSS - CURRENTS IN VOCATIONAL TRAINING	Dr. Vandana Mohanty Colonel (Dr.) J. Satpathy Ms. Dhanashree Shinde	98
19.	TEXT MINING: APPLICATION OF STEMMING ALGORITHMS FOR INFORMATION RETRIEVAL IN HEALTHCARE WITH SPECIAL REFERENCE TO VIRAL INFECTIVE DISEASES	Dr. Ashwini Manish Brahme Dr. Shivaji D. Mundhe	105
20.	STUDY OF THE FINANCIAL IMPLICATIONS OF PSB MERGERS IN INDIA	Mr. Yogesh K. Nakhale Dr. Gauri Prabhu	111
21.	A POST IMPLEMENTATION STUDY OF SUGARCANE IRRIGATION MANAGEMENT SYSTEM	Dr. Nilam Jadhav Dr. Shivaji Mundhe	119
22.	A COMPARATIVE ANALYSIS OF CAPITAL STRUCTURE ADJUSTMENT WITH RESPECT TO PARTIAL CAPITAL STRUCTURE ADJUSTMENT IN THE SELECTED BSE LISTED INDIAN MANUFACTURING COMPANIES	Vikas R. Adhegaonkar Dr. E. B. Khedkar	124
23.	ENVIRONMENTAL ANALYSIS: TO ANALYZE VARIOUS ENVIRONMENTAL FACTORS AFFECTING ESSITY AND IDENTIFY THE KEY DRIVERS AMONGST THEM.	Tejas Gujar Dr. Bharati Jadhav	128
24.	CUSTOMER REVIEWS SENTIMENTS ANALYSIS USING NATURAL LANGUAGE PROCESSING (NLP) AND DEEP LEARNING.	Dr. Sachin Misal Dr. Shivaji Mundhe	134
25.	BEHAVIOUR OF INVESTOR ON INVESTMENT DECISION WITH SPECIAL REFERENCE TO MUTUAL FUND VS. TRADITIONAL INVESTMENT AVENUES- A REVIEW	Prof. Mahesh Mahankal Dr. Prabha Singh	139
26.	FACE IDENTIFICATION IN A GROUP BASED ON LBP ALGORITHM AND NEURAL NETWORK CLASSIFIER FOR CLASS ATTENDANCE	Narayan Kulkarni Dr. H. S. Fadewar	145

27.	FACTORS AFFECTING CONSUMER'S BUYING BEHAVIOR TOWARDS ORGANIZED RETAILING WITH REFERENCE TO STAR BAZAAR IN PUNE CITY	Dr. Pushpraj Wagh	151
28.	VEHICLE REGISTRATION NUMBER RECOGNITION USING MACHINE LEARNING	Dr. Sachin Misal Mr. Sunil Joshi	158
29.	INNOVATION- AN INSTRUMENT FOR SUSTAINABLE ECONOMY	Singha Roy Ekta	164
30.	AUTOMATIC IRIS RECOGNITION SYSTEM IN HCI	Ganesh K. Awasthi Dr. H. S. Fadewar	168
31.	VOCATIONAL EDUCATION AND TRAINING: IS IT REALLY EFFECTIVE?	Dr. Vandana Mohanty Dr. Ashutosh Zunjur	175
32.	AN EMPIRICAL STUDY ON TECHNIQUES TO CREATE A POSITIVE LEARNING ENVIRONMENT	Dr. Sonali Dharmadhikari Dr. Shweta Joglekar	182
33.	EFFECT OF BONUS ISSUE ON PRICES OF COMPANIES AT STOCK EXCHANGE	Prof. Mahesh Mahankal Mr. Rahul Bari	187
34.	TREND ANALYSIS OF FINTECH USAGE IN PUNE CITY DURING COVID-19	Mrs. Poorva Pachpore Dr. Gauri Prabhu	191
35.	A STUDY OF AWARENESS REGARDING DIABETES MELLITUS AMONG RURAL COMMUNITY OF AHMEDNAGAR DISTRICT	Dr. Gorakshanath T. Gund Dr. Prashant Radhakrishna Tambe	196
36.	VALIDATING CONSUMER ETHNOCENTRISM SCALE (CETSCALE)	Dr. Ashutosh Zunjur Dr. Joe Lopez	202
37.	NLP FOR SOCIAL MEDIA DATA PROCESSING	Miss. Ankita D. Garud Prof. Prashant N. Wadkar	208
38.	A COMPARATIVE STUDY OF WIDELY USED VIRTUAL CLASSROOM TECHNIQUES	Dr. Nilam Jadhav Dr. Kedar Marulkar	214
39.	LINKING BETWEEN KNOWLEDGE AND TALENT MANAGEMENT IN AN IT FIRMS – LITERATURE REVIEW	Ms. Deepti M. Yadav Dr. Bholu Sarang S.	219
40.	STUDY OF E-MARKETING PRACTICES OF SELECTED SMARTPHONE BRANDS FOR PCMC REGION	Prof. Amarnath Gupta Ganesh Kalshetty	226



NLP FOR SOCIAL MEDIA DATA PROCESSING

Miss. Ankita D. Garud
Prof. Prashant N. Wadkar*

ABSTRACT

As we entered in a 21th century, we got connected to all over the world by the technology called "Internet". Internet has made major impact on our lives and one of the big impact it made is it introduced us with "Social media". The world full of busy as bee, people found their comfort on social media. Instead of meeting in person, people started meeting on social media. Now days people use social media as platform to express their views on any topic. There are some popular social medias such as Facebook, Twitter, YouTube, Instagram etc. there are billions of users using these apps to share their strong opinions by sharing, liking and commenting on the posts. Hence lots of data is being created. This data can be used for a future predictions. We can check people's views on a posts ad using that we can make a prediction on that topic. We have to use some analytical methods to make correct prediction. To use analytical methods, we have to retrieve that data from social media and have to convert in machine understandable form. For this situation "Natural Data Processing" also called as NLP technique is used. It is easy way to get data from social media. By using NLP we can use huge amount of data which is already has been created by humans on social media and can make future predictions. It is used to extract relevant information from a large volume of data. NLP is one of the most promising avenues for the social media data processing .The researcher highlights on this paper to study the role of NLP in social media data processing using Facebook data comments. Data used for this study is retrieved from comments under 'Fau-G' Game's post on Facebook SNS .

Keywords: Social media, Reviews, NLP, Facebook, Game reviews, Comments, Data Pre-processing classification Algorithm, Facepager, NLTK, Python, Sentiment Analysis, Machine Learning.

INTRODUCTION

Social media analytics holds the ability to collect and find meaning in data collected from social channels. It is used to support trade, business or commerce decisions and measure the performance. One can take actions based on those conclusions through social media. The research paper is focused on Natural Language processing of secondary data, which has been preprocessed and further prepared for implementing NLP and hence analysis. For the research the Fau-G Indian developed game's review data has been used.

BACKGROUND

1. NLP

Natural language processing commonly known as NLP is a subfield of artificial intelligence which is concerned with the computers and human language interactions. It deals with how to program computers to process and analyze large amounts of natural language data. The main focus is to make computer capable to understand the contents of documents, including the contextual nuances of the language within them. We can extract information and insights

*Student, ASM's CSIT, Pune, India

**Asst. Professor, ASM's IIBR, Pune, India

contained in the documents and categorize them. Natural Language Processing is the automatic manipulation of natural language, like speech and text which we implemented in this research.

Python

Python, the dynamic programming language is used for this research, which made the coding and debugging extremely quick. There are extensive open source libraries available for this version of python

NLTK

Natural Language Toolkit, which is also known as NLTK is the python's library which provides the base for classification of data. NLTK library is used for processing text, tagging, tokenization as well as classification. It plays a vital role in transforming the text data or dataset in the tweets/reviews/comments/posts into a format that can be used to extract sentiments.

The various functionalities and python coding is used in pre-processing of data/dataset and hence the data which made available from the social media made fit for mining and extracting features. It also helps in machine learning algorithms which can be used to train the model or classifier. We also can calculate and check the accuracy of different classifier using NLTK.

In this research the researcher has used python programming language and anaconda as a platform. The NLP is used for text to a sentiment positive or negative which further helped us to know the opinion of people and to the further analysis.

OBJECTIVE

The main objective of this research study work is to carry out the natural language processing technique to process data extracted from popular social media site Facebook and to analyze this data with NLP.

1. To perform NLP on Fau-G Indian developed game's review data.
2. People's opinions about this game,

negative or positive approach towards the game, etc.

SIGNIFICANCE OF THE STUDY:

Social Media plays a key role in connecting people all over the world and developing relationships with the help of Facebook, Twitter, YouTube etc. Facebook is one of the popular website widely used globally. It seems that there are 2.45 billion users worldwide and 320 million users from only India. It allows expressing their feelings, emotions, opinions, thoughts, reviews, etc with the online posts and about any of the topic in their human natural language. Not only this, the people are able to create, share, like and give comment on any of the posts. Python found to be best for NLP. For processing natural language, NLTK provided us large amount of corpora which helps in training classifiers. It plays a vital roll in performing all NLP methodology like tokenizing, part-of-speech, tagging, stemming, lemmatizing, parsing and doing analysis of our Fau-G datasets.

Working on large dataset was a challenging task, but with the use of NLTK it made easy to classify the dataset and give more precise results. The aim of this research is to perform analysis on Fau-G game reviews data. Launching of Fau-G game was a first step towards Atmanirbhar Bharat, developing games by own country people. Youngsters were looking forward and excited to try new game. They all had very high expectations from the game for graphics and other things compared to Pub-G game. Public opinions for this game have mined from Facebook and then classified into whether the opinion or comment is positive or negative by using sentiment analysis. This research outcome will help in, to study opinions of people on this game.

RESEARCH METHODOLOGY ADOPTED

There are infinite users of Facebook SNS with communication bridges like groups, pages, posts. The new game "Fau-G" which is launched on 26th January (Republic day of India) used for the present study, this secondary data which is

already available on Facebook in the form of post comments has been collected by **Facepager app**. The data is regarding reviews about "Fau-G" game. Data then pre-processed and applied the NLP and hence the sentiment analysis.

A. Data Collection:

1) Facebook Data Extraction:

To use Facebook SNS we must first have a Facebook account.

The app called as 'Facepager' which is used to retrieve data from many social networking sites such as Facebook, YouTube, and Twitter etc. is used for getting data by which we easily retrieved Facebook comments.

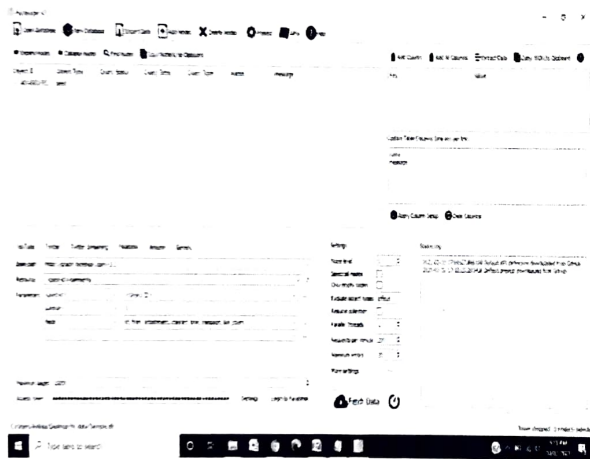


Fig 1 . Facepager app UI

1. Open Facepager app ---> login to Facebook account
2. Create New Database ---> Add Node
3. Insert post Id as object Id
4. Set resource : /<post-id>/comments
5. Set parameters : <post-id>
6. Set maximum page limit
7. Click on 'Fetch data'.
8. Data will be stored in database. Click on Export data
9. .csv (comma separated values) file will be generated.

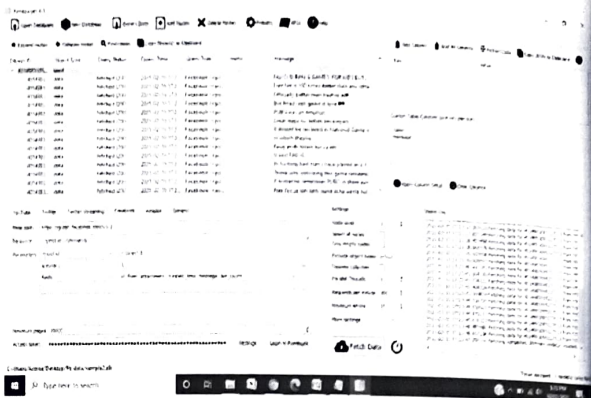


Fig 2. Illusion of Data extracted from Facepager

Comment Type	RESULT
Original comment	FAU G & BAN G GAMES FOR KIDS BUT FREE FIRE FOR LEGENDS ❤️❤️
Processed comment	fau g ban g games kids free fire legends

TABLE 1 SAMPLE OF COMMENTS AND PROCESSED COMMENTS

2) Facebook Data Storage:

Researcher started getting data from Facebook and to store that data so that we can use it for further processing. We exported collected data in .csv (comma separated values). file generated which consists of comments which are extracted from Facebook. CSV separated each field with a comma, which was then easy to access the particular field that consists of text.

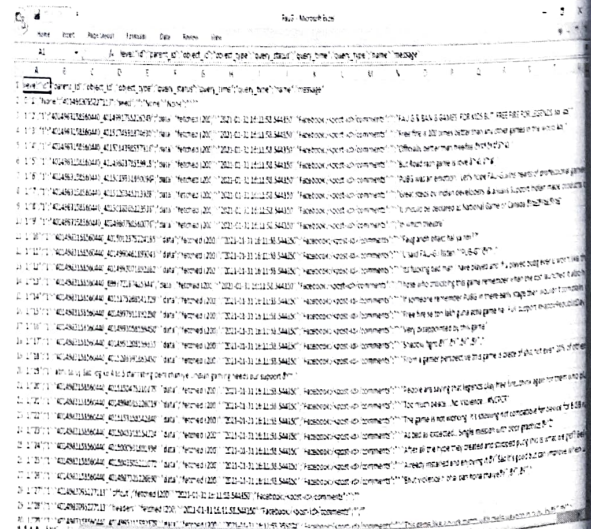


Fig 3. .csv file contains extracted comments

(0.0, 0.0)	459
(-0.4, 0.4)	95
(0.4, 0.8)	15
(0.5, 0.5)	14
(0.45, 0.65)	9
(0.25, 0.3333333333333333)	8
(0.0, 0.4)	8
(0.35, 0.55)	7
(0.8, 0.75)	5
(-0.6999999999999999, 0.6666666666666666)	5
(0.0499999999999999, 0.45)	4
(0.7, 0.6000000000000001)	4
(0.0499999999999999, 0.5)	4
(0.2, 0.3)	4
(1.0, 0.3)	4
(-0.2, 0.2)	4
(-0.4, 0.6)	4
(0.8, 0.7)	3
(0.14999999999999997, 0.5)	3
(0.3333333333333333, 0.7666666666666666)	3
(0.3, 0.35)	3
(0.0, 0.6000000000000001)	3
(-0.4000000000000001, 0.4000000000000001)	2
(-0.4, 0.5)	2
(0.6, 1.0)	2
(-0.1, 0.35000000000000003)	2
(0.15, 0.65)	2
(-0.03333333333333336, 0.4666666666666666)	2
(-0.5, 0.5)	2

Fig 6. Categorizing sentiments

It has been observed out of 804 rows of dataset the 459 rows neutral polarity and only 14 are with positive polarity and rest of other are with variations. It depicts that most of people while giving random comments, are not bother about positive of negative type of comments.

TABLE 2 SENTIMENTS BIFURCATIONS

Neutral	Positive	Other	Total
459	14	331	804

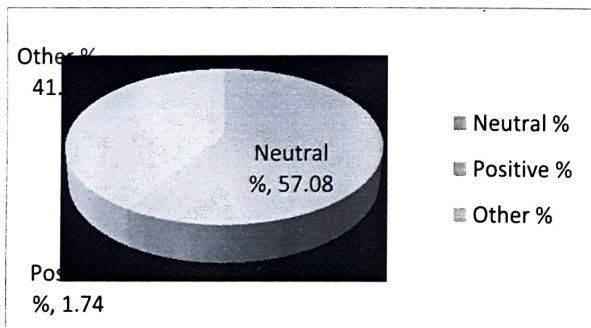


CHART 1. SENTIMENTS POLARITY IN %

CONCLUSION:

NLP is used to process gathered social network data which is in human language. It is used to preprocess, clean and classify data which

can be given as input to perform analysis using unsupervised machine learning. It is helpful to know people's opinion. The opinion can be positive or negative. The parts of speech are used as feature to extract the sentiment of the text. To identify sentiment from parts of speech, adjective plays an important role. It is difficult to identify sentiment when words having adjective and adverb are used together for the opinion.

Facebook comments data contain more emojis, special symbols and regional languages hence it is need to be cleaned and process in more sufficient way. I found natural processing technique more reliable to perform such tasks. It is not necessary that NLP can only be used for Facebook comments or just for game review. It can be used for any purpose of natural language data. It can be used for marketing, finance, reviewing etc. And the way which researcher implemented coding while analyzing data the most of time spent on pre-processing of data and doing exploratory analysis. Later on extracting the sentiments form dataset. It has been observed that most of people while giving random comments they are not bother about positive of negative comment. The 41% sentiments were neutral and 1.74% were positive and 57% were other as depicted in CHART 1. So for further study we can proceed for creating machine learning model which will help us for predicting the sentiments.

REFERENCES:

A. Book reviews

- Natural Language Processing with Python by Edward Loper, Steven Bird, Ewan Klein published by "O'Reilly Media, Inc.", 2009
- Fluent Python: Clear, Concise, and Effective Programming by Luciano Ramalho published by O'Reilly Media, Incorporated, Inc. ISBN: 9781492056355.
- Statistical Methods for Practice and Research; A S Gaur & S S Gaur, Response Books, 2006.

B. Journals, Articles and News.

- "Computer Speech and Language" by R.K. Moore, ISSN: 0885-2308.

C.D. Manning, P. Raghavan and H. Schütze, "Introduction to Information Retrieval", Cambridge University Press, pp. 234-265, 2008.

E. Loper and S. Bird, "NLTK: the Natural Language Toolkit", Proc. ACL-02 Workshop on Effective tools and methodologies for teaching natural language processing and computational linguistics ,vol. 1,pp. 63-70, 2002

Study of Sentiment Analysis from Social Media using Python by Prof. Prashant N. Wadkar and Hidayat Peerjade ASM's International International Conference on "Ongoing Research in Management & IT,

Pune in the proceeding **MARKETING MANAGEMENT & INFORMATION TECHNOLOGY** ISBN : 978-93-89835-81-6, JANUARY, 2020

C. Web References:

- <https://medium.com/analytics-vidhya/data-cleaning-in-natural-language-processing-1f77ec1f6406>
- <https://www.bgr.in/gaming/top-5-made-in-india-mobile-games-faug-real-cricket-20-wwe-racing-showdown-ludo-king-indian-air-force-a-cut-above-94100>
- www.Kaggle.com

