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:

25

Prashant Wadkar Yashaswi Education Society-International Institute of Management Science

9 PUBLICATIONS 0 CITATIONS

SEE PROFILE

Control of the object control of the object of the ob	Cl es unorra	are uninung ac in/Anne1/Hear/MohA/Sea	rzbliet				* 5 +
UGC-CARE List UGC-CARE List Vou searched for "2229-3620". Total Journals : 1 Search: <u>Search:</u> <u>Search:</u> <u>Search:</u> <u>Search:</u> <u>Search:</u> <u>Search:</u> <u>Publicher</u> <u>Publicher</u> <u>Shodh Sanchar Bulletin (print only)</u> Shodh Sanchar Educational and Research Foundation <u>2229-3620</u> NA from September-2019 to April-2021 Discontinued from April 202 Showing 1 to 1 of 1 entries <u>Previous</u> <u>1</u> N	- ugees	анскаприленски укррз у озстучном узек	n n Lise				
UGC-CARE List Searcher for "2229-3620". Total Journals : 1 Searcher ("2229-3620". Total Journal Title 1 Juncational Title 1 1 Shodh Sanchar Bulletin (print only) Previous 1 Previous 1	≡ UGC-0	CARE List					
You searched for "2229-3620". Total Journals : 1 Search:	UGC-CA	RE List					
Sn.No. Journal Title Publishee 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 202 Showing 1 to 1 of 1 entries Image: Coverage years	You searched	i for "2229-3620". Total Journals : 1				1	Search:
1 Shodh Sanchar Bulletin (print only) Shodh Sanchar Educational and Research Foundation 2229-3620 NA from September-2019 to April-2021 Discontinued from April 202 Showing I to I of 1 entries	SnNo.	Journal Title	Publisher	ISSN	E-ISSN	UGC-CARE coverage years	Details
Showing I to I of I entries	12						
	1	Shodh Sanchar Bulletin (print only)	Shodh Sanchar Educational and Research Foundation	2229-3620	NA	from September-2019 to April-2021	Discontinued from April 202
	1 Showing 1 to	Shodh Sanchar Bulletin (print only)	Shodh Sanchar Educational and Research Foundation	2229-3620	NA	from September-2019 to April-2021	Discontinued from April 202 Previous 1 Ne
	1 Showing I to	Shodh Sanchar Bulletin (print only) 1 of 1 entries	Shodh Sanchar Educational and Research Foundation	2229-3620	NA	from September-2019 to April-2021	Discontinued from April 202 Previous 1 Nr
	1 Showing I to	Shodh Sanchar Bulletin (print only) 1 of 1 entries	Shodh Sanchar Educational and Research Foundation	2229-3620	NA	from September-2019 to April-2021	Discontinued from April 202 Previous 1 Ne
	1 Showing I to	Shodh Sanchar Bulletin (print only) 1 of 1 entries	Shodh Sanchar Educational and Research Foundation	2229-3620	NA	from September-2019 to April-2021	Discontinued from April 202 Previous 1 Ne
	1 Showing 1 to	Shodh Sanchar Bulletin (print only) 9 I of I entries	Shodh Sanchar Educational and Research Foundation	2229-3620	NA	from September-2019 to April-2021	Discontinued from April 202 Previous 1 Ne
	1 Showing I to	Shodh Sanchar Bulletin (print only) 1 of 1 entries	Shodh Sanchar Educational and Research Foundation	2229-3620	NA	from September-2019 to April-2021	Discontinued from April 202 Previous 1 Ne
	1 Showing 1 to	Shodh Sanchar Bulletin (print only)	Shodh Sanchar Educational and Research Foundation	2229-3620	NA	from September-2019 to April-2021	Discontinued from April 202 Previous 1 No
	1 Showing I to	Shodh Sanchar Bulletin (print only)	Shodh Sanchar Educational and Research Foundation	2229-3620	NA	from September-2019 to April-2021	Discontinued from April 202 Previous 1 Ne
	1 Showing I to	Shodh Sanchar Bulletin (print only)	Shodh Sanchar Educational and Research Foundation	2229-3620	NA	from September-2019 to April-2021	Discontinued from April 202 Previous 1 Ne
Copyright © 2024 Savitribai Phule Pune University. All rights reserved. Disclaimer	1 Showing 1 to	Shodh Sanchar Bulletin (print only) 9 I of I entries 4 Savitribai Phule Pune University. All rip	Shodh Sanchar Educational and Research Foundation	2229-3620	NA	from September-2019 to April-2021	Discontinued from April 202: Previous 1 Ne

GOVT. OF INDIA RNI NO.: UPBIL/2015/62096 UGC Approved Care Listed Journal

SIC

ISSN 2229-3620

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



r	ઝ	CONTENTS						
	S.	TITLE	NAME OF AUTHORS	PAGE				
	No.			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) MEDIAADVERTISING 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 SKILLINGTHE SKILLS FORFUTURE EMPLOYMENT CHALLENGE-2030	Prof. Manoj Sathe Dr. Sudarshan Pawar	24				
	6.	INVESTORS PREFERENCES TOWARDS GOLD AS AN INVESTMENT OPTION WITH REFERENCE TO PUNEDISTRICT.	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				

COR			8
10	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 DISEASEPANDEMIC 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: APPLICA TION 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: TOANALYZE 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 A TTENDANCE	Narayan Kulkarni Dr. H. S. Fadewar	145

27.	FACTORS AFFECTING CONSUMER'S BUYING BEHA VIOR 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. A wasthi Dr. H. S. Fadewar	168
31.	VOCATIONAL EDUCATION ANDTRAINING: 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 A WARENESS 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. Bhola Sarang S.	219
40.	STUDY OF E-MARKETING PRACTICES OF SELECTED SMARTPHONE BRANDS FOR PCMC REGION	Prof. Amarnath Gupta Ganesh Kalshetty	226

ISSN - 2229-3620 UGC CARE LISTED JOURNAL



SHODH SANCHAR Bulletin January-March, 2021 Vol. 11, Issue 41

Page Nos. 208-213

AN INTERNATIONAL BILINGUAL PEER REVIEWED REFEREED RESEARCH JOURNAL

NLP FOR SOCIAL MEDIA DATA PROCESSING

D 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 artificia intelligence which is concerned with the computers and human language interactions. I 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

**Asst. Professor, ASM's IIBR, Pune, India	
*Student ASN's CSIT Dung India	_

ntained in the documents and categorize them. atural Language Processing is the automatic anipulation of natural language, like speech and xt which we implemented in this research.

Python

Python, the dynamic programming nguage is used for this research, which made the sting and debugging extremely quick. There are tensive open source libraries available for this ersion of python

NLTK

Natural Language Toolkit, which is also nown as NLTK is the python's library which ovides the base for classification of data. NLTK prary is used for processing text, tagging, kenization as well as classification. It plays a tal role in transforming the text data or dataset in e tweets/reviews/comments/posts into a format at can be used to extract sentiments.

The various functionalities and python oding is used in pre-processing of data/dataset ad hence the data which made available from the ocial media made fit for mining and extracting atures. It also helps in machine learning gorithms which can be used to train the model or assifier. We also can calculate and check the ocuracy of different classifier using NLTK.

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

JECTIVE

The main objective of this research study ork is to carry out the natural language occessing technique to process data extracted om popular social media site Facebook and to alyze 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 help of the developing relationships with 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 а 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 exited 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

11 • Issue 41 • January to March 2021 SHODH SANCHAR BULLETIN 209

BI-LINGUAL INTERNATIONAL RESEARCH JOURNAL

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'.
- Data will be stored in database. Click on Export data
- 9. .csv (comma separated values) file will be generated.

	a mage of Procession	. a. perc.					•	the later and in cases of Antonican Bins and
e 10+16	in the bas	Coon how	Sec 1.0	and a second sec			· .	
WORK with								
4:54(E) AND	Setting # () 2	224 12	Location and		Tap C D Barry & Lan	ET FOR HETELS,		
31.531 444	Nacket (11)	10.00	The part of the		147-14-1-1 17-17	CORPTIAL PL OF		
47548R	Carlos Carlo	And the ball of	fagears rec		particular i trade inte-			
1-10. 141	10.00	2001 Log 11 1 1	for the second second		for wear of the part			
8-045 Yes	Constant Const	Br. (1.11.11)	farmer or		interest products		6.00	the Contract State of the state
1.412 044	and a second sec	THE ACT IN MICH.	Income of		I down i he want	a list and family		
4.991 849	the base of the	\$254 JZ 50 PT 5	farme in		t which Parties			
A	Address The		farmer an		faut mit hour be			
4. M. 4.	and and the second s	Mark 161 To 111	Instant in		tractat d			
6.16 10	and the second	EPA DOL	In start in		It furthers had then	in a chine of a		
1-10 100	telebert (1)	manual for the	farment int		here all anti-	The later of the later		
	Interfact (2)	2010/01/17	Louis new Year		I kinging proving	P.S. other set		
10110	below of the	400 2 10 17 2	Indian in		free fact on the sector			Con Course
a ing in	e ingen (name	region (or	-			tribup		Tana ing
1.000 (m. 1.000) 1.000 (m. 1.000)	and moved	erador (or				tribup This and West of water		Terre ing An of the State of State and State An Official Control of State and State
1.000 1.000 1.00 1.000 1.000 1.000 1.000 1.000 1.000	ar stragang – Canalona gang meto () ar (n nadar (ar				terbage Table and Same of Access (instance) subsci		Terrer Las Autor de la construction de las de las de las de las de las Autor de las de
taa tulaa ta aaa taa iyo iyoorta uurr yeete iyo aayoo yoota	ar isagang - Fandrah gang mensis aris					teritorge Votin ande Sameri af negati forst strafte stades Forlande aftern forme	1	
նար հետար նա ապան նախությացքն ներ չթողեն վար առրությունն առուղել	andragangi fanton gang metoloj anto	e engan (en				Serbage Vote and Servir al value fors straft radio. Foliote algori tam	2	There is a finite set of the set
Saa halay far asaa kaya ngayar k asar yeyar kala asara yeshal asara i aasa	eringengi fantan gang mening ang			18.75		Arthug Mithug Mithung Mithung Mithung Mithung Mithung Mithung Mithung Mithung Mithung Mithung Mithung Mithung Mithung Mithung Mithung Mithung Mithung Mithung Mithung Mithung Mithung Mithung Mithung Mithung Mithung Mithung Mithung Mithung Mithung Mithung Mithung Mithung Mithung Mithung Mithung Mithung Mithung Mithung Mithung Mithung Mithung Mithung Mithung Mithung Mithung Mithung Mithung Mithung Mithung Mithung Mithung Mithung Mithung Mithung Mithung Mithung Mithung Mithung Mithung Mithung Mithung Mithung Mithung Mithung Mithung Mithung Mithung Mithung Mithung Mithung Mithung Mithung Mithung Mithung Mithung Mithung Mithung Mithung Mithung Mithung Mithung Mithung Mithung Mithung Mithung Mithung Mithung Mithung Mithung Mithung Mithung Mithung Mithung Mithung Mithung Mithung Mithung Mithung Mithung Mithung Mithung Mithung Mithung Mithung Mithung Mithung Mithung Mithung Mithung Mithung Mithung Mithung Mithung Mithung Mithung Mithung Mithung Mithung Mithung Mithung Mithung Mithung Mithung Mithung Mithung Mithung Mithung Mithung Mithung Mithung Mithung Mithung Mithung Mithung Mithung Mithung Mithung Mithung Mithung Mithung Mithung Mithung Mithung Mithung Mithung Mithung Mithung Mithung Mithung Mithung Mithung Mithung Mithung Mithung Mithung Mithung Mithung Mithung Mithung Mithung Mithung Mithung Mithung Mithung Mithung Mithung Mithung Mithung Mithung Mithung Mithung Mithung Mithung Mithung Mithung Mithung Mithung Mithung Mithung Mithung Mithung Mithung Mithung Mithung Mithung Mithung Mithung Mithung Mithung Mithung Mithung Mithung Mithung Mithung Mithung Mithung Mithung Mithung Mithung Mithung Mithung Mithung Mithung Mithung Mithung Mithung Mithung Mithung Mithung Mithung Mithung Mithung Mithung Mithung Mithung Mithung Mithung Mithung Mithung Mithung Mithung Mithung Mithung Mithung Mithung Mithung Mithung Mithung Mithung Mithung Mithung Mithung Mithung Mithung Mithung Mithung Mithung Mithung Mithung Mithung Mithung Mithung Mithung Mithung Mithung Mithung Mithung Mithung Mithung Mithung Mithung Mithung Mithung Mithung	-12-	
Saan tulaan tur aanaan ungu ngagarta aunar gangtar (dan aunara) aunara aunara) aunara	na stangangi (finakang ganga mata si arat k k k k k k k k k k k k k k k k k k k	(m 				teringe Velle and Same of selas fors myle same forset algorithm forsets (dense forsets (dense forsets)		
Saan tulaa tu aaaa ulaa ngagarka ausar gangtarida ausara ausara ausara ausara	an straggengi – Frenderst generations (s.) an S			10.00		Eritage Total and Sandri di vargat Frankri di vargat Frankri di vargat Frankri di Vanda Frankri di Vanda		
Saa halay har aaa oho nyayih aany yesta den angen hada anna har aana	en (separa) - final en genera mercis (anti- anti- si - si - si - si - si - si - si - si					tering Selfing Selfing Selfing and Selfing and Selfing and Selfing and Selfing and Selfing and Selfing and Selfing and Selfing and Selfing and		
San Sulay Se man one oper S serve you a de nergy Sould and And						Antiger Selfinger Selfinger Senter of suppli- ficts angle spectrum Facebook Spectrum		
Saa halay he maa ay og gy h ar y stild de were kere kere ar	er (Yagenig) (* Andrei gener servis (artis i i i i i i i i i i i i i i i i i i					Antique Sector and Sector and Sector and Sector Sector angles sades Principal angles for Principal Sector Principal Sector Pr		
taa haba he aaat uha ngan he aan yen suta h aan yen suta aan yen suta aan yen suta aan yen suta aan yen suta aan yen suta	er ihregenig (* menori unders service) with L L L L L L L L L			100		Artista Vali and Serie of water Fried attack Fried attack Fried attack Report and Research of rough Research of rough Research of rough		

Fig 2. Illusion of Data extracted from Facepager

Comment	RESULT
Туре	
Original	FAU G & BAN G GAMES
comment	FOR KIDS BUT FREE FIRE
	FOR LEGENDS ♥♥
Processed	fau g ban g games kids free
comment	fire legends
TINTE 1 CARE	D O D O O O O O O O O O O O O O O O O O

TABLE 1 SAMPLE OF COMMENTS AND PROCESSEE 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 was 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.

		Rui - Monathing			- ;
" tote bort Augesteaut forma	ill Data Raves time				9
AL · / // leveling	fipeert gi'deet citoet see'	own state" been time" to	uen nice l'hame' mez	94	
9 2 5 4	E E E E	i anteni i territore d	: 8 \	5 8 4	X 5 *
Here''s area is tober is tober a	THE OWN SALE THE TOWN	he'' 's te' 'netwe'			
T "hare" (CHEORED' 11" Wed""	"YON" NON"	and the stands			
177.177404940256040_4049805025	08/7365 "Hotes (2011-32) 0-224	1253 5467 "Newton red	d abigment (******	51 \$44 \$ \$40 ET FOR \$25 B	C. HELME ON TELEPIST
1.3 "1" CLARIC (LIGHOL) CLIS (MISSING	43017 feasi Mechan (2011 202) 42-02 5	115154590 "Restook.kg	t originates' "The	141 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1	of core frames is all indices.
17.17 40460 400 400 400 400	10"" NO. "HOME DE ' 20-0-01	ELLSI SERIO "Restrocker	to "remainer."	out other net restauted	942-01
1.5.1.404961005040 4146637555	9.5";"ben" "fetthed (207"" "202) Ct-11 2	till 51 Stats" "Heatoor - St	a comments	Roudings range to real Paul	¥.
1.6.1.40459.2256040 4230393399	1997 'mai "fecter uz? " 200-0-01	ALL SE SALLST FREEDOK, CO	a channets" "N	הטר לישו רכתופי אנשוג ה	LOUNT NET DOBUS
177.7740460529640 4051234013	305" "State" (States 1000" "300 CI-EL1	CLISE SALEY "Repetook / co	durinanent" be	e cooleve notion of loops	S A HART S LOTOR POST MAN I
UTIT CARDING COURSES	BIF familites an inches an	1251 MAIN Tacebook - So	downers U	culo se declared zi hational i	Sue : Own Sizura is
179 TO ALASSINGIA ALASSINGIA	0767an "Network" (22-0-00	t u SI S4037 "Facebook, -SX	diation mediate	101710-010	
1 10 11 101450150560440 4015013552	24185 "data"/ fermed (200 "1021-11 11	1911-1853435 ¹¹² 409000-57	18-12-12-12-12-12-12-12-12-12-12-12-12-12-	43 YO MAC 16 12 10 17	
1712 71 YELAND 19296440 4019904618	5043 ¹¹ data"/herched (2001/12/21-01-33	the most states of the second states	1""(eremo:-b.so	udit. 6 150 "9.65"	e.:
TUTT CARDINGAN AUSSING	228211 data1/14ccred (202110221-01-0	18-11-58-544450" Newcoscolog	of Common String	tuoing such an investigation	TE TO DATE NOT THE TO
A LUD TO MUSELLINGE ENTERIES	5144", "Sela "Netzina 1200 ""2021-01-11.0	STATIST MARINE MARROOK - SO	deciment' "In	the site models and party in	URDAR SURVEY IN CAL
174"T"CR88118:6640 (01175265)	ALES "dra", Terred 200 TELLAS 3	121083465° (408000)	on a bonnet " "	Komeane remember Außa in t	AGRU DE 24. HOLE IS
17377 424583356644 4245979110	1226 "data" "fermen (20) "(221-0-0)	a line season second	R""," Cremerci - Do Sci	en frei se ter feit gena seine Be	and as you proport indication
118 1 CASOLINKAN CLASSING	5450 "drif"/femed.000 (1200-0-0	te 10 % SAUST THEREON A	CEAD Comments	BY DISCONTEED, THE BITE	
111711 60480306440 40480285	5413 'data' 'wmed 120 " Til-II-II	141118 SHERE REPORTED	C. (Section of the section of the se	Sec. 428 8 8 9 2	
CONTRACTOR COMPLEX	ALL COLUMN THE PARTY CARD WAS A	14 11-93 SARSY TREBERED OF	as a comment "	or a gamer perspective Trd p	me 5 260 2162 1010 101
Littin in sy is gutestim	Regional change under commanded as	u support 811			
1172717 (C.86199844) (C.11545)	10478 "1864" (etmes 120 200-00-00	1911-19 States" Residencia	an a lanner a sa	NOR PERMIT	in the bears have been
177 1 KTANI1196441 KTANI011	2019 'MAT NET 45 201 ' 2014 10	HE LEGE SHEET ("HERENDED")	ויייניוארצי פאנט	STUDING & IDEA	NOT
S 172111 REAMOUNT RELEASE	ANE ' MU' 'WE'ND ' TUI-01-01	1611-19 SALST "HORODOLIN	C	a press to sorry 1100	sure of concepts where
T.B.J. J. M. WEITHRAM, M. WORKIN,	SUDF "Sifa", "Waters (200" "2015-0-0	1011-91-SARSY FROM TAX	ana amin'ny	DEC IN COLOCIEL, SAVIN DES	at up sto perior .
1 14 T TOANLESSON CLOSED	0.99 data Teches (200 1223-0-31	IELLAR SAILS, "HEREBOOK	201-0-01000-0-02	ter pi tre nyce trey bened a	20 STORED P. 19 10 10 10 10 10
1 B T GLOBERSON CARES	2077 "Mail" Nored (201-201-201-201-	HELM SHIDS FERROR S	as diament'	NO REPORT OF	ST Sectores and
1.3.1. CONTRACT CONTROL	268 "ana", ferr ed (20" "211-11-11	11110 SAIN "428000'S	at plannets'"''	Any officer of a partitional f	A461 2. 2.
S 1 77 T CARENSETTER STOL (NE	CT 45 (20) "201-01-01 1-01-25 54450"	Notaci-conner	5		
CLOSED CLOSED THESE	WIRE DO NOT ALL STREET	התה לו הנט וכנולאלי "	57.0°.0		
C LAUT C. MULDANT, C. BOULD	וונובורירי יינרואשריאל "וואין לאלה	······································	on a longer and	SEPARAL MADE NO	1 Selsados
and the second second second second	and the second sec		and the second	and the second se	ant

Fig 3. .csv file contains extracted comments

BI-LINGUAL INTERNATIONAL RESEARCH JOURNAL

RE – PROCESSING OF DATA

Data which is collected from Facebook as not fit for extracting features. Comments sted under post were mostly consist of emojis, ecial characters, extra spaces, time stamps, RL's, stop words, ,regional languages etc. To ake this data fit for mining the researcher then e-processed this dataset by making use of rious functions of NLTK. For pre- processing e data researcher has firstly extracted the main comments which was cebook stored in lessage' column of the dataset, then removed all npty spaces and rows. By using inbuilt method e removed stop words (you, and, the, an, her is, he, of, to etc.) and hash tags, repeated words, RL's, etc. Once we are done with it, then we are ady with pre-processed reviews which then used r further processing.

	-	benear ra	oldect_iq	object_type	query_status	Bully_Dime		D.A.The	
¢	۲	None	4714983095227113	1000	Lish:	lanne	Here	Taski	Hari
,	2	1	A014983156568445_4014991735236349	C81	967 ~9 (298)	0001-02-12 20:48 30 305750	Facetrics insect identifients	Nati	FAU C & BAYLG GAMES FOR KIDS BUT FREE FIRE F
,	3	1	4014383158556440_4215174591874530	cata	REITed (200)	2021-02-14 20:49 38 395169	Facebook/houst id+locmmignts	hab	Free fire is 100 brows hedler than any other 6
1	÷	,	4014363158569440_4015014308557316	045	Mithec (200)	2021-02-14 25:40 30 395169	Facebook - pesi-	tagN	Chowly beter tran teafre ⊥₩
1	٢	,	4014983158550440_4014968378558918	544	Wates (200)	2031-02-14 20 45 30 396139	Facebook input-	NaN	Bid Rhied rash game

Fig 4. Raw data before preprocessing

Python code used to import and read .csv file:

port numpy as np

port pandas as pd

port re

port nltk

tk.download('wordnet')from learn.datasets import load_files

tk.download('stopwords')

port pickle

om nltk.corpus import stopwords

port matplotlib.pyplot as plt

om textblob import TextBlob

_data=pd.read_csv("sample.csv",delimi

using head command we can view first ve rows of dataset i.e by

0. 11 • Issue 41 • January to March 2021 SHODH SANCHAR BULLETIN 211

fb_data.head()

Python code for removing unwanted stuff and stopwards from raw data:

fb_data['reviews.text'] = fb_data['reviews.text'].str.replace('[^a-zA-Z]',' ')

stop = stopwords.words('english')

fb_data['reviews.text'] =
fb_data['reviews.text'].apply(lambda x: "
".join(x for x in x.split() if x not in stop))

Applying Porter Stemer for steaming the words

from nltk.stem import PorterStemmer

st = PorterStemmer()

fb_data['reviews.text'] =
fb_data['reviews.text'].apply(lambda x: "
".join([st.stem(word) for word in
x.split()]))

B. Getting polarity of sentiments

def senti(x):

return TextBlob(x).sentiment

fb_data['senti_score'] fb_data['reviews.text'].apply(senti)

fb_data.senti_score.head()

_

C. Fig 5. Sentiment Score

It has been observed the two scores, the first score is the sentiment polarity which tells either the sentiment is positive or negative and the second score showed the subjectivity score to tell how subjective the text is.

D. fb_data['senti_score'].value_counts()

459 (0.0, 0.0)95 (-0.4, 0.4)15 (0.4, 0.8) 14 (0.5, 0.5) 9 (0.45, 0.65) 8 (0.25, 0.33333333333333333) 8 (0.0, 0.4)(0.35, 0.55) (0.8, 0.75) 7 5 5 4 (0.04999999999999999, 0.45) 4 (0.7, 0.600000000000000) 4 (0.04999999999999999, 0.5) 4 (0.2, 0.3) 4 (1.0, 0.3)(-0.2, 0.2)4 4 (-0.4, 0.6) (0.8, 0.7)3 3 (0.14999999999999997, 0.5) 3 (0.333333333333333, 0.766666666666666) 3 (0.3, 0.35) 3 2 (-0.40000000000001, 0.4000000000000)) (-0.4, 0.5) 2 (0.6, 1.0) 2 (-0.1, 0.35000000000000000) 2 (0.15, 0.65) 2 (-0.0333333333333336, 0.46666666666666666) 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





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. In 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/datacleaning-in-natural-language-processing-1f77ec1f6406
- https://www.bgr.in/gaming/top-5-made-inindia-mobile-games-faug-real-cricket-20-wweracing-showdown-ludo-king-indian-air-forcea-cut-above-94100
- www.Kaggle.com

