



Sentiment analysis of Social Media Data

Piniseti Vishal,

PG Scholar, MCA, Department of Computer Application, School of Computing, Mohan Babu University, Tirupati, India.

Peddinti Neeraja

Assistant Professor, Department of Computer Application, School of Computing, Mohan Babu University, Tirupati, India.

Piniseti Srikanth

Data Engineering, University of Missouri, Kansas City, USA

Abstract - Big data has been around for quite long period of time, and yet most industries now understand their ability to use business intelligence to significantly enhance their business processes by collecting all of the information that goes into their processes. Facebook is a social media platform that has been available for use for over a decade. It has over 2 billion active monthly users. The company has been present in the headlines for its data privacy issues, but the company's data provides a wealth of information on interpersonal interactions and what they are interested in. This project's goal is to explore what can be done with Facebook data analysis. It will also look at some of the tools available to help make sense of it all. Among the insights Facebook data provides is how people and their interests change over time. For example, if you want to promote your company, it would be worthwhile to look at what people are interested in. You could look at the keywords that are trending each year or month. This might assist you in making a decision. Where in society your marketing dollars will help you most effectively reach people and convert them into paying customers. To explore this, we need an indication of what is popular right now. The best way to get this information is to use a tool such as Hadoop MapReduce, Spark, Zeppelin etc. Through our project we intend to carry out analysis on a preferably large dataset. So, we have chosen the dataset obtained from several Facebook users. By carrying out certain operations, we intend to harness their data and use it to identify new opportunities.

Keywords: *Intrusion detection systems (IDS), IoT security, wireless sensor networks, universal IDS.*

1. INTRODUCTION

Facebook is the world's most popular social media network. It was developed to facilitate communication between individuals and share information. However, it has now become a strong promotional instrument that enables businesses to interact with their customers in a more effective way. The data generated by Facebook is important for many businesses as it helps them to enhance their marketing campaigns additionally gain insight into customer behaviour and provide better customer service. In this project we will be using Bigdata tools for Facebook data analysis and how to use the knowledge gained during this examination for better understanding the customers and hence providing them with a better experience. Bigdata tools for Facebook data analysis: Facebook offers an assortment of tools for businesses to get insights about their customers. Among these tools is Insights, which is the central hub that provides analytics and insights on different types of social media content. Another tool is the "People You May Know" feature which allows users to search people they may know. To ensure that obtain marketing insights from Facebook, you need to use several bigdata tools such as Hadoop, Spark etc.

2. RELATED WORK

The study listed below has been completed. which are linked to our Facebook data analysis:

- World Bank's mixed initiative social media analytics: Bigdata tools are accustomed to discuss and analyze the connection between public trust and social protest.
- Sentiment analysis on Facebook groups lexicon based strategy:
This essay addresses how cluster Analysis may be employed to ascertain the sentiment behind each post, comment, and like. If sentiment analysis is carried out on the vast amount of data, business values can be ascertained from these objects and events.

- An analytical tool for visualizing emotion in social media text streams: This document is about the visual graphical approach of analyzing the textual data.
- Social Media Big Data (SMBD) Advances in Research and Development Trends: Knowledge Mapping Analysis Using Cite Space: This essay examines about the research trends in social media.
- Crohn's Disease Patients' Attitudes: This study's objective was to demonstrate how methods taken from an assortment of disciplines, such as computer engineering, mathematical modelling, statistics, data collection, and social sciences, may be applied to examine data.

DATASET

In this project, we attempted to find datasets with maximum size, but we were Capable of obtaining datasets larger than 100 MB. The necessary data is gathered from the Netvizz tool. In addition, Tovek Tools were applied to content analysis and data research processing. Data from tools is later appended. The pictures below show The quantity of columns existing inside the dataset.

```

SanityCheck.py: 17
File Edit View Insert RunTime Tools Help Alt+command

+ Code + Test

[1] pip install pyspark

[2] Linking to libevent: https://github.com/openssl/openssl...
Requirement already satisfied: pyjosh.18.9.5 in /usr/local/lib/python3.11/site-packages (from pyspark) (0.18.9.5)

[3] from pyspark.sql import SparkSession
from pyspark.sql import Row
from pyspark.sql import functions
import pandas as pd
import os

# Main pyspark.sql.functions import col,liter, when, count

[4] os.chdir('~/content/Drive/BigDataProject/realProject')

[5] data = pd.read_csv('facebook.csv')
col = data.columns.tolist()

[6] # Create a SparkSession (the config bit is only for Windows!)
spark = SparkSession.builder.appName("SanityCheck").getOrCreate()
# Get the raw data
lines = spark.sparkContext.textFile('facebook.csv')

[7] def parseInput(line):
    fields = line.split(',')
    return Row(*fields + [""])

[8] from pandas.core.base import AbstractDataframe
from pandas.io.parsers.pyarrow_parser import count_empty_value
if __name__ == '__main__':
    ["userid", "age", "dob_day", "dob_year", "dob_month", "gender", "tenure", "friend count", "friendships_initiated", "likes", "likes_received", "mobile_likes", "mobile_likes_received",
     "www_likes", "www_likes_received", "created_time", "description", "likes", "message", "page_id", "post_id", "post_lang", "react_likes", "react_likes", "react_likes", "react_likes",
     "www_likes_received"]
    # Convert it to a RDD of Row objects with (value)
    rdd = lines.map(parseInput)
    # Convert that to a DataFrame
    df = spark.createDataFrame(rdd)
    # Column name of null values
    counts = df.filter(df["value"]=="").count()
    # Print them out
    print("Null counts: %s" % counts)
    # Stop the session
    spark.stop()

userid : 0
age : 0
dob_day : 0
dob_year : 0
dob_month : 0
gender : 175
tenure : 0
friend count : 0
friendships_initiated : 0
likes : 0
likes_received : 0
mobile_likes : 0
mobile_likes_received : 0
www_likes : 0
www_likes_received : 0
    
```

Fig-2

```

userid : 0
age : 0
dob_day : 0
dob_year : 0
dob_month : 0
gender : 175
tenure : 0
friend count : 0
friendships_initiated : 0
likes : 0
likes_received : 0
mobile_likes : 0
mobile_likes_received : 0
www_likes : 0
www_likes_received : 0
    
```

Fig-3

**Data Processing and Visualization:
Gender Graph:**

We are plotting a gender graph using the Facebook data and getting insights about the greatest number of users as per the gender.

2. METHODOLOGY

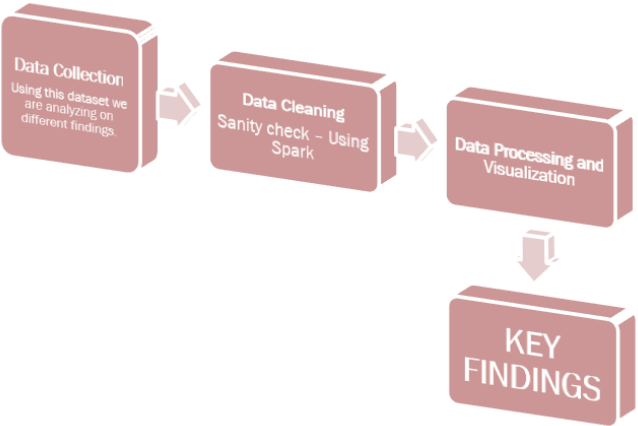


Fig-1

To prepare the data for analysis, we will retrieve it. Spark will then be used to clean the data. We will employ various techniques for each kind of information as we process and visualize it.

ANALYSIS AND RESULTS

Data Cleaning:

In this case, we're looking for any null values in the dataset. We used spark to accomplish this.

```

File Edit View Insert Cell Kernel Widgets Help

317595 Surveillance footage released by the Palm Beach...
317596 They'll need Miami voters to approve the plans...
317597 Police say the woman first struck at the Macy'...
317598 The government's offer is limited to those sta...
317599 The 38-year old electrical contractor had been...
Name: message, Length: 317600, dtype: object

In [20]: #finding number of males and females using facebook
label = dataset.groupby('gender').size()
label.plot(kind="barh")
plt.xlabel('gender')
plt.ylabel('count')
plt.xticks(rotation=90)
plt.title("Facebook Gender Graph")
plt.show()
    
```

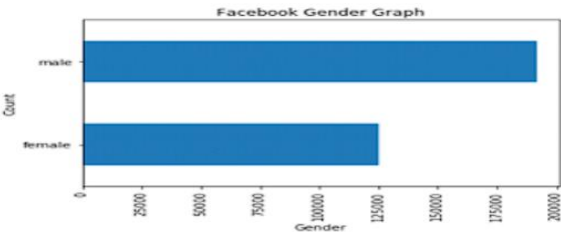


Fig-4
Age Groups:

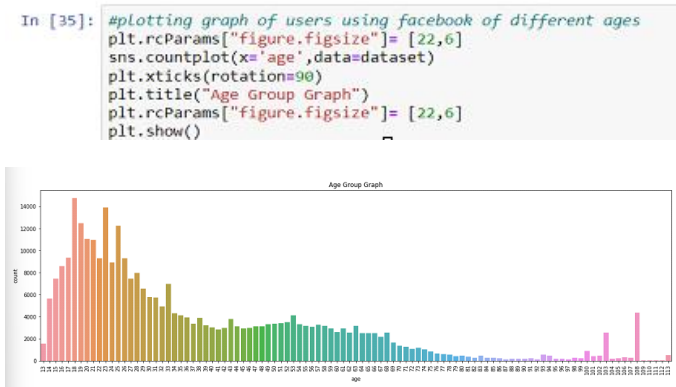


Fig-5

Here we are plotting a graph as per the age group Likes from Gender:

Blue dots for male and orange dots for females and x axis represents Ages and y axis represents likes so in above graph we are plotting LIKES graph using gender.

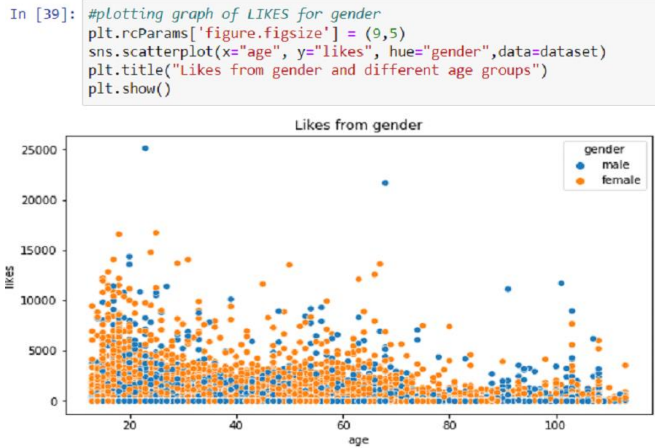


Fig-6

Facebook users born in different years:

We are plotting graph of Facebook users born in different years.

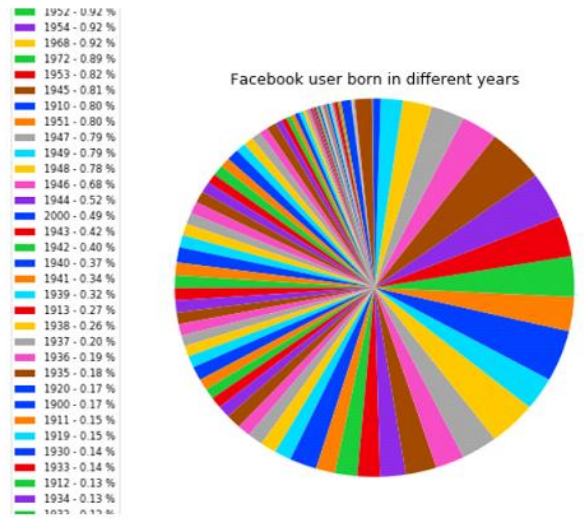
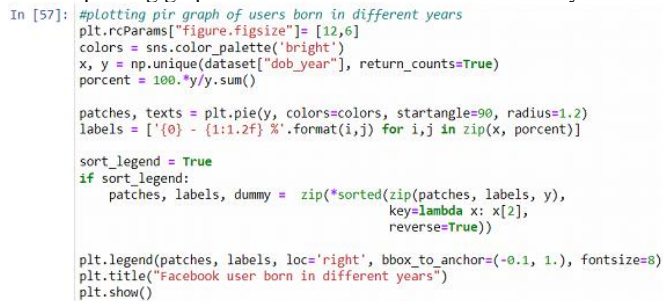


Fig-7

Mobile users like graph:

We have divided users into age groups like 10 20, 20 30 and then plotting graph of LIKES from various age groups and these likes are given using MOBILE platform.

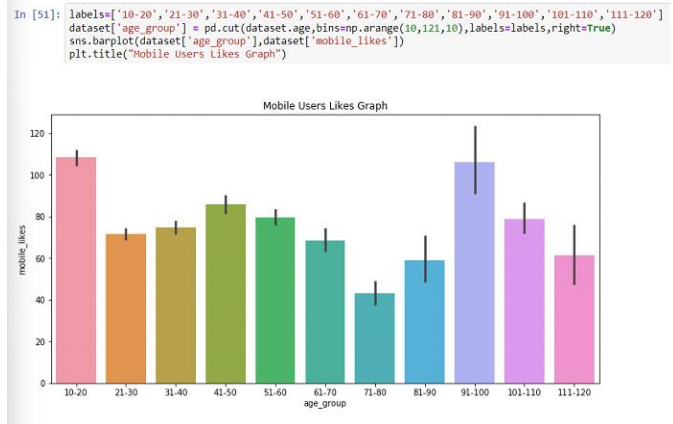


Fig-8

Web users likes graph:

WEB users of different ages giving likes from website.

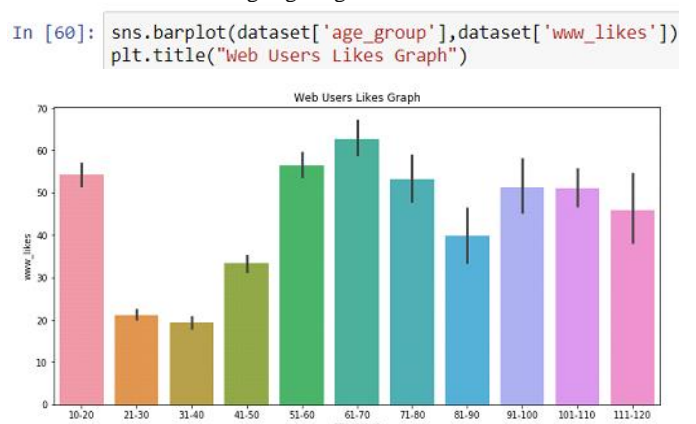


Fig-9

Frequency of likes found in dataset:

We are plotting frequency of likes where x axis represents likes from 1 to 5 and y axis represents users count.

```
In [80]: #frequency of different restaurant found in dataset
names, count = np.unique(dataset['likes'], return_counts=True)
names = names[0:5]
temp = count[0:5]
data = []
i=4
while i >= 0:
    data.append(temp[i])
    i = i - 1
plt.plot(names, data)
plt.xticks(rotation=90)
plt.rcParams["figure.figsize"]= [12,6]
plt.title("Frequency of Likes found in Dataset")
plt.show()
```

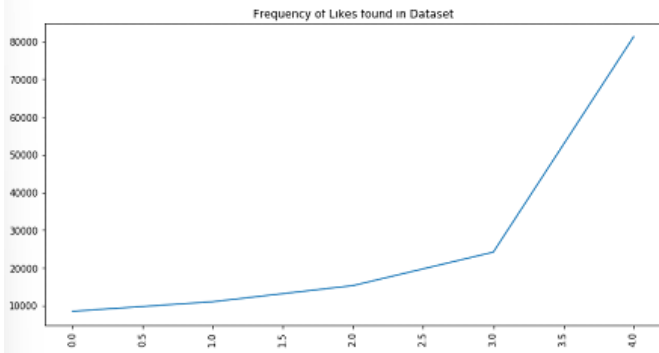


Fig-10

Frequency of dislikes found in dataset:

```
In [81]: #frequency of different restaurant found in dataset
print("Frequency Table for dislikes")
names, count = np.unique(dataset['react_sad'], return_counts=True)
plt.plot(names[0:5], count[0:5])
plt.xticks(rotation=90)
plt.rcParams["figure.figsize"]= [12,6]
plt.title("Frequency of Dislikes found in Dataset")
plt.show()
```

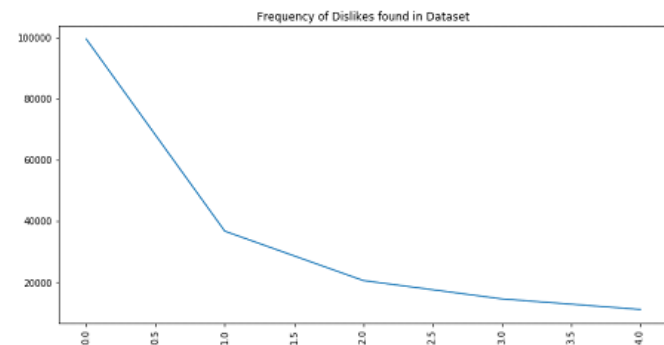


Fig-11

Friends count for gender:

We are plotting friend's graph of males and females

```
In [88]: #top 10 segment categories with highest sale
max_segment_category_sales = dataset.sort_values(['friend_count'], ascending=False)
max_segment_category_sales = max_segment_category_sales[0:16]
plt.xticks(rotation=90)
sns.violinplot(x='gender',y='friend_count', data=max_segment_category_sales, palette="rainbow")
plt.rcParams["figure.figsize"]= [12,6]
plt.title("Friend Counts for Gender")
plt.show()
```

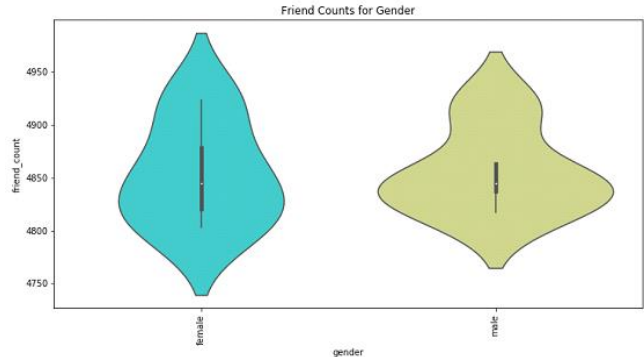


Fig-12

User behavior graph:

Plotting USER behaviours graph like HAPPY, SAD, ANGRY and WOW.

```
In [95]: ct = dataset[['react_wow', 'react_haha', 'react_angry', 'react_sad']]
wow = ct['react_wow'].sum()
happy = ct['react_haha'].sum()
angry = ct['react_angry'].sum()
sad = ct['react_sad'].sum()
data = pd.DataFrame([wow,happy,angry,sad], columns=['Wow','Happy','Angry','Sad'])
display(data)
height = [wow,happy,angry,sad]
bars = ('Wow','Happy','Angry','Sad')
y_pos = np.arange(len(bars))
plt.bar(y_pos, height)
plt.xticks(y_pos, bars)
plt.title("User Behaviour Graph")
plt.show()
```

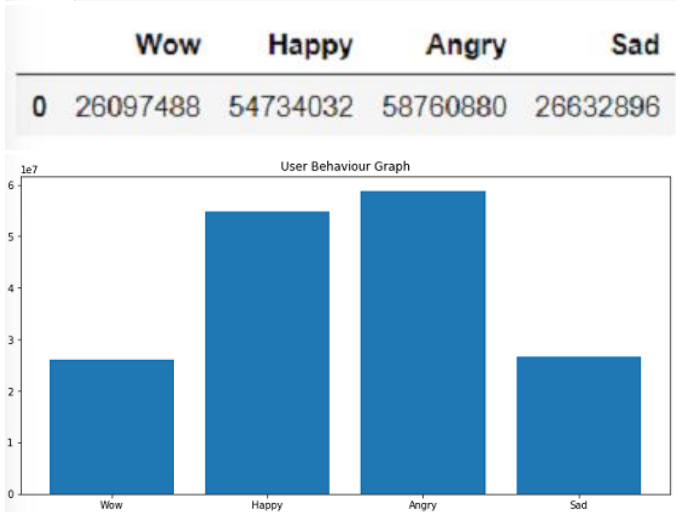


Fig-13

3. CONCLUSION

In the current digital era, Facebook is among the most often utilized apps worldwide. Every minute, Facebook's servers receive a large amount of data. This information comprises an extensive array of informational content, such as news updates, entertainment, and education. Users of this site have a variety of options. Investigating the potential applications of Facebook data analysis is the aim of this project.

As a component of our project, analysis was done, ideally on a sizable dataset. As a result, we chose to utilize the information that was gathered from various Facebook users. Many businesses value the information provided by Facebook since it is beneficial to them to improve their marketing initiatives, acquire understanding of customer behavior, and improve customer service.

REFERENCES

- [1]. Combining collaborative categorization of emotion and filtering to provide better movie suggestions. In The 5th International Conference Proceeding on Multi Disciplinary Trends in Artificial Intelligence, MIWAT'11, pages 38-50, Berlin, Heidelberg, 2011. Springer Verlag
- [2]. Vohra, S. M., & Teraiya, J. B. (2013) A comparative study sentiment analysis techniques. Journal Jikrce
- [3]. H A. Go, R. Bhayani, and L. Huang Using remote supervision to classify sentiment on Twitter, 2009
- [4]. Hassan, Sundus, Muhammad Rafi, and Muhammad Shahid Shaikh --"Contrasting svm and guileless bayes classifiers for text order with wikitology as information improvement." 2011 IEEE fourteenth International. IEEE, 2011.
- [5]. M. M. Mostafa --"More than words: Social organizations' content digging for purchaser brand slants," Volume 40 of Expert Systems with Applications no. 10.
- [6]. Rana, Shweta, and Archana Singh --"Relative investigation of conclusion direction utilizing SVM and Naive Bayes methods." Next Generation Computing Technologies (NGCT), International Conference IEEE, 2016.