Social Media Analytics for Business Decision Making

Project report submitted in partial fulfillment of the requirement for the degree of Bachelor of Technology

in

Computer Science and Engineering/Information Technology

By

Ajay Tyagi (191525)
Devbrat Srivastava (191540)

Under the supervision of

Dr. Ruchi Verma

to

Department of Computer Science & Engineering and Information Technology

Jaypee University of Information Technology Waknaghat, Solan-173234, Himachal Pradesh

CANDIDATE'S DECLARATION

I hereby declare that the work presented in this report entitled **" Social Media Analytics for Business Decision Making "** in partial fulfillment of the requirements for the award of the degree of **Bachelor of Technology** in **Computer Science and Engineering/Information Technology** submitted in the department of Computer Science & Engineering and Information Technology**,** Jaypee University of Information Technology Waknaghat is an authentic record of my own work carried out over a period from August 2015 to December 2015 under the supervision of **Dr. Ruchi Verma** (Assistant Professor , CSE & IT).

I also authenticate that I have carried out the above-mentioned project work under the proficiency stream **Data Science.**

The matter embodied in the report has not been submitted for the award of any other degree or diploma.

(Student Signature)
Devbrat Srivastava, 191540

(Student Signature)
Ajay Tyagi, 191525

This is to certify that the above statement made by the candidate is true to the best of my knowledge.

(Supervisor Signature)

Dr. Ruchi Verma

Assistant Professor

CSE & IT

Dated:

PLAGIARISM CERTIFICATE

JAYPEE UNIVERSITY OF INFORMATION TECHNOLOGY, WAKNAGHAT

PLAGIARISM VERIFICATION REPORT

Date: …………………………..

Type of Document (Tick): PhD Thesis | M.Tech Dissertation/ Report | B.Tech Project Report | Paper

Name:______________________________**Department:**_________________**Enrolment No** _________

Contact No.______________________________**E-mail.**_______________________________________

Name of the Supervisor: ____

Title of the Thesis/Dissertation/Project Report/Paper (In Capital letters): ________________________

____

____

UNDERTAKING

I undertake that I am aware of the plagiarism related norms/ regulations, if I found guilty of any plagiarism and copyright violations in the above thesis/report even after award of degree, the University reserves the rights to withdraw/revoke my degree/report. Kindly allow me to avail Plagiarism verification report for the document mentioned above.

Complete Thesis/Report Pages Detail:

- Total No. of Pages =
- Total No. of Preliminary pages =
- Total No. of pages accommodate bibliography/references =

(Signature of Student)

FOR DEPARTMENT USE

We have checked the thesis/report as per norms and found **Similarity Index** at .... (%). Therefore, we are forwarding the complete thesis/report for final plagiarism check. The plagiarism verification report may be handed over to the candidate.

(Signature of Guide/Supervisor) **Signature of HOD**

FOR LRC USE

The above document was scanned for plagiarism check. The outcome of the same is reported below:

Copy Received on	**Excluded**	**Similarity Index (%)**	**Generated Plagiarism Report Details** (Title, Abstract & Chapters)	
	• All Preliminary Pages • Bibliography/Images/Quotes • 14 Words String		Word Counts	
Report Generated on			Character Counts	
		Submission ID	Total Pages Scanned	
			File Size	

Checked by
Name & Signature **Librarian**

………

Please send your complete thesis/report in (PDF) with Title Page, Abstract and Chapters in (Word File) through the supervisor at plagcheck.juit@gmail.com

ACKNOWLEDGEMENTS

Firstly, We express our heartiest thanks and gratefulness to almighty God for His divine blessing makes it possible to complete the project work successfully.

We are really grateful and wish our profound indebtedness to Supervisor **Dr. Ruchi Verma,** Assistant Professor (SG)**,** Department of CSE Jaypee University of Information Technology, Waknaghat. Deep Knowledge & keen interest of my supervisor in the field of **Machine Learning** to carry out this project. His endless patience, scholarly guidance, continual encouragement, constant and energetic supervision, constructive criticism, valuable advice, reading many inferior drafts and correcting them at all stages have made it possible to complete this project.

We would like to express our heartiest gratitude to **Dr. Ruchi Verma,** Department of CSE, for his kind help to finish my project.

We would also generously welcome each one of those individuals who have helped us straightforwardly or in a roundabout way in making this project a win. In this unique situation, we might want to thank the various staff individuals, both educating and non-instructing, which have developed their convenient help and facilitated my undertaking.

Finally, we must acknowledge with due respect the constant support and patients of our parents.

Project Group No. – 109

Ajay Tyagi
(191525)

Devbrat Srivastava
(191540)

<u>Table of Content</u>

List of Abbreviations

LR : Logistic Regression

KNN : K Nearest Neighbor Classifier

CART : Decision Tree Classifier

RF : Random Forest Classifier

NB : Gaussian NB

SVM : Support Vector Machine

XGBoost : XGB Classifier

List of Figures

List of Tables

ABSTRACT

This essay discusses social media's growth, user demographics, and advantages. Many social media platforms are now being used efficiently, and the data from these platforms also aids in revealing hidden facts through analysis. In social computational science, which aids people in decision-making, data analysis is crucial. Big data analytics on social media is often known as social media data analytics or just social media analytics, and it offers analysts great opportunity to come up with original concepts and recognize new patterns. The many forms of social media analysis, as well as the widely used technologies for social media analytics, are also covered in this study. Also mentioned are the numerous research difficulties associated with studying social media.

To find patterns, trends, and insights in the acquired data, several statistical and machine learning techniques are applied. Sentiment analysis, network analysis, text analysis, and machine learning algorithms are a few examples of these techniques. Last but not least, data visualization entails displaying the analyzed data in an aesthetically pleasing and understandable manner using graphs, charts, and other visual aids.

The advantages of social media analytics include strengthening overall business performance, raising consumer interaction, raising brand awareness, spotting market trends, and improving customer service. Businesses that primarily rely on social media to reach their clients and increase brand awareness may find social media analytics to be particularly helpful. Businesses may determine their target market, assess the success of their social media efforts, and make data-driven choices by analyzing social media data.

As a result, social media analytics is an effective tool for businesses to learn more and make defensible choices based on social media data. Businesses may use social media analytics to improve consumer interaction, raise brand awareness, and improve overall business performance with the correct tools and strategies.

Chapter 1
INTRODUCTION

1.1 Introduction

In order to gain knowledge and make wise business decisions, social media analytics is the practice of gathering, analyzing, and interpreting data from social media platforms. Social media analytics is becoming a crucial tool for businesses to gauge the success of their social media marketing initiatives and interact more deeply with their customers. This is because of the rise of social media platforms like Facebook, Twitter, LinkedIn, Instagram, and others.

Businesses can track their social media presence, gauge engagement, and spot trends in real-time thanks to social media analytics. Businesses may learn more about the tastes, actions, and opinions of their customers by analyzing social media data. This can help them develop better goods and services, boost client satisfaction, and foster brand loyalty.

Social media analytics data can be utilized to identify potential customers, improve customer service, and develop focused marketing initiatives. Additionally, it can be used to gauge consumer sentiment, keep tabs on business reputation, and evaluate the efficacy of social media operations.

Analytics for social media are not just for companies with big marketing resources. By employing free or inexpensive social media monitoring tools to track their social media performance and find areas for improvement, small businesses and startups can also gain from social media analytics.

In conclusion, social media analytics is a potent tool that enables organizations to understand the behavior, preferences, and viewpoints of their customers. Businesses may strengthen their marketing strategies, increase consumer happiness, and build their brands by utilizing social media analytics.

1.2 Problem Statement

- The most frequent self-question is, "How do other people feel about us?"
- While conversing, everyone is curious to know what the other person is thinking about them; yet, it is impossible to judge another person correctly.
- In order to use sentiment analysis in discussion, this paper offers a method.
- The data received from WhatsApp chats are first pre-processed, and then sentiment analysis is done to each chat.

1.3 Objectives

Social media analytics data can be utilized to identify potential customers, improve customer service, and develop focused marketing initiatives. Additionally, it can be used to gauge consumer sentiment, keep tabs on business reputation, and evaluate the efficacy of social media operations.

Analytics for social media are not just for companies with big marketing resources. By employing free or inexpensive social media monitoring tools to track their social media performance and find areas for improvement, small businesses and startups can also gain from social media analytics.

In conclusion, social media analytics is a potent tool that enables organizations to understand the behavior, preferences, and viewpoints of their customers. Businesses may strengthen their marketing strategies, increase consumer happiness, and build their brands by utilizing social media analytics.

Enhancing customer engagement: By identifying customer pain points and potential areas for improvement, social media analytics assist firms in enhancing customer engagement. Businesses can strengthen their ties with their customers by reacting to customer comments and participating in social media discussions.

By monitoring mentions, sentiment, and feedback on social media platforms, social media analytics can assist organizations in keeping track of their brand reputation. Businesses can improve their brand reputation and foster customer loyalty by responding to negative criticism and interacting with customers.

Making data-driven decisions: Social media analytics offers companies data-driven perceptions that can be applied to decision-making. Businesses can generate data-driven judgements about product development, marketing tactics, and customer service by examining social media data.

1.4 Methodology

1.4.1 Data

- Social Media Platforms
- News data
- Public data

1.4.2 Analytics

- Dashboard
- Data Analysis
- Data visualization

1.4.3 Facilities

- Data storage
- Computational facility

1.5 Organizations

Like any other company strategy, social media marketing is most effective when your objectives and strategies are supported by actual facts. Social media data analytics offer facts that enables you to comprehend what is effective. More importantly, you'll be aware of what isn't working so you may adjust your plan and make the proper business decisions. We may customize our social media marketing plan for each social network by gathering data from social media. Strategies can be modified specifically based on geography and demography.

Chapter 2
LITERATURE SURVEY

[1] Ruhi, Umar – Social media analytics is a young and developing field that can assist businesses in developing and putting into practice measuring strategies for drawing conclusions from social media interactions and gauging the performance of their own social media activities. In the end, a social media analytics program that is effective can help firms enhance their performance management program across all business functions. The adoption, implementation, and institutionalization of approaches and procedures for a successful social media analytics program, however, are still challenges for corporations. In order to help organizations, link their social media program, procedures, and technology with the overall strategic objectives of the firm, this paper gives a business intelligence viewpoint on social media analytics. The paper presents the conceptual foundations of social media analytics in order to achieve this.

[2] Gupta, Bhumika – Sentiment analysis is an approach to analyzing data to get the emotions it contains. Twitter Sentiment Analysis applies sentiment analysis to data from Twitter (tweets) to extract the sentiments sent by users. Research in this area has grown steadily over the past decades. The reason for this is the difficult format of tweets, which is difficult to process. The tweet format is so small that it introduces a whole new dimension of problems, such as the use of slang and abbreviations. In this article, we review several articles on sentiment analysis research on Twitter, describe the methods employed and the models applied, and describe a general Python-based approach.

[3] Mohana, R. S – Users can send and receive messages known as "tweets" on the well-known social media network Twitter. It allows people to express their ideas and viewpoints on a variety of topics. Various parties, including consumers and advertisers, conduct sentiment analysis on these tweets for product insights and market research. Additionally, current improvements in machine learning methodologies have increased the precision of sentiment analysis forecasts. In this study, we used a variety of machine learning techniques to do sentiment analysis on "tweets." attempts to classify tweets as being favorable or bad based on their polarity. Overall sentiment should be used to score tweets that exhibit both positive and negative trends. The Kaggle dataset, which was crawled and categorized as either positive or negative, was used in this study.

[4] Patnana, Divya Sai, Gone Hitesh, and IpsitaSahu N. Suresh Kumar – Most machine learning models require tables and charts to be drawn in order to analyze results and make predictions. Analyzing results is a basic need for companies that need to analyze their product development objectives. In the current work, a logistic regression model is developed to describe the social network advertising dataset. A basic implementation of this logistic regression model can predict whether a user is ready to purchase a product. In my current work, I develop a logistic regression model in Python to assess accuracy and make predictions.

[5] Hota, Soudamini, and Sudhir Pathak – Emotion is the Latin word for feeling. Sentiment analysis, also known as opinion mining, is the process of analyzing data from online sources such as microblogging sites, social media, online news articles, user reviews, etc. to ascertain how people feel about particular occasions, groups, and products. It is a type of data mining that draws judgments about things like people, brands, etc. The categorizing of emotions into many classes is done in this piece of work. The suggested methodology, which is based on the ANN classification algorithm, outperforms the current methodology, which is based on the SVM classification algorithm. The most popular microblogging platform, Twitter, provided the data for the investigation. Using Python's Tweepy, source data was collected from Twitter. For feature extraction, an n-gram modelling technique was employed.

[6] Sudira, Hanif, Alifiannisa Lawami Diar, and Yova Ruldeviyani – Numerous researchers are using text data to do sentiment analysis on data generated by social media and internet trends to assess user satisfaction with particular brands. increase. In order to quantify user satisfaction with digital payment services in Indonesia (Go-Pay, Ovo, LinkAja), this study looked at Instagram sentiment analysis results using Instagram comments as the text data type. I was. Naive Bayes, K-Nearest Neighbors (KNN), and Customer Satisfaction Theory were used to develop survey models utilizing sentiment analysis categorization methods. The findings of this study, which used 3800 training and 200 test sets with 20k fold cross-validation, demonstrate that GO-PAY has a wide variety of clients, including both satisfied and dissatisfied ones.

[7] Roshini, T – Social media is one of the most important aspects of our daily lives. What exactly is social media to you? Social media are simply websites or applications used to create and share content on social networks. A recent study found that the average person spends about 142 minutes a day on social media. That number may seem low, but it could be even higher when you consider how many people are addicted to social media. In recent years, the average amount of time spent on social media has increased from 100 minutes to 142 minutes per day. People all over the world spend part of his day on his social media platforms, but it's hard to tell if such platforms are a plus or a minus for humanity. Most people argue that social media is a complete waste of time, but recent studies have concluded that people who use social media experience less stress. A woman who used social media several times in her day was 21% less stressed than a woman who had no interest in social media at all. However, there is much debate about its adverse effects in humans as well. One of the most common is the fact that people become so engrossed in social media that they forget its value or even how to interact with someone face-to-face. I wasn't particularly interested in the impact of social media, but wanted to know what kinds of social media platforms people prefer. Given that we live in a digital age where all data on the internet can be easily tampered with, we wanted to know how safe people feel on each social media platform. With that in mind, I decided to learn more about how people approach social media platforms. ⌋

[8] Aufar, Mohammad, Rachmadita Andreswari, and Dita Pramesti – People frequently share their views and opinions on social media. Social media encourages all interested parties to take part, freely provide their opinions, remark, and freely exchange information. YouTube is still a popular social media platform for him in the neighborhood. YouTube is frequently used to advertise items since it has so much video material. A public review of a Nokia product is the case study of a researcher. Nokia was one of the top consumer stocks and a significant product on the worldwide market, but it declined in 2013. Sentiment towards numerous Nokia goods was broken down into positive, negative, and neutral attitudes. made by categorizing This study will allow you to decide whether the product quality is high overall. However, comments with a majority are classified as neutral.

[9] Bahrawi, Nfn – Every day, the internet, including forums, blogs, social media, and review sites, receives billions of text-based records. With the use of sentiment analysis, unstructured data may be made more organized and converted into useful information. Data may show how people feel or think about many things, including goods, services, charity, politics, and other things. Natural language processing's (NLP) area of sentiment analysis develops textual opinion recognition and extraction techniques. The objective is to extract the sentiment or "feeling" from a group of words or sentences. To obtain the text for sentiment analysis, also known as "opinion mining," a data mining procedure is unavoidably necessary.

[10] Karthika, P., R. Murugeswari, and R. Manoranjithem – Sentiment analysis is the process of extracting emotions or viewpoints from a text. Reviews on social media produce a significant quantity of emotional data. From user reviews, sentiment analysis is performed to ascertain the customer's viewpoint. Online shopping is becoming more and more common because of its convenience, affordability, and quick delivery. We use various brand ratings and reviews to see how customers genuinely feel about our items in the cutthroat e-commerce industry of today. The feedback environment is made to assist customers in making the best product choices and to direct businesses in enhancing product features in response to consumer desire. Customers have trouble locating precise ratings for the particular features of the product they wish to buy.

[11] Bhowmik, Pankaj – Social Media (SM) becomes the next level of journalism as it reflects all the viral content, social issues and current conditions around us. SM site Facebook reached the most popularity in Bangladesh. Today, people prefer to share their opinions on issues on SM than on any other platform. With these results in mind, this study (Bangladesh perspective) proposed a system to analyze Facebook data and classify and recognize the most frequent problems people face at any given time. The issues are grouped into 12 major classes dealing with socio-economic aspects of Bangladesh. A dataset containing public Facebook posts (and comments) is manually collected and tagged with classes. With efficient data preparation, a feature vector is constructed with TF-IDF weights, then chi-square is applied to select the most strongly correlated features. Four machine learning algorithms are used as supervised classifiers. B. Logistic Regression, Support Vector Machines, Multinomial Naive Bayes, Random Forest. In addition, hyperparameter tuning of these classification algorithms is performed using grid search and 10-fold cross-validation is applied to select the best candidate models. Finally, the selected logistic regression model is built with a classification accuracy guarantee of >93%. A classification report shows important information about the growth of the problem.

[12] Fan, Weiguo, and Michael D. Gordon – This essay is intended for (social science) academics who wish to examine the wide variety of social media now accessible. gives a thorough review of the software applications for wikis, blogs, newsgroups, chats, and news feeds as well as social networking platforms. An introduction to sentiment analysis, data cleansing, storage, and social media scraping are also given for completeness. The study presents a methodology and a critique of social media tools in addition to being largely an overview. Due to the availability of web-based application programming interfaces (APIs) supplied by Twitter, Facebook, and news services, analyzing social media, particularly Twitter feeds for sentiment analysis, has grown in importance as a research and commercial endeavor. As a result, data services, scraping, and analytics software tools have "exploded."

[13] Hama Aziz, Roza Hikmat, and Nazife Dimililer - One of the most difficult and significant study areas in text mining is sentiment analysis. Detecting the polarity of emotions on social media and categorizing them as good or negative is the first step in sentiment analysis. The majority of earlier research on sentiment analysis concentrated on designing classifiers for particular data sets using supervised machine learning methods and feature extraction techniques. To accurately categorize the emotion polarity of the text, it is crucial to select the most suitable classification technique. He therefore suggests a novel ensemble classifier strategy in this paper that integrates numerous feature sets with ensemble classification by integrating several weak base classifiers into a single ensemble classifier. Among these feature sets are Words in a Bag.

Chapter 3
SYSTEM DESIGN & DEVELOPMENT

3.1 Analytical

3.1.1 Content Metrics

You may examine the material you post on social media and see particular trends by tracking content analytics. Maybe just 25% of your posts are video-based, while maybe 75% of them are image-based. How does this strategy function?

You can instantly discover a split of shared original content and content generated by others by adhering to the 50/50 guideline for content. B. Industry-specific articles that you link to in your posts or reshared posts that influencers create for your product.

This category also includes a wide range of additional information, such as client testimonials, reviews, guest blogs, publications, and industry infographics. You may attract the interest of prospective influencers, show off your knowledge, and generate favorable sentiments for your company by posting this sort of material.

3.1.2 Timing metrics

The best times to publish are those when the target audience is most engaged on social media platforms, either during the day or on specific days of the week. Businesses can determine the optimal times to release information in order to increase engagement by examining engagement rates for various posting hours.

Time spent on site: This statistic reveals how long, on average, customers stay on a company's website after clicking a social network link. Businesses may determine which social media networks are bringing the most interested visitors to their website by tracking time spent there.

Response time: This indicator gauges how quickly a company responds to questions or comments from clients on social media sites. Businesses can enhance customer satisfaction by minimizing response times.

Post frequency is a metric that identifies how many posts a company makes on social networking sites in a specific time frame. Businesses can determine the best posting frequency to increase engagement by examining post frequency and engagement rates.

Time to publish: This indicator gauges how quickly a company posts material on social media after it has been developed. Businesses may increase their agility and reactivity on social media platforms by cutting the time it takes to publish.

Seasonal patterns: This statistic examines the relationship between social media participation rates and seasonal patterns, such as holidays or significant events. Businesses may provide content that resonates with their target audience and encourages participation by recognizing seasonal trends.

3.1.3 Audience Metrics

Age, gender, location, education level, and other demographic information are included in this measure to help firms determine the characteristics of their target audience. Businesses can better focus their social media marketing to appeal to their target audience by looking at demographic data.

The size of the audience is determined by the number of fans or followers a company has on social media. Businesses may understand how their social media presence is developing over time and spot chances for audience expansion by analyzing audience numbers.

Measures the degree to which a target audience actively interacts with a company's social media material. Likes, remarks, shares, and other forms of interaction are included. Businesses can identify the content categories that engage their target audience the best by analyzing audience engagement.

Influencers are the most influential users in a company's social media audience, according to this statistic. Influencers are people with a significant following who may assist businesses in reaching a larger audience. Businesses can develop focused campaigns that take advantage of social media influencers' ability to reach a larger audience by identifying influencers.

Customer behavior: This metric tracks the interests, preferences, and purchase habits of customers on social media sites. Businesses can design tailored social media campaigns that increase engagement and revenue by studying client behavior.

Audience sentiment: This indicator assesses the attitude of online discussions about a company or its goods. It aids companies in understanding client mood and locating problem areas.

3.1.4 Listening Metrics

Mentions: This indicator counts the instances in which a company is discussed on social media. Businesses can determine their social media reach and track client sentiment by looking at mentions.

Share of voice is a metric that quantifies how much of a company's industry- or niche-specific social media conversation it contributes to. Businesses can find opportunities to expand their share of the conversation and raise brand awareness by comparing their share of voice to that of their rivals.

Sentiment analysis: This statistic assesses the mood of online discussions about a company or its goods. It aids companies in understanding client mood and locating problem areas. Customers who express positive sentiment are satisfied, whereas those that express negative sentiment have problems that need to be fixed.

Hashtag performance assesses the effectiveness of hashtags associated with a company or its goods. Businesses may determine which hashtags are driving the greatest engagement by analyzing hashtag performance and incorporating those into next social media efforts.

Engagement of social media influencers who discuss or endorse a company or its products is measured using this metric. Businesses may find the most productive influencers to work with and monitor the results of influencer marketing campaigns by analyzing influencer engagement.

Brand reputation is a metric that assesses a company's general online reputation. Businesses can improve their brand image by monitoring brand reputation to find areas of strength and weakness and make data-driven decisions.

3.1.5 Competitor metrics

Follower growth is a metric that tracks how quickly a competitor's social media audience is expanding over time. Businesses may discover which rivals are gaining momentum on social media platforms and create strategies to compete successfully by analyzing follower growth.

Engagement rate gauges the quantity of likes, comments, shares, and other engagements that a competitor's social media post is receiving. Businesses may detect which competitors are connecting with their target audience by analyzing engagement rates, and then design plans to increase interaction on their own social media channels.

Share of voice quantifies the percentage of social media conversations about a particular subject or business that bring up a rival. Businesses can determine which rivals are more effective by looking at share of voice

Content analysis: This statistic examines the many kinds of content, such as text, photographs, videos, and infographics, that rivals post on social media sites. Businesses can establish strategies to produce similar content by analyzing the content of their competitors to determine which forms of content most engage their target audience.

Sentiment analysis: This metric gauges the attitude of online discussions involving a rival or its goods. Businesses can use sentiment analysis to determine which rival brands are viewed favorably or unfavorably by their target market and then devise plans to enhance their own brand perception.

This indicator tracks the amount of money a rival spends on social media advertising. Businesses may determine which competitors are spending a lot of money on social media advertising by analyzing advertising expenditure.

3.1.6 Engagement metrics

Likes and responses are metrics that track how frequently people click on the "like" or "reaction" buttons in response to a company's social media content. High amounts of likes and reactions show that the material is engaging and resonating with the target audience.

Comments: Comments are a metric that counts the number of times that people leave comments on a company's social media posts. High amounts of comments show that the audience is engaging with the material and having conversations about it.

Clicks are a metric that counts the instances in which customers connect with a company's social media content by clicking a link to a landing page or website. High clickthrough rates show that the content is bringing visitors to the website and producing leads or sales.

Influencer engagement is a metric that assesses the degree of interaction between influencers or brand ambassadors and the social media content of a company. High levels of influencer interaction show that the article's message is getting across to influential people in the field and creating buzz among their followers.

Engagement rate: The percentage of users who interact with a company's social media material through actions like likes, comments, shares, and clicks is measured by engagement rate, a metric.

3.1.7 Social traffic metrics

Referral traffic counts the number of people who go to a website of a company via social media. Businesses may determine which social media sites are directing the most traffic to their website by analyzing referral traffic.

The percentage of people that click on a company's social media posts and visit the website after doing so is known as the "click-through rate" (CTR). A high CTR shows that the content is increasing website traffic and generating leads or sales.

The percentage of visitors that leave a website after just reading one page is known as the "bounce rate." High bounce rates are a sign that the website content might not be interesting or relevant to the intended audience. Businesses may pinpoint areas for development by examining bounce rates.

Time on site: This statistic reveals how long, on average, customers stay on a company's website after clicking a social network link. Businesses may determine which social media networks are bringing the most interested visitors to their website by tracking time spent there.

Conversion rate: The percentage of visitors to a website that complete a desired activity, such completing a purchase or filling out a form, is known as the conversion rate. Businesses may determine which social media channels are most successful at producing leads or sales by looking at conversion rates.

Revenue: The amount of money a company makes from social media traffic is measured by revenue. Businesses may determine which social media channels are most successful in generating money by analyzing revenue.

3.1.8 Branding metrics

Brand mentions: This statistic counts the instances in which a company's name is discussed on social media. Businesses may assess the success of their brand marketing activities and spot chances to interact with customers and address their feedback by monitoring brand mentions.

Brand sentiment is a statistic that assesses the general attitude people have towards a company's brand on social media, whether that attitude be good, negative, or neutral. Businesses may assess the success of their branding initiatives and pinpoint areas for development by analyzing brand sentiment.

Reach: The total number of individuals who have viewed a company's social media material is measured by reach. Businesses may assess the efficacy of their social media content by analyzing reach, and they can also spot possibilities to increase the exposure of their brand.

Share of voice is a metric that compares a company's brand mentions in online conversations to those of its rivals. Businesses may assess their brand's market position and spot possibilities to take a bigger market share by monitoring share of voice.

Engagement rate: The percentage of people who like, comment on, or share a company's social media material is known as the engagement rate. Businesses may assess the success of their social media content by looking at engagement rates, and they can also spot chances to increase interaction with their target market.

Brand loyalty: Brand loyalty gauges how devoted consumers are to a company's name and goods. Businesses may assess the success of their efforts to promote their brands by analyzing brand loyalty.

3.1.9 Social media management metrics

Follower growth rate: This indicator gauges how quickly a company's followers on social media are increasing over time. Businesses may evaluate the success of their social media content and audience engagement efforts by looking at follower growth rate.

Engagement rate: The percentage of followers who like, comment on, and share a brand's social media material is known as the engagement rate. Businesses may determine the efficacy of their content and modify their social media strategy by looking at engagement rate.

Reach: The quantity of distinct individuals who have viewed a company's social media material is measured by reach. Businesses may determine how well they are reaching their target audience by analyzing reach, and then modify their content distribution strategy as necessary.

Impressions: Impressions track the total number of times consumers have viewed a company's social media material. Businesses may determine how well they are spreading their information by analyzing impressions, and then modify their content strategy appropriately.

Response time: This indicator gauges how quickly a company responds to questions or comments from clients on social media sites. Businesses may enhance engagement rates and boost customer satisfaction by speeding up response times.

Brand sentiment assesses the general tone of social media talks regarding a company's brand, including whether they are good, negative, or neutral. Businesses may enhance their brand impression by analyzing brand sentiment and modifying their social media strategy as necessary.

3.1.10 Goal and summary metrics

Reach: The number of users who have viewed a company's social media material is measured by reach. This measure is essential for assessing how well social media efforts do at raising brand awareness.

User involvement with a company's social media material, such as likes, comments, and shares, is measured as engagement. High engagement rates show that the material is creating interest and resonating with the target audience.

Followers: The number of people who subscribe to a company's social media accounts is measured by its followers. This statistic is crucial for assessing how social media audiences are expanding and how well social media initiatives are doing at raising brand recognition.

Impressions: Impressions track how frequently a company's social media material is seen. This measure is essential for assessing how well social media initiatives do at raising brand awareness and piquing interest.

Sentiment: Sentiment analyses the general emotional undertone of user comments and responses to social media material from businesses. favorable sentiment shows that the target audience is connecting with the material and developing favorable sentiments for the brand.

ROI: The financial return on a company's social media effort is measured by return on investment, or ROI. Businesses may assess the success of their social media strategy in producing income by comparing social media cost with revenue from social media campaigns.

3.2 Computational

3.2.1 Reach

Organic reach: The number of users who have viewed a company's social media material without any paid advertising is referred to as organic reach. The effectiveness and relevancy of the material, the frequency of publishing, and audience participation all have an impact on organic reach.

sponsored reach: Paid reach counts the number of users who have viewed a company's social media posts as a result of sponsored advertising on social media, for example. Targeting criteria, ad placement, and ad content are just a few examples of the variables that affect paid reach.

Impressions: Impressions count the instances in which consumers have seen a company's social media material. If a person views the social media material of a company more than once, the user may create numerous impressions.

Unique reach: The quantity of distinct people who have viewed a company's social media material is measured by unique reach. Unique reach sheds light on how well social media efforts perform in terms of reaching new consumers and raising brand recognition.

Reach rate: The proportion of a company's social media audience that has viewed a certain piece of content is measured by reach rate. The timing of publishing, the caliber and applicability of the material, and audience participation all have an impact on reach rate.

Viral reach: Viral reach counts the users who have viewed a company's social media posts as a consequence of other users sharing or reposting them. Viral reach offers perceptions into how well social media strategies work to generate word-of-mouth advertising and

3.2.2 Impressions

Impressions are the total number of times a specific piece of content, such a tweet, post, or video, has been seen to a user. Impressions just show that the information has been viewed rather than necessarily implying that the user has interacted with it.

Reach vs. impressions: Although reach and impressions relate to somewhat distinct measures, they are frequently used synonymously in social media analytics. While impressions relate to the total number of times the material has been displayed on a user's screen, reach refers to the total number of unique users that have viewed a certain piece of content. In other words, if a person views the information more than once, the user may credit for numerous impressions.

Importance: Businesses should be aware of impressions since they may use them to gauge how well their social media ads are reaching their intended demographic. High impressions might show that the brand is generating interest in the target demographic and that the content is connecting with them. Low impressions could be a sign that the content needs to be updated since it isn't reaching the target demographic.

Limitations: User engagement with the material, such as like, commenting, or sharing, is not always shown by impressions. Additionally, bot traffic and people that rapidly scroll through the material without really seeing it might exaggerate the number of impressions. Therefore, organizations should combine impressions with other data, including engagement and click-through rates, to have a better understanding of Social Media Analytics.

3.2.3 Audience growth rate

The percentage rise or decrease in the number of fans or followers on a company's social media account over a specific time period is known as the audience growth rate.

Calculation: Companies may use the calculation (New Followers - Lost Followers) / Total Followers x 100 to get their audience growth rate. The audience growth rate would be ((200-50)/1,000) x 100 = 15% if a company had 1,000 followers at the beginning of the month, 200 new followers, and 50 followers lost.

Importance: The audience growth rate is a crucial indicator for businesses since it shows how well their social media strategy is working to draw in and keep followers. A strong audience growth rate shows that the target audience is engaging with the company's social media material.

The quality of social media content, frequency of posting, interaction with followers, usage of hashtags and keywords, and overall social media strategy are some of the variables that might impact an organization's audience growth rate.

How to increase audience growth rate: Companies can increase audience growth rate by creating high-quality social media content that appeals to the target audience, posting frequently and consistently, interacting with followers, utilizing pertinent hashtags and keywords, and continuously evaluating and fine-tuning their social media strategy.

3.2.4 Engagement rate

Total Engagements / Total Followers or Impressions x 100 = Engagement Rate

The engagement rate would be determined as follows, for instance, if a company had 10,000 Instagram followers and a post earned 1,000 likes and 100 comments:

$((1{,}000 + 100) / 10{,}000) \times 100 = 11\%$ is the engagement rate.

A high percentage of engagement shows that the material is interesting and engaging with the target audience, which may raise brand visibility, website traffic, and eventually income. A low engagement rate could be a sign that the material is not connecting with the intended audience and has to be changed in order to be more effective.

3.2.5 Amplification rate

The number of shares or retweets a company's social media material receives in relation to the number of followers is measured by a social media analytics indicator called amplification rate. It is a crucial statistic for assessing how well social media initiatives and content work to raise brand recognition and broaden a company's online presence. Amplification rate in social media analytics is calculated as follows:

Amplification rate is computed as a percentage by multiplying the total number of shares or retweets a piece of information receives by the total number of followers. The equation reads as follows:

(Total Shares or Retweets / Total Followers) $\times$ 100 equals the amplification rate.

An improved brand recognition and website traffic might result from content that is engaging the target audience and producing a high amplification rate. A low amplification rate could be a sign that the message is not getting through to the intended audience and has to be changed in order to be more successful.

Businesses may raise their amplification rate in a number of ways, including by producing shareable and educational material that benefits their target audience, leveraging pertinent hashtags to increase exposure, and interacting with their followers to promote shares and retweets. Businesses may assess the success of their social media strategy and make data-driven choices to enhance their social media performance by monitoring amplification rate in social media analytics.

3.2.6 Virality rate

The virality rate is computed by dividing the total number of interactions with a post by the total number of views, then multiplying the result by 100 to produce a percentage. The equation reads as follows:

(Total Engagement / Total Reach) times 100 equals the virality rate.

The virality rate would be determined as follows, for instance, if a company had 10,000 Facebook followers and a post earned 1,000 likes, comments, and shares with a reach of 20,000 people:

(1,000 / 20,000) x 100 = 5% is the virality rate.

When a piece of content has a high virality rate, it is highly shareable and connecting with the intended audience, which expands its reach and raises brand exposure. A low virality rate might mean that the material is not connecting with the intended audience and has to be changed in order to be more effective.

Businesses may raise their virality rate in a number of ways, including by producing highly shareable and educational material that benefits their target audience, leveraging pertinent hashtags to increase exposure, and interacting with their followers to promote shares and comments. Businesses may assess the efficacy of their social media strategy and make data-driven decisions to enhance their social media performance by measuring virality rate in social media analytics.

3.2.7 Video Views

The number of times a video has been seen on a social media network is counted as video views, a social media analytics indicator. It is a crucial indicator that companies should monitor in order to assess how well their video marketing and content are doing. The way social media analytics calculate video views is as follows:

Depending on the social media platform, multiple metrics may be used to measure video views. For instance, on Facebook, a user must watch at least three seconds of a video before it counts as seen, but on YouTube, a user must watch at least 30 seconds of a video before it counts as viewed, or the full video must be watched, whichever comes first.

By visiting the insights or analytics area of their social media platform, businesses may measure video views in social media analytics. In addition to other data like viewer demographics, engagement rate, and retention rate, they may check the overall number of views for a single video. Businesses may assess the success of their video content by examining these indicators and can enhance the performance of their social media channels by making data-driven choices.

Businesses may optimize their video content to boost video views on social media by producing interesting and educational movies, employing captions or subtitles, including interactive components like polls or quizzes, and marketing their videos through targeted social media advertising.

3.2.8 Completion rate for videos

The percentage of viewers that watched a video through to the end or up to a certain point is measured by the completion rate for videos, a social media analytics indicator. It is a crucial indicator for assessing how well advertising and video content engage the target audience and raise brand recognition. The social media metrics for video completion rate looks like this:

The number of individuals who watched a video through to the end or up to a certain point is divided by the total number of people who started watching it to obtain the completion rate, which is then multiplied by 100 to produce a percentage. The equation reads as follows:

(Number of People Who Completed the Video / Total Number of People Who Started Watching the Video) x 100 represents the completion rate for videos.

The completion rate of a video, for instance, would be determined as follows if it had 10,000 views on YouTube and 5,000 of those viewers watched the entire thing: A video's completion rate is equal to (5,000/10,000) × 100, or 50%.

A high completion rate for videos shows that the material is interesting and resonates with the intended audience, which raises user engagement and brand recognition. A low completion rate may be a sign that the video's message is not getting through to the intended audience and that changes are needed to increase effectiveness.

Businesses may increase their video completion rates in a number of ways, including by producing high-quality video content that benefits their target audience, tailoring the duration of the video to the platform and audience preferences, and advertising the video to the right people. Businesses may assess the success of their video marketing efforts and make data-driven decisions to enhance their social media performance by monitoring the completion rate for videos in social media analytics.

3.2.9 Customer satisfaction (CSAT) rate

A social media analytics indicator called customer satisfaction rate gauges how satisfied customers are with a company's goods, services, or customer assistance on social media platforms. It enables companies to monitor how consumers view their brand and pinpoint areas where customer service needs to be improved. Customer satisfaction rating in social media analytics is calculated as follows:

The mood of consumer comments on social media sites is often analyzed to determine the level of customer happiness. Sentiment analysis is the technique of examining the voice and feelings of customer reviews, whether they are favorable, unfavorable, or neutral, in order to comprehend how people generally feel about a company. Natural language processing (NLP) algorithms are used by social media analytics platforms to analyze and categories client comments.

Businesses may determine the customer satisfaction rate once the sentiment of customer feedback has been identified by dividing the amount of positive customer feedback by the total number of customer feedback, multiplying by 100 to obtain a percentage. The equation reads as follows:

(Positive Feedback / Total Feedback) times 100 equals the customer satisfaction rate.

For instance, the customer satisfaction rate would be determined as follows if a company got 200 client comments on social media and 150 of them were favorable:

Customer Service Rate is (150 / 200) times 100, or 75%.

An organization that consistently meets or exceeds client expectations has a high customer satisfaction rating, which may boost patronage, encourage effective word-of-mouth advertising, and promote customer retention.

Businesses may discover areas for improvement and take action to address customer complaints and enhance customer satisfaction by measuring customer satisfaction rate in social media analytics.

3.2.10 Net promoter score (NPS)

A social media analytics statistic called Net Promoter Score (NPS) gauges customer loyalty and satisfaction by asking customers how likely they are to suggest a company's goods or services to others. Businesses may track their performance and enhance customer experience using the NPS score as a useful indicator. NPS in social media analytics functions as follows:

Businesses commonly ask consumers one question to get the NPS score: "On a scale of 0 to 10, how likely are you to recommend our product/service to a friend or colleague?" They are divided into one of three groups based on the customer's response:

Customers who give a product or service a score of 9 or 10 are regarded to be loyal, pleased customers who are likely to suggest it to others.
Passives: clients who provide a score of 7 or 8 and who are deemed satisfied but not necessarily devoted clients and who would hesitate to suggest the good or service.
Detractors are consumers who provide a score between 0 and 6 and are deemed disgruntled and dissatisfied; they may prevent others from utilizing the product or service.
The NPS score ranges from -100 to +100 and is determined by subtracting the percentage of critics from the percentage of promoters. The equation reads as follows:

Promoters: % Detractors: % NPS Score

The NPS score would be determined, for instance, as follows if a company polls 100 consumers and finds that 60 are promoters, 20 are passives, and 20 are detractors.

NPS Score is (60/100) times 100 - (20/100) x 100, which equals 40.

An organization with a high NPS score has satisfied and loyal customers, which can boost brand loyalty and encourage repeat business. In order to boost customer happiness and loyalty, a company may need to improve its customer experience if its NPS is low.

Businesses may analyze their performance over time and pinpoint areas for customer experience improvement by measuring NPS levels in social media analytics. Additionally, businesses can raise their NPS score by keeping an eye on and responding to customer feedback on social media platforms.

3.2.11 Click through rate (CTR)

The success of a company's social media efforts and content in generating traffic to its website or other landing pages is gauged by the click-through rate (CTR), a social media analytics indicator. CTR measures the proportion of viewers that clicked on a link or call-to-action button inside a post or advertisement compared to the total number of viewers. The social media analytics click-through rate works as follows:

By dividing the number of clicks on a link or call-to-action button by the number of views or impressions of the post or advertisement, and then multiplying the result by 100, the click-through rate is computed. The equation reads as follows:

(Total Clicks / Total Impressions) x 100 is the CTR formula.

The CTR would be determined as follows, for instance, if a company runs a Facebook advertisement that obtains 10,000 impressions and 100 clicks:

CTR = (100 / 10,000) x 100 = 1%

A high click-through rate suggests that the article or advertisement is successful in bringing visitors to the company's website or landing page. A poor click-through rate may be a sign that the material or advertisement needs to be improved, maybe with stronger writing or a more obvious call to action.

Businesses may increase their click-through rate in a number of ways, including by writing clear and persuasive ad language, utilising arresting imagery, and providing a clear call-to-action button. Businesses may assess the success of their social media campaigns by monitoring click-through rate in social media analytics.

3.2.12 Conversion rate

A social media analytics indicator called conversion rate counts the number of visitors who visit a company's website or landing page after clicking on a social media post or advertisement. For companies that rely on social media to increase website traffic and revenue, it is a crucial measure. Here is how social media analytics' conversion rate functions:

A social media post or advertisement's conversion rate is computed by dividing the number of individuals who took the desired action by the total number of clicks, then multiplying the result by 100 to get a percentage. The equation reads as follows:

(Number of Conversions / Total Clicks) x 100 = Conversion Rate

The conversion rate would be determined as follows, for instance, if a company received 10,000 hits on a social media ad and 100 individuals carried out the required action (such making a purchase or completing a lead form):

Conversion Rate is equal to (100/10,000) × 100, or 1%.

In contrast, a low conversion rate might mean that the campaign is not connecting with the target demographic or that there are problems with the landing page or website. A high conversion rate suggests that the social media campaign is successfully generating website traffic and sales. Businesses may assess the success of their social media initiatives and make data-driven choices to enhance their social media performance and boost conversions by analyzing conversion rate in social media analytics.

Businesses may concentrate on producing highly targeted social media campaigns with obvious calls to action, appealing imagery and language, and optimizing the landing page or website for the desired action in order to increase conversion rates. They may also make modifications to increase the success of their ads by using social media analytics to see which ones are leading to the most conversions.

3.2.13 Cost per click (CPC)

The cost of each click produced by a company's social media advertising campaign is measured using the social media analytics statistic known as cost per click (CPC). It is an important indicator for assessing the potency and return on investment of social media advertising. Here is how social media analytics' cost per click functions:

The cost per click is computed by dividing the entire cost of an advertising campaign on social media by the total number of clicks the campaign produced. The equation reads as follows:

Total Cost / Total Clicks (CPC)

The cost per click would be computed as follows, for instance, if a company conducts a Facebook advertising campaign that costs $100 and receives 200 clicks:

CPC = $100 / 200 = $0.50

A lower cost per click (CPC) means that the advertising campaign is producing clicks at a greater ROI. A high CPC may be a sign that the advertising campaign is not effectively generating clicks and that changes are needed to increase effectiveness.

Businesses may increase their CPC in social media advertising in a number of ways, including by optimizing ad targeting, utilizing extremely relevant and compelling ad content, utilizing relevant keywords, and establishing reasonable bid pricing. Businesses may assess the success of their social media advertising campaigns and make data-driven decisions to enhance their advertising performance by measuring CPC in social media analytics.

3.2.14 Cost per thousand impressions (CPM)

A social media advertising statistic called cost per thousand impressions (CPM) calculates the price of showing an advertisement to 1,000 users of a social media network. In social media advertising, CPM is a commonly used statistic that is especially helpful for evaluating the cost of advertising across various platforms or campaigns. The CPM in social media analytics works as follows:

The cost per thousand impressions (CPM) is computed by dividing the entire cost of a social media advertising campaign by the total number of impressions it produced. The equation reads as follows:

CPM is equal to 1,000 times (total cost / total impressions).

The CPM would be computed as follows, for instance, if a company paid $1,000 on a Facebook advertising campaign that resulted in 100,000 impressions:

CPM = ($1,000 / 100,000) x 1,000 = $10

In other words, this Facebook advertising campaign costs $10 per 1,000 impressions.

Because it enables companies to assess the effectiveness and cost-effectiveness of their social media advertising campaigns, CPM is useful in social media analytics. Businesses can spot patterns in the cost of advertising on various platforms or using various tactics by measuring CPM over time. Then, with the help of this information, marketing campaigns can be improved while spending less overall.

In order to obtain a more complete picture of the success of a social media advertising campaign, it is vital to keep in mind that CPM is only one indicator in social media advertising and should be assessed alongside other metrics like click-through rate, conversion rate, and return on investment.

3.2.15 Social share of voice (SSoV)

Social share of voice (SSoV) is a social media analytics statistic that compares a brand's proportion of online discussions or mentions to that of its rivals in a certain market or sector. It is employed to assess a company's social media presence and reputation in comparison to its rivals. The social share of voice in social media analytics works as follows:

In order to calculate social share of voice, divide the total number of online mentions or talks about all brands in the same market or industry by the total number of mentions or conversations about all brands in that market or industry, then multiply the result by 100 to get a percentage. The equation reads as follows:

(Brand Mentions / Total Industry Mentions) x 100 = Social Share of Voice

The social share of voice would be determined as follows, for instance, if a certain brand received 500 mentions in a given month and the whole industry received 5,000 mentions within the same time period:

(500 / 5,000) x 100 = 10% is the social share of voice.

A strong social share of voice demonstrates that the business is actively engaging in social media conversations and creating a lot of buzz, which may raise brand recognition and enhance consumer engagement. A low social share of voice might mean that a brand is not as well-known or well-liked as its rivals, necessitating changes to the social media strategy.

Businesses can increase their social share of voice in a variety of ways, including by producing highly shareable and educational content that benefits their target market, interacting with their followers to promote conversations and mentions, and using social listening tools to track and examine online discussions about their brand and rivals. Businesses may assess the success of their social media strategy and make data-driven choices to enhance their social media presence and reputation by monitoring social share of voice via social media analytics.

3.2.16 Social sentiment

A social media analytics statistic called "social sentiment" gauges the general emotional tone and mood of online discussions centered on a given brand, item, or subject. It is determined by scrutinizing the wording, keywords, and context of social media postings and comments, and can be either good, negative, or neutral. What social sentiment looks like in social media analytics is as follows:

The language of social media postings and comments are examined to determine the social sentiment, which is then expressed as a sentiment score for each piece of material. The language, tone, and context of social media material are identified by sentiment analysis tools using natural language processing (NLP) and machine learning algorithms to ascertain whether the information is good, negative, or neutral.

If a company receives 100 mentions on social media, for instance, sentiment analysis systems may evaluate each mention's language, tone, and context and award it a sentiment score. The total social sentiment for the brand or product may then be calculated by averaging the sentiment scores.

A brand or product with a high positive sentiment score will likely have a favorable overall perception, whereas one with a high negative sentiment score would likely have a bad overall perception. The general emotion is neither good nor negative, according to a score of neutral sentiment.

Businesses may utilize social sentiment in social media analytics in a variety of ways to enhance their social media performance, including:

Monitoring social sentiment in real-time to spot possible problems or emergencies and act fast in the case of bad emotion.

Tracking the success of marketing and social media operations by analyzing social sentiment over time.

Spotting patterns and trends in social sentiment may help you understand the preferences, requirements, and behavior of your customers.

Businesses may assess the success of their social media strategy and make data-driven choices to improve their social media performance and boost their overall brand reputation by monitoring social sentiment through social media analytics.

3.3 Experimental

Additionally, social media sites give users the chance to easily perform real-world field studies. The influence of the therapy on subsequent social media activity is examined in this scenario where individuals are randomly allocated to receive (or not receive) a treatment. By completely eliminating demand effects, this method enables researchers to preserve causal conclusions while maintaining ecological plausibility. This is due to the fact that Subject 4464 uses social media in a real-world manner and is typically not aware that he is a participant in an experiment. This method also enables researchers to analyze people who choose not to participate in conventional investigative tests, such as conspiracy theorists. To demonstrate what is possible and how it may be done, consider the following scenarios. Examining the relationship between account attributes and a user's propensity to engage with an experimental account is one option. One experiment, for instance, had a politically diverse sample of Twitter users follow Republican or Democratic researcher accounts. were given to users at random, and cross-party accounts were roughly three times more likely to be tracked. Studying how processing impacts later social media activity is another option (e.g., 4464 sharing, liking different types of content, etc.). Delivering experimental therapies by private messaging is one method, which has the benefit of guaranteeing that only individuals with a certain treatment condition get that therapy. Platform-specific restrictions apply while sending private messages. For instance, Twitter only permits users who follow your account to send her private messages. As a result, in the experiment (Pennycook, Epstein, et al., 2021), we increased the number of people who followed us by following many Twitter users who had recently published political news articles. About 10% of these people followed us back. Her accounts were found by researchers. These followers of the researcher received a secret Direct Message from her account. We randomly assigned followers to different treatment appointments to allow for causal inference of; This treatment improved the quality of the news sources that the content user retweeted within her 24 hours. Her message asked them to rate the accuracy of one of her non-political articles. Ads may also be used to give individualized treatments, but it is frequently challenging to identify who has seen which ads and when. To improve ad delivery, random attribution can be employed. Platform algorithms have the ability to undermine it. Researchers can interact with individuals publicly in addition to privately by tagging them in posts and tweets, and by tweeting back to them. This strategy has the advantage of not restricting treatable users to those who follow the researcher's account, at least on Twitter, but it may expose users for one ailment to therapy for another, according to the researcher. There are negatives. In one such experiment (Munger, 2017), tweets were responded to with racial slurs using her accounts with various follower counts and races. The most effective sanctions for lowering subsequent usage of racial slurs among sanctioned users were those from white male high-follower accounts (in-group) researchers. In a different one of his tests, he found tweets with links to bogus news articles that fact-checkers had disproved, and the researcher account responded to these tweets with the appropriate information. The impact of public corrections was comparable to that of following high-quality, partisan remarks by users, in contrast to the effect of direct communications with an accuracy advantage (Pennycook, Epstein, et al., 2021). It was done to promote harmful information and vastly boost the number of retweets of the tweet. Additionally, you may mix private and public contacts. Using a non-profit advocacy organization that urged users to sign and circulate petitions as an example, Coppock et al. (2016) compared the efficacy of private and public messaging and discovered that the former was much more successful. They also discovered that people who followed users who shared petitions were more likely to sign than people who followed users who did not share petitions, empirically proving the peer effect. However, social media experiments raise a number of useful and analytical issues that are not normally addressed in conventional research trials. These topics are covered in the online supplementary material.

3.4 Mathematical

Calculating Value of Impressions: If you want to calculate how much worth are the impressions you gained on social media, this is the way how to do it:

Impression Value = Total of all the impressions gained on Social Media (Twitter, Facebook, YouTube (Total views of all the videos), impressions from all other social media properties + Website Traffic X (multiply by) Average Industry (the amount that you must have paid or pay for your CMP campaigns for your brand) CPM(Cost per thousand Impression)

Formulas for Facebook and Twitter

Facebook

Specific Post Engagement Rate = No. of Likes + Comments + Shares on particular day / Total No. of fans on day when you posted X 100

Average Post Engagement Rate = No. of Likes + Comments + Shares on particular day / Total no. of Posts on that day / Total No. of fans on day when you posted X 100

Facebook Content Reach = Brand Reach + Creating Story x Total Friends Users who Engaged with your post

Weekly Page Engagement Rate = Total No. of Likes + Comments + Shares in a given week / Total No. of Fans during that given day

Weekly Community Interaction Rate for Facebook = Total No. of Likes + Comments + Shares by others to Posts made by Facebook users on Brands Page in that week/ Total no. of posts made by users in that week / No. of Fans of Brands Page in that week

Twitter

Specific Tweet Engagement Rate = Total of No. of Replies & RTs on that Tweet/ Total Number of followers on day when you tweeted X 100

Average Tweet Engagement Rate = Total of No. of Replies & RTs on that Tweet/ Total no. of Tweets on that day / Total Number of followers on day when you tweeted X (multiply by) 100

Twitter Content Reach = Brand Reach + Shares (RTs & Quotes) x Twitter Followers of Users who Retweeted your Tweet

Weekly Community Interaction Rate for Twitter = Total No. of RTs + Quote + Replies by others to Tweets made by Twitter users mentioning Brand’s Twitter Handle in that week/ Total no

3.6 Algorithms

Social media algorithms are ways to order postings in a user's newsfeed based on relevancy rather than when they were published. Based on the chance that a user will really desire to see a piece of material; social networks order the content that appears first in users' feeds. Most social media feeds used to show content in reverse chronological order until algorithms were introduced. In other words, postings from accounts the user is following that have been most recently updated were shown first. Facebook and Twitter both provide the ability to arrange the feed in chronological order. By default, social media algorithms use your use to determine what material to show you. For instance, Facebook and Twitter let you give priority to postings from your close friends and family in your feeds.

Chapter 4
EXPERIMENTS & RESULT ANALYSIS

4.1 Techniques

- **Natural Language Processing** – The interaction between computers and human (natural) languages is the focus of the field of (NLP), which combines computer science, artificial intelligence, and linguistics. More specifically, it is a computer process that generates natural language output or extracts meaningful information from natural language input.
- **Message Analysis** - Measuring different text message qualitative and quantitative aspects (unstructured data). These qualities include novelty, relevance, and feeling.
- **Opinion Extraction** - Opinion Extraction, also known as Sentiment Extraction or Opinion/Sentiment Extraction, is a field of study that aims to create automated systems to extract human opinion from documents published in natural language.
- **Scraping** is the process of gathering unstructured text data from social media and other websites. also known as web harvesting, web data extraction, and site scraping.
- **Sentiment Analysis** - Sentiment analysis is the practise of using text analysis, computational linguistics, and natural language processing to find and extract subjective information from sources.
- **Text Analysis** - consists of information retrieval (IR), lexical analysis to look at word frequency distributions, pattern recognition, annotation and tagging, information extraction, data mining techniques like linkage and association analysis, visualization, and predictive analytics.

4.2 Tracking performance

You may view just the stats that matter at any given time with the aid of customizable dashboards. Dashboards make connections and offer information that informs strategy. In order to prevent data gaps, piece together puzzles, and attempt to understand what the heck is going on, we export data from Facebook, Instagram, and other sources and examine each piece independently. Then try producing reports that concisely illustrate the most important data insights that affect your strategy. hard times Dashboards include all of the data that has been collected. Connect the dots, and the narrative your data presents is beautifully translated and delivered in reports that are suitable for sharing and presentations. nice timing.

4.3 Analytical method

The volume of user-generated data, the technological options, and the expertise for processing this type of data all continue to expand. This review suggests such a definition in order to progress the area of study because concise and understandable definitions are uncommon. In social media, data's timeliness is crucial. It is impossible to replicate or verify studies conducted on antiquated platforms like Google+ and Foursquare that have not gained traction in the European market. Additionally, precise and reproducible method descriptions are necessary if research is to be replicated. Even if a particular area of study interest may be found, not all authors are fully aware of the rationale behind creating a novel approach or the best framework for analyzing social media data. No. The developing subject of social media analytics needs further study. Future study techniques will also look into practical ways to make the identification, collection, processing, and analysis of social media data public. We also suggest emphasizing the moral and legal implications of exploiting social media data more.

4.4 Experimental method

Randomized field experiments employing social media sites have already stirred up plenty of controversy despite their relatively recent history. Public unwillingness to evaluate the effectiveness of (Meyer et al. 2019). It could be best to randomly assign two options for a thorough review. This type of test is added to the enormous quantity of data that the platform gathers about its users, documenting their behavior, preferences, and interactions with one another, in the world of precisely individualized social media. can. Experiments with social media users have been linked to previous abuses of power by scientists and researchers due to a lack of public knowledge of how this data is utilized and with whom it is shared. What has been done may not be shocking (for a broad discussion of experimental ethics, see, for instance, Teele's chapter in this volume). However, not every massive social media experiment has resulted in popular criticism. Meanwhile, a widely reported study on Facebook emotional contagion has caused a huge uproar and discussion about informed consent on social media (Kramer, Guillory, & Hancock 2014; Verma 2014). The likelihood that Facebook posts had a negative or positive emotional valence was altered at random by a research team made up of both Facebook scientists and outside academics (standard Measured using the dictionary-based His scoring method). If these changes were reflected in the emotional content of users' own posts, the researchers were particularly interested. However, Facebook's well publicized randomization experiment from two years ago resulted in hundreds of thousands of votes being cast in his 2010 midterm elections. Media coverage of this was primarily favorable (Bond et al. 2012). These varied responses, which are inherently anecdotal, might be the deciding factor in whether research makes a breakthrough advance in knowledge or merely repeats old, sinister scientific experiments. Due to these factors, several experimental investigations on the effects of social media are carried out in controlled settings using simulated stimuli. 7 These "off-platform" investigations utilized lab Participants or those being questioned are frequently shown extremely realistic simulations of situations. Results include hypothetical intent to share social media messages and perceived news headline accuracy.

4.5 Output at various stages

	Choice	A_follower_count	A_following_count	A_listed_count	A_mentions_received	A_retweets_received	A_mentions_sent	A_retweets_sent	A_posts	A_network_featu
0	0	228	302	3	0.583979	0.100503	0.100503	0.100503	0.362150	
1	0	21591	1179	228	90.456506	25.798292	5.709329	1.111159	5.176620	
2	0	7310	1215	101	25.503644	9.556347	5.361519	0.591206	3.589718	
3	0	20	7	2	7.690824	0.277306	1.331508	0.100503	2.830627	
4	1	45589	862	2641	148.854279	36.998884	27.881768	3.333492	23.861282	
...	...	...	...	...	...	...	...	...	...	
5495	0	41765	185	1356	1529.643058	282.858500	76.809514	5.392171	104.438625	
5496	1	112	243	5	1.445174	0.100503	0.100503	0.100503	0.603177	
5497	0	15385	673	747	55.993546	22.321945	6.946233	0.341936	6.503977	
5498	0	265258	209	551	631.915946	457.648550	5.460985	0.100503	7.498126	
5499	0	628	921	6	3.943848	0.618590	4.769930	0.870136	2.953237	

5500 rows × 23 columns

Fig.1 Train Dataset

```
Twitter_Train.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 5500 entries, 0 to 5499
Data columns (total 23 columns):
 #   Column               Non-Null Count  Dtype  
---  ------               --------------  -----  
 0   Choice               5500 non-null   int64  
 1   A_follower_count     5500 non-null   int64  
 2   A_following_count    5500 non-null   int64  
 3   A_listed_count       5500 non-null   int64  
 4   A_mentions_received  5500 non-null   float64
 5   A_retweets_received  5500 non-null   float64
 6   A_mentions_sent      5500 non-null   float64
 7   A_retweets_sent      5500 non-null   float64
 8   A_posts              5500 non-null   float64
 9   A_network_feature_1  5500 non-null   int64  
 10  A_network_feature_2  5500 non-null   float64
 11  A_network_feature_3  5500 non-null   float64
 12  B_follower_count     5500 non-null   int64  
 13  B_following_count    5500 non-null   int64  
 14  B_listed_count       5500 non-null   int64  
 15  B_mentions_received  5500 non-null   float64
 16  B_retweets_received  5500 non-null   float64
 17  B_mentions_sent      5500 non-null   float64
 18  B_retweets_sent      5500 non-null   float64
 19  B_posts              5500 non-null   float64
 20  B_network_feature_1  5500 non-null   int64  
 21  B_network_feature_2  5500 non-null   float64
 22  B_network_feature_3  5500 non-null   float64
dtypes: float64(14), int64(9)
memory usage: 988.4 KB
```

Fig.2 Information about the dataset

	Choice	A_follower_count	A_following_count	A_listed_count	A_mentions_received	A_retweets_received	A_mentions_sent	A_re
count	5500.000000	5.500000e+03	5.500000e+03	5500.000000	5.500000e+03	5500.000000	5500.000000	
mean	0.509455	6.498840e+05	1.265895e+04	5952.453273	2.666032e+03	1032.371839	6.011873	
std	0.499956	2.028787e+06	4.900867e+04	17339.141191	2.916543e+04	10954.953223	9.519797	
min	0.000000	1.600000e+01	0.000000e+00	0.000000	1.005034e-01	0.100503	0.100503	
25%	0.000000	2.663750e+03	3.220000e+02	85.000000	3.453649e+00	0.716816	0.359534	
50%	1.000000	4.558900e+04	7.780000e+02	932.000000	4.876542e+01	14.029113	2.299666	
75%	1.000000	3.927380e+05	2.838000e+03	6734.000000	3.498196e+02	118.704407	7.198330	
max	1.000000	3.654319e+07	1.165830e+06	549144.000000	1.145219e+06	435825.874241	76.809514	

Fig.3 Calculating statistical data about the dataset

```
# Testing the correlation

correlationA = Twitter_Train .iloc[:,1:12].corr(method = 'pearson')
correlationA

ax = sns.heatmap(
    correlationA,
    vmin=-1, vmax=1, center=0,
    cmap=sns.diverging_palette(20, 220, n = 200),
    square=True,
)
ax.set_xticklabels(
    ax.get_xticklabels(),
    rotation=45,
    horizontalalignment='right'
)
```

```
[Text(0.5, 0, 'A_follower_count'),
 Text(1.5, 0, 'A_following_count'),
 Text(2.5, 0, 'A_listed_count'),
 Text(3.5, 0, 'A_mentions_received'),
 Text(4.5, 0, 'A_retweets_received'),
 Text(5.5, 0, 'A_mentions_sent'),
 Text(6.5, 0, 'A_retweets_sent'),
 Text(7.5, 0, 'A_posts'),
 Text(8.5, 0, 'A_network_feature_1'),
 Text(9.5, 0, 'A_network_feature_2'),
 Text(10.5, 0, 'A_network_feature_3')]
```

Fig.4 Testing correlation of A

```
correlationB = Twitter_Train.iloc[:,12:23].corr(method='pearson')
correlationB

ax = sns.heatmap(
    correlationB,
    vmin=-1, vmax=1, center=0,
    cmap=sns.diverging_palette(20, 220, n = 200),
    square=True,
)
ax.set_xticklabels(
    ax.get_xticklabels(),
    rotation=45,
    horizontalalignment='right'
)
```

```
[Text(0.5, 0, 'B_follower_count'),
 Text(1.5, 0, 'B_following_count'),
 Text(2.5, 0, 'B_listed_count'),
 Text(3.5, 0, 'B_mentions_received'),
 Text(4.5, 0, 'B_retweets_received'),
 Text(5.5, 0, 'B_mentions_sent'),
 Text(6.5, 0, 'B_retweets_sent'),
 Text(7.5, 0, 'B_posts'),
 Text(8.5, 0, 'B_network_feature_1'),
 Text(9.5, 0, 'B_network_feature_2'),
 Text(10.5, 0, 'B_network_feature_3')]
```

Fig.5 Testing correlation of B

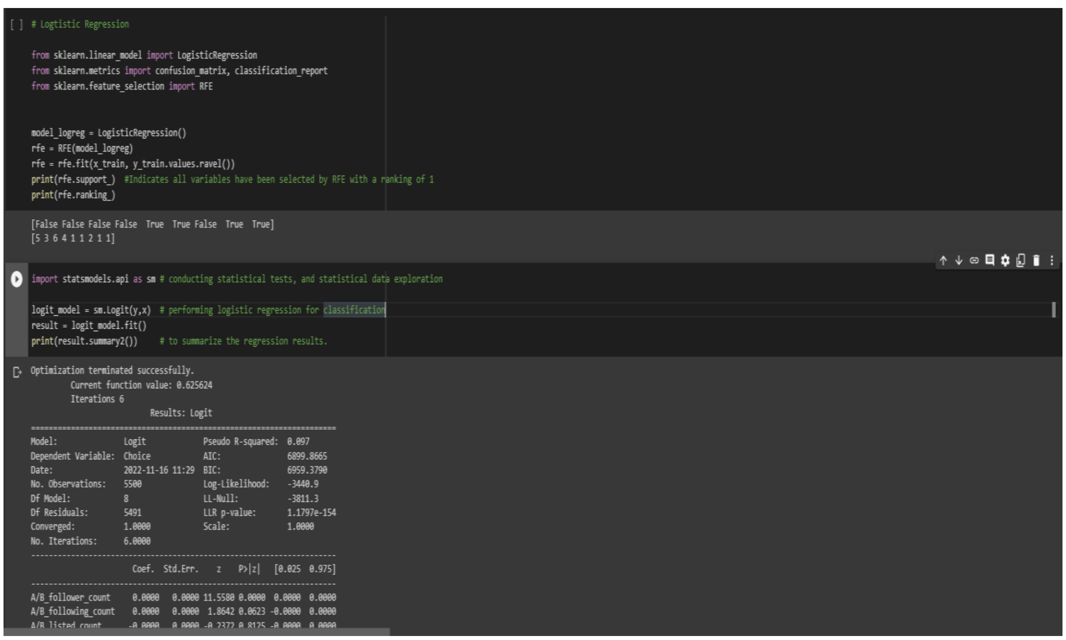
```
from sklearn.linear_model import LogisticRegression
from sklearn.neighbors import KNeighborsClassifier
from sklearn.tree import DecisionTreeClassifier
from sklearn.ensemble import RandomForestClassifier
from sklearn.naive_bayes import GaussianNB
from sklearn.svm import SVC
from xgboost import XGBClassifier
from sklearn.model_selection import cross_val_score
from sklearn.model_selection import StratifiedKFold


models = []
models.append(('LR', LogisticRegression(solver='liblinear', multi_class='ovr')))
models.append(('KNN', KNeighborsClassifier()))
models.append(('CART', DecisionTreeClassifier()))
models.append(("RF", RandomForestClassifier()))
models.append(('NB', GaussianNB()))
models.append(('SVM', SVC(gamma='auto')))
models.append(('Xgboost', XGBClassifier()))

# evaluate each model in turn
results = []
names = []
for name, model in models:
    kfold = StratifiedKFold(n_splits=10, random_state=1, shuffle=True) # Provides train/test indices to split data in train/test sets
    cv_results = cross_val_score(model, x_train, y_train, cv = kfold, scoring='accuracy') # cross-validation configuration,returning list of scores calculated for each fold.
    results.append(cv_results)
    names.append(name)
    print('%s: %f (%f)' % (name, cv_results.mean(), cv_results.std()))
```

```
LR: 0.708831 (0.014483)
KNN: 0.737922 (0.016281)
CART: 0.709351 (0.026295)
RF: 0.755584 (0.018961)
NB: 0.547273 (0.013196)
SVM: 0.523117 (0.006989)
Xgboost: 0.772987 (0.021831)
```

Fig.6 Spot checking various algorithms

Model	Mean	Standard deviation
Logistic Regression	0.708831	0.014483
KNN	0. 737922	0. 016281
Decision Tree (CART)	0. 709351	0. 026295
Random Forest Classifier	0. 755584	0. 018961
Gaussian NB	0. 547273	0. 013196
SVM	0. 523117	0. 006989
XG Boost	0. 772987	0. 021831

Table.1 Spot checking algorithms



```
# Logtistic Regression

from sklearn.linear_model import LogisticRegression
from sklearn.metrics import confusion_matrix, classification_report
from sklearn.feature_selection import RFE


model_logreg = LogisticRegression()
rfe = RFE(model_logreg)
rfe = rfe.fit(x_train, y_train.values.ravel())
print(rfe.support_)  #Indicates all variables have been selected by RFE with a ranking of 1
print(rfe.ranking_)

[False False False False  True  True False  True  True]
[5 3 6 4 1 1 2 1 1]

import statsmodels.api as sm # conducting statistical tests, and statistical data exploration

logit_model = sm.Logit(y,x)  # performing logistic regression for classification
result = logit_model.fit()
print(result.summary2())     # to summarize the regression results.

Optimization terminated successfully.
         Current function value: 0.625624
         Iterations 6
                         Results: Logit
=====================================================================
Model:              Logit            Pseudo R-squared:  0.097
Dependent Variable: Choice           AIC:               6899.8665
Date:               2022-11-16 11:29 BIC:               6959.3790
No. Observations:   5500             Log-Likelihood:    -3440.9
Df Model:           8                LL-Null:           -3811.3
Df Residuals:       5491             LLR p-value:       1.1797e-154
Converged:          1.0000           Scale:             1.0000
No. Iterations:     6.0000
---------------------------------------------------------------------
                      Coef.  Std.Err.    z    P>|z|   [0.025  0.975]
---------------------------------------------------------------------
A/B_follower_count    0.0000   0.0000 11.5580 0.0000  0.0000  0.0000
A/B_following_count   0.0000   0.0000  1.8642 0.0623 -0.0000  0.0000
A/B_listed_count     -0.0000   0.0000 -0.2372 0.8125 -0.0000  0.0000
```

Fig.7 Applying logistic regression and summarizing results

```
# Removing variables with p-values greater than 0.05 to improve the model

columns_to_keep = ['A/B_follower_count', 'A/B_retweets_received', 'A/B_mentions_sent',
                   'A/B_network_feature_2', 'A/B_network_feature_3']
x = x_train[columns_to_keep]
y = y_train

logit_model = sm.Logit(y,x)
result = logit_model.fit()
print(result.summary2())
```

```
Optimization terminated successfully.
         Current function value: 0.618079
         Iterations 6
                          Results: Logit
=====================================================================
Model:              Logit            Pseudo R-squared: 0.108
Dependent Variable: Choice           AIC:              4769.2079
Date:               2022-11-16 11:29 BIC:              4800.4870
No. Observations:   3850             Log-Likelihood:   -2379.6
Df Model:           4                LL-Null:          -2667.5
Df Residuals:       3845             LLR p-value:      2.7700e-123
Converged:          1.0000           Scale:            1.0000
No. Iterations:     6.0000
---------------------------------------------------------------------
                       Coef.  Std.Err.    z    P>|z|   [0.025  0.975]
---------------------------------------------------------------------
A/B_follower_count     0.0000   0.0000 10.5261 0.0000  0.0000  0.0000
A/B_retweets_received -0.0000   0.0000 -2.2968 0.0216 -0.0000 -0.0000
A/B_mentions_sent      0.0001   0.0000 15.0594 0.0000  0.0001  0.0001
A/B_network_feature_2  0.0220   0.0023  9.7371 0.0000  0.0176  0.0265
A/B_network_feature_3 -0.0000   0.0000 -8.0290 0.0000 -0.0000 -0.0000
=====================================================================
```

Fig.8 Improving the model

```
# Logistic Regression Model Fitting

model_logreg.fit(x_train, y_train)

# Predicting test results

y_pred = model_logreg.predict(x_test)
```

```
# Calculating the accuracy of our model

print("Training accuracy of Logistic Regression model:", model_logreg.score(x_train, y_train))
print("Testing accuracy of Logistic Regression model:", model_logreg.score(x_test, y_test))
```

```
Training accuracy of Logistic Regression model: 0.7103896103896103
Testing accuracy of Logistic Regression model: 0.7260606060606061
```

The difference between training and testing accuracy indicates how much the model has overfitted. The Logistic Regression model has a low overfitting value of 0.02 which indicates the model has not trained data too well to negatively impact the performance.

Fig.9 Fitting logistic regression and calculating accuracy

```
[ ] # Confusion Matrix

    from sklearn.metrics import confusion_matrix, classification_report

    confusion_matrix = confusion_matrix (y_test,y_pred)
    print(confusion_matrix)

    [[674 146]
     [306 524]]
```

The confusion matrix visualises the actual and predicted values of the data thereby depicting the performance of a supervised algorithm. Here, it can be inferred that there are 1,198 correct predictions and 452 incorrect predictions.

```
[ ] # Classification report will summarise our model by computing precision, recall, f-measure and support

    print(classification_report(y_test, y_pred))

                  precision    recall  f1-score   support

               0       0.69      0.82      0.75       820
               1       0.78      0.63      0.70       830

        accuracy                           0.73      1650
       macro avg       0.73      0.73      0.72      1650
    weighted avg       0.74      0.73      0.72      1650
```

In social media analytics, high precision means that an algorithm returned more relevant results than irrelevant, while high recall means that an algorithm returned most of the relevant results. In the logistic regression model, there is a 74% precision that there are more relevant values and a 73% recall that most of the relevant values are included in our model.

Fig.10 Confusion matrix and classification report of Logistic regression

	Precision	Recall	F1-score	Support
0	0.69	0.82	0.75	820
1	0.78	0.63	0.70	830
Accuracy			0.73	1650
Macro Avg.	0.73	0.73	0.72	1650
Weighted Avg.	0.74	0.73	0.72	1650

Table.2 Classification report of Logistic report

```
# Random Forest Classifier

from sklearn.ensemble import RandomForestClassifier

model_rf = RandomForestClassifier()
model_rf.fit(x_train, y_train) #Train the model on training data

# Predicting the test results

y_pred = model_rf.predict(x_test)
```

```
# Calculating the accuracy of our model

print("Training accuracy of Random Forest model:", model_rf.score(x_train, y_train))
print("Testing accuracy of Random Forest model:", model_rf.score(x_test, y_test))
```

```
Training accuracy of Random Forest model: 0.9935064935064936
Testing accuracy of Random Forest model: 0.7727272727272727
```

Fig.11 Applying random forest and calculating accuracy

```
# To interpret the model and report the results, feature importances is used to quantify how much a particular variable improves predctions

feature_importances = pd.DataFrame(model_rf.feature_importances_,
                                   index = x_train.columns,
                                   columns = ['importance']).sort_values('importance', ascending = False)
feature_importances
```

	importance
A/B_mentions_sent	0.153849
A/B_network_feature_3	0.146153
A/B_listed_count	0.138348
A/B_follower_count	0.132385
A/B_retweets_sent	0.122169
A/B_posts	0.089116
A/B_retweets_received	0.079057
A/B_network_feature_2	0.075243
A/B_following_count	0.063680

```
# Select features with importance larger than 0.08

columns_to_keep = ['A/B_mentions_sent', 'A/B_listed_count', 'A/B_follower_count',
                   'A/B_network_feature_3', 'A/B_retweets_sent', 'A/B_posts']

x = x_train[columns_to_keep]
y = y_train

model_rf = RandomForestClassifier() #New random forest classifier for the most important features
model_rf.fit(x_train, y_train) #Train the model on training data

# Predicting the test results

y_pred = model_rf.predict(x_test)
```

Fig.12 Improving predictions

```
[ ] # Calculating the accuracy of the model

    print("Training accuracy of Random Forest model:", model_rf.score(x_train, y_train))
    print("Testing accuracy of Random Forest model:", model_rf.score(x_test, y_test))
```

```
Training accuracy of Random Forest model: 0.9937662337662337
Testing accuracy of Random Forest model: 0.7696969696969697
```

The difference between training and testing accuracy indicates how much the model has overfitted. The Random Forest model has a low overfitting value of 0.22 which indicates the model has not trained data too well to negatively impact the performance.

```
[ ] # Confusion Matrix

    from sklearn.metrics import confusion_matrix, classification_report

    conf_matrix = confusion_matrix (y_test,y_pred)
    print(conf_matrix)
```

```
[[622 198]
 [182 648]]
```

The confusion matrix visualises the actual and predicted values of the data thereby depicting the performance of a supervised algorithm. Here, it can be inferred that there are 1,274 correct predictions and 376 incorrect predictions.

Fig.13 Calculating accuracy and confusion matrix

```
[ ] # Classification report will summarise our model by computing precision, recall, f-measure and support

    print(classification_report(y_test, y_pred))
```

```
              precision    recall  f1-score   support

           0       0.77      0.76      0.77       820
           1       0.77      0.78      0.77       830

    accuracy                           0.77      1650
   macro avg       0.77      0.77      0.77      1650
weighted avg       0.77      0.77      0.77      1650
```

In social media analytics, high precision means that an algorithm returned more relevant results than irrelevant, while high recall means that an algorithm returned most of the relevant results. In the Random Forest model, there is a 78% precision that there are more relevant values and a 78% recall that most of the relevant values are included in our model.

Fig.14 Calculating classification report of Random Forest

	Precision	Recall	F1-score	Support
0	0.77	0.76	0.77	820
1	0.77	0.78	0.77	830
Accuracy			0.77	1650
Macro Avg.	0.77	0.77	0.77	1650
Weighted Avg.	0.77	0.77	0.77	1650

Table.3 Classification report of Random Forest

```
# K-Nearest Neighbors

from sklearn.neighbors import KNeighborsClassifier

model_knn = KNeighborsClassifier(n_neighbors = 5, metric = 'euclidean')
model_knn.fit(x_train, y_train)

y_pred = model_knn.predict(x_test)
```

```
# Calculating training and testing of our model

print("Training accuracy of K-Nearest Neighbors model:", model_knn.score(x_train, y_train))
print("Testing accuracy of K-Nearest Neighbors model:", model_knn.score(x_test, y_test))
```

```
Training accuracy of K-Nearest Neighbors model: 0.8184415584415584
Testing accuracy of K-Nearest Neighbors model: 0.7412121212121212
```

The difference between training and testing accuracy indicates how much the model has overfitted. The KNN model has a low overfitting value of 0.07 which indicates the model has not trained data too well to negatively impact the performance.

Fig.15 Applying K-Nearest neighbor and calculating accuracy

```python
# Confusion Matrix

from sklearn.metrics import confusion_matrix, classification_report

confusion_matrix = confusion_matrix(y_test, y_pred)
print(confusion_matrix)
```

```
[[584 236]
 [191 639]]
```

The confusion matrix visualises the actual and predicted values of the data thereby depicting the performance of a supervised algorithm. Here, it can be inferred that there are 1,223 correct predictions and 427 incorrect predictions.

```python
# Classification report will summarise our model by computing precision, recall, f-measure and support

print(classification_report(y_test, y_pred))
```

```
              precision    recall  f1-score   support

           0       0.75      0.71      0.73       820
           1       0.73      0.77      0.75       830

    accuracy                           0.74      1650
   macro avg       0.74      0.74      0.74      1650
weighted avg       0.74      0.74      0.74      1650
```

In social media analytics, high precision means that an algorithm returned more relevant results than irrelevant, while high recall means that an algorithm returned most of the relevant results. In the KNN model, there is a 74% precision that there are more relevant values and a 74% recall that most of the relevant values are included in our model.

Fig.16 Confusion matrix and classification report of K-nearest neighbor

	Precision	Recall	F1-score	Support
0	0.75	0.71	0.73	820
1	0.73	0.77	0.75	830
Accuracy			0.74	1650
Macro Avg.	0.74	0.74	0.74	1650
Weighted Avg.	0.74	0.74	0.74	1650

Table.4 Classification report K-nearest neighbor

```
# XGBoost Classifier

from xgboost import XGBClassifier
from xgboost import plot_importance

model_xgb = XGBClassifier(max_depth = 2,
                          objective = 'binary:logistic',
                          eta = 0.3
                         )
model_xgb

model_xgb.fit(x_train, y_train)
print(model_xgb)

y_pred = model_xgb.predict(x_test)
```

```
XGBClassifier(eta=0.3, max_depth=2)
```

```
# Calculating training and testing of our model

print("Training accuracy of XGBoost model:", model_xgb.score(x_train, y_train))
print("Testing accuracy of XGBoost model:", model_xgb.score(x_test, y_test))


plot_importance(model_xgb, max_num_features=9)
pyplot.show()
```

```
Training accuracy of XGBoost model: 0.7958441558441558
Testing accuracy of XGBoost model: 0.7824242424242425
```

Fig.17 Applying XGBoost and calculating accuracy

```
# Select the top 5 features and re-run the model

from sklearn.metrics import confusion_matrix, classification_report

cols = ['A/B_follower_count','A/B_network_feature_3','A/B_listed_count',
        'A/B_network_feature_2','A/B_retweets_received' ]
x = x_train[cols]
y = y_train


model_xgb = XGBClassifier() #New xgb classifier for the most important features
model_xgb.fit(x_train, y_train) #Train the model on training data

#Predicting the test results

y_pred = model_xgb.predict(x_test)

plot_importance(model_xgb, max_num_features=5)
pyplot.show()
```

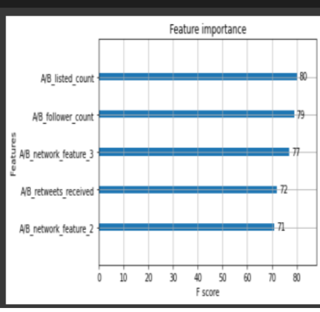


Fig.18 Improving the model

```
[ ] # Calculating the accuracy of the model

    print("Training accuracy of XGBoost model:", model_xgb.score(x_train, y_train))
    print("Testing accuracy of XGBoost model:", model_xgb.score(x_test, y_test))
```

```
Training accuracy of XGBoost model: 0.8166233766233766
Testing accuracy of XGBoost model: 0.7866666666666666
```

The difference between training and testing accuracy indicates how much the model has overfitted. The XGBoost model has a low overfitting value of 0.03 which indicates the model has not trained data too well to negatively impact the performance.

```
# Confusion Matrix

from sklearn.metrics import confusion_matrix, classification_report

confusion_matrix = confusion_matrix(y_test, y_pred)
print(confusion_matrix)
```

```
[[636 184]
 [168 662]]
```

The confusion matrix visualises the actual and predicted values of the data thereby depicting the performance of a supervised algorithm. Here, it can be inferred that there are 1,298 correct predictions and 352 incorrect predictions.

Fig.19 Calculating accuracy and confusion matrix

```
[ ] # Classification report will summarise our model by computing precision, recall, f-measure and support

    print(classification_report(y_test, y_pred))
```

```
              precision    recall  f1-score   support

           0       0.79      0.78      0.78       820
           1       0.78      0.80      0.79       830

    accuracy                           0.79      1650
   macro avg       0.79      0.79      0.79      1650
weighted avg       0.79      0.79      0.79      1650
```

In social media analytics, high precision means that an algorithm returned more relevant results than irrelevant, while high recall means that an algorithm returned most of the relevant results. In the XGBoost model, there is a 79% precision that there are more relevant values and a 79% recall that most of the relevant values are included in our model.

+ Code + Text

I tried 4 models on the dataset: Logistic Regression, Random Forest, K-Nearest Neighbors and XGBoost. On comparing these models, XGBoost has the highest training and testing accuracy of 81% and 79% respectively. Further, on analyzing the confusion matrix, XGBoost model has the highest number of correctly predicted values (1,298). Lastly, according to the classification report, XGBoost model has the highest precision, recall and f-measure (79%) indicating that the model selected more relevant values than irrelevants values as well as the most relevant values in the dataset have been included in our model. Therefore, the key predictors of social influence according to XGBoost model are: A/B_listed_count, A/B_follower_count, A/B_network_feature_3, A/B_retweets_received and A/B_network_feature_2.This data can be used by companies in choosing most influencial people on media to partner with.

Fig.20 Classification report of XGBoost

	Precision	Recall	F1-score	Support
0	0.79	0.78	0.78	820
1	0.78	0.80	0.79	830
Accuracy			0.79	1650
Macro Avg.	0.79	0.79	0.79	1650
Weighted Avg.	0.79	0.79	0.79	1650

Table.5 Classification report XG Boost

Chapter 5
CONCLUSIONS

5.1 Conclusions

The quote from a well-known online strategist perfectly sums up the benefits and significance of viewing social media analytics as a measuring framework that integrates tactical requirements and strategic goals, or as a business intelligence activity. Organizations must connect their measurement frameworks to high-level business goals like revenue creation, cost reduction, or operational excellence as they get more serious about gauging the success of their social media program. A practical way for businesses to maximize the return on their efforts in such projects is to include a social media analytics program into the broader business intelligence plan. A BI mindset for social media analytics has the ability to deliver real-time feedback and usable insights to aid companies in their decision-making processes, as is clear from the suggestions in the existing literature and significant results from our expert panels. By directly connecting social media analytics to their broader business objectives, such an approach would also assist firms to better the tactical execution of social media initiatives like increasing engagement across numerous channels and platforms. We anticipate that the debate in this paper will be used as a springboard for businesses taking part in various social media efforts to develop a central guiding principle. In order to further our research, we incorporated important results from an expert panel's social media analysis of preliminary findings from the literature to create a more formal investigative action plan. Additional empirical research, qualitative interviews, and quantitative research methods are all part of our goal. A clear identification of the organizational, technological, and behavioral elements that affect the acceptance and implementation of social analytics initiatives in enterprises is the predicted result of our study. Our mission is to use research to develop practical methods for delivering social media analytics that are connected to strategic business goals so that businesses may more effectively give actionable insights.

5.2 Future Scope

The panelists in both online expert sessions concurred that the use of big data and the developing discipline of data science are crucial to the future of social media analytics. Businesses will soon be able to access the vast amounts of structured data stored within their internal enterprise systems and continually employ social media monitoring tools to extract unstructured data from the social web thanks to unprecedented access to data and computational power. In order to gather and process information about their users' digital footprints, social networking sites like Facebook are setting up data centers all over the world. This information could one day be used to support business decision-making processes, particularly in the areas of marketing and brand management. The key, according to the experts on the panel, is to effectively use big data by overcoming its fragmented character and by reassembling pertinent information and connecting it with other internal sources of data in a time-sensitive manner to enable real-time business intelligence. Additionally, businesses will need to develop efficient strategies for separating signal from noise and determining which data and KPIs are most important to their operations. An organization can then strive toward developing a successful analytics program and offering strategic insights. Numerous speakers also shared their opinions on the use of predictive analytics methods using social media data. Along with descriptive metrics tools, several software companies also offer predictive and prescriptive analytics solutions. These traits may offer hints as to what steps a corporation should take in response to other (perhaps insufficient) Social Media Metrics Scorecard components. A variety of technologies are promising, enabling analysts to discover new market niches, forecast public opinion, and recognize his prospective customers as champions. Predictive analytics, as a cutting-edge technology that enables businesses to anticipate rather than respond to market situations and become more proactive in enhancing customer relationships and pursuing development prospects, was generally viewed favorably by the panelists.

5.3 Applications Contributions

Customer network value estimation, bounce rate analysis, recommender engine, website search analysis, revenue analysis, recommender performance analysis, conversion analysis, and segment-specific analysis are all tools used by practitioners and researchers in the field of social media analytics. We are always looking at things like key performance indicator (KPI) analysis. Crowdsourcing, social network analytics, opinion mining, and sentiment mining are some of the topics that are becoming more and more crucial. Let's examine a few social media analytics applications. Identifying prominent people who have an impact on group behavior is one of the most crucial topics (Scheer & Stern, 1992). Let's say someone uploads some stuff on Facebook. Some posts made by specific persons constantly receive a lot of attention, while others essentially go unnoticed. In order to find these "hot" users of social media networks, statistics can make a significant impact. This assists in locating important decision-makers in private social networks. Therefore, the close friends and family who are the target audience for advertisements aimed at these key influencers readily accept them. Our analytical study primarily focuses on identifying the dynamics among these important social media influencers. Clique identification is just another aspect of this social interaction (Bron & Kerbosch, 1973). A clique in social networks is a collection of users who communicate with one another often. Imagine a small group of these people who communicate often on a social media site like Facebook. Depending on their shared interests, their interactions may be limited or close-related (Dunbar

& Spoors, 1995). Offers accepted by one member of a clique within a network may also be accepted by other members of the clique. But only if other members of the category are also prospective customers.

REFERENCES

[1] Ruhi, Umar. "Social media analytics as a business intelligence practice: Current landscape & future prospects." *Umar Ruhi (2014)," Social Media Analytics as a Business Intelligence Practice: Current Landscape & Future Prospects", Journal of Internet Social Networking & Virtual Communities* 2014 (2014).

[2] Gupta, Bhumika, et al. "Study of Twitter sentiment analysis using machine learning algorithms on Python." *International Journal of Computer Applications* 165.9 (2017): 29-34.

[3] Mohana, R. S., et al. "Twitter based sentiment analysis to predict public emotions using machine learning algorithms." *2021 Third International Conference on Inventive Research in Computing Applications (ICIRCA)*. IEEE, 2021.

[4] Patnana, Divya Sai, Gone Hitesh, and IpsitaSahu N. Suresh Kumar. "Logistic Regression Analysis on Social Networking Advertisement."

[5] Hota, Soudamini, and Sudhir Pathak. "KNN classifier based approach for multi-class sentiment analysis of twitter data." *Int. J. Eng. Technol* 7.3 (2018): 1372-1375.

[6] Sudira, Hanif, Alifiannisa Lawami Diar, and Yova Ruldeviyani. "Instagram sentiment analysis with naive bayes and KNN: exploring customer satisfaction of digital payment services in Indonesia." *2019 International Workshop on Big Data and Information Security (IWBIS)*. IEEE, 2019

[7] Roshini, T., et al. "Social media survey using decision tree and Naive Bayes classification." *2019 2nd International Conference on Intelligent Communication and Computational Techniques (ICCT)*. IEEE, 2019.

[8] Aufar, Mohammad, Rachmadita Andreswari, and Dita Pramesti. "Sentiment analysis on youtube social media using decision tree and random forest algorithm: A case study." *2020 International Conference on Data Science and Its Applications (ICoDSA)*. IEEE, 2020.

[9] Bahrawi, Nfn. "Sentiment analysis using random forest algorithm-online social media based." *Journal of Information Technology and Its Utilization* 2.2 (2019): 29-33.

[10] Karthika, P., R. Murugeswari, and R. Manoranjithem. "Sentiment analysis of social media network using random forest algorithm." *2019 IEEE international conference on intelligent techniques in control, optimization and signal processing (INCOS)*. IEEE, 2019.

[11] Bhowmik, Pankaj, et al. "Analysis of Social Media Data to Classify and Detect Frequent Issues Using Machine Learning Approach." *2020 2nd International Conference on Advanced Information and Communication Technology (ICAICT)*. IEEE, 2020.

[12] Fan, Weiguo, and Michael D. Gordon. "The power of social media analytics." *Communications of the ACM* 57.6 (2014): 74-81.

[13] Hama Aziz, Roza Hikmat, and Nazife Dimililer. "SentiXGboost: enhanced sentiment analysis in social media posts with ensemble XGBoost classifier." *Journal of the Chinese Institute of Engineers* 44.6 (2021): 562-572.

APPENDICES

```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
%matplotlib inline
%pylab inline
import seaborn as sns
from pandas import read_csv
from matplotlib import pyplot
```

Fig.21 Importing required libraries

```
from sklearn.linear_model import LogisticRegression
from sklearn.neighbors import KNeighborsClassifier
from sklearn.tree import DecisionTreeClassifier
from sklearn.ensemble import RandomForestClassifier
from sklearn.naive_bayes import GaussianNB
from sklearn.svm import SVC
from xgboost import XGBClassifier
from sklearn.model_selection import cross_val_score
from sklearn.model_selection import StratifiedKFold
```

Fig.22 Important machine learning algorithms

	importance
A/B_mentions_sent	0.153849
A/B_network_feature_3	0.146153
A/B_listed_count	0.138348
A/B_follower_count	0.132385
A/B_retweets_sent	0.122169
A/B_posts	0.089116
A/B_retweets_received	0.079057
A/B_network_feature_2	0.075243
A/B_following_count	0.063680

Fig.23 Finding important features

In social media analytics, high precision means that an algorithm returned more relevant results than irrelevant, while high recall means that an algorithm returned most of the relevant results. In the XGBoost model, there is a 79% precision that there are more relevant values and a 79% recall that most of the relevant values are included in our model.

I tried 4 models on the dataset: Logistic Regression, Random Forest, K-Nearest Neighbors and XGBoost. On comparing these models, XGBoost has the highest training and testing accuracy of 81% and 79% respectively. Further, on analyzing the confusion matrix, XGBoost model has the highest number of correctly predicted values (1,298). Lastly, according to the classification report, XGBoost model has the highest precision, recall and f-measure (79%) indicating that the model selected more relevant values than irrelevants values as well as the most relevant values in the dataset have been included in our model. Therefore, the key predictors of social influence according to XGBoost model are: A/B_listed_count, A/B_follower_count, A/B_network_feature_3, A/B_retweets_received and A/B_network_feature_2.This data can be used by companies in choosing most influencial people on media to partner with.

Fig.24 Result