CBCS SCHEME

USN

MMCA311F

Third Semester MCA Degree Examination, Dec.2025/Jan.2026

Social Media Analytics

Time: 3 hrs. Max. Marks: 100

Note: 1. Answer any FIVE full questions, choosing ONE full question from each module.
2. M : Marks , L: Bloom's level , C: Course outcomes.

			M	L	C
		Module – 1			
Q.1	a.	Discuss the KPI metrics of social metrics analytics.	10	L2	CO1
	b.	Compare Facebook and Instagram based on user engagement and content type.	10	L2	CO1
		OR			
Q.2	a.	List the differences between Structured and Unstructured data.	10	L2	CO1
	b.	Explain how do we extract twitter data using twitter API.	10	L2	CO1
		Module – 2			
Q.3	a.	Discuss the different methods of preprocessing the text.	10	L2	CO2
	b.	Describe Word2Vec in detail and analyze its significance in word representation.	10	L3	CO2
		OR			
Q.4	a.	Write python code to apply SVM algorithm for sentiment classification	10	L2	CO2
	b.	Discuss the fundamentals of Sentiment analysis.	10	L2	CO2
		Module – 3			
Q.5	a.	What is Centrality? Explain Degree Centrality in detail.	10	L2	CO3
	b.	Explain community detection method in detail	10	L2	CO3
		OR			
Q.6	a.	What is meant by influencer identification and user engagement analytics	10	L3	CO3
	b.	Describe how do you perform time series analysis	10	L2	CO3
		Module – 4			
Q.7	a.	Discuss the steps of Dashboard creation using Tableau.	10	L3	CO4
	b.	Explain the importance of visualizing data using Network Maps with an example.	10	L2	CO4
		OR			
Q.8	a.	What are word clouds? Demonstrate its usage with a code snippet.	10	L3	CO4
	b.	Briefly describe how heat maps help in analyzing customer sentiment data.	10	L2	CO4
		Module – 5			
Q.9	a.	Explain the ethical Considerations in Social Media Analytics?	10	L2	CO5
	b.	Discuss how social media analytics can improve brand monitoring and crisis management.	10	L2	CO5
		OR			
Q.10	a.	Discuss emerging trends in social media analytics, highlighting the role of AI-driven social insights.	10	L2	CO5
	b.	Explain the role of social media in Ad performance and customer engagement.	10	L2	CO5

1.a. Discuss the KPI metrics of Social Media analytics.

Impressions

Impressions refer to the number of times your content has been shown on someone's screen.

Reach

While impressions refer to the number of times an ad is displayed, reach refers to the number of users who have seen a piece of content.

Engagement

Engagement is the action a user takes with content, and it could be anything from:

● A click-through to a website or ad

● A social media or website article share

● A comment on a blog or social post

● A click of a “Shop Now” button on an Instagram photo

● A click on a hyperlink in an article that guides a user’s customer journey

Sentiment Scoring:

Sentiment scoring is a metric used to quantify the sentiment or emotion expressed in qualitative data like customer feedback or social media interactions. It depicts the level of emotion analysis as positive, negative or neutral.

1.b. Compare facebook and Instagram based on user engagement and content type.

Aspect	Facebook	Instagram
Primary Content Type	Text posts, links, images, videos, groups, events	Images, short videos, Reels, Stories, carousels
Nature of Content	Informational, community-based, discussion-oriented	Visual, creative, lifestyle-oriented
User Engagement Style	Comments, shares, group discussions, long interactions	Likes, shares, saves, story replies, quick interactions
Engagement Rate	Moderate; engagement spread across posts, groups, and pages	High; strong engagement on visual and short-form video content
Audience	Wider age range,	Younger audience, strong presence of

Aspect	Facebook	Instagram
Demographics	strong presence of adults and professionals	Gen Z and Millennials
Content Consumption	Users spend time reading posts, news, and discussions	Users prefer fast, visually appealing, scroll-based content
Brand Engagement	Effective for brand communities, customer support, announcements	Highly effective for influencer marketing and brand storytelling
Ad Performance	Suitable for detailed ads, lead generation, and links	Best for visual ads, product discovery, and impulse engagement

2.a. List the differences between Structured and Unstructured data.

Aspect	Structured Data	Unstructured Data
Technology	Stored in relational database tables (RDBMS)	Stored as character-based or binary data
Transaction Management	Mature transaction management with concurrency control	No transaction management or concurrency control
Version Management	Versioning supported at tuple, row, and table levels	Versioned as a whole
Flexibility	Rigid and schema-dependent	Highly flexible with no predefined schema
Scalability	Schema scaling is difficult	Highly scalable
Robustness	Very robust and reliable	Less robust

Aspect	Structured Data	Unstructured Data
Query Performance	Supports structured queries with complex joins	Supports only basic or textual queries

2.b. Explain how do we extract twitter data using Twitter API.

- Twitter might be described as a real-time, highly social microblogging service that allows users to post short status updates, called tweets, that appear on timelines.
- Tweets may include one or more entities in their (currently) 280 characters of content and reference one or more places that map to locations in the real world.
- An understanding of users, tweets, and timelines is particularly essential to effective use of Twitter's API
- Tweets are the essence of Twitter, and while they are notionally thought of as short strings of text content associated with a user's status update, there's really quite a bit more metadata
- In addition to the textual content of a tweet itself, tweets come bundled with two additional pieces of metadata that are of particular note: entities and places.
- Tweet entities are essentially the user mentions, hashtags,URLs, and media that may be associated with a tweet, and places are locations in the real world that may be attached to a tweet.
- Note that a place may be the actual location in which a tweet was authored, but it might also be a reference to the place

  described in a tweet.

Accessing Twitter Data:

- i)we can use Twitter API
- ii)we can use some of the tools like Flume
- We can customize #tag for data. Provides data for a week.
- It provides username,timestamp,content of the tweet,URL,retweets.
- Twitter has taken great care to craft an elegantly simple RESTful API that is intuitive and easy to use.
- A particularly beautiful Python package that wraps the Twitter API and mimics the public API semantics almost one-to-one is twitter.
- Like most other Python packages, you can install it with pip by typing

  pip install twitter in a terminal.

- One popular alternative is tweepy.
- Before we make any API requests to Twitter, we need to create an application at

https://dev.twitter.com/apps.

- Creating an application is the standard way for devel-

  opers to gain API access and for Twitter to monitor and interact with third-party platform developers as needed.

- To use twitter API we must apply for a Twitter developer account and be approved in order to create new apps.
- Creating an app will also create a set of authentication tokens that will let us programmatically access the Twitter platform.

```
import twitter

# Go to http://dev.twitter.com/apps/new to create an app and get values

# for these credentials, which you'll need to provide in place of these

# empty string values that are defined as placeholders.

# See https://developer.twitter.com/en/docs/basics/authentication/overview/oauth

# for more information on Twitter's OAuth implementation.

CONSUMER_KEY = ''

CONSUMER_SECRET = ''

OAUTH_TOKEN = ''

OAUTH_TOKEN_SECRET = ''

auth = twitter.oauth.OAuth(OAUTH_TOKEN, OAUTH_TOKEN_SECRET,

CONSUMER_KEY, CONSUMER_SECRET)

twitter_api = twitter.Twitter(auth=auth)

# Nothing to see by displaying twitter_api except that it's now a

# defined variable

print(twitter_api)
```

Ex 2: Retrieving Topics

Example 1-2. Retrieving trends

```
# The Yahoo! Where On Earth ID for the entire world is 1.

# See http://bit.ly/2BGWJBU and

# http://bit.ly/2MsvwCQ
```

WORLD_WOE_ID = 1

US_WOE_ID = 23424977

Prefix ID with the underscore for query string parameterization.

Without the underscore, the twitter package appends the ID value

to the URL itself as a special case keyword argument.

world_trends = twitter_api.trends.place(_id=WORLD_WOE_ID)

us_trends = twitter_api.trends.place(_id=US_WOE_ID)

print(world_trends)

print()

print(us_trends)

3.a. Discuss the different methods of preprocessing the text.

Text preprocessing is the process of transforming raw text into a structured and analyzable form for NLP tasks.

Tokenization:

- Tokenization is the process of splitting text into smaller units called tokens, such as words, subwords, or sentences.

  Ex: Text: "NLP makes machines understand language."

  Tokens: ["NLP", "makes", "machines", "understand", "language", "."]

  Tokenization will

- Enables word-level analysis
- Simplifies downstream NLP tasks
- Helps in vectorization and embedding creation

  Types of Tokenization

- **Word Tokenization:** Splits text into words
- **Sentence Tokenization:** Splits text into sentences
- **Subword Tokenization:** Splits words into smaller meaningful units (used in BERT, GPT)
  import nltk

  nltk.download('punkt')

  from nltk.tokenize import word_tokenize, sent_tokenize

  text = "NLP makes machines understand language. It's fascinating!"

print("Word Tokens:", word_tokenize(text))

print("Sentence Tokens:", sent_tokenize(text))

Stopword Removal:

- In **NLP(Natural Language Processing)**, stop words are the words that are filtered out before or after processing text data, such as "is", "and", "a" etc. These words do not add meaning to the text and can be removed to improve the efficiency.
- The **Natural Language Toolkit (NLTK)** is the python library that provides the easy to use interface and the tools for text processing such as tokenization and stop word removal.

Stemming:

- Stemming is the process of reducing words to their root form by chopping off prefixes or suffixes — often without considering linguistic correctness.

Example:

- “playing” → “play”
- “flies” → “fly”

Lemmatization:

- Lemmatization reduces words to their lemma (dictionary form) using vocabulary and morphological analysis. Unlike stemming, it returns meaningful root words, which is often part of image preprocessing for AI tasks that combine vision and language data.

Example:

- “running” → “run”
- “better” → “good”

Stemming:

- Reduces dimensionality of the dataset
- Speeds up processing in large text corpora
- Useful when exact meaning is less critical (e.g., search engines)

Popular Stemming Algorithms

- Porter Stemmer (most common in NLP)
- Snowball Stemmer (improved version of Porter)
- Lancaster Stemmer (more aggressive)

`

Why Use Lemmatization?

- Preserves semantic meaning

- More accurate than stemming
- Necessary for linguistic tasks like part-of-speech tagging
  Ex:

```
from nltk.stem import WordNetLemmatizer

# Initialize the lemmatizer

lemmatizer = WordNetLemmatizer()

# Lemmatize words

print(lemmatizer.lemmatize("running", pos="v")) # Output: run

print(lemmatizer.lemmatize("better", pos="a")) # Output: good

print(lemmatizer.lemmatize("geese", pos="n")) # Output: goose
```

3.b. Describe word2vec in detail and analyze its significance in word representation.

•provides Feature Representation

•Uses a neural network model to learn word associations from a large corpus of text.

•A 2 layer neural network to generate word embeddings given in a text corpus.

•Once trained such a model can detect synonymous words or suggest additional words for a particular sentence

Word2vec is needed to provide

•Preserves relationship between words.

•Deals with addition of new words in the vocabulary.

•Better results in lots of deep learning applications

CBOW Working:

•We generate one hot word vectors corresponding to the context

•These vectors are embedded using n dimensions say 300

$$\left(v_{c-m} = \gamma x^{(c-m)}, v_{c-m+1} = \gamma x^{(c-m+1)}, \ldots, v_{c+m} = \gamma x^{(c+m)}\right)$$

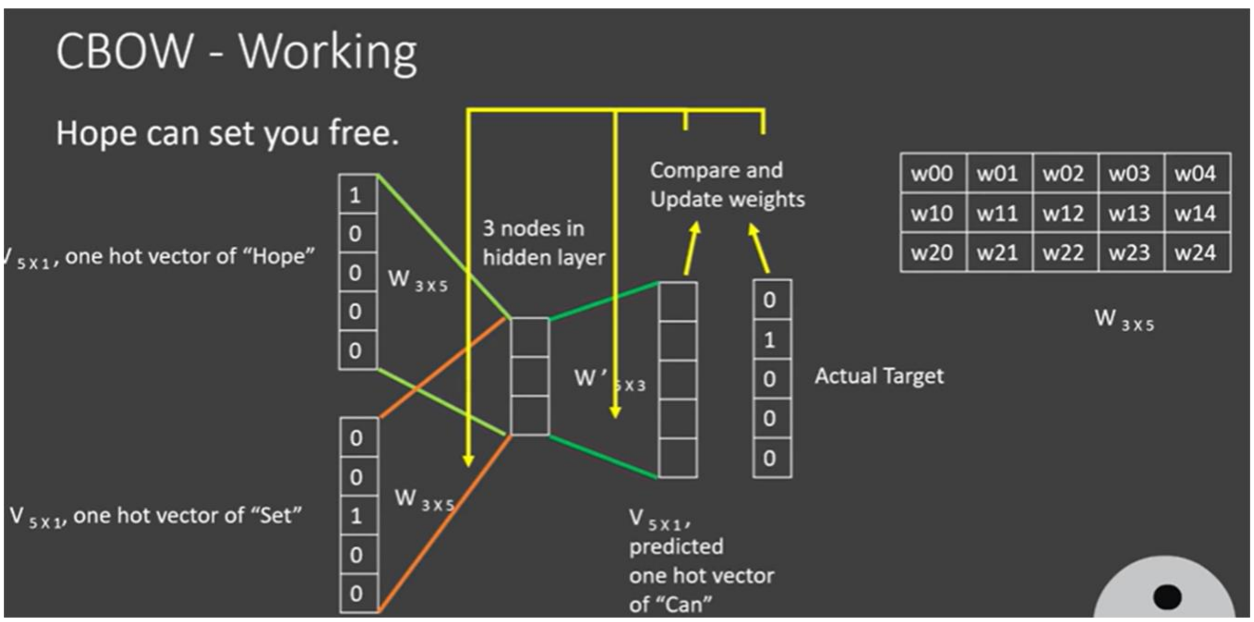
•The context vectors are averaged before using in prediction.

$$\hat{v} = \frac{v_{c-m} + v_{c-m+1} + \ldots + v_{c+m}}{2m}$$

•Generate a score vector z= U^V

•Turn the scores into probabilities using softmax(z)

•Match the the probabilities generated with true probabilities



Skip Gram

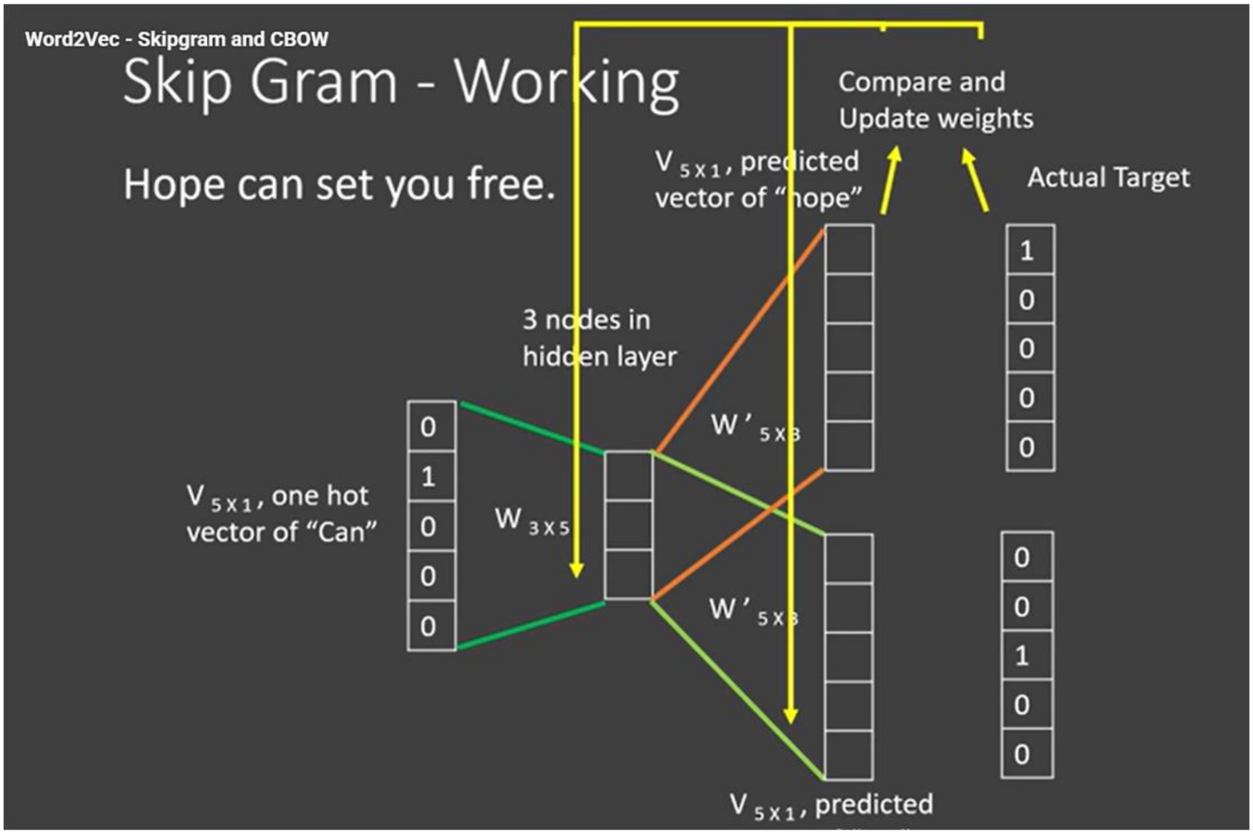
•It is required to predict if candidate word c is a neighbor of a given target word t

•The target word t and a neighboring context word c are treated as positive examples.

•Now other words in the lexicon are sampled randomly to obtain negative examples.

•Then logistic regression is used to train a classifier to distinguish the two types of cases.

•The learned weights are used as embeddings.



4.a. Write python code to apply SVM algorithm for sentiment classification.

```
# Create sample dataset
data = pd.read_csv('sample.csv')
df = pd.DataFrame(data)
# Split the dataset
X_train, X_test, y_train, y_test = train_test_split(
    df["text"], df["label"],
    test_size=0.3,
    random_state=42
)
# Convert text to TF-IDF features
vectorizer = TfidfVectorizer(stop_words="english")
X_train_tfidf = vectorizer.fit_transform(X_train)
X_test_tfidf = vectorizer.transform(X_test)
# Train SVM model
svm_model = LinearSVC()
svm_model.fit(X_train_tfidf, y_train)
# Predict sentiments
y_pred = svm_model.predict(X_test_tfidf)
# Evaluate the model
```

```
print("Accuracy:", accuracy_score(y_test, y_pred))
print("\nClassification Report:\n")
print(classification_report(y_test, y_pred))
```

4.b. Discuss the fundamentals of Sentiment analysis.

Sentiment analysis, or opinion mining, is the process of analyzing large volumes of text to determine whether it expresses a positive sentiment, a negative sentiment or a neutral sentiment. Sentiment analysis involves several stages — from cleaning the text data to classifying it into positive, negative, or neutral categories.

The main steps are:

1 **Text Preprocessing**
2 **Feature Extraction**
3 **Sentiment Classification**

1. Text Preprocessing

Before analyzing sentiments, the raw text data must be cleaned and standardized.
This step improves the quality and accuracy of the model.

a. Cleaning the text

Remove punctuation: Characters like , . ! ? do not affect sentiment directly.
Example: *"I love this movie!"* → *"I love this movie"*
Remove stop words: Common words (like *is, the, and, of*) that don't contribute to sentiment.
Remove special characters/numbers: To focus only on meaningful text.

b. Tokenization

Splitting text into smaller units called **tokens** (words or phrases).
Example:
"I love this movie" → ["I", "love", "this", "movie"]

c. Normalization

Lowercasing: Converts all words to lowercase → ensures uniformity.
Example: *"Love"* and *"love"* are treated the same.
Stemming: Reduces words to their **root form** by chopping endings.
Example: *"loved", "loving"* → *"love"*
Lemmatization: Converts words to their **dictionary base form** (more accurate).
Example: *"better"* → *"good"*, *"running"* → *"run"*

2. Feature Extraction

After preprocessing, the text is converted into numerical format so that machine learning models can process it.

Techniques:

Bag of Words (BoW):

Represents text as a set of word counts (frequency of each word).
Ignores grammar and order.
Example:
"I love movies" → {I:1, love:1, movies:1}
TF-IDF (Term Frequency – Inverse Document Frequency):
Weights words based on how important they are to a document.
Common words (like "the") get less weight; rare but important words get more.
Word Embeddings (Semantic Representations):

Captures meaning and context of words.
Each word is represented as a dense vector.
Examples:
Word2Vec and GloVe: Capture semantic similarity.
BERT: Context-aware embeddings; considers the position and meaning of words in sentences.
3. Sentiment Classification
Once the text is represented numerically, models are trained to classify the sentiment as positive, negative, or neutral.
Common Models:
Traditional Machine Learning Models:
Naïve Bayes: Uses probability based on word frequencies.
Logistic Regression: Assigns weights to features and predicts the sentiment.
SVM (Support Vector Machine): Finds boundaries separating different sentiment classes.
Deep Learning Models:
LSTM (Long Short-Term Memory): Captures sequence and context in text (useful for long sentences).
BERT (Bidirectional Encoder Representations from Transformers):
Understands both left and right context.
achieves state-of-the-art performance in sentiment analysis tasks.

5.a. What is centrality? Explain Degree Centrality in detail.

Centrality -Centrality defines how important a node is within a network.

Degree centrality transfers the same idea into a measure. The degree centrality measure ranks nodes with more connections higher in terms of centrality. The degree centrality Cd for node vi in an undirected graph is

Cd(vi) = di where di is the degree (number of adjacent edges) of node vi

. In directed graphs, we can either use the in-degree, the out-degree, or the combination

as the degree centrality value:

$$\begin{aligned} C_d(v_i) &= d_i^{\text{in}} && \textit{(prestige)}, \\ C_d(v_i) &= d_i^{\text{out}} && \textit{(gregariousness)}, \\ C_d(v_i) &= d_i^{\text{in}} + d_i^{\text{out}}. \end{aligned}$$

When using in-degrees, degree centrality measures how popular a node Prominence or is and its value shows prominence or prestige. When using out-degrees, it measures the gregariousness of a node. When we combine in-degrees and out-degrees, we are basically ignoring edge directions.

Simple normalization methods include normalizing by the maximum possible degree,

$$C_d^{\text{norm}}(v_i) = \frac{d_i}{n-1},$$

where n is the number of nodes. We can also normalize by the maximum degree,

$$C_d^{\max}(v_i) = \frac{d_i}{\max_j d_j}.$$

Finally, we can normalize by the degree sum,

$$C_d^{\text{sum}}(v_i) = \frac{d_i}{\sum_j d_j} = \frac{d_i}{2|E|} = \frac{d_i}{2m}.$$

5.b. Explain Community detection method in detail.
Balanced Communities

Balanced communities arise from **graph-based clustering**, where a social network is represented as a graph and partitioned into groups (communities). A common approach is to apply **graph cuts**, where the graph is divided into subsets by removing edges.

A **minimum cut** minimizes the number (or total weight) of edges removed. However, minimum cuts often produce **highly unbalanced partitions**, such as one node forming a community while all others form another. Such partitions are not meaningful in social network analysis.

To overcome this, **balanced cuts** are preferred. Two important formulations are:

- **Ratio Cut**: Normalizes the cut size by the number of nodes in each partition.
- **Normalized Cut**: Normalizes the cut size by the total degree (volume) of each partition.

These approaches encourage **communities of comparable size**, leading to more natural groupings. Since optimizing these cuts is NP-hard, **spectral clustering** is used as an approximation, leveraging eigenvectors of the **graph Laplacian** to find balanced communities.

D**ense Communities**

Dense communities are groups where **interactions among members are frequent**. The defining property is **high internal edge density**.

Graph density is defined as the ratio of existing edges to the maximum possible edges. A graph is considered **γ-dense** if it contains at least $\gamma|V|^2$ edges.

- A **clique** is a fully connected dense community.
- A **quasi-clique** (or γ-clique) is a connected dense subgraph that is not fully connected but still has strong internal connectivity.

Dense communities are particularly important in **social media analytics**, where sufficient interaction is needed for reliable statistical analysis.

Robust Communities

Robust communities are designed to be **resilient to failures or removals**.

A community is considered:

 k-vertex connected if at least k nodes must be removed to disconnect it. **k-edge connected** if at least k edges must be removed to disconnect it.

Such communities maintain connectivity even under disruptions, making them useful for analyzing **stable social groups**, activist networks, or resilient communication structures.

Modular Communities

Modular communities are defined based on **modularity**, which measures how different a community structure is from a random graph with the same degree distribution.

Modularity compares:

- The actual number of edges within communities
- The expected number of edges if connections were random

High modularity indicates:

- Strong internal connectivity
- Weak external connectivity

Community detection algorithms aim to **maximize modularity**, often using spectral methods on the **modularity matrix**. Modular communities are useful for identifying **interest-based or topic-based groups** in social networks.

Hierarchical Communities

Hierarchical communities capture the **multi-level structure** of social networks, where communities exist within larger communities.

Hierarchical clustering can be:

- **Agglomerative** (bottom-up): merging small communities
- **Divisive** (top-down): splitting large communities

A well-known method is the **Girvan–Newman algorithm**, which:

- Computes **edge betweenness** (number of shortest paths passing through an edge)
- Iteratively removes edges with high betweenness
- Gradually reveals community hierarchies

The output is often represented using a **dendrogram**, showing nested community structure.

6.a. What is meant by influencer identification and user engagement analytics.

Influencer Identification
Influencer identification refers to the process of finding users on social media who have the ability to influence the opinions, behaviors, or decisions of others. These users usually have a large follower base, high interaction rates, or occupy central positions in social networks. Techniques such as network analysis (degree, betweenness, and eigenvector centrality), content analysis, and interaction patterns (likes, shares, retweets, mentions) are used to identify influencers. Influencer identification is important for marketing, awareness campaigns, information diffusion, and detecting opinion leaders in online communities.

User Engagement Analytics
User engagement analytics focuses on measuring and analyzing how users interact with content on social media platforms. It includes metrics such as likes, comments, shares, retweets, replies, click-through rates, and time spent on content. Engagement analytics helps understand user interest, participation level, and content effectiveness. By analyzing engagement patterns, organizations can evaluate campaign performance, improve content strategies, and understand audience behavior over time.

6.b. Explain how do you perform time series analysis.

Time series analysis is a statistical technique used to analyze **data points collected over time** at regular intervals (daily, monthly, yearly). The primary objective is to understand underlying patterns and **forecast future values**.

Time series in social media is needed as Detect emerging events.

- Predict virality patterns.
- Identify cyberbullying waves.
- Track misinformation spread.
- Forecast engagement and user activity. **Types of data:**
- Stationary: constant mean & variance.
- Non-stationary: changes in mean/variance over time.
- Stationarity is required for many classical models.

ARIMA (AutoRegressive Integrated Moving Average) is one of the most widely used models for time series forecasting. It combines three components to model temporal dependence.

ARIMA(p, d, q):

- **p (AutoRegressive order)**: Number of past values used
- **d (Integrated order)**: Number of times data is differenced to achieve stationarity
- **q (Moving Average order)**: Number of past forecast errors used

Model Components:

- **AutoRegressive (AR)**: Relationship between an observation and previous observations
- **Integrated (I)**: Differencing to remove trend and make data stationary

- **Moving Average (MA)**: Relationship between an observation and past errors

Steps in ARIMA Modeling

1. Visualize the time series
2. Check stationarity (ADF test)
3. Apply differencing if needed
4. Identify p and q using ACF and PACF plots
5. Train the ARIMA model
6. Forecast future values
7. Evaluate model performance

7.a. Discuss the steps of Dashboard creation using Tableau.

1. Create a New Dashboard

A dashboard is created in Tableau in a manner similar to creating a new worksheet. At the bottom of the Tableau workbook, the user clicks on the **New Dashboard** icon. This opens a blank dashboard canvas where different visualizations (worksheets) and objects can be placed and arranged

2. Add Sheets to the Dashboard

Once the dashboard is created, existing **worksheets (views)** are added to it. From the **Sheets** list on the left side, the required views are dragged and dropped onto the dashboard area. If a visualization needs to be replaced, the existing sheet in the dashboard can be selected, and another sheet can be swapped in using the **Swap Sheets** option

3. Add Interactivity to the Dashboard

Interactivity is a key strength of Tableau dashboards. A sheet can be enabled as a filter by selecting the **Use as Filter** option. This allows user selections in one visualization to dynamically filter data in other views within the dashboard.

Additionally, Tableau supports **dashboard actions**, such as filter actions, highlight actions, navigation actions, and URL actions, enabling richer user interaction and exploratory analysis

.**4. Add Dashboard Objects**

In addition to worksheets, Tableau allows the inclusion of various **dashboard objects** to enhance layout, usability, and visual appeal. These objects are dragged from the **Objects** pane into the dashboard. Common objects include:

- **Horizontal and Vertical Containers** for structured layout and responsive resizing
- **Text Objects** for titles, headings, and descriptions
- **Image Objects** to add logos or visuals, with optional URL links
- **Web Page Objects** to embed web content

- **Blank Objects** to manage spacing and alignment
- **Navigation Objects** to move between dashboards or sheets
- **Download Objects** to export dashboards as PDF, PowerPoint, or images
- **Extension Objects** and **Pulse Metric Objects** to extend dashboard functionality

5. Set Options for Dashboard Objects

Each object added to the dashboard can be customized. By selecting an object and opening its shortcut menu, users can modify properties such as size, formatting, background, borders, and tooltips. These options help improve clarity, accessibility, and user experience

6. Work with Images, Navigation, and Download Objects

Tableau provides detailed configuration options for specific objects:

- **Image Objects** can either embed image files or link to web-based images. Linked images are preferred for large or animated images to improve performance. Images can also include clickable URLs and alternative text for accessibility.
- **Navigation and Download Objects** can be customized with text or images as buttons, tooltip descriptions, and formatting styles to clearly indicate their purpose, such as navigating to another dashboard or exporting data

7. Copy and Reuse Dashboard Objects

Tableau allows copying and pasting of dashboard objects within the same dashboard or across different dashboards and workbooks. However, certain elements such as sheets within dashboards, filters tied to specific sheets, and dashboard titles cannot be copied directly

7.b. Explain the importance of visualizing data using network maps with an example.

Network graphs visualize the relationship between words in a text.

```
text = """Network graphs are powerful tools for visualizing relationships between entities in a dataset.
In text analysis, they can be used to represent relationships between words, phrases, or other elements"""

# Preprocessing
text = re.sub(r'\W+', ' ', text.lower())
tokens = word_tokenize(text)
tokens = [word for word in tokens if word not in stopwords.words('english')]

# Generate bigrams
bigram_list = list(bigrams(tokens))
bigram_counts = Counter(bigram_list)

# Create the bigram network
G = nx.Graph()
for (word1, word2), freq in bigram_counts.items():
```

```
    G.add_edge(word1, word2, weight=freq)

# Draw the network
plt.figure(figsize=(14, 10))
pos = nx.spring_layout(G, k=0.5)
edges = G.edges(data=True)
weights = [edge[2]['weight'] for edge in edges]

nx.draw(G, pos, with_labels=True, node_size=3000, node_color='skyblue', edge_color='gray',
width=weights, font_size=10)
plt.title('Bigram Network Graph')
plt.show()
```

8.a. What are word clouds? Demonstrate its usage with a code snippet.

A **word cloud** is a **text visualization technique** where words are displayed in varying **sizes and colors based on their frequency or importance** in a text corpus.

- **Larger words** appear more frequently or have higher importance.
- **Smaller words** appear less frequently.

```
from wordcloud import WordCloud

import matplotlib.pyplot as plt


# Sample text

text = """

Data visualization is an interdisciplinary field that deals with the graphic representation of data.

It is a particularly efficient way of communicating when the data is numerous as for example a
time series."""


# Generate the word cloud

wordcloud = WordCloud(width=800, height=400, background_color='white').generate(text)


# Display the word cloud using matplotlib

plt.figure(figsize=(10, 5))

plt.imshow(wordcloud, interpolation='bilinear')
```

plt.axis('off') # Remove axes

plt.show()

8.b. Briefly describe how heat maps help in analyzing customer sentiment data.

A **heat map** is a data visualization technique that uses **color intensity** to represent the magnitude of values in a dataset.

Importance: Quick Pattern Recognition, Easy Comparison Across Dimensions, Efficient

Analysis of Large Datasets, Identification of Trends and Anomalies, Better Decision-Making Support.

Ex: import numpy as np

import seaborn as sns

import matplotlib.pyplot as plt

Generate random data

data = np.random.randint(1, 100, (10, 10))

Create a heatmap

sns.heatmap(data)

plt.show()

9.a. Explain the ethical considerations in Social Media Analytics.

Social media analytics involves collecting and analyzing vast amounts of user-generated data to derive insights. While this provides significant benefits, it also raises serious **privacy and ethical challenges**. Addressing these challenges responsibly is essential to protect user rights and maintain trust.

1. Unauthorized Data Collection

Data is often collected from social media platforms without users' explicit knowledge or permission, especially through scraping techniques.

Organizations should collect data only through authorized APIs and follow platform policies. Data collection should be lawful, justified, and limited to stated purposes.

2. Lack of Informed Consent

Users are frequently unaware of how their data is being analyzed or used, as consent mechanisms are unclear or hidden in lengthy terms and conditions.

Adopt **informed and explicit consent**, clearly explaining what data is collected, why it is used, and how long it will be retained.

3. Exposure of Personally Identifiable Information (PII)

Social media data may contain sensitive information such as names, usernames, locations, images, or contact details.

Apply **data anonymization and pseudonymization** techniques to remove or mask personal identifiers before analysis.

4. Data Re-identification Risk

Even anonymized datasets can sometimes be re-identified when combined with other datasets.

Use strong anonymization methods, limit data sharing, and avoid combining datasets in ways that increase re-identification risks.

5. Surveillance and User Profiling

Continuous monitoring of user activity can lead to intrusive surveillance and unfair profiling.**Follow** data minimization **principles by collecting only necessary data and avoiding excessive tracking of individuals.**

6. Data Breaches and Security Risks

Large-scale social media datasets are vulnerable to cyberattacks and data leaks.Ensure **secure data storage and access control** using encryption, authentication, and role-based access mechanisms.

7. Bias and Ethical Misuse

Biased analytics or AI models can lead to discrimination, manipulation, or unfair targeting of specific groups.

Implement **ethical AI and bias mitigation** by auditing models, using diverse datasets, and ensuring fairness in analysis.

8. Transparency and Accountability

Lack of clarity about how analytics systems operate reduces trust and accountability. Maintaintransparency and accountabilityby documenting data practices, model decisions, and enabling audits.

9. Compliance with Privacy Regulations

Failure to comply with legal frameworks can result in legal and ethical consequences. Ensure compliance withprivacy regulationssuch as GDPR, CCPA, and national data protection laws, respecting user rights like data access and deletion.

9.b. Explain how social media analytics can improve brand monitoring and crisis management.

Social media analytics plays a crucial role in helping organizations monitor brand reputation and manage crises effectively. By continuously analyzing social media data, organizations can identify potential risks early, respond promptly, and protect brand image.

Role in Brand Monitoring

Real-time brand mention tracking enables organizations to continuously monitor when and

where their brand is being discussed across social media platforms. This helps in understanding brand visibility and public attention.

Sentiment analysis analyzes user opinions and emotions associated with brand mentions. It helps determine whether public perception is positive, negative, or neutral, providing insights into brand reputation.

Trend and hashtag monitoring helps identify emerging topics, popular hashtags, and discussions related to the brand. This allows organizations to align their marketing strategies with current trends.

Audience engagement analysis evaluates how users interact with brand content through likes, comments, and shares. This helps assess campaign effectiveness and audience interest.

Competitor benchmarking enables comparison of brand performance with competitors in terms of engagement, sentiment, and reach, helping organizations identify strengths and weaknesses.

Influencer identification helps detect influential users who impact brand perception and reach, enabling strategic collaborations and reputation management.

Role in Crisis Management

Early detection of negative sentiment spikes helps identify potential crises at an early stage before they escalate.

Identification of viral negative content and misinformation allows organizations to detect harmful or misleading content spreading rapidly across platforms.

Root-cause analysis of issues helps trace the origin, timeline, and reasons behind negative events or public backlash.

Real-time alerts and dashboards provide instant notifications and visual summaries, enabling quick decision-making and rapid response during crises.

Evaluation of crisis response effectiveness measures how audience sentiment and engagement change after corrective actions are taken.

Post-crisis analysis helps organizations learn from incidents, improve response strategies, and strengthen future crisis preparedness.

10.a. Discuss emerging trends in social media analytics hioghlighting the role of AI driven social insights

Social media analytics has rapidly evolved due to the explosive growth of user-generated content and advances in Artificial Intelligence (AI). Modern analytics systems no longer focus only on measuring past performance but increasingly support prediction, personalization, automation, and strategic intelligence. The major recent trends and the contribution of AI-driven social insights are discussed below.

1. From Descriptive to Predictive & Real-Time Analytics

Earlier, social media analytics was largely **descriptive**, focusing on metrics such as likes, shares, comments, and follower counts to explain *what happened*. Recent trends emphasize **predictive and real-time analytics**, where AI models analyze historical and streaming data to forecast future outcomes.

Machine learning and time-series models predict content virality, audience growth, engagement spikes, and campaign performance in real time. This enables organizations to take proactive actions, such as adjusting campaigns instantly or responding early to emerging issues.

2. AI-Powered Social Listening & Sentiment Analysis

Social listening has become more sophisticated with the use of AI-driven **Natural Language Processing (NLP)**. Instead of manually tracking keywords, AI systems automatically analyze large volumes of unstructured text across platforms.

Deep learning models (e.g., BERT-based models) detect sentiment, emotions (anger, joy, fear), sarcasm, and hate speech with high accuracy. This helps brands understand public perception, monitor reputation, and detect crises or misinformation at an early stage.

3. Hyper-Personalization & Content Optimization

A key trend in social media analytics is the shift toward **hyper-personalized content delivery**. Generic content strategies are being replaced by data-driven personalization. AI analyzes user behavior, interests, engagement patterns, and past interactions to recommend personalized content, optimal posting times, and suitable platforms. This improves user engagement, customer satisfaction, and advertising effectiveness.

4. Automated & AI-Generated Content

Another emerging trend is the use of **generative AI** in content creation and optimization.

Social media analytics tools are increasingly integrated with AI content generators. AI systems generate captions, hashtags, images, and short videos based on engagement insights. Automation reduces manual effort, speeds up content production, and ensures alignment with audience preferences identified through analytics.

5. Competitive Analysis & Trend Detection

Modern social media analytics goes beyond analyzing a single brand to include **competitive intelligence** and market-level trend detection.

AI will help to continuously monitors competitor activity, trending hashtags, topics, and influencer behavior. Pattern recognition and clustering algorithms identify emerging trends and coordinated campaigns, enabling organizations to benchmark performance and adapt strategies quickly.

10.b. Explain the role of social media in advertisement performance and customer engagement

Social media has become a powerful platform for digital marketing, enabling organizations to promote their products and services effectively while engaging customers in meaningful ways. By leveraging data analytics, interactivity, and targeted communication, social media significantly enhances both advertisement performance and customer engagement.

1. Precise Audience Targeting

Social media platforms allow advertisers to target audiences based on demographics, interests, location, behavior, and online activity. This ensures that advertisements reach the most relevant users, increasing click-through rates and conversion rates. Precise targeting reduces wastage of advertising budgets and improves campaign effectiveness.

2. Cost-Effective Advertising

Compared to traditional advertising media such as television and print, social media advertising is more cost-effective. Businesses can run campaigns with flexible budgets using pay-per-click or pay-per-impression models. Even small businesses can reach large audiences at lower costs, achieving better return on investment (ROI).

3. Real-Time Performance Tracking

Social media platforms provide real-time analytics on advertisement performance, including impressions, reach, clicks, and engagement levels. Advertisers can continuously monitor campaign performance and make immediate adjustments to improve results. This real-time feedback enables quick decision-making and better campaign control.

4. Data-Driven Optimization

Social media analytics tools help analyze user interactions and campaign outcomes. Based on data insights, advertisers can optimize ad creatives, messaging, audience segments, and posting schedules. Techniques such as A/B testing further enhance the effectiveness of advertisements through data-driven improvements.

5. Personalized Content Delivery

Social media enables personalized advertising by delivering content tailored to individual user preferences and behaviors. Personalized ads and recommendations increase relevance, leading to higher engagement and better customer experience. AI-driven personalization strengthens the connection between brands and users.

6. Influencer and Community Marketing

Influencer marketing plays a significant role in enhancing advertisement reach and credibility. Influencers help brands connect with their followers in an authentic way. Additionally, brand communities on social media encourage interaction, discussions, and peer influence, further boosting engagement.

7. Improved Customer Feedback and Support

Social media platforms allow customers to provide instant feedback, reviews, and queries. Brands can respond quickly to customer concerns, improving satisfaction and trust. This twoway communication strengthens customer relationships and enhances brand image.

8. Brand Awareness and Loyalty

Consistent advertising and engagement on social media increase brand visibility and recall. Interactive campaigns, personalized communication, and prompt responses foster trust and loyalty among customers. Over time, this leads to long-term customer relationships and brand advocacy.