Module 1

1. What is predictive analytics? (5 mks)
2. Challenges of social media analytics.
3. Explain briefly seven layers of social media analytics with an example.
4. Why it is important for business managers to understand and mine social media data?
5. What is social media analytics, and how it is different from traditional business analytics?
6. Explain the social media analytics cycle.
7. What ethical issues should be considered when mining social media data?

1. What is predictive analytics? (5 mks)

Predictive analytics is the process of analyzing current and historical data to **make predictions about future events, behaviors, or outcomes**. It helps organizations **anticipate what might happen**, so they can plan better and make smarter decisions.
It uses a combination of:

- **Statistics** (to find trends and relationships in data),
- **Machine learning** (to learn from data patterns automatically),
- **Artificial intelligence** (to improve prediction accuracy), and
- **Mathematical models** (to calculate likely future scenarios).

By studying how things happened in the past and what's happening now, predictive analytics helps estimate future results — from **a few seconds ahead** to **years into the future**.

Key Uses of Predictive Analytics:

- **Fraud Detection:** Banks use it to spot unusual transactions.
- **Customer Behavior:** Companies predict what products customers are likely to buy.
- **Inventory Planning:** Retailers forecast stock needs for festive or peak seasons.
- **Weather Forecasting:** Agencies predict floods, storms, or temperature changes.
- **Healthcare:** Doctors use it to predict patient health risks.

Example:
A **video streaming platform** like Netflix uses predictive analytics to recommend shows or movies based on what you've already watched. It learns your interests and predicts what you'll want to watch next.

Why is Predictive Analytics Important?

- Helps reduce risks
- Improves decision-making
- Saves time and money
- Personalizes user experience

2. Challenges of social media analytics.

1. Volume and Velocity

Explanation:

Social media platforms generate massive amounts of data at high speeds (real-time updates, likes, shares, comments). Processing such huge and fast-changing data requires high computational power and efficient algorithms.

Example:

Twitter users post over 500 million tweets per day. Analyzing trending topics in real-time (e.g., during an election or natural disaster) can overwhelm standard processing tools.

2. Diversity Challenge

Explanation:

Social media content is extremely diverse—users differ in language, culture, demographics, and intentions. Additionally, posts include text, emojis, images, audio, and video, making it harder to interpret content consistently.

Example:

An Indian user may post a meme in Hindi using Hinglish (a mix of English and Hindi), while another in Japan posts in Japanese with localized slang. Analyzing sentiment across such varied content is complex.

3. Unstructuredness

Explanation:

Most social media data is unstructured—text without a clear format, including slang, sarcasm, abbreviations, and multimedia. This makes traditional structured data analysis techniques ineffective.

Example:

A tweet like “LOL this gov is totally lit #electionday” expresses sarcasm or strong emotion but is difficult for algorithms to interpret correctly without contextual understanding.

4. Noise and Misinformation

Explanation:

A significant portion of social media data includes spam, bots, fake news, or misleading content. This ‘noise’ distorts analysis and can mislead conclusions.

Example:

During a crisis, automated bots might spread false information about rescue operations, leading to public confusion and poor policy decisions.

5. Privacy and Ethical Concerns
Explanation:
Analyzing user data on social platforms raises concerns about consent, privacy, and ethical use—especially when dealing with personal or sensitive content.
Example:
Using Facebook friend graphs to study mental health trends without user consent may violate ethical standards, even if data is technically "public."

6. Temporal Dynamics
Explanation:
Social networks evolve over time—new users join, old users leave, relationships form and dissolve. Analysis must account for these changes to be meaningful.
Example:
A user might be very influential in 2020 but become inactive in 2022. Static analysis might falsely assume they still hold influence.

7. Ambiguity in Relationships and Context
Explanation:
Relationships in social networks can be ambiguous. A "like" can indicate support, sarcasm, or mere acknowledgment. Similarly, context changes meaning.
Example:
A comment like “Nice job!” on a political post might be genuine or sarcastic, depending on the commenter’s stance and tone, which is hard to detect computationally.

8. Scalability and Computational Complexity
Explanation:
SNA often involves operations on graphs with millions of nodes (users) and edges (connections). Ensuring algorithms scale effectively is a major challenge.
Example:
Running a community detection algorithm on the entire Facebook friendship graph is computationally intensive and not feasible on average hardware.

3. Briefly explain the seven layers of social media data. Support your answer with examples.

- Social media has a **minimum seven layers of data**
- Each layer carries potentially valuable information and insights that can be harvested for business intelligence purposes.
- Out of the seven layers, some are visible or easily identifiable (e.g., text and actions) and other are invisible (e.g., social media and hyperlink networks).
- The following are seven social media layers

1. Text Layer
Explanation: The text layer consists of user-generated content such as comments, tweets, blog posts, product reviews, and status updates. Text analysis (or text mining) extracts hidden insights, patterns, and sentiment from these textual elements for business intelligence.
Example: A company launching a new smartphone can analyze tweets and Facebook comments to gauge public sentiment about its features and identify common concerns.

2. Network Layer
Explanation: This layer involves connections and relationships between users, such as followers, friends, and interactions on social media. Social Network Analysis (SNA) identifies influential users, communication patterns, and information flow.
Example: A social media platform like LinkedIn uses network analytics to recommend connections by analyzing mutual friends and shared professional interests.

3. Actions Layer
Explanation: Actions include interactions such as likes, comments, shares, clicks, and views. Analyzing these actions helps measure the popularity and influence of content or products on social media.
Example: A company can analyze Facebook likes and Twitter mentions to determine how well a new marketing campaign is performing and adjust strategies accordingly.

4. Hyperlink Layer
Explanation: Hyperlinks connect web pages and social media posts to external resources. Analyzing hyperlink patterns helps understand website traffic, content reach, and user behavior.
Example: An online news website tracks which article links receive the most clicks to determine the most engaging topics and optimize future content.

5. Mobile Layer
Explanation: This layer focuses on mobile social media interactions, including device type, operating system, and user engagement with mobile apps. It helps optimize app performance and user experience.
Example: A fitness app like Fitbit analyzes mobile user data to determine which devices are most commonly used and prioritizes updates accordingly.

6. Location Layer
Explanation: Location-based data reveals where users access social media, helping businesses tailor their marketing strategies to specific geographic areas.
Example: A ride-sharing app like Uber uses location data to identify high-demand areas and adjust driver availability in real time.

7. Search Engine Layer
Explanation: Search engine data includes how users find content through search queries, keywords, and search trends. This layer helps businesses optimize their content for better visibility.
Example: An e-commerce website like Amazon analyzes search queries to understand which products are trending and adjust inventory and advertisements accordingly.

Each of these layers plays a crucial role in extracting insights from social media, enabling businesses to make data-driven decisions.

Example : A Restaurant's Social Media Marketing Campaign

- The restaurant posts a description of a new dish across its social media platforms, including appealing visuals and hashtags to attract attention. "Introducing our signature Spicy Thai Noodles—bursting with authentic flavors! 🍜 #FoodieHeaven“ ….(**Text, image….Multimedia**)
- Geotargeted ads ensure the post appears to users within 10 miles of the restaurant…..(**Location**)
- The post reaches a broader audience ..*A customer tags friends on Instagram to check out the dish*…..(**Network**)
- A search for "Spicy Thai Noodles near me" leads the customer to the restaurant’s website and reviews….. (**Search Engine**)
- The post contains a clickable link leading to the restaurant’s website, menu, or reservation system. Reserve your table now: www.ourrestaurant.com/reservations… (**Hyperlink**)
- A customer browsing Instagram clicks the link to book a table during their lunch break…. (**Mobile)**
- After a good experience, Users comment, "Looks delicious! What’s the price?" or share it on their own feeds….(**Action**)

4. Why it is important for business managers to understand and mine social media data?

1. Latest Trends in Society: The data available from social media platforms can give important insights regarding society and user behavior. It is a process that starts with identifying the target audience and ends with digging into what they are passionate about. Businesses may analyze the keywords, search results, comments, and mentions to identify the current trend, and a deeper study of behavior change can also help in predicting future trends. This data is very useful for businesses to make informed decisions when the stakes are high.

2. Sentiment Analysis: Sentiment Analysis is the process of identifying positive or negative sentiments portrayed in information posted on social media platforms. Businesses use Social Media Mining to identify the same sentiments associated with their brand and product lines. When combined with social media monitoring, sentiment analysis can help you analyze your brand image and bring negative aspects of the business to your attention. With this information, you can address the negative sentiments and prioritize them so that they can be addressed properly to improve the customer experience.

3. Keyword Identification: Keywords are those words that reveal the behavior of users and highlight the frequently used and popular terms related to their products. Social Media Data Mining can be highly effective in finding these keywords. The process is as basic as scanning the list of the most frequent words or phrases used by customers to search for or define your product. Using these keywords to define your product in digital media and implementing SEO can yield its pretty good results. Your product will rank higher, and by implementing frequent and popular terms, you can make your product listings better.

4. Competitor Analysis: Analyzing competitor behavior on social media during the launch of a product will help you define a trend and use it to your advantage. Posts by competitor employees and management regarding hiring may give you an idea of the expansion of business or even a subtle change in operations will help you to be proactive. Having an idea of when to stay on your toes is advantageous in highly competitive industries.

5. What is social media analytics, and how it is different from traditional business analytics?

- Social media analytics is the process of **extracting valuable insights** from **social media data to perform decision making.**
- By analyzing social media data, **businesses can increase brand loyalty, generate leads, drive traffic, and make forecasts.**
- Social media analytics can also be used to increase awareness of a brand and drive users to a website for the latest news and information.
- Social media growth, tools, and big data also present huge opportunity to market products, enhance brand loyalty, network with customers, crowd-source ideas, drive sales, and mine business insights.
- The purpose of social media analytics is to enable **informed decision making** by leveraging social media data.

Social Media Vs Traditional Business Analytics

- The main difference between social media analytics and traditional business analytics is the **source, type, and nature of the data** being mined.
- Social media data is **diverse, high volume, real-time, and stored in third-party databases** in semistructured and unstructured formats, while traditional business data is mostly stored in **databases and spreadsheets in machine-readable** i.e. **Structured** format.
- Social media data is **socialized in nature and originates from the public internet**, while traditional business data is **bureaucratic and formal in nature and is controlled by organizations**.
- Social media data gains value when shared widely, while traditional business data is valuable when kept within a company for competitive advantage.

Social Media Vs Traditional Business Analytics

Social Media Analytics	Business Analytics
Semistructured and unstructured data	Structured data
Data is not analytical friendly	Data is analytical friendly
Real-time data	Mostly historical data
Public data	Private data
Stored in third-party databases	Stored in business-owned databases
Boundary-less data (i.e., Boundary within the Internet)	Bound within the business intranet
Data is high volume	Data is medium to high volume
Highly diverse data	Uniform data
Data is widely shared over the Internet	Data is only shared within organizations
More sharing creates greater value/impact	Less sharing creates more value
No business control over data	Tightly controlled by business
Socialized data	Bureaucratic data
Data is informal in nature	Data is formal in nature

6. Explain the social media analytics cycle.

The social media analytics process involves six steps to mine desired business insights from raw social media data.
The process begins with defining business goals and objectives, and continues until these objectives are fully satisfied.

Step 1 - Identification

- The identification stage of social media analytics involves **finding the right sources of data to analyze** in order to gain valuable business insights.
- The **data** should be aligned with the **business's objectives** and can come from both official business-owned platforms, such as social media accounts and blogs, and nonofficial platforms such as Google search trends or Twitter search stream data.
- It is important to consider the business objectives when identifying the sources and types of data to be analyzed.

Step 2 - Extraction

- The extraction stage of social media analytics involves using **appropriate methods and tools to gather data from identified sources.**
- The type (e.g., text, numerical, or network) and size of data will determine the method and tools suitable for extraction.
- This can include manual data collection for small-scale data and automated extraction using APIs (application programming interfaces) for larger data sets.
- It is important to **consider privacy and ethical issues** when mining data from social media platforms and to have a clear social media privacy policy in place to ensure that data handling and extraction practices do not violate user privacy.
- Specialized tools may be needed to extract certain types of data, such as social network and hyperlink network data.

Step 3 - Cleaning

- The cleaning step in social media analytics involves **removing unwanted data** from the collected data set.
- This can involve processes such as **cleaning, coding, filtering, clustering, and natural language processing** to remove irrelevant data.
- Both **automated and manual techniques** may be used for cleaning, depending on the type of data and the desired level of accuracy.

Step 4- Analyzing

- The analyzing stage of social media analytics involves using clean data to **identify valuable insights for the business.**
- The approach and techniques used will depend on the type of data being analyzed and the tools and algorithms employed.
- It is important **to maintain the integrity of the data** while **extracting meaningful insights** and to have a good understanding of the capabilities of the tools being used.

Step 5 - Visualization

- The visualization step in social media analytics involves **creating visual representations of the results of the analysis**
- Visualization can help reveal hidden patterns, relationships, and trends in complex and large data sets
- Different types of types of visualizations include **charts, graphs, trees, heat maps, 3D views**
- **Effective visualization** is important for effectively communicating the results of the analysis to top management
- **Text analytics** can result in a word concoccurrence cloud; **hyperlink analytics** will provide visual hyperlink networks; and **location analytics** can produce interactive maps.

Step 6 - Interpretation

- Interpret and translate analytics results into a meaningful business problem.
- Two strategies or approaches used are:

  1) Producing easily consumable analytical results --training data scientists and analysts to produce interactive and easy-to-use visual results

  2) Improving analytics for understanding the consumption capabilities

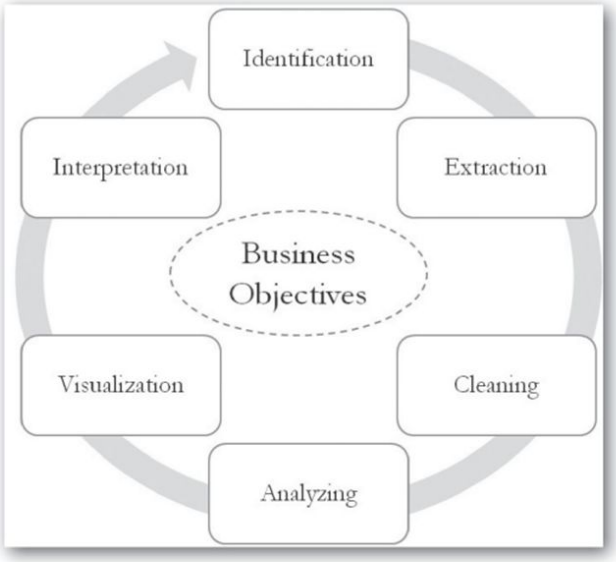


Figure 3. Social media analytics cycle

7. What ethical issues should be considered when mining social media data?

When mining social media data, several ethical issues must be considered to ensure responsible and lawful use of the collected information:

1. **Privacy Concerns** – Users may not be aware that their data is being collected and analyzed. Even publicly available data can include sensitive personal information. Ensuring proper anonymization and obtaining user consent are critical.

2. **Informed Consent** – Many users share content without explicitly consenting to its use for data mining. Ethical data collection practices should involve clear disclosures and, where possible, user consent.

3. **Data Security** – Handling large amounts of social media data increases the risk of breaches and unauthorized access. Strong security measures must be implemented to protect user information.

4. **Bias and Discrimination** – Algorithms analyzing social media data may reinforce biases present in the dataset, leading to unfair outcomes, such as discrimination in law enforcement, hiring, or social profiling. Ensuring fairness and transparency in AI models is essential.

5. **Misuse of Data** – Data collected from social media can be misused for surveillance, political manipulation, targeted advertising, or misinformation campaigns. Ethical guidelines should be followed to prevent harmful applications.

6. **Transparency and Accountability** – Organizations and researchers should be transparent about their data collection methods and intended use. Users should have the ability to access and understand how their data is being used.

7. **Legal Compliance** – Different regions have varying laws, such as GDPR (Europe) and CCPA (California), which regulate data collection and user rights. Compliance with these regulations is necessary to avoid legal consequences.

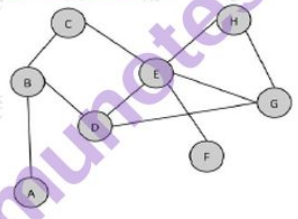
8. **Psychological Impact** – The mining of personal conversations, emotions, and behaviors can lead to unintended psychological harm, such as profiling individuals for mental health conditions without consent. Ethical considerations must include user well-being.

To address these issues, ethical frameworks and guidelines should be established to ensure social media data mining is conducted responsibly, balancing innovation with privacy, fairness, and respect for user rights.

Module 2

1. How degree distribution is plotted for the graph? Show degree distribution of the following graph. (May 24)
2. Centralization in social media analytics with example.
3. Define centrality and its types. How is it computed?
4. Qs.2.a May 2023
5. Consider the hypothetical scenario within the social media network LinkedIn, where Ramesh, an engineering graduate, maintains connections with 10 of his friends. Ramesh is represented as a node in the LinkedIn network, with each connection reflecting a professional link to a friend. The structure of this network, shaped by these connections, can be analyzed using various network measures.Explain following terms in context of above case study: Degree distribution, density of connections, tie strength and hubs
6. Compare different social media analytics tools available in the market and explain their strengths and weakness.
7. Describe density, bridge, hub of a social network with example. (5mks)
8. Define Clique and Cluster with example
9. What are the different scale issues that occur in network visualization?(5M) (gpt se liya....ppt mai mil nai raha tha)

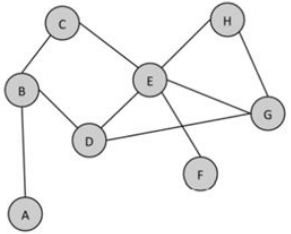
Q.2 a. Answer the following questions about this graph. 10



i. How many nodes are in the network?
ii. How many edges are in the network?
iii. Is this graph directed or undirected?
iv. Create an adjacency list for this graph.
v. Create an adjacency matrix for this graph.
vi. What is the length of the shortest path from node A to node F?
vii. What is the largest clique in this network? How many cliques of that size are there?
viii. How many connected components are there in this network?
ix. Estimate the density of the graph?
x. Are there any hubs in the network? If so, which node (s) and why is it a hub?

1. **List the 5 people you are closest to on facebook. Turn this list into a network by listing all the connections between these people.**
   - Is your network directed or undirected?
   - What do the edges represent (friendship, family relationship, close relationships, acquaintances, etc.)?
   - Give the adjacency list for the network.
   - Give the adjacency matrix for the network.
   - Are there any singletons?

Exercise



1. Answer the following questions about this graph.

a. How many nodes are in the network?

b. How many edges are in the network?

c. Is this graph directed or undirected

d. Create an adjacency list for this graph.

e. Create an adjacency matrix for this graph

f. What is the length of the shortest path from node A to node F?

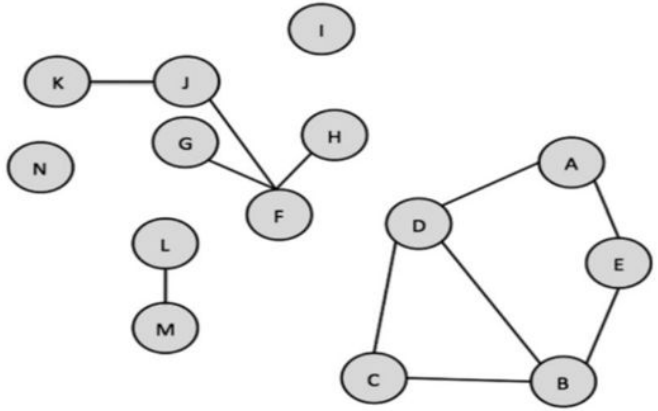
g. What is the largest clique in this network? How many cliques of that size are there?

h. How many connected components are there in this network?

i. Draw the 1.5 ego network for node E (without including node E in the graph). How many singletons are in the ego network?

j. Are there any hubs in the network? If so, which node(s) and why is it a hub?

2. Consider this graph



a. How many singletons are there in the network? List them.
b. What is the largest connected component?
c. Are there any bridges in the network? If so, where are they?
d. Create an adjacency list for the network.
e. Create an adjacency matrix for the network.

Q. How degree distribution is plotted for the graph? Show degree distribution of the following graph. (May 24)

1. **Understand Degree of a Node**:
    - The **degree** of a node is the number of connections (edges) it has.
    - In **directed graphs**, this is split into:
        - **In-degree**: Number of edges coming into a node.
        - **Out-degree**: Number of edges going out from a node.
2. **Count the Degrees**:
    - Go through every node in the graph and **count how many connections each node has** (its degree).
    - Example: If Node A connects to 3 others, its degree is 3.
3. **Create a Frequency Table**:
    - Count **how many nodes have each degree**.
    - Example: 5 nodes have degree 2, 3 nodes have degree 3, etc.
    - Include **degrees that occur zero times** (e.g., no node has degree 6).
4. **Plot the Distribution**:
    - On a graph/chart:
        - The **x-axis** represents the **degree values** (0, 1, 2, 3, ...).
        - The **y-axis** shows the **number of nodes** with that degree.
    - This chart is called the **degree distribution plot**.

5. Interpret the Distribution:

- This plot helps understand the structure of the network:
  - Are most nodes low-degree (few connections)?
  - Are there hubs with very high degree?
- For example, in a **Facebook friendship graph**, this might show how many friends people have on average.

6. Applications:

- Degree distribution helps analyze:
  - Social media networks (e.g., Facebook, Twitter).
  - Web links.
  - Trust and tie strength in networks.
- It also supports identifying influencers and understanding **network robustness**.

Centralization in social media analytics with example.

- **Centralization uses the distribution of a centrality measure to understand the network as a whole.**

- Since there are different centrality measures (e.g., betweenness, closeness, etc.), there are different centralization measures for a graph

- Centralization is computed by looking at the sum of the differences in centrality between the most central node and every other node in the network, and dividing this by the maximum possible difference in centrality that could exist in the graph
- Let C(n) be the centrality of node n, using whatever centrality measure we choose.
- Say n is the most central node. We want to find the difference in centrality between n and every other node in the network, and add those up.
- If there are N nodes in the network, the formula for this is:

$$\sum_{i=1}^{N} C(n^*) \quad C(n_i)$$

- Then, we want to divide this by the sum of the maximum possible differences between n and every other node.
- However, this maximum possible centrality will change depending on which centrality measure we are using. Denote this by using the same formula with max in front.

$$\max \sum_{i=1}^{N} C(n^*) \quad C(n_i)$$

- Now, we can compute centralization. It is equal to the sum of the differences (the first formula) divided by the maximum possible sum of differences (the second formula):

$$\frac{\sum_{i=1}^{N} C(n^*) \quad C(n_i)}{\max \sum_{i=1}^{N} C(n^*) \quad C(n_i)}$$

Why It Matters in Social Media:

- Helps identify **influencers**, **key opinion leaders**, or **information hubs**.
- Useful for **marketing**, **content spreading**, **viral campaigns**, and understanding **information flow**.
- It also shows if the network is **vulnerable**—for example, if removing one central node breaks the whole network apart.

Example in Social Media:

Imagine a Twitter network:

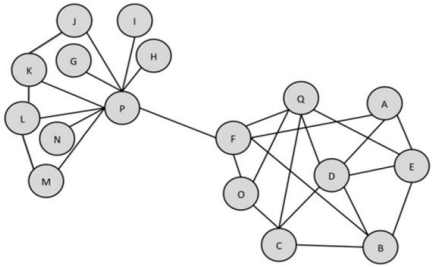
- One user (like a celebrity or brand) has **thousands of followers**, while most others have only a few.
- That user would have a **high centrality**, and the network would be **highly centralized**.
- If that one user shares content, it spreads widely. But if they leave, the network's reach is weakened.

3. Define centrality and its types. How is it computed?

- Centrality is one of the core principles of network analysis. It measures how “central” a node is in the network.
- This is used as an estimate of a node’s importance in the Network
- In a Facebook friendship network, for example, the users who occupy the central position are better positioned to control the flow of social media content
- In network analysis, one or more of these measures may be reported in order to gain a better perspective on the network. A node may appear highly central with one measure but have low centrality with another. That does not mean one measure is incorrect, though; they are simply different ways of describing nodes. The interpretation of the centrality measures is left to a human analyst.

Degree centrality :

- The degree centrality of a node is simply its degree... the number of edges the node has.
- **The higher the degree, the more central the node is.**
- Node P has the highest degree centrality of 9. Meanwhile, node F has a relatively low degree centrality of 5. Many other nodes have that same centrality value or higher (e.g., node D has a degree centrality of 5).



Closeness centrality :

- **Closeness centrality indicates how close a node is to all other nodes in the network.**
- It is calculated as the average of the shortest path length from the node to every other node in the network.
- The benefit of closeness centrality are that it indicates nodes as more central if they are closer to most of the nodes in the graph.

Closeness Centrality of Node D :

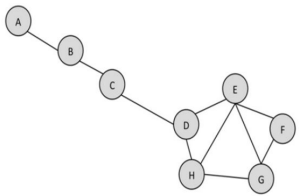


Table 3.1 The Shortest Path Lengths from D to each Other Node in the Network

Node	Shortest Path from D
A	3 (D C B A)
B	2
C	1
E	1
F	2
G	2
H	1

- The table shows each node and the length of the shortest path from D.
- The average of those shortest path lengths is:
  (3 + 2 + 1 + 1 + 2 + 2 + 1) ÷ 7 = 12 ÷ 7 = 1.71

Closeness Centrality of Node A :

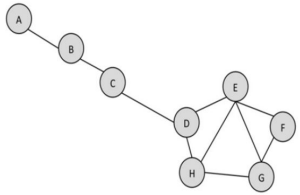


Table 3.2 The Shortest Path Length from node A to Every Other Node in the Network

Node	Shortest Path from A
B	1
C	2
D	3
E	4
F	5
G	5
H	4

- The table shows each node and the length of the shortest path from A.
- The average of those shortest path lengths is:
  (1 + 2 + 3 + 4 + 5 + 5 + 4) ÷ 7 = 24 ÷ 7 = 3:43

- In the case of closeness centrality, or average shortest path length, lower values indicate more central nodes.
- Thus, since node D's closeness centrality is 1.71 and node A's is 3.43, node D is more central by this measure.

Betweenness centrality :

- Betweenness centrality measures how important a node is to the shortest paths through the network.
- **Betweenness centrality measures the number of times a node lies on the shortest path between other nodes.**
- In directed networks, betweenness can have a different meaning. A user with high betweenness may be followed by many others who don't follow the same people as the user. This would indicate that the user is well-followed.
- Betweenness centrality can in fact be used in optimizing **City Transport** planning.
- in a **Telecommunication Network**, a node with higher betweenness centrality would have more control over the network, because more information will pass through that node.
- In online **Social Networks** a high betweenness centrality coincides with nominations of closest friends (i.e., strong interpersonal ties), because it reflects social capital investments into the relationship when distant social circles (e.g., family and university) are bridged.

- To compute betweenness for a node N,
  - we select a pair of nodes and find all the shortest paths between those nodes.
  - Then we compute the fraction of those shortest paths that include node N.
  - Sum up all of them
- If there were five shortest paths between a pair of nodes, and three of them went through node N, then the fraction would be 3 ÷ 5 = 0.6.
- We repeat this process for every pair of nodes in the network.
- We then add up the fractions we computed, and this is the betweenness centrality for node N.

Eigenvector centrality :

- Eigenvector centrality **measures a node's importance while giving consideration to the importance of its neighbors.**
- It is determined by performing a matrix calculation to determine what is called the principal eigenvector using the adjacency matrix.
- For example, a node with 300 relatively unpopular friends on Facebook would have lower eigenvector centrality than someone with 300 very popular friends (like Barak Obama).
- Eigenvector centrality is at the core of **Google's PageRank algorithm**, which they use to rank web pages.
- It is sometimes used **to measure a node's influence** in the network.

Chapter 2-Point 2.pptx - Google Slides

For example of Betweenness centrality refer this ppt

7. Describe density, bridge, hub of a social network with example.

Density

- **Definition**:
  Density measures **how connected** the nodes in a network are. It shows the **proportion of actual connections** to all possible connections.
- **Formula**:

$$\text{Density} = \frac{\text{Number of actual edges}}{\text{Number of possible edges}}$$

For undirected networks:

$$\text{Possible edges} = \frac{n(n-1)}{2}$$

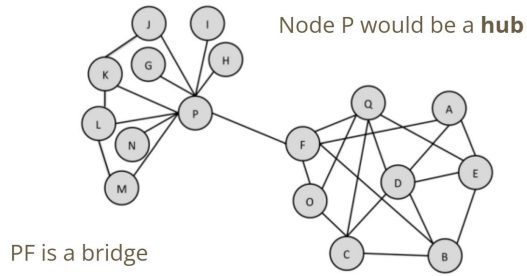
- **Example**:

  In a Facebook group of 8 people:
  - If everyone is friends with everyone (28 connections), density = 1 (a **clique**).
  - If only 5 connections exist, density = 5 / 28 ≈ 0.18, meaning **low density**.

- **Use**:
  High density means **strong internal communication** (like a tight friend group). Low density suggests **fewer connections**, possibly a larger but loosely connected community

Bridges and hubs

- **A bridge i**s an edge that connects two otherwise separate groups of nodes in the network. Formally, a bridge is an edge that, if removed, will increase the number of connected components in a graph.
- **Use**:
  Bridges are important for **spreading information** across otherwise disconnected groups. They play a big role in **network cohesion**.

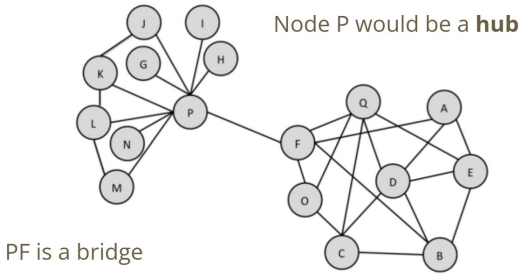


Hub - The term is used to refer to the most connected nodes in the network. E.g. Node P would be a hub because social media, hubs can be used to identify individuals or groups that have a large number of connections or followers within the platform.

- They are often considered as **key influencers** within the network due to their large number of connections, and they can be used to understand how information, ideas and influence spread across the platform.

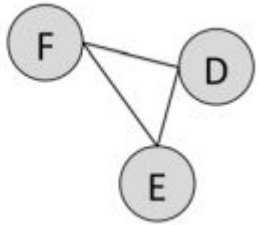
- **Measuring popularity:** By analyzing the number of connections or followers of individuals within a social network, researchers can measure their popularity within the platform. This can include analyzing the number of likes, shares, and comments they receive, as well as their level of engagement with others.

Bridges and hubs



Define Clique and Cluster with example

- All nodes in a group are connected to one another. When this happens, it is called a clique.
- **Clique analysis** can be used for Community Detection.
- **Community Detection:** Cliques can be used to identify communities within a social network. These communities can be defined by shared interests, demographics, or behaviors. By identifying these communities, social media analysts can gain a better understanding of the social structure of the network.



9. What are the different scale issues that occur in network Visualization?

1. Node Overcrowding (Visual Clutter)

- **Issue**: In large networks, nodes are drawn so close together that they overlap.
- **Impact**: It becomes difficult to distinguish individual nodes or see patterns.
- **Example**: In a social media graph with thousands of users, all the profile dots may cluster in one spot, making the network unreadable.

2. Edge Crossings and Overlap

- **Issue**: When many edges (connections) exist, they cross each other, forming a tangled mess.
- **Impact**: Makes it hard to follow the relationships between nodes.
- **Example**: In a transport network connecting cities, if all routes are shown without optimization, it may look like a spiderweb of confusion.

3. Loss of Detail (Small Node Visibility)

- **Issue**: Important but less-connected nodes may shrink or disappear when scale increases.
- **Impact**: The visualization hides valuable data about smaller actors in the network.
- **Example**: A new Twitter user with few followers may be invisible in a graph dominated by celebrities.

4. Scalability of Tools

- **Issue**: Visualization tools or software may not handle large graphs efficiently.
- **Impact**: Long loading times, crashing, or incomplete rendering.
- **Example**: Trying to visualize an entire company's email network in a basic tool like Gephi may crash the system.

5. Label Overlap and Missing Labels

- **Issue**: As nodes increase, labels often overlap or vanish.
- **Impact**: Makes it hard to read node names or values.
- **Example**: In a research citation network, paper titles may overlap or not be shown at all if too many nodes are present.

6. Layout Distortion

- **Issue**: Graph layout algorithms may distort the network structure to fit everything in.
- **Impact**: False impressions about node importance or community structure.
- **Example**: A group of tightly connected nodes may be shown far apart, making them look unrelated.

7. Zooming and Navigation Difficulties

- **Issue**: Large graphs become difficult to explore manually.
- **Impact**: Users get lost, can't zoom into meaningful areas, or miss patterns.
- **Example**: In a healthcare network showing disease spread, users may struggle to zoom in on the patient zero region.

Module 3

1. What is text analytics, and why it is useful? Explain the steps in text analysis with an example.
2. Explain **Social Media Action Analytics,** Common Social Media Actions and Actions Analytics Tools.
3. Explain tools of Hyperlink Analytics.
4. Briefly discuss in-links, out-links, and co-links. (5 mks)
5. Briefly list and define different actions performed by social media users.
6. Discuss and differentiate social media texts
7. What is hyperlink analysis and its underlying assumptions (5 mks)
8. Differentiate between static and dynamic text (5 mks)
9. Explain the four main purposes of social media text analytics.
10. Why it important to measure actions performed by social media users?
11. What are hyperlinks, and why they are important?
12. What is hyperlink environment analysis?
13. What is link impact analysis?

1. What is text analytics, and why it is useful? Explain the steps in text analysis with an example.

Text Analytics is the process of analyzing and understanding written or spoken language using computer algorithms. It helps extract valuable insights, patterns, and information from extensive textual data, enabling computers to interpret human language effectively.

What Can Text Analytics Do?

1. **Extracting Meaning**: Identify themes and trends in large text datasets.
2. **Sentiment Analysis**: Determine emotions (positive, neutral, or negative) in text.
3. **Topic Modeling**: Detect main subjects in documents automatically.
4. **Entity Recognition**: Identify people, places, and organizations.
5. **Text Classification**: Categorize text into predefined groups (e.g., spam filtering, review classification).

Text analytics is useful because it helps organizations and individuals extract meaningful insights from vast amounts of unstructured text data. Some key benefits include:

1. **Improved Decision-Making** – Businesses can analyze customer feedback, social media trends, and market research to make informed decisions.
2. **Efficiency & Automation** – Automates the analysis of large text datasets, saving time and reducing manual effort.
3. **Sentiment Analysis** – Helps companies understand public opinion, customer satisfaction, and brand perception.
4. **Fraud Detection** – Identifies suspicious activity in financial transactions, emails, and reviews.
5. **Healthcare Advancements** – Analyzes medical records, patient feedback, and clinical notes for better healthcare insights.
6. **Legal & Compliance** – Assists in summarizing legal documents, identifying compliance risks, and improving regulatory adherence.
7. **Enhanced Customer Experience** – Personalized recommendations and responses based on user interactions and preferences.

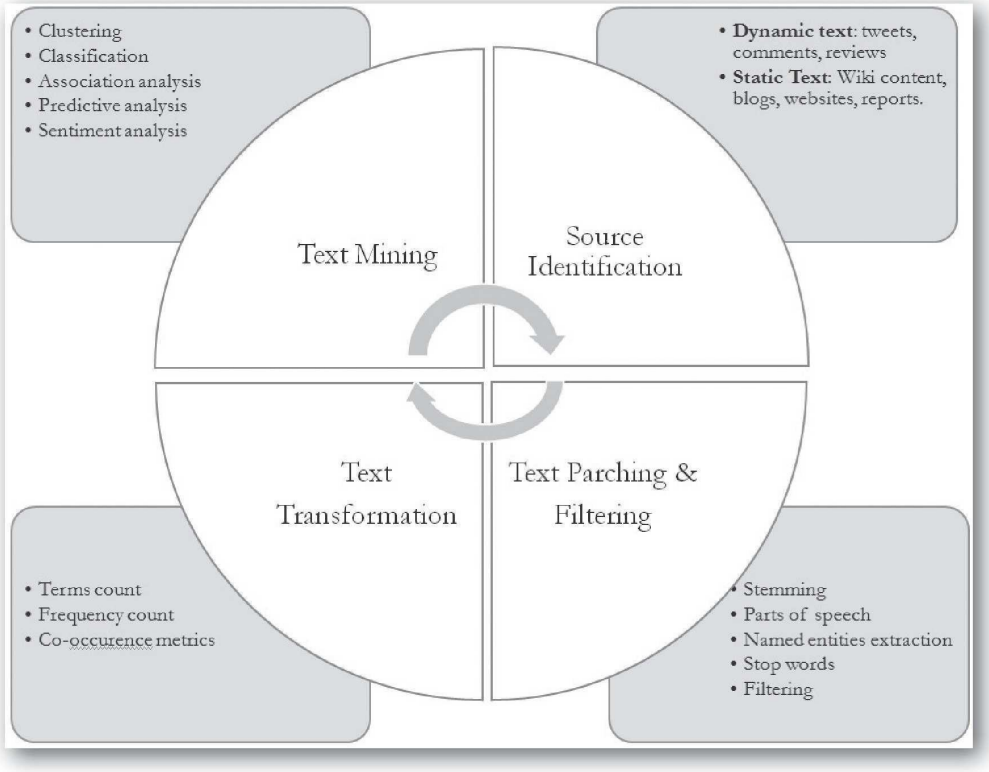
Real-World Example of Text Analytics

Consider a company receiving thousands of customer reviews online. Manually analyzing these reviews is impractical. Text analytics automates the process, identifying patterns and sentiments. For instance:

- Words like *love*, *great*, *excellent* indicate positive sentiment.
- Words like *disappointed*, *issues*, *poor* indicate negative sentiment.

Applications of Text Analytics

1. **Sentiment Analysis**: Analyzing customer reviews and social media opinions.
2. **Customer Feedback Analysis**: Identifying areas for product/service improvement.
3. **Social Media Monitoring**: Understanding public reactions and trends.



Steps in text analysis

Steps in text analysis

1. Source Identification

- The first step involves identifying and collecting text data from various sources.
- **Dynamic text sources** include **tweets, comments, and reviews**, which continuously change over time.
- **Static text sources** include **blogs, websites, reports, and wiki content**, which remain relatively stable.
- Proper source identification ensures **relevant and diverse data collection** for analysis.

2. Text Parsing & Filtering

- Raw text data often contains **irrelevant information** that needs to be cleaned.
- This step involves:
  - **Stemming** – Reducing words to their root form (e.g., "running" → "run").
  - **Parts of speech tagging** – Identifying **nouns, verbs, adjectives, etc.**
  - **Named entity recognition (NER)** – Extracting important names, places, and dates.
  - **Stop words removal** – Eliminating common words like **"the," "is," "and"** that don't add value.
  - **Filtering** – Removing unnecessary symbols, numbers, or special characters.

3. Text Transformation

- After filtering, the cleaned text is converted into **structured numerical data** for further analysis.
- This step includes:
  - **Term Count** – Counting the occurrences of individual words.
  - **Frequency Count** – Measuring how often specific words appear.
  - **Co-occurrence Metrics** – Identifying relationships between words appearing together frequently.
- These transformations help in detecting **patterns, trends, and relationships** in the data.

4. Text Mining

- The final step applies **machine learning and statistical techniques** to extract valuable insights.
- Common text mining techniques include:
  - **Clustering** – Grouping similar texts together.
  - **Classification** – Categorizing text into predefined groups.
  - **Association Analysis** – Identifying relationships between words or phrases.
  - **Predictive Analysis** – Using past data to predict future trends.
  - **Sentiment Analysis** – Determining if the text expresses **positive, negative, or neutral** emotions.
- These techniques help in **decision-making, customer feedback analysis, trend detection, and more**.

2. Explain Social Media Action Analytics, Common Social Media Actions and Actions Analytics Tools.

Social Media Action Analytics refers to the process of analyzing and interpreting the actions performed by users on social media platforms—such as likes, shares, comments, mentions, and clicks—to evaluate the effectiveness of social media content and campaigns. These actions are symbolic reactions that carry emotional and behavioral signals. By tracking and analyzing these actions, businesses can understand user engagement, sentiment, and the overall impact of their content, helping them refine their strategies, increase brand visibility, and improve marketing outcomes. This analysis can also help in identifying content trends, measuring campaign performance, and making data-driven decisions for future planning.

Common Social Media Actions include:

- **Likes**: A way for users to express positive reactions to content. Often used as a basic metric for popularity.
- **Dislikes**: Allow users to show disapproval, commonly seen on platforms like YouTube.
- **Views**: Count how many times content has been seen by users. This is useful to measure visibility and reach.
- **Visitors, Visits, Revisits**: A visitor is a user who accesses a page; a visit (or session) can include multiple page views. Revisits are repeat visits by the same user.
- **Clicking**: Tracks when users click on links, images, or other elements. This is used to measure interest and guide improvements in content or design.
- **Mentions**: Occurs when users refer to a brand or person using a tag (e.g., @username), indicating visibility and engagement.
- **Tagging**: Adding descriptive labels or keywords to posts or people. Helps with content organization and discovery.
- **Hovering**: Mouse-over actions can indicate user interest and attention span, even if no clicks occur.
- **Pinning**: Saving and sharing content via visual boards, especially on platforms like Pinterest or Tumblr.
- **Check-ins**: Users sharing their real-time location, useful for local engagement and offering location-based services.
- **Embeds**: Integrating social media content (like videos or posts) into other websites or blogs.
- **Endorsements**: Approval-based actions like endorsing a skill on LinkedIn, which enhance professional credibility.
- **Uploading and Downloading**: Refers to the act of users adding content to or retrieving content from platforms, such as posting a photo or downloading a file.

Action Analytics Tools that help track and analyze these actions include:

- **Facebook Insights**: Offers detailed metrics on reach, engagement, and audience demographics for Facebook pages.
- **Google Analytics**: While primarily for websites, it can be used to track social media referral traffic and user behavior.
- **Hootsuite**: Manages and analyzes social media activity across multiple platforms with various plan options.
- **SocialMediaMineR**: A tool in R that collects data on likes, shares, and pins for URLs across platforms.
- **Lithium**: Provides tools for social media marketing, analytics, and customer engagement.
- **Klout** (now discontinued but historically important): Measured social media influence on a 1–100 scale.
- **Topsy** (also discontinued): Specialized in analyzing real-time social media conversations.
- **TweetReach**: Measures the reach and impressions of Twitter content, including hashtags.
- **Kred**: Evaluates the influence of Twitter accounts based on engagement and reach.
- **Hashtagify**: Analyzes hashtag performance and influence.
- **Twtrland**: Offers insights and visualizations into social media user behavior.
- **TweetStats**: Provides analytics for Twitter, such as tweet frequency and engagement patterns.

By leveraging these tools and understanding user actions, companies can fine-tune their social media strategy to better engage their audience and achieve business goals.

3. Explain tools of Hyperlink Analytics.

Hyperlink Analytics involves the extraction, analysis, and interpretation of hyperlinks such as in-links (incoming links), out-links (outgoing links), and co-links (common links between pages). These hyperlinks act as the pathways of social media and web traffic, and analyzing them can reveal insights such as traffic patterns, content authority, relationships between websites, and sources of incoming and outgoing traffic. Hyperlink analytics has been applied in diverse areas including ranking universities, analyzing scholarly and political networks, understanding the blogosphere, and measuring business competitiveness.

Hyperlinks are clickable references that connect users to web resources such as documents, websites, or files. They may function within a single document (interlinking) or across multiple documents and websites (intralinking). For instance, a hyperlink in a tweet might direct users to an external website, contributing to traffic generation and connectivity analysis.

The following are some popular tools used in hyperlink analytics:

- **Webometric Analyst**: A web impact analysis tool that supports various types of hyperlink analysis on social media and web platforms. It helps in evaluating academic and business websites by measuring online visibility and connectivity.

- **VOSON (Virtual Observatory for the Study of Online Networks)**: Available at http://www.uberlink.com, VOSON enables the construction and analysis of hyperlink networks. It supports network visualization and helps researchers map relationships between websites.

- **Open Site Explorer**: A backlink analysis tool that helps users research and compare competitors' backlinks, identify high-authority pages, and view social engagement data associated with links.

- **Link Diagnosis**: Available at http://www.linkdiagnosis.com, this is a free online tool used to analyze and diagnose backlinks. It provides detailed reports on anchor text, PageRank, and other SEO-related metrics.

- **Advanced Link Manager**: Offers extensive link analysis capabilities such as tracking link-building progress, analyzing domain quality, visualizing backlink evolution, and crawling websites. More information is available at http://www.advancedlinkmanager.com.

- **Majestic**: Found at https://majestic.com, Majestic is one of the most comprehensive link intelligence tools. It provides detailed data on backlinks, referring domains, trust flow, citation flow, and link profile history.

These tools are essential for SEO professionals, researchers, and analysts aiming to understand how websites are connected and how online visibility and influence are shaped through hyperlink structures.

4. Briefly discuss in-links, out-links, and co-links. (5 mks)

Hyperlink environment analysis involves examining the hyperlinks connected to a particular website or a group of websites. These hyperlinks—**in-links**, **out-links**, and **co-links**—help analyze how websites are connected and how web traffic flows between them. This analysis reveals important insights such as website popularity, influence, and similarity.

1. In-Links

In-links (also known as backlinks) are hyperlinks from other websites that point to a specific website. They are often used to measure the **authority or popularity** of a website.

Example:

If website A is linked by website B, C, and D, then A has three in-links. These in-links suggest that other websites find A valuable or relevant.

2. Out-Links

Out-links are hyperlinks that a website creates, pointing to other websites. These represent the external resources that a website refers its users to.

Example:

If website A links to website X, Y, and Z, then A has three out-links. These links show which other sources A considers useful or related.

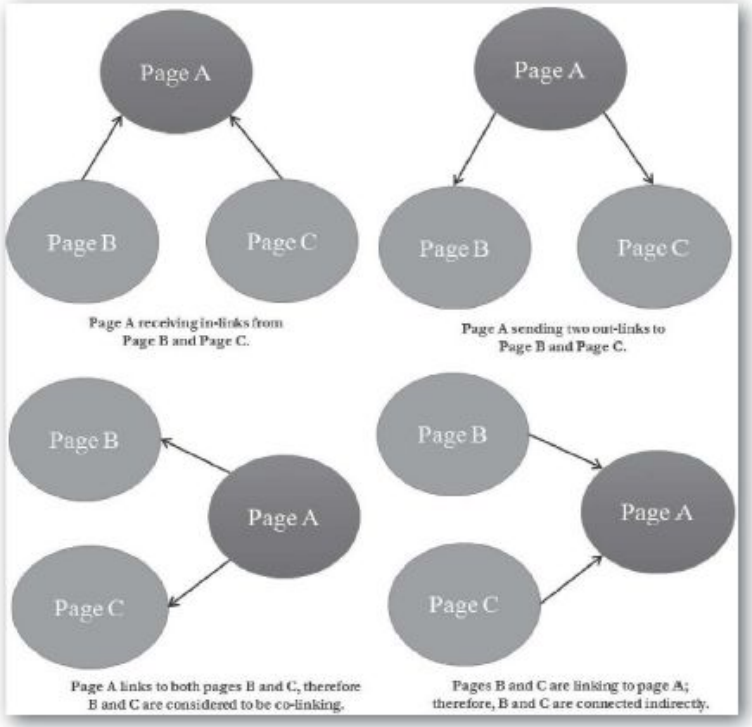


Figure 9. Different types of hyperlinks

3. Co-Links

Co-links refer to a situation where **two different websites are both linked by the same third-party websites**. Co-links are used to measure **similarity or thematic relatedness** between websites.

Example:

If both website A and website B are linked by websites X, Y, and Z, then A and B are said to be co-linked. A high number of co-links indicates that A and B may deal with similar topics or belong to the same community.

Network Types in Hyperlink Environment Analysis:

- **In-link and Out-link Networks:**
  Nodes represent websites, and the directed edges represent either in-links or out-links between them.

- **Co-link Networks:**
  Nodes are websites, and an edge between two nodes represents the number of shared inbound links from the same websites. Tools like **Webometric Analyst** can be used to construct co-link networks and analyze web similarity patterns.

Briefly discuss in-links, out-links, and co-links.

1. In-Links (Inbound Links or Backlinks)

- **Definition**: In-links are links that come from other websites to your webpage. These are also called **backlinks** because they link **back** to your site.
- **Importance in SEO**:
  - Backlinks act as a **vote of confidence** from other sites, meaning search engines like Google view them as a signal of credibility and authority.
  - Websites with **high-quality backlinks** (from authoritative sources) tend to rank higher in search results.
- **Example**:
  - Imagine you write a blog post about "Best JavaScript Frameworks."
  - If a popular tech website links to your article in one of their posts, you receive an **in-link (backlink)** from that site.
  - If multiple high-authority sites link to your page, **search engines will consider your content valuable**, boosting its ranking.

2. Out-Links (Outbound Links)

- **Definition**: Out-links are links that point from **your webpage to an external website**. These are opposite of in-links.
- **Importance in SEO**:
  - Outbound links help search engines **understand the context** of your content.
  - Linking to **relevant and authoritative sources** can improve user trust and add credibility to your content.
  - However, **too many outbound links** or linking to **low-quality sites** can negatively impact SEO.
- **Example**:
  - Suppose you write an article on "Machine Learning Basics" and **link to a research paper from MIT** about neural networks.
  - That **MIT research paper gets an inbound link**, while your page creates an **outbound link**.

3. Co-Links (Co-Citation Links)

- **Definition**: Co-links occur when two different webpages are **linked to by a common third-party website**, even if they **do not** link to each other directly.
- **Importance in SEO**:
  - Co-linking helps search engines identify **topic relevance and relationships** between websites.
  - It influences how search engines group and rank pages based on **shared mentions**.
  - Even if two websites **don't directly link to each other**, a **third-party site linking to both** can indicate a strong **topical connection**.
- **Example**:
  - Suppose there are **three websites**:
    - Website A (Tech News Blog)
    - Website B (AI Research Website)
    - Website C (MIT Research Paper)
  - If the MIT Research Paper **links to both** Website A and Website B, then Website A and B **are co-linked** even if they do not directly link to each other.
  - Search engines may consider both websites as **related to AI research**, improving their ranking for AI-related queries.

5. Briefly list and define different actions performed by social media users.

1. Content Creation: Users generate and share original content like text posts, images, videos, and blogs. Empowering users to express their creativity, share experiences, or provide information on various topics. Content creation fosters community engagement and helps platforms grow by offering diverse user-generated content.

2. Content Sharing: Users distribute existing content to their followers or communities. It amplifies content, allowing it to reach a broader audience. Sharing helps spread ideas, entertainment, or news and contributes to community-building.

3. Liking and Reacting: Users express their opinions using likes, reactions, or upvotes. Likes and reactions serve as an instant form of feedback to the content creators, offering a way to gauge the content's reception. They also help surface popular content through engagement algorithms.

4. Commenting: Users provide feedback, opinions, or engage in discussions. Commenting fosters deeper interactions, allows for conversations between users, and strengthens the sense of community. It provides users a chance to ask questions, share opinions, or connect with others over shared interests.

5. Follow/Subscribe: Users connect with others to receive updates and content. Following allows users to keep up with the content creators or personalities they are interested in, while also providing creators with an audience to engage with. It's a vital mechanism for building online communities.

6. Messaging: Users engage in private conversations. Messaging enables users to engage in personal, private conversations and share content, information, or opinions outside of public feeds. It creates a more intimate connection between users.

7. Live Streaming/Video Calls: Users broadcast live videos or engage in real-time video communication.Live streaming allows for real-time engagement with an audience, enabling interactive experiences such as Q&A sessions, product launches, tutorials, or casual chats. Video calls, on the other hand, are great for personal interactions, remote work, or virtual events.

Common Social Media Actions include:

- **Likes**: A way for users to express positive reactions to content. Often used as a basic metric for popularity.
- **Dislikes**: Allow users to show disapproval, commonly seen on platforms like YouTube.
- **Views**: Count how many times content has been seen by users. This is useful to measure visibility and reach.
- **Visitors, Visits, Revisits**: A visitor is a user who accesses a page; a visit (or session) can include multiple page views. Revisits are repeat visits by the same user.
- **Clicking**: Tracks when users click on links, images, or other elements. This is used to measure interest and guide improvements in content or design.
- **Mentions**: Occurs when users refer to a brand or person using a tag (e.g., @username), indicating visibility and engagement.
- **Tagging**: Adding descriptive labels or keywords to posts or people. Helps with content organization and discovery.
- **Hovering**: Mouse-over actions can indicate user interest and attention span, even if no clicks occur.
- **Pinning**: Saving and sharing content via visual boards, especially on platforms like Pinterest or Tumblr.
- **Check-ins**: Users sharing their real-time location, useful for local engagement and offering location-based services.
- **Embeds**: Integrating social media content (like videos or posts) into other websites or blogs.
- **Endorsements**: Approval-based actions like endorsing a skill on LinkedIn, which enhance professional credibility.
- **Uploading and Downloading**: Refers to the act of users adding content to or retrieving content from platforms, such as posting a photo or downloading a file.

6. Discuss and differentiate social media texts

Social media texts can be broadly categorized into two types based on their nature, frequency of update, and length: **Dynamic Text** and **Static Text**. These two types serve different purposes and exhibit distinct characteristics.

1. Dynamic Text

Definition:

Dynamic social media text refers to user-generated content that is created and updated frequently in real time. It typically consists of short, informal, and rapidly evolving interactions.

Characteristics:

- Short in length (usually a few words to a few sentences).
- Created and consumed in real-time.
- Frequently updated or changed.
- Highly conversational and interactive in nature.
- Often reflects immediate user opinions, sentiments, or reactions.

Examples:

- **Tweets:** Short messages (up to 280 characters) posted on Twitter, often including hashtags, mentions, or links.
- **Comments:** User responses on social platforms (e.g., Facebook, YouTube, Instagram).
- **Discussions:** Back-and-forth conversations in forums or chat threads.
- **Conversations:** Real-time interactions, often in messaging apps or live chats.
- **Reviews:** Brief feedback or ratings provided by users on platforms like Amazon, Google Reviews, etc.

2. Static Text

Definition:

Static social media text refers to content that is more formal, longer in length, and less frequently updated. It is typically designed to provide structured and persistent information.

Characteristics:

- Large in size (can be several paragraphs to multiple pages).
- Not updated or changed frequently.
- Structured and often reviewed or curated before publishing.
- Designed for informative or archival purposes rather than quick interaction.

Examples:

- **Wiki Content:** Collaborative encyclopedic entries like those on Wikipedia.
- **Blog Pages:** Detailed posts on specific topics, often authored by individuals or organizations.
- **Word Documents:** Shared text files on cloud platforms like Google Docs.
- **Corporate Reports:** Formal documents such as annual reports or business strategies.
- **Emails:** Textual communication via electronic mail, sometimes used in social CRM systems.
- **News Transcripts:** Official write-ups of news bulletins or broadcasts.

Aspect	Dynamic Text	Static Text
Length	Short	Long
Update Frequency	Frequently updated	Infrequently updated
Nature	Real-time, conversational	Informative, structured
Examples	Tweets, comments, chats	Blogs, wiki pages, reports, emails
Interactivity	Highly interactive and reactive	Less interactive, more one-way

7. What is hyperlink analytics and its underlying assumptions? (5 mks)

- Hyperlink analytics involves **extracting, analyzing, and interpreting hyperlinks**.
- The **number and quality of hyperlinks to a website** are believed to **reflect its importance or value.**
- Hyperlink analytics can reveal **Internet traffic patterns and sources of incoming or outgoing traffic to and from a website.**
- Hyperlink analysis has been used to study a variety of topics, including **ranking universities, understanding the blogosphere, and measuring business competitiveness**.
- Hyperlink analysis has some limitations, including that it does not provide insight into the type or amount of traffic flowing among websites.
- Hyperlinks within a website between pages are mostly for navigational purposes and are given low importance by search engine ranking algorithms.

Underlying Assumptions of Hyperlink Analytics

1. **Links Represent Authority & Importance**
    - The more inbound links (backlinks) a page has, the more important or authoritative it is.
    - Example: Google's **PageRank Algorithm** assumes that a page linked by many authoritative sites is more relevant.
2. **Link Structures Reflect Real-World Relationships**
    - Websites link to each other based on relevance, credibility, and trust.
    - Networks formed by hyperlinks can reveal communities, hierarchies, and influence.

3. Anchor Text Provides Context

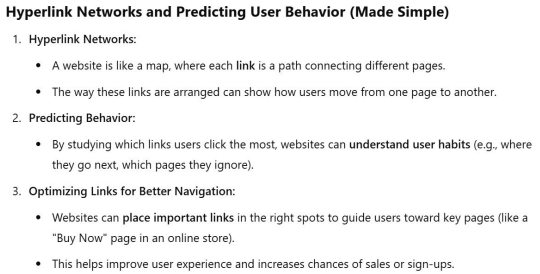
- The text used in a hyperlink (anchor text) provides insight into the content of the linked page.
- Example: If multiple sites link to a page using "best programming tutorials," that page may rank higher for related searches.

4. Reciprocity and Link Propagation

- Many websites engage in reciprocal linking (A links to B, and B links back to A).
- Links tend to spread in a chain reaction; if a well-linked page links to a new page, the new page gains visibility.

5. Websites Can Be Categorized Based on Link Patterns

- Spam websites often exhibit unnatural link-building patterns.
- High-quality websites have organic, contextually relevant links.



Hyperlink Networks and Predicting User Behavior (Made Simple)

1. **Hyperlink Networks:**
   - A website is like a map, where each **link** is a path connecting different pages.
   - The way these links are arranged can show how users move from one page to another.
2. **Predicting Behavior:**
   - By studying which links users click the most, websites can **understand user habits** (e.g., where they go next, which pages they ignore).
3. **Optimizing Links for Better Navigation:**
   - Websites can **place important links** in the right spots to guide users toward key pages (like a "Buy Now" page in an online store).
   - This helps improve user experience and increases chances of sales or sign-ups.

6. Hyperlink Networks Can Predict Behavior

- User navigation patterns can be inferred by analyzing the hyperlink structure of a website.
- Websites can optimize internal linking to guide users to important pages (e.g., conversion pages in e-commerce).

7. Decay of Link Influence Over Time

- Older links may lose influence if they are not maintained or if the linking site becomes irrelevant.
- Search engines may prioritize fresh, frequently updated links over static ones.

8. Differentiate between static and dynamic text (5 mks)

Aspect	Static Text	Dynamic Text
Nature	Long-form, structured, formal	Short-form, spontaneous, informal
Update Frequency	Updated rarely	Updated frequently in real-time
Length	Usually long (several paragraphs or pages)	Usually short (a few words or sentences)
Purpose	To inform, document, or educate	To engage, converse, or respond quickly
Interactivity	Low	High (likes, replies, shares)
Creation Process	Often reviewed or edited before publishing	Created and posted instantly by users
Examples (Type)	Wiki articles, blog pages, corporate reports, Word documents, news articles	Tweets, Instagram comments, YouTube live chats, Facebook posts, Reddit replies
Example Text	“"Introducing our new line of summer dresses! Shop now and get 20% off!" - This post would display the same information to everyone who sees it, with no variations depending on the user.	"Hey [User Name], we noticed you recently viewed our blue jeans! Get an extra 15% off with code 'BLUEJEANS15' today!" - This post would change based on the user's browsing history, offering a personalized discount code tailored to their interests.

Static social media text is a fixed piece of content that remains unchanged for all users, while dynamic social media text can adapt and change based on factors like user interaction, location, or current events, providing a personalized experience for each individual; for example, a static post might be a simple product description with an image, whereas a dynamic post could display a personalized discount code based on a user's previous purchases on the same platform.

Example of Static Social Media Text:

- "Introducing our new line of summer dresses! Shop now and get 20% off!" - This post would display the same information to everyone who sees it, with no variations depending on the user.

Example of Dynamic Social Media Text:

- "Hey [User Name], we noticed you recently viewed our blue jeans! Get an extra 15% off with code 'BLUEJEANS15' today!" - This post would change based on the user's browsing history, offering a personalized discount code tailored to their interests.

Key Differences:

- **Consistency:** Static text stays the same across all users, while dynamic text can vary depending on individual factors.
- **User Interaction:** Static text doesn't usually respond to user actions, while dynamic text can change based on user clicks, location, or other interactions.
- **Implementation:** Static text is usually simple to create, while dynamic text often requires more complex programming to integrate user data and adapt content.

9. Explain the four main purposes of social media text analytics.

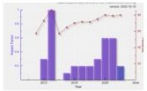
Sentiment Analysis
Sentiment analysis involves categorizing social media text as positive, negative, or neutral. It is often used to understand how customers feel about a product, service, or issue.
Tools like Semantria use algorithms to identify sentiment-bearing phrases in text and assign them a score based on a logarithmic scale.
Scores are then combined to determine the overall sentiment of a document or sentence.
Sentiment analysis can provide valuable insights into the emotions and opinions of social media users.

Sentiment Analysis Process
POS Tagging – Identifies parts of speech (nouns, adjectives, verbs, etc.).
Sentiment Detection – Recognizes sentiment-bearing phrases (e.g., “terrible service”).
Scoring – Assigns sentiment values on a -10 to +10 scale.
Overall Calculation – Combines phrase scores using log odds ratio, yielding a final document sentiment score (-2 to +2).
For example, to calculate the sentiment of a phrase such as “terrible service,” Semantria uses search engine queries similar to the following: “(Terrible service) near (good, wonderful, spectacular)” “(Terrible service) near (bad, horrible, awful)”
Each result is added to a hit count; these are then combined using a mathematical operation called “log odds ratio” to determine the final score of a given phrase.

Intention Mining

- Intention mining involves **discovering users' intentions (such as desire, wish, or intention to buy)** from natural language social media text.
- Companies can use intention mining **to identify potential customers** and **service existing customers** who have issues with a product.
- Examples of intention-bearing phrases include **"buy," "purchase,"** and **"quit."**
- Tools like **Semantria** can be used to mine intentions from social media text.
- **Intention mining** can provide valuable insights into the **needs and desires of social media users.**

Trends Mining



- Trends mining, also known as predictive analytics, uses large amounts of historical and real-time social media data to predict future events.
- It involves **identifying patterns and trends** in social media data to **improve products, services, or customer satisfaction.**
- Techniques used in trends mining include **machine learning, data mining, and social network analysis.**
- Predictive analytics using business data has been used in various industries, but social media predictive analytics is still an emerging practice.
- Trends mining can provide **valuable insights** for businesses and organizations looking to anticipate the needs and desires of their customers.

Concept Mining

- Concept mining is a method for **extracting ideas** and **concepts** from documents.
- It is used to **classify, cluster,** and **rank** these ideas.
- Concept mining is different from text mining, which focuses on extracting specific information rather than broader ideas and concepts.
- Examples of documents that can be analyzed using concept mining include social media text, web pages, and news transcripts.

10. Why it important to measure actions performed by social media users?

1. **Understanding User behaviour**: Measuring actions performed by social media users is crucial for understanding their behavior. By tracking likes, shares, comments, and time spent on content, businesses can gain deep insights into audience preferences and engagement patterns. This data allows them to identify which types of content resonate most with users, enabling more strategic decision-making. Understanding user behavior also helps platforms refine their algorithms to enhance the overall user experience, making social media interactions more engaging and personalized.

2. **Optimising Content Strategy**: Optimizing content strategy is another major reason for tracking user actions. By analyzing what works and what doesn't, businesses and content creators can tailor their posts to maximize engagement. A/B testing different content types, captions, or posting times helps determine the most effective strategies. This ensures that each post is crafted based on real audience feedback rather than guesswork, leading to higher interaction rates and better overall performance on social media.

3. **Enhancing User Experience**: Enhancing user experience is a direct benefit of measuring social media interactions. Platforms can use engagement data to provide personalized recommendations, making content more relevant to individual users. When users see content that aligns with their interests, they are more likely to spend time on the platform, leading to better retention rates. Additionally, tracking interactions helps in identifying pain points, such as which types of content users ignore or report, allowing platforms to improve their overall design and features.

4. **Monetization and Ad Tracking**: Monetization and ad targeting are also greatly enhanced through user activity tracking. Social media platforms analyze user behavior to deliver more relevant advertisements, improving the chances of conversions. Businesses can leverage this data to run highly targeted campaigns, ensuring that their ads reach the right audience segments. This leads to higher engagement rates, better return on investment (ROI), and improved efficiency in digital marketing efforts.

11. What are hyperlinks, and why they are important?

- Hyperlinks are **references to web resources** that users can access by clicking on them.
- They can **link resources within a document (interlinking) or among documents (intralinking).**
- **Hyperlinks** can represent relationships such as **trust, bonding, and authority** between organizations.
- They can also **indicate content similarity** between websites.
- Hyperlinks serve as a symbolic means of communication and validation between organizations.

Why Are Hyperlinks Important?

1. Easy Navigation

- Allows users to move between web pages, sections, or external resources **effortlessly**.
- Example: Clicking "Contact Us" takes users to the contact page.

2. Improves User Experience (UX)

- Provides quick access to **relevant information**, making websites **more interactive and user-friendly**.
- Example: Wikipedia articles contain many links for further reading.

3. Enhances SEO (Search Engine Optimization)

- **Inbound Links (Backlinks):** Improve a page's ranking on search engines by **showing credibility**.
- **Outbound Links:** Help search engines understand content relevance and quality.

4. Supports Content Organization

- Helps in structuring information by **interlinking related topics**, making it easier to find and understand.
- Example: A blog post about "AI in Healthcare" linking to another post about "Machine Learning Basics."

types of hyerlinks add kar dena

12. What is hyperlink environment analysis?

Pehele toh what is hyperlink analytics likh dena

- Hyperlink environment analyses deal with a particular website or set of websites.
- Hyperlinks (i.e., out-links, in-links, and co-links) of a website are **extracted and analyzed** to identify the sources of Internet traffic.
- Hyperlinks environment networks can take two forms:
    - **1) co-links networks or**
    - **2) in-links and out-links networks.**

Co-Link Networks

- In co-links environment networks, nodes are websites and links that represent similarity between websites, as measured by co-link counts.
- With the Webometric Analyst tool, one can construct a co-link network diagram among a set of websites .

in-links and out-links networks.

- In-links and out-links hyperlink environment networks are constructed based on in-links and out-links from a website or set of websites.
- Nodes will be websites and links will present in-links and out-links

Co-Links

Co-links have two dimensions. First, if two websites receive a link from a third website, they are considered to be connected indirectly. For example, page A links to both pages B and C, therefore B and C are considered to be co-linking, or connected indirectly

In-Links

Represents all the links **pointing to** a webpage from other sites.

- Indicates a page's **authority and credibility** (more backlinks = higher trust).
- Used in **SEO** to improve search rankings.
- Example: A tech blog getting backlinks from Forbes, TechCrunch, and Wired.

Out-Links

Represents all the links a webpage **points to** on other sites.

- Helps in **providing references and enhancing content value**.
- Can impact SEO positively if linking to **relevant and authoritative** sources.
- Example: A research paper citing Wikipedia, IEEE, and government sites.

Ppt mai content nahi tha iseleye

13. What is link impact analysis?

Pehele toh what is hyperlink analytics likh dena

- Link impact analysis investigates the impact of a website's URL in terms of **citations or mentions** it receives on the web.
- **Statistics** about web pages that mention the URL of a given website are collected and analyzed.
- It is assumed that a **frequently cited URL is more important,** so measuring the web impact of URLs can provide an idea about the importance of a website.

Module 4

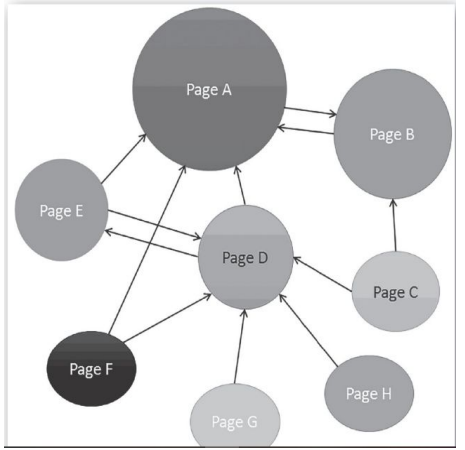
1. What is search engine analytics? (5 mks)
2. List all the location analytics tools and also explain working of every tool.
3. What is Location analytics? Explain its significance in context of social media analytics? Explain its benefit
4. Explain two main categories of search engine analytics/Illustrate main categories od SEA
5. What is search engine optimization? What are different methods to do it?
6. What is purpose of search engine optimization? (5 mks)
7. Sources of location data
8. Explain the two main categories of location analytics.
9. Discuss privacy concerns related to location analytics.
10. What is a Search Engine? Discuss the different types of Search Engines
11. Explain the concepts of Web Scraping and Web Crawling
12. Write Short Notes on
    a. Functions of a Search Engine
    b. SEO
    c. Search Engine Trend Analytics
    d. Google Trends

1. What is search engine analytics? (5 mks)

Search engine analytics involves **analyzing and interpreting data** from search engines to understand and improve a website's performance in search results. By tracking various data points, website owners can optimize their content and marketing strategies to achieve better visibility and attract more visitors.

Search engine analytics helps in:

- Understanding website traffic patterns
- Analyzing keywords that bring visitors to a site
- Tracking user behavior and pages visited
- Monitoring a website's position in search results
- Evaluating the effectiveness of SEO efforts
- Guiding future search engine optimization (SEO) strategies



Types of Search Engine Analytics

There are two main types of search engine analytics:

1. **Search Engine Optimization (SEO) Analytics**
2. **Search Engine Trend Analysis**

1) Search Engine Optimization (SEO) Analytics

SEO analytics is the process of **analyzing and improving a website's ranking** on a search engine results page (SERP). A **Search Engine Results Page (SERP)** is the list of results displayed by a search engine in response to a user's query.

SERPs consist of two types of results:

- **Organic Results** – Ranked based on relevance to the user's search query (unpaid).
- **Non-Organic Results** – Paid advertisements that appear on the search results page.

Since a **higher SERP ranking** increases website visibility and potential customer engagement, businesses invest heavily in SEO techniques.

PageRank Algorithm and Website Ranking

Google's **PageRank algorithm** plays a key role in determining a website's ranking on the SERP.

It evaluates a website's ranking based on:

- **The number of incoming links (in-links)** from other sites
- **The quality of these in-links** (links from reputable websites have more value)

For example:

- If **Website B** receives in-links from a high-authority website like **cnn.com**, it will rank higher than another website with many low-quality in-links.
- Even if another website has more links overall, **Google prioritizes link quality over quantity**.

Tools like **Open SEO Stats** can be used to check a website's PageRank and other relevant information such as traffic, hyperlink status, and speed of the page.

2) Search Engine Trend Analysis

Search engine trend analysis focuses on understanding and analyzing keyword trends to predict user behavior and market demand.

Google Trends for Search Analytics

Google Trends is a powerful tool for analyzing search patterns and predicting trends over time. It provides insights into:

- How people search for a particular **brand, product, or topic**.
- When **interest spikes** in specific products or services.
- Which **keywords drive the most traffic** to websites.
- How a **brand compares with competitors** in terms of search volume.

Real-World Applications of Search Trend Analysis

- Businesses use **Google Trends** to predict market demand and optimize their marketing strategies.
- It has been used in **healthcare** to track flu outbreaks by analyzing search queries related to symptoms.
- **News organizations** monitor trending searches to identify breaking news topics.

Types of Analytics Provided by Google Trends

1. **Year in Search** – A summary of the top search trends of the year.
2. **Trending Searches** – Displays what is currently trending worldwide.
3. **Trending on YouTube** – Highlights the most popular YouTube searches.
4. **Top Charts** – Ranks the most-searched topics in various categories.
5. **Explore** – Allows users to analyze keyword search trends over time.
6. **Subscription** – Provides alerts on trending topics.

2. List all the location analytics tools and also explain working of every tool.

1. Google Fusion Tables

Google Fusion Tables was a web-based service that allowed users to upload tabular data, visualize it, and overlay it on Google Maps.

Working: Users could upload data (e.g., spreadsheets or CSV files), geocode location columns (like city, address), and generate interactive maps with markers, heatmaps, or polygons to represent different attributes. It was mainly used for sharing and collaborating on geographic data online. *(Note: It was discontinued in December 2019.)*

2. Agos

Agos is a platform focused on geo-tagging and community-based reporting for addressing climate change and disaster risk reduction.

Working: Community members can report local climate-related or disaster events by submitting geo-tagged data (e.g., photos, text). The platform collects, maps, and visualizes this data to support risk planning, community engagement, and early warning systems.

3. Tweepsmap

Tweepsmap is a tool that maps Twitter followers based on their location.

Working: By analyzing the location data from Twitter profiles, it visualizes the geographic distribution of followers on a map. It also provides demographics, engagement metrics, and audience insights to help users understand and target their social media strategy effectively.

4. Trendsmap

Trendsmap is a real-time visualization tool that displays trending topics on Twitter geographically.

Working: It fetches trending hashtags and keywords from Twitter and displays them on a global or regional map, allowing users to see what topics are trending in specific locations in real time. It's widely used in journalism, marketing, and social media analysis.

5. Followerwonk

Followerwonk is a Twitter analytics tool that provides detailed information about followers, including location, activity times, and influence.

Working: Users can analyze their followers' geographic locations, bio keywords, and tweet timing to optimize posting schedules and build targeted follower growth strategies.

6. Esri's GIS Software

Esri provides powerful Geographic Information System (GIS) software for spatial data analysis.

Working: Esri's tools allow users to collect, store, analyze, and visualize geospatial data. It supports a wide range of functions including mapping, layering different datasets, querying relationships between variables, and generating spatial insights. It is extensively used in urban planning, environmental monitoring, public health, and business logistics.

3. What is Location Analytics? Explain its significance in the context of social media analytics. Explain its benefits.

Location Analytics, also known as **spatial analysis** or **geo-analytics**, is the process of collecting, analyzing, and visualizing geographic data to gain insights into the relationships between location, people, and events. It integrates data with a geographic component (such as coordinates, city names, or zip codes) to detect patterns, trends, and behaviors that are associated with specific places.
In simple terms, it is about understanding **"where"** things happen and **"why"** they happen in those locations. With the rise of GPS-enabled devices and geo-tagged content on social media, location analytics has become an essential tool for deriving meaningful conclusions from location-based data.

Significance in Social Media Analytics
In the domain of social media, location analytics plays a critical role in understanding **user behavior, content trends, and engagement patterns** across different geographic regions. Many social media posts (like tweets, Instagram posts, Facebook check-ins, etc.) are embedded with location metadata, either automatically through GPS or manually through user input.
Using location analytics on social media data can help in the following ways:

1. **Understanding Audience Geography**
   Brands and organizations can identify where their audience is located, which cities or countries show the highest engagement, and how user interests vary by region. For example, **Tweepsmap** allows you to geo-locate your Twitter followers by country, state, or city.

2. **Location-Based Targeting**
   Marketers can create location-specific campaigns and deliver personalized ads or content to users based on their physical location, increasing the relevance and effectiveness of promotions.

3. **Analyzing Regional Trends and Issues**
   Researchers can use geo-tagged posts to understand regional concerns, such as health issues, political opinions, or public sentiment on current events.

4. **Disaster Response and Public Safety**
During natural disasters or emergencies, geo-tagged social media data can help in real-time tracking of affected regions, coordinating relief efforts, and issuing location-specific alerts.

5. **Movement and Navigation Patterns**
Businesses and governments can analyze location data to study how people move through cities, which routes are most frequented, and where bottlenecks occur.

6. **Real-Time Monitoring**
Tools like **Trendsmap** help visualize real-time trending topics on Twitter by geographic location, enabling newsrooms, governments, and businesses to stay updated on public opinion and emergent events.

Benefits of Location Analytics

1. **Hyperlocal Intelligence**
   Organizations can tailor products and services to suit the needs of local markets, communities, and even neighborhoods. This helps improve customer satisfaction and loyalty.

2. **Enhanced Decision-Making**
   By analyzing patterns and trends across different locations, decision-makers can make informed choices regarding resource allocation, campaign planning, or service delivery.

3. **Improved Customer Segmentation**
   Location analytics allows segmentation not just by demographics but also by geographic behavior. This enables more accurate targeting and understanding of customer preferences.

4. **Real-World Context to Digital Data**

Location data bridges the gap between digital interactions and physical environments, offering a complete picture of user behavior.

5. Actionable Insights and Forecasting

The ability to anticipate needs or trends in certain areas empowers organizations to act proactively rather than reactively.

Example Use Cases

- **Urban Analytics**: A well-known example is the study of New York City using geo-located Twitter data, often called the "heartbeat of the city" (França et al., 2015). Researchers tracked patterns in tweet volume and sentiment throughout the day to understand how the city's energy changed by location and time.

- **Healthcare and Policy**: Consider an analysis of cancer patients' blog posts from platforms like Reddit. By linking posts with location data:

    - Healthcare providers can identify regions where users are discussing specific symptoms or treatments.

    - Sentiment analysis can show areas with high emotional stress or satisfaction with care.

    - Governments can direct awareness campaigns or resources to areas with low engagement or negative sentiment.

- **Marketing and Business Strategy**: Retailers can use customer check-ins and location-tagged reviews to understand where customers visit most frequently and adjust store placements or local promotions accordingly.

4. Explain two main categories of search engine analytics

5. What is search engine optimization? What are different methods to do it?

6. What is purpose of search engine optimization? (5 mks)

7. Sources of location data

Sources of Location Data
Location data comes from various sources, which help businesses, social media platforms, and researchers analyze geographic trends and user behaviors. These sources can be broadly classified into **technical sources** and **text-based sources**.

1) Technical Sources of Location Data

Postal Addresses:
- The most traditional form of location data, often used in **customer databases, business directories, and mailing lists**.
- Helps in **geocoding**, which converts addresses into latitude-longitude coordinates.

Latitude & Longitude Coordinates:
- Provides precise geographic locations.
- Collected through GPS, mobile devices, or mapping services.
- Used for **navigation, location tracking, and geospatial analysis**.

GPS-Based Location Data:
- Collected through **smartphones, fitness trackers, and vehicle navigation systems**.
- Essential for **real-time tracking, route optimization, and location-based services** like Uber and Google Maps.

IP-Based Location Data:
- Extracted from internet activity; each device connected to the internet has an IP address that can estimate location.
- Useful for **content localization, fraud detection, and cybersecurity**.

2) Text-Based Sources of Location Data

Social Media & Forums:

- Many social media platforms (Twitter, Facebook, Instagram) and online forums (Reddit, health blogs) contain **user-generated location data**.
- Users often **tag locations** in their posts or include location details in their bios.
- Businesses use tools like **Tweepsmap** to analyze follower locations.

User Profiles & Online Registrations:

- Websites and apps collect location data from user profiles when users register or update their accounts.
- Examples: **E-commerce platforms, travel websites, and dating apps**.

Mentions in Text Content (Text Mining & NLP):

- People often mention places in their **blog posts, tweets, and discussions**.
- **Natural Language Processing (NLP)** techniques can extract and geocode these locations.
- Example: Analyzing **cancer patients' blog posts** to map discussions around hospitals, treatment centers, and local support groups.

8. Explain the two main categories of location analytics.

Location analytics can be divided into two broad categories:

1. **Business Data-Driven Location Analytics**
2. **Social Media Data-Driven Location Analytics**

1) Business Data-Driven Location Analytics

This type of location analytics focuses on using geographical data to analyze **business trends and customer behavior**. It helps businesses make informed decisions, such as where to open a new store or how to target specific customers.

Applications of Business-Driven Location Analytics:

- **Powerful Intelligence:** Advanced mapping techniques like:
  - **Clustering** (grouping similar data points)
  - **Heat Mapping** (highlighting high-activity areas)
  - **Data Aggregation** (combining multiple data sources)
  - **Color-Coded Mapping** (using colors to differentiate data)
- **Geo-Enrichment:** Businesses can enhance maps with **customer demographics, spending habits, and lifestyle data** to understand where their loyal customers are and how they behave.
- **Collaboration and Sharing:** Business teams can easily share and discuss location-based insights using interactive maps and tools like Google Fusion Tables, which helps create maps, tables, and charts.

Example:

A retail company can analyze customer purchase locations to decide where to open a new store. If heat maps show that most purchases come from a particular neighborhood, opening a store there would be a smart decision.

2) Social Media Data-Driven Location Analytics

This category focuses on analyzing location data from social media platforms to understand user behavior, preferences, and trends.

Applications of Social Media-Driven Location Analytics:

- **Recommendation Purposes:** Social media platforms use location data to suggest content, places, and events based on users' locations.
- **Customer Segmentation:** Tools like **Tweepsmap** analyze Twitter followers by location (city, state, country), helping businesses tailor their marketing efforts.
- **Targeted Advertisements:** Businesses can show ads based on a user's real-time location (e.g., promotions at a nearby restaurant).
- **Information Requests:** Users can search for nearby services like restaurants, ATMs, or hospitals based on their current location.
- **Alerts & Notifications:** Location data is used for important alerts such as:
    - Sales & promotion alerts
    - Traffic congestion updates
    - Weather warnings
- **Search and Rescue:** In emergency situations, authorities can use location data to **track missing people** or provide disaster relief.
- **Navigation:** Apps like **BE-ON-ROAD** provide **offline GPS navigation**, helping users find locations without the internet.

Example:

During a flood, social media location data can help rescue teams **locate stranded people** and send real-time alerts about safe zones.

Both **business-driven** and **social media-driven** location analytics help in different ways—businesses use it to make smarter decisions, while social media platforms use it to improve user experience and emergency responses.

Explain the two main categories of location analytics.

- Based on its scope, location analytics can be broadly classified into two categories:
  1) Business data-driven location analytics
  2) Social media data- driven location analytics.

Business data-driven location analytics:

- Business data-driven location analytics refers to the use **of spatial data analysis tools** to gain insights into business operations, customer behavior, and market trends. This type of location analytics uses data from a **variety of sources**, including sales data, customer data, and geographic data, to identify patterns and trends that can inform business decisions.
- Examples of business data-driven location analytics include analyzing customer traffic patterns in a retail store, optimizing delivery routes for a logistics company, or identifying areas with high demand for a particular product or service.
- Business data-driven location analytics can help organizations improve operational efficiency, increase revenue, and enhance customer experience.

Applications of Business Data-Driven Location Analytics

- Powerful Intelligence
- Geo-Enrichment
- Collaboration and Sharing

Social media data-driven location analytics:

- Social media data-driven location analytics: Social media data-driven location analytics refers to the use of spatial data analysis tools to gain insights into social media user behavior and sentiment in a geographic context. This type of location analytics uses data from social media platforms, such as Twitter, Facebook, and Instagram, to identify patterns and trends in user behavior related to specific locations.
- Examples of social media data-driven location analytics include analyzing sentiment towards a particular brand or product in different geographic regions, identifying areas with high levels of social media activity around a particular topic or event, or monitoring the spread of misinformation on social media related to public health concerns

2nd answer

9. Discuss privacy concerns related to location analytics.

While location analytics provides valuable insights for businesses and services, it also raises **important privacy concerns** that need to be addressed.

1) Privacy and Anonymization
User Transparency & Control:

- People should know when and how their location data is being tracked.
- Users should have options to **turn off location tracking or delete stored data**.

Anonymization of Data:

- Companies should ensure that location data is stored **without identifying individuals** to protect privacy.
- **Example:** Instead of tracking "John Smith at 123 Main St," businesses should track anonymous location trends.

Risk of Misuse:

- If location data falls into the wrong hands, it could be used for **unauthorized tracking or targeted cyber threats**.
- **Example:** Hackers gaining access to real-time location data of individuals.

2) Regulatory Compliance
Legal Protections & Regulations:

- Governments enforce **laws to protect users' location privacy** (e.g., GDPR in Europe, CCPA in California).
- **Example:** Companies must get **explicit consent** before collecting location data.

Historical Location Data & Law Enforcement:

- Authorities may request location history for investigations, raising concerns about **unreasonable searches**.
- The legal system must balance **privacy rights and security needs**.

Disclosure Control:

- Businesses must set clear policies on **who can access location data and for what purpose**.
- **Example:** A rideshare app should **only** share a rider's location with the driver, not third parties.

Conclusion

To **protect user privacy**, businesses and governments must implement **strong policies**, ensure **data anonymization**, and comply with **legal regulations**. Transparency, user control, and ethical data usage are key to maintaining trust in location analytics.

10. What is a Search Engine? Discuss the different types of Search Engines

A **search engine** is a software program that helps users find relevant information on the internet by using **keywords** or **phrases**. Given the **billions of websites** online, search engines play a crucial role in organizing and retrieving the most relevant results efficiently.
Search engines work by scanning, indexing, and ranking web pages based on various factors such as **relevance, popularity, and content quality**.

Types of Search Engines
Search engines can be classified based on how they store and retrieve data. The three main types are:

1) Crawler-Based Search Engines

- Crawler-based search engines create and update their databases automatically using **web crawlers (bots or spiders)**.
- These crawlers scan the web, collect data, and store it in an index.
- When a user enters a search query, the engine retrieves and ranks results based on relevance.

How Crawler-Based Search Engines Work

1. **Web Crawling** – The search engine sends bots to scan web pages and collect data.
2. **Indexing** – The collected data is categorized and stored in a database for quick retrieval.
3. **Searching & Ranking** – When a user searches, the engine queries the index and ranks results based on factors like **keywords, backlinks, and content quality**.

Examples of Crawler-Based Search Engines: Google (Googlebot), Bing (Bingbot), Yahoo (Yahoo! Slurp), DuckDuckGo (DuckDuckBot)

2) Directory-Based Search Engines

- These search engines rely on **human editors** to review and categorize websites into directories.
- Website owners can **submit** their site for inclusion, and editors verify the quality before listing.
- Unlike crawler-based engines, directory-based engines do not scan the entire web automatically.

Examples of Directory-Based Search Engines: Yahoo Directory (Discontinued), Open Directory Project (DMOZ), LookSmart

3) Meta Search Engines

- Meta search engines do not maintain their own database but fetch results from multiple search engines.
- When a user submits a query, the meta search engine sends it to various individual search engines and compiles the best results.
- These engines **eliminate duplicate results** and display the most relevant ones in an integrated manner.

Examples of Meta Search Engines: Dogpile, Metacrawler. Mamma

Meta search engines **save time** by aggregating results from different sources and providing users with comprehensive search results.

Classification Based on Scope

Search engines can also be classified based on their **scope and reach**:

1) Local Search Engines

- These are **embedded within a specific website** and only search content available on that website.
- Example: **Amazon's CloudSearch**, site-specific search engines used on e-commerce platforms.

2) Global Search Engines

- These are used to search content across the entire internet.
- Examples: **Google, Bing, Yahoo**

11. Explain the concepts of web scraping and web crawling

- Web scraping is the process of extracting structured data from web pages using automated tools or scripts.
- In social media analysis, web scraping helps collect publicly available data like posts, comments, hashtags, likes, and user interactions.
- This data is analyzed to derive insights into user sentiment, trending topics, and audience engagement.
- Businesses use web scraping to monitor customer feedback on platforms like Twitter and Facebook, allowing for strategy adjustments.
- Researchers use web scraping to study public opinion on topics like politics, health crises, or consumer behavior.
- Social media platforms often implement strict anti-scraping policies.
- Measures like CAPTCHAs, IP blocking, and API rate limits are used to prevent unauthorized data extraction.
- Ethical concerns arise when scraping personal user data without consent.
- Adherence to legal guidelines like GDPR and platform-specific terms of service is necessary.

Web crawling, or **web spidering**, is the process of systematically navigating web pages by following hyperlinks to collect and organize data.

Unlike **web scraping**, which focuses on extracting specific pieces of information, **crawling** is used to discover and index a vast number of interconnected web pages.

In **social media analysis**, web crawlers help researchers and analysts:

- Track content updates.
- Monitor discussions across different platforms.
- Identify emerging trends in online conversations.

Key application in social media: **Network analysis**, where crawlers:

- Map user interactions.
- Identify influential accounts.
- Analyze the spread of information, such as viral posts or misinformation.
- Example: A crawler can follow retweet chains on Twitter to study how news spreads.

Crawling is also useful for **competitive intelligence**, allowing companies to monitor brand mentions,
Track customer sentiment across multiple platforms.

12 a. Short note on functions of search engine

1. Crawling

- Search engines use web crawlers (also known as spiders or bots) to browse the internet and discover new or updated content.
- Crawlers start with a set of known URLs and follow links within those pages to discover more.
- They gather various types of content, including text, images, and videos.

2. Indexing

- After crawling, the gathered data is processed and stored in a massive database known as the search engine index.
- The indexing process involves analyzing the content of web pages, extracting keywords, metadata, and links.
- Search engines apply algorithms to categorize and structure the information efficiently for faster retrieval.
- Duplicate, broken, or low-quality pages may not be indexed to improve search quality.

3. Ranking

- When a user searches for a query, the search engine ranks pages based on multiple ranking factors.
- Ranking is determined by algorithms that consider elements like:
    - **Keyword relevance** – How well the page content matches the search query.
    - **Backlinks** – The number and quality of external websites linking to a page.
    - **Page authority** – Credibility of the website, often influenced by domain age, traffic, and content quality.
    - **User engagement** – Metrics like click-through rates (CTR), bounce rates, and dwell time.
    - **Freshness** – How recent and frequently updated the content is.

4. Retrieval (Fetching Search Results)

- When a user submits a query, the search engine searches its index to find the most relevant results.
- It applies various ranking algorithms to determine which web pages best match the user's intent.
- The retrieval process considers different search types, such as text-based, voice search, and image search.
- Search engines also personalize results based on user location, search history, and preferences.

5. Displaying Results

- The final step is displaying the search results in an organized manner on the Search Engine Results Page (SERP).
- Results may include:
    - **Organic results** – Web pages ranked based on SEO (Search Engine Optimization) factors.
    - **Paid ads** – Sponsored links from advertisers (e.g., Google Ads).
    - **Featured snippets** – Direct answers to queries displayed at the top of the page.
    - **Knowledge panels** – Information boxes with structured data about people, places, or topics.
    - **Image and video results** – Multimedia search results relevant to the query.
- The goal is to provide users with the most useful, quick, and accurate answers to their queries.

12 b. Short note on SEO

Search Engine Optimization (SEO) refers to techniques used to improve a website's ranking in **Search Engine Results Pages (SERPs)**. SERPs display two types of results:

1. **Organic Results** – Appear based on relevance to a user's query.
2. **Nonorganic Results** – Include paid advertisements.

SEO helps websites rank higher in search results, increasing visibility and attracting more visitors. A key factor in SEO is **Google's PageRank algorithm**, which ranks websites based on the quality and number of incoming links (**in-links**). Websites with high-quality in-links from authoritative sources rank higher.SEO is essential for businesses and marketers as higher rankings in SERPs lead to increased website traffic, which can convert into customers.

Key Aspects of SEO:

- **Types of Search Results**
  - **Organic Search Results** – Appear naturally based on website relevance and SEO strategies.
  - **Nonorganic Search Results** – Paid advertisements that appear above or beside organic results.

- **PageRank Algorithm**
  - Developed by **Google**, it ranks websites based on **in-links** (incoming hyperlinks).
  - High-quality **in-links** from authoritative sites (e.g., cnn.com) improve ranking more than a large number of low-quality links.
  - Websites with strong **PageRank** appear higher in search results.

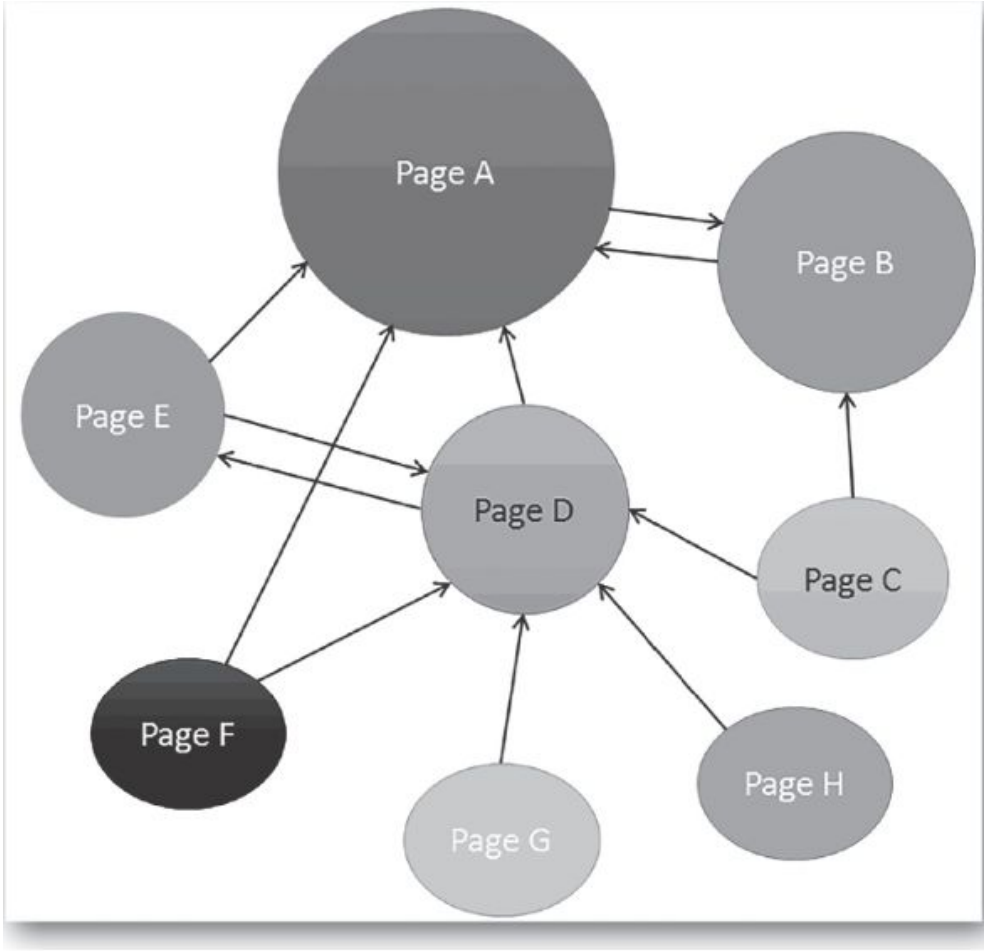


Figure 12. PageRank algorithm ranking example

- **SEO Strategies**
  - **Keyword Optimization** – Using relevant words that users search for.
  - **Quality Backlinks** – Getting links from reputable websites.
  - **Website Speed & User Experience** – Faster, mobile-friendly websites rank better.

- **Importance of SEO**
  - Helps websites appear in the **top 10 search results**, as users focus on top-ranked pages.
  - **More visibility = More traffic = Higher potential revenue**.

12 c. Search Engine Trend Analytics

Search Engine Trend Analytics involves analyzing and understanding the keywords that people use in search engines. It provides insights into customer interests, behaviors, and emerging trends, making it a valuable tool for businesses and marketers. By studying search trends, companies can optimize their marketing strategies and predict future demands.

Importance of Search Trend Analytics: Search engine data acts as a **gateway into consumer behavior**, revealing what users are looking for. This information is beneficial for:

- **Social marketers** to understand customer needs.
- **Businesses** to track demand for products/services.
- **Researchers** to analyze trends in various sectors like finance and healthcare.

Google Trends as a Tool for Search Analytics: One of the most widely used tools for search trend analytics is **Google Trends**. This tool analyzes massive amounts of search engine data to identify worldwide interests and predict trends.

Applications of Google Trends:

1. **Financial Sector** – Helps detect early warning signs of **stock market movements** by analyzing search behavior related to financial terms.
2. **Health Sector** – Used to monitor **flu epidemics** and predict disease outbreaks by studying search trends related to symptoms and treatments.
3. **Marketing & Business** – Helps businesses understand when interest in their **brand, products, or services** spikes, allowing better campaign planning.

Key Questions Answered by Search Trend Analytics: Search trend analysis helps businesses and marketers answer:

- **How do people search for your brand?** → Identifying popular search terms related to a brand.
- **When does interest in your products or services increase?** → Finding seasonal or event-driven trends.
- **Which keywords drive more traffic?** → Understanding which search terms bring users to a website.
- **Which regions are interested in your brand?** → Locating geographical demand for products or services.
- **What are trending topics on the internet?** → Tracking viral content and emerging discussions.
- **How are your competitors performing?** → Comparing search trends of competing brands.

12 d. Google Trends

Google Trends is a tool provided by Google that analyzes the popularity of search queries across various regions and time periods. It helps users track the volume of searches related to specific keywords, providing insights into what people are interested in at any given moment. By offering data on trends, Google Trends allows users to see how search interest fluctuates, making it valuable for businesses, marketers, and content creators to understand public behavior and tailor their strategies accordingly.

- Google Trends helps in identifying trending topics and keywords across regions and time periods, allowing social media analysts to track shifts in public interest.
- By analyzing search queries, it provides insights into the interests of different demographics, which can inform content strategies on social media platforms.
- Social media managers can use Google Trends data to create content based on popular and rising trends, increasing engagement and reach.
- It allows comparisons between search volumes of multiple keywords, helping to determine the most relevant and popular topics for content creation.
- It provides real-time insights into search trends, allowing social media campaigns to respond quickly to emerging topics and discussions.
- Social media strategies can be tailored to specific locations based on geographical search data from Google Trends, enhancing localized engagement.
- Identifying trending keywords or topics can also help in finding potential influencers aligned with popular discussions, aiding influencer marketing decisions.
- By analyzing historical data, trends can be predicted, helping in proactive content creation and marketing campaigns.

12 d. Google Trends is a powerful search engine analytics tool that analyzes global search patterns based on keyword usage. It helps businesses, researchers, and marketers understand public interest and predict trends.

Key Features of Google Trends:

1. **Search Trend Analysis** – Tracks how people search for brands, products, or topics over time.
2. **Trending Topics** – Identifies popular searches, trending keywords, and interests across different regions.
3. **Business Insights** – Helps determine when interest spikes in products or services.
4. **Competitor Analysis** – Compares search trends of competing brands.
5. **Industry Applications** – Used in finance (stock market predictions) and healthcare (tracking disease outbreaks).

Types of Analytics in Google Trends:

- **Year in Search** – Highlights the most searched topics in a given year.
- **Trending Searches** – Shows the top searches daily for a specific region.
- **Trending on YouTube** – Lists popular YouTube videos.
- **Top Charts** – Ranks real-world topics, people, and products by search volume.
- **Explore** – The most detailed feature, analyzing search trends over time, across regions, and by related searches.

Google Trends normalizes search data on a scale of 0–100, making it easier to compare interest levels over time. It is widely used for market research, content strategy, and business decision-making.

Module 5

1. Explain the steps needed to formulate a social media strategy. (5 mks)
2. What is social media risk? Explain the four steps in social media risk management/Social Media Risks Management Framework.
3. Automated, **Traditional and Social recommender systems.**
4. Differentiate between Traditional Media and Social Media
5. Explain common social media risks mitigation strategies
6. Specify the significance of social media KPI
7. What is a social media based recommendation and how does it differ from a traditional recommendation system
8. Types of social media risks
9. Discuss the parameters used to measure success of a Social Media Platform
10. What are some of the most common types of traditional recommendation systems, such as collaborative filtering or content-based filtering?
11. What are different threats to privacy on social media?
12. Trust-based collaborative filtering is an improvement over traditional similarity-based recommendation methods. Discuss two key advantages of using trust-weighted similarity in recommendation systems.
13. Provide an example of a real-world application where trust scores are used to enhance recommendations (e.g., social media, e-commerce, travel platforms). Explain how trust influences the recommendations in that scenario.
14. In a collaborative filtering-based recommender system, the Pearson correlation coefficient is commonly used to measure user similarity. Explain how trust scores can enhance this method and modify recommendations.
15. ROI (Return Of investment) Analysis from social media activities and campaigns

22. **Explain how the Pearson correlation coefficient is used in collaborative filtering for recommender systems. Using the given movie ratings below, calculate the similarity between Alice and Bob using the Pearson correlation formula. Interpret the result and explain how it helps in making recommendations.**

User	Star Wars	Jaws	Wizard of Oz	The Godfather 2001
Alice	5	4	3	3	1
Bob	3	5	2	5	1

(Hint: Use the formula below for Pearson correlation coefficient)

21. Using the **given movie ratings and trust scores**, compute the **trust-weighted Pearson correlation** between Alice and both Bob & Chuck. Based on your calculations, determine **which user will have a greater influence on Alice's recommendations and why**.

User	Star Wars	Jaws	Wizard of Oz	The Godfather 2001
Alice	5	4	3	3	1
Bob	3	5	2	5	1
Chuck	4	3	2	2	2

User	Trust Score with Alice (0, 1)
Bob	0.6
Chuck	0.9

(Hint: Use the formula below for Pearson correlation coefficient, adjusted for trust weighting)

$R_{trust} = r \times T$

where:

- R is the standard **Pearson correlation coefficient** between Alice and another user.
- T is the **trust score** Alice has assigned to that user.
- R_{trust} is the **trust-weighted similarity score** used for recommendations.

Key Tasks:

1. **Calculate the standard Pearson correlation** between Alice & Bob, and Alice & Chuck.
2. **Adjust these values using the given trust scores**.
3. **Interpret the results**: Which user should influence Alice's recommendations more? How does trust change the way recommendations are made?

1. Explain the steps needed to formulate a social media strategy. (5 mks)
The purpose of formulating social media strategy is to create rules and procedures to align your social media engagement with business goals.

STEPS IN FORMULATING A SOCIAL MEDIA STRATEGY :
The following steps will lead to the formulation of a sound social media strategy.

- **Define objectives:** Identify the specific goals that the business wants to achieve through social media, such as increasing brand awareness, generating leads, or improving customer satisfaction.

- **Identify target audience:** Determine the demographics and interests of the business's target audience, including age, gender, location, and social media habits.

- **Research competition:** Analyze the social media presence and strategies of the business's competitors to understand what is and is not working in the industry.

- **Choose social media platforms:** Select the social media platforms that are most relevant to the business's target audience and objectives.

- **Create a content calendar**: Plan and schedule the types of content that the business will post on social media, including text, images, videos, and links.

- **Engage with followers**: Monitor and respond to comments and inquiries from followers, and encourage user-generated content and interactions.

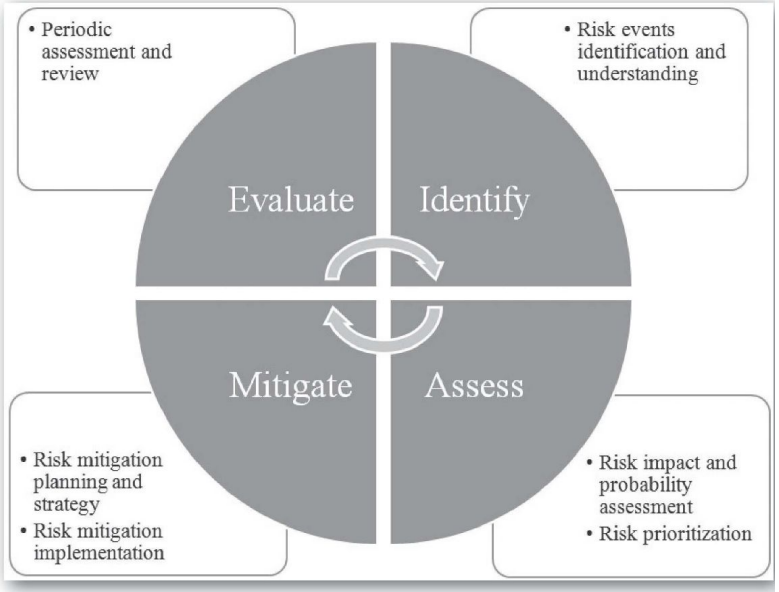
- **Analyze and adjust**: Use social media analytics tools to track the performance of the business's social media efforts and make adjustments as needed to improve results.

2. What is social media risk? Explain the four steps in social media risk management

Social media risk refers to the potential negative consequences that organizations or individuals may face due to their activities, presence, or interactions on social media platforms. These risks can affect a company's **reputation, data security, compliance, employee behavior, and customer trust**.

Common social media risks include:

- **Reputation damage** from negative reviews or viral backlash
- **Data breaches** through phishing or oversharing of confidential information
- **Legal and regulatory violations** (e.g., violating copyright, consumer protection laws)
- **Fake news or misinformation** that can mislead or defame
- **Employee misconduct** (e.g., inappropriate posts by staff)



Managing social media risks is essential for individuals and organizations to protect their data, reputation, and security. The Social Media Risk Management Framework helps identify, assess, mitigate, monitor, and review potential risks to ensure a safe and responsible online presence.

1. Risk Identification

- Identify potential threats, such as hacking, phishing scams, impersonation, and information leaks.
- Recognize how these risks can affect personal safety, business reputation, or financial security.

2. Risk Assessment

- Evaluate how likely each risk is to occur and how severe its impact could be.
- Prioritize the most critical risks and focus on preventing major threats.

3. Risk Mitigation

- Implement security measures to prevent or reduce risks.
- Use strong passwords, enable two-factor authentication, and educate users about online threats.
- Set up privacy settings and establish guidelines for safe social media use.

4. Risk Monitoring

- Continuously track social media activities to detect and respond to new threats.
- Monitor accounts for suspicious activity, unauthorized access, or harmful content.

5. Risk Review

- Regularly review and update risk management strategies to keep up with changing threats.
- Improve security measures based on past incidents and new developments in online safety.

3. Automated, Traditional and Social recommender systems.

Recommender systems are tools designed to suggest relevant items to users based on their preferences, behavior, or social influence. There are three primary types of recommender systems: **Traditional**, **Automated**, and **Social**.

1. Traditional Recommender Systems
Definition:
Traditional recommender systems rely on human judgment, expertise, and manual processes to suggest items. These systems are often used in physical settings or where human interaction plays a key role.
Characteristics:

- Based on expert opinions or user reviews
- Can involve recommendations from staff or friends
- Often used in libraries, bookstores, or retail stores
- Tend to be less personalized but more human and trustworthy

Example:
A librarian recommending a book based on your reading history or preferences.

2. Automated Recommender Systems
Definition:
Automated systems use algorithms and data (such as user behavior, ratings, and preferences) to provide item suggestions automatically, without human intervention.

Characteristics:

- Highly personalized and data-driven
- Use techniques like collaborative filtering, content-based filtering, or hybrid models
- Efficient and scalable for large user bases
- Common in online platforms like Amazon, Netflix, and Spotify

Example:
Netflix suggesting shows based on what you've previously watched and rated.

3. Social Recommender Systems
Definition:
Social recommender systems use social data and networks—such as friends' preferences, likes, shares, and interactions—to make recommendations.
Characteristics:

- Leverage social media activity and peer influence
- Encourage discovery through community-driven insights
- Focus on shared tastes and trends within social groups

Example:
Facebook suggesting events your friends are attending, or Spotify showing playlists your friends liked.

Type	Basis of Recommendation	Personalization	Source of Trust	Example
Traditional	Human expertise, reviews	Low to Medium	Human interaction	A book recommended by a friend
Automated	Algorithms, user behavior	High	System accuracy	Netflix suggestions
Social	Friends' preferences, social data	Medium to High	Peer influence	Spotify showing friend's playlist

4. Differentiate between Traditional Media and Social Media

Difference Between Traditional Media and Social Media

Aspect	Traditional Media	Social Media
Definition	One-way communication using offline channels like TV, radio, and newspapers.	Interactive digital platforms where users create, share, and engage with content.
Communication	One-way (broadcasting information to the audience).	Two-way (users can interact, comment, and share content).
Content Creation	Controlled by professionals (journalists, publishers, broadcasters).	Anyone can create and share content instantly.
Speed of Distribution	Slow (requires time for production and publication).	Instant (real-time sharing and updates).
Audience Reach	Broad but limited to specific regions or demographics.	Global and easily accessible to anyone with an internet connection.
Cost	Expensive (printing, broadcasting, and advertising costs are high).	Low-cost or free (posting and sharing content is often free).
Interactivity	Minimal (no direct interaction with the audience).	Highly interactive (likes, comments, shares, and live discussions).
Customization	One-size-fits-all content for a broad audience.	Personalized content based on user preferences and behavior.
Examples	Newspapers, magazines, radio, television, billboards.	Facebook, Instagram, Twitter, YouTube, TikTok.

5. Explain common social media risks mitigation strategies

6. Specify the significance of social media KPI

Significance of Social Media KPIs
Social Media Key Performance Indicators (KPIs) are crucial metrics that help businesses track the success of their social media efforts. They align social media activities with overall business goals such as increasing brand awareness, driving website traffic, improving engagement, and boosting conversions. By regularly monitoring KPIs, businesses can make data-driven decisions, refine their strategies, and maximize the impact of their marketing efforts.

1. Reach
Definition: Reach measures the number of unique users who see your content, including both followers and non-followers.
Significance:

- Indicates how far your content is spreading across platforms.
- Helps assess brand visibility and the effectiveness of hashtags or shares.

Improvement Tips:

- Use trending hashtags and collaborate with influencers.
- Share content that encourages re-sharing and tagging.

2. Engagement
Definition: Engagement tracks how users interact with your content — through likes, comments, shares, and clicks.
Significance:

- Reflects how well your content resonates with the audience.
- Higher engagement often boosts visibility via platform algorithms.

Improvement Tips:

- Create interactive content like polls and questions.
- Respond promptly to comments to build community.

3. Traffic

Definition: Traffic measures the number of users visiting your website via social media platforms.

Significance:

- Indicates which platform drives the most interest in your business.
- Helps identify the effectiveness of social media campaigns in leading users to your website.

Improvement Tips:

- Use strong CTA buttons like “Learn More” or “Sign Up.”
- Share links across posts, bios, and stories consistently.

4. Conversion Rate

Definition: The percentage of users who complete a desired action (purchase, sign-up, etc.) after clicking your social media link.

Significance:

- Directly ties social media performance to tangible business outcomes.
- A high rate means your campaigns are compelling and effective.

Improvement Tips:

- Offer attractive incentives like discounts.
- Use optimized landing pages and test various content types.

Tracking and analyzing social media KPIs provides vital insights into what’s working and what needs improvement. These metrics not only help justify marketing investments but also guide strategic decisions for achieving long-term success in a highly competitive digital space.

7. What is a social media-based recommendation system and how does it differ from a traditional recommendation system?

It is an algorithm that suggests relevant products to users based on a variety of factors. Sometimes, when you search for a certain product on a website you notice that you start receiving several suggestions of similar products, there is a system behind this. It is generally used to target potential users more efficiently and improve the user experience by suggesting new items, saving users' time, and narrowing down the set of choices.

Feature	Social Media-Based Recommendation System	Traditional Recommendation System
Data Source	Social interactions, likes, comments, shares, friends' activity, and engagement trends	User purchase history, ratings, browsing behavior, and explicit preferences
Recommendation Type	People, posts, ads, trending topics, influencers, groups	Products, movies, books, music, articles
Personalization Basis	Network-driven (friends' preferences, social trends)	Individual-driven (past interactions and preferences)
Algorithm Approach	Social graph analysis, trust-based filtering, collaborative filtering, content-based filtering	Collaborative filtering, content-based filtering, hybrid methods
Influence Factor	Peer influence and social engagement impact recommendations	Personal preferences and past behavior determine recommendations
User Behavior Adaptation	Highly dynamic, updates in real-time based on social trends	Updates periodically based on browsing and purchase history
Diversity of Recommendations	High—recommends diverse content based on real-time discussions and social interactions	More structured—focuses on items similar to past behavior
Trust-Based Filtering	Uses trust scores from social connections to refine suggestions	Uses ratings and user-item interactions without social trust factors
Cross-Platform Influence	Integrates multiple social platforms (e.g., Facebook suggesting Twitter posts)	Usually restricted to a single platform (e.g., Amazon suggesting products on Amazon)
Example	Facebook suggesting "People You May Know" or trending posts based on friends' activity	Netflix recommending movies based on past viewing history

Aspect	Traditional Media	Social Media
Reach	Wider reach as it includes TV, radio, and print, accessible even without the internet.	Limited to internet users; only those with online access can engage.
Cost	Expensive to advertise as it requires purchasing airtime or ad space.	More cost-effective; allows targeted advertising and campaign adjustments.
Control	More control over the message, as it is one-way communication.	Less control due to two-way communication, allowing audience interaction and information spread beyond the intended audience.
Speed	Slower turnaround time due to production and distribution processes.	Faster content creation and distribution, allowing instant sharing.
Engagement	Lower engagement as it is primarily a one-way communication model.	Higher engagement as it enables two-way interaction, comments, and feedback.

A company pays thousands of dollars for a 30-second TV ad during prime time.	A startup runs a low-cost Instagram ad targeting specific demographics.

8. Types of social media risks

1. Cyberbullying
Cyberbullying involves using social media platforms to intimidate, harass, or demean others. It often takes the form of hurtful comments, rumors, or public shaming. Victims, especially teens, may suffer from anxiety, depression, and social withdrawal. This risk highlights the darker side of digital communication.

2. Uploading Inappropriate Content
Users sometimes post photos or videos that are embarrassing, offensive, or overly provocative. Such content, even if posted for fun, can harm personal reputation or cause professional consequences if seen by employers or the public. What's shared online can quickly become permanent and widely spread.

3. Oversharing Personal Information
Many users unintentionally share sensitive data like their birth date, location, or daily routines. This opens the door to identity theft, stalking, and scams. For example, posting about a vacation while away from home may alert criminals to an empty house.

4. Imposter Accounts
Fake accounts are created to impersonate individuals, celebrities, or brands, often for malicious reasons like scamming followers or spreading misinformation. These accounts can mislead the public and damage reputations if not quickly identified and removed.

5. Privacy Settings Mismanagement
Default privacy settings on social media profiles often expose more than users realize. Without adjusting these settings, personal posts, contact info, or sensitive business data can become public, making users vulnerable to unwanted attention or data misuse.

6. Phishing Attacks and Scams
Phishing scams use fake messages or links to trick users into revealing login credentials, banking details, or other personal info. These attacks are common on social platforms and can lead to major data breaches, especially when work-related accounts are compromised.

7. Productivity Loss
Excessive personal use of social media during work hours can reduce employee efficiency and focus. Organizations may suffer from missed deadlines or reduced output if staff are distracted by scrolling, posting, or engaging with online content during business hours.

8. Reputation Damage
A single inappropriate post, comment, or customer complaint can go viral and damage a person's or brand's reputation. Since social media content spreads fast, even an honest mistake can lead to public backlash, loss of customers, or long-term credibility issues.

9. Discuss the parameters used to measure success of a Social Media Platform

The success of a social media platform is measured using various key metrics that analyze user engagement, interaction, and influence. Below are some important parameters used to assess social media performance:

1. Counts (Followers, Likes, Views, and Shares)

- Measures the number of followers, likes, video views, and shares a post receives.
- Helps understand the reach and popularity of content on the platform.

2. Social Sharing

- Tracks how often users share posts, articles, or media with their followers.
- High social sharing indicates strong audience engagement and content relevance.

3. Engagement Rate

- Calculates the percentage of people who interact with posts through likes, comments, and shares.
  A high engagement rate means the content is resonating well with the audience.

4. Interaction

- Measures direct interactions such as comments, replies, and discussions.
- More interactions show active audience participation and interest in the content.

5. Referral Rates

- Tracks how much traffic is being directed from social media to websites or other platforms.
- A higher referral rate means social media is effectively driving users to external content.

6. Importance and Influence of Users

- Identifies key influencers and active users who contribute to spreading content.
  Engaging with influential users can boost brand credibility and reach.

10. What are some of the most common types of traditional recommendation systems, such as collaborative filtering or content-based filtering?

1. Explain collaborative and content based filtering in recommendation system

Collaborative Filtering is a way to recommend things based on how similar users or items are. Instead of looking at the features of the items themselves (like content-based filtering does), it looks at how users interact with those items, suchas their ratings or what they've bought before. The basic idea is that if two users have liked or bought similar things in the past, they will probably have similar preferences in the future.

Collaborative Filtering components:

1. Utility Matrix: This is a table where the rows represent users, and the columns represent items (like movies, products, etc.). The entries in the table show how a user interacted with an item (such as a rating or a purchase). This table is usually sparse because most users only interact with a few items.

2. User-User Similarity: By looking at the rows (user preferences) in the table, we can compare how similar two users are. If two users have interacted with items in a similar way, we can say they likely have similar tastes. This is calculated using a method like cosine similarity.

3. Item-Item Similarity: We can also compare the columns (item preferences) to find items that have been rated or purchased similarly by many users. This helps identify items that are alike based on how users interact with them.

4. Recommendation Generation:
   - User-based Collaborative Filtering: Recommend items that people similar to the target user liked but the target user hasn't interacted with yet.
   - Item-based Collaborative Filtering: Recommend items that are similar to the ones the target user has already interacted with.

Step 2: Measuring Similarity (User-Based Filtering)
We calculate the **similarity** between users based on their ratings of common items. For example, we can use cosine similarity, Jaccard similarity, or Pearson correlation to measure how close two users are.
•**User A and User C** have both rated Products 1 and 3. Since their ratings are similar for these products (both rated Product 1 highly and Product 3 moderately), they would be considered **similar users**.

Step 3: Making Recommendations
Based on similarity:
•**User-Based Filtering**: Since User A and User C are similar, the system may recommend Product 4 to User A because User C has rated Product 4 and User A hasn't interacted with it yet.
•**Item-Based Filtering**: We can also recommend products based on the similarity of items. For example, if **Product 1** and **Product 3** are frequently rated similarly by different users, the system may recommend Product 1 to users who have shown interest in Product 3.

Step 1: Utility Matrix Creation

A utility matrix is created where:

- Rows represent users (User A, User B, User C).
- Columns represent products (Product 1, Product 2, Product 3).
- Entries in the matrix show whether a user has purchased, viewed, or rated a product (e.g., ratings from 1 to 5).

	Product 1	Product 2	Product 3	Product 4	Product 5
User A	5		4		2
User B		3		5	1
User C	4		4	2	

Here:

- User A has rated Products 1, 3, and 5.
- User B has rated Products 2, 4, and 5.
- User C has rated Products 1, 3, and 4.

Step 4: Cluster Users and Items
In cases where the matrix is sparse (users haven't rated many items), the system might group or cluster users and items.
•**Clustering Users**: Users who have similar preferences may be grouped together. For example, a group of users who frequently buy high-end electronics might be clustered, and new products in this category will be recommended to the group.
•**Clustering Items**: Items like books, gadgets, or clothes that are frequently bought together or rated similarly by different users can be clustered. This allows for recommending entire clusters of products instead of individual ones.
Example Output for User A:
•**User-Based Recommendation**: Recommend **Product 4** to User A because similar users (e.g., User C) have rated it.
•**Item-Based Recommendation**: Recommend **Product 2** because it is similar to Product 5, which User A has interacted with.

Content-based filtering recommends items by comparing their attributes (content) to what the user has interacted with previously. It focuses on the features of the items rather than user interactions.

Assigning attributes

Content-based filtering assigns specific attributes to items in a database, enabling the algorithms to categorize each product and analyze user behavior. For instance, Amazon analyzes attributes like titles, descriptions, and product features to create a comprehensive item profile for each product.

When a user shows interest in a smartphone with certain specifications—such as a high-resolution camera or long battery life—the system recommends similar phones with matching attributes. It might also recommend phone cases, car mounts, and other accessories that match the user's preferences, based on the similarity between the user vector and various item profiles.

In this way, Amazon uses content-based filtering to tailor recommendations to individual preferences. This both highlights items users may like and enables the company to improve revenue through cross-sells, up-sells, and increased engagement.

Building a user profile

User profiles are central to content-based recommender systems. They encompass the user's interactions with database objects—such as purchases, searches, and user ratings—and their attributes. These user-item interactions form the foundation of the user vector, which represents the user's preferences in the system.

In these profiles, attributes that appear in multiple interactions receive higher weighting, indicating their importance to the user's preferences. This process involves constant user feedback, typically through ratings and other forms of explicit feedback, to refine the weighting of different items.

The system then creates a model reflecting each user's likes and dislikes based on their past activities and weighted by attribute importance. Each database object is scored for its similarity to this user profile, often using techniques like cosine similarity, ensuring tailored recommendations.

Example: Suppose you've listened to Billie Eilish's "Happier Than Ever," Dua Lipa's "Don't Start Now," and Olivia Rodrigo's "Drivers License."

A recommender system might deduce you enjoy songs by contemporary female pop artists with themes of self-reflection and empowerment. Expect to receive suggestions for similar tracks by these and other artists, like Ariana Grande's "thank u, next."

11. What are different threats to privacy on social media?

1. Data Harvesting & Tracking

Social media platforms collect vast amounts of user data, including personal details, browsing behavior, and engagement patterns. This data is often used for targeted advertising and analytics. However, if misused, it can lead to:

- Unwanted personalized ads
- Profile tracking by third parties
- Sale of user data to unknown entities

Example: Social media companies may track the posts you like, the pages you visit, and even the time spent on content to refine their advertising strategies.

Prevention: Regularly review privacy settings and limit the data shared with platforms.

2. Identity Theft

Personal details shared on social media—such as full name, date of birth, and location—can be exploited by criminals to impersonate individuals for fraud or financial crimes.

Example: Attackers may use stolen personal information to open bank accounts, apply for loans, or commit fraud in the victim's name

Prevention: Avoid sharing sensitive personal details publicly and enable two-factor authentication (2FA) for accounts.

3. Location Tracking & Geotagging

Many social media platforms allow users to share their live location, which can expose them to stalking, burglary, or personal threats.

Example: Posting vacation photos with location tags can inform criminals that a person's home is unoccupied.

Prevention: Disable geotagging on social media posts and limit location-sharing permissions for apps.

4. Third-Party App Risks

Many social media users connect third-party apps (e.g., quiz apps, games, or productivity tools) to their accounts, which may request access to private data. Some of these apps misuse or leak this data.

Example: The Cambridge Analytica scandal involved a quiz app collecting Facebook users' data and using it for political profiling.

Prevention: Revoke permissions for unnecessary third-party apps and avoid granting excessive access to applications.

12. Trust-based collaborative filtering is an improvement over traditional similarity-based recommendation methods. Discuss two key advantages of using trust-weighted similarity in recommendation systems.

Traditional similarity-based recommendation methods rely on comparing users' or items' interactions, behaviors, or attributes to find similarities and generate recommendations. These methods typically use metrics like cosine similarity or Pearson correlation to measure how alike two users or items are, recommending items based on those similarities. However, these methods can be limited by sparse data or biases in user preferences.

Trust-based collaborative filtering, on the other hand, incorporates trust relationships between users. Instead of just using behavior-based similarities, it weighs recommendations based on the level of trust between users. This method assumes that a user is more likely to trust and act on recommendations from someone they have a positive relationship with, leading to more personalized and reliable recommendations.

1. Better Personalization

- **Trust-based recommendations enhance relevance**: Trust-based collaborative filtering takes into account the relationships between users, meaning that recommendations are influenced by users who are trusted. If a user trusts another user (e.g., based on shared interests or past interactions), the recommendations made by that trusted user carry more weight. This ensures that the system provides more personalized suggestions that align better with the user's preferences.
- **Addresses cold-start problem**: In traditional similarity-based methods, when a user has few interactions, the system struggles to find accurate recommendations. However, in trust-based filtering, even with minimal interaction data, a user can still receive accurate suggestions by relying on trusted users with similar interests or preferences. This improves the recommendation experience for new users or users with sparse data.

2. Reduced Noise and Bias

- **Mitigates irrelevant or noisy recommendations**: Traditional methods often rely on aggregating all user interactions to measure similarity, which can lead to noisy recommendations. For example, if a user has a diverse range of interactions but no clear pattern, similarity-based methods might recommend irrelevant items. In contrast, trust-based systems focus on recommendations from trusted users, which are more likely to be relevant to the target user, reducing the impact of noisy or irrelevant suggestions.

- **Reduces bias from large, heterogeneous datasets**: Traditional similarity-based methods can be biased by the majority of users or popular trends, meaning they may favor widely liked items, even if they are not a good fit for a specific user. Trust-based filtering mitigates this bias by prioritizing recommendations from users who have similar preferences and trust each other, offering a more refined and accurate set of suggestions that align with a user's true interests, rather than general popularity trends.

13. Provide an example of a real-world application where trust scores are used to enhance recommendations (e.g., social media, e-commerce, travel platforms). Explain how trust influences the recommendations in that scenario.

A real-world application where trust scores are used to enhance recommendations is **E-commerce platforms** such as **Amazon** or **eBay**. In these platforms, trust scores influence recommendations by considering user reviews, ratings, and past purchase behaviors, along with direct trust relationships between users or sellers and buyers.

1. Trust in Product Reviews and Ratings:

- **Reputation-Based Reviews**: On platforms like Amazon, user reviews play a significant role in shaping purchasing decisions. However, not all reviews are created equal. Trust scores are integrated into the review system where reviews from highly trusted users (those with a high reputation score or frequent, relevant reviews) carry more weight than reviews from new or unknown users. For example, a user may trust reviews from top-rated reviewers who consistently provide helpful and detailed feedback, ensuring that the recommendations they see are of high quality and relevance.

- **Sentiment-Based Trust**: Trust is also incorporated by analyzing the sentiment of reviews from users that have similar tastes. For instance, if a customer consistently enjoys a particular genre of products (like eco-friendly or tech gadgets), reviews from other users with similar preferences are weighted more highly. Thus, the system can prioritize products with reviews from people who have a trusted history of recommending products the user likes

2. Trust in Seller and Brand:

- **Seller Trust and Buyer History**: E-commerce platforms like Amazon also incorporate trust scores based on buyer-seller relationships. For example, if a user has previously bought products from a particular seller and had a positive experience (e.g., high-quality products, good customer service, fast shipping), their trust score with that seller increases. Future product recommendations will prioritize listings from those trusted sellers, as the platform recognizes that the buyer has had satisfactory transactions with them in the past.

- **Repeat Purchase Behavior**: Trust scores are built over time, with each interaction increasing or decreasing trust. For instance, if a user buys from a particular brand repeatedly and has a history of leaving positive feedback, the platform will increase the weight given to that brand's products when suggesting new items. This personalized trust-based recommendation ensures that the user gets products that align with their past preferences.

3. Trust-Based Filtering in Deals and Discounts:

- **Trusted Offers**: E-commerce platforms use trust-based filtering to enhance deal recommendations. For example, discounts or limited-time offers from trusted sellers or brands can be highlighted more prominently for a user. If a user has a positive history with certain brands, they are more likely to receive alerts or recommendations about sales or special offers from those specific brands. This makes the promotional content more relevant and increases the likelihood of conversions.

14. In a collaborative filtering-based recommender system, the Pearson correlation coefficient is commonly used to measure user similarity. Explain how trust scores can enhance this method and modify recommendations.

In a **collaborative filtering-based recommender system**, the **Pearson correlation coefficient (PCC)** is widely used to measure the similarity between users based on their past ratings. While effective, PCC has limitations, such as susceptibility to noisy data, lack of reliability in sparse datasets, and an inability to consider external trust relationships

1. Improving Recommendation Accuracy

- Trust scores represent the **degree of reliability** a user has in the system (e.g., based on reputation, past behavior, or explicit trust relationships).
- By weighting the Pearson correlation coefficient with trust scores, recommendations can prioritize highly trusted users, reducing the influence of outliers or malicious users.

2. Handling Data Sparsity

- In many recommender systems, users rate only a small fraction of items, leading to sparse user-item matrices.
- Trust-based filtering can supplement similarity-based approaches by inferring preferences through **trusted neighbors**, even if they have few overlapping ratings.

3. Reducing the Cold Start Problem

- New users often lack sufficient interaction history for meaningful similarity calculations.
- Trust scores allow recommendations to be based on **trusted users' preferences**, rather than requiring direct rating overlaps.

4. **Dynamic Adaptation**

- Trust scores can be **updated over time** based on user interactions, feedback, and observed behavior.
- This makes the recommendation system more **adaptive and resilient** to evolving user behavior patterns.

15. ROI (Return Of investment) Analysis from social media activities and campaigns

ROI (Return on Investment) analysis from social media activities and campaigns helps businesses assess the financial effectiveness of their social media efforts. It compares the revenue generated against the costs of campaigns, allowing companies to determine whether their social media activities are profitable. By tracking metrics such as conversions, leads, and sales, businesses can optimize their strategies, allocate resources efficiently, and improve the overall return on investment.

- ROI analysis helps measure the profitability and effectiveness of social media activities by comparing the revenue generated to the cost of campaigns.
- It provides insights into the financial returns from social media investments, helping businesses determine whether their social media efforts are yielding positive outcomes.
- By tracking key metrics like conversions, leads, and sales, ROI analysis evaluates the impact of social media campaigns on business objectives.
- It helps businesses justify the allocation of resources by showing how social media campaigns contribute to the bottom line.
- ROI analysis helps measure customer acquisition costs, allowing businesses to assess whether the cost of gaining a customer through social media is sustainable.
- It assists in determining the most profitable platforms and strategies, allowing businesses to focus their resources on high-performing social media channels.
- Social media ROI analysis aids in understanding the long-term value of customer relationships created through social media efforts.
- By analyzing ROI, businesses can refine their social media strategies, optimize campaigns, and ensure better resource allocation for future activities.

Module 6

1. Differentiate among social media, Web 2.0, and social network sites.
2. Discuss various privacy attributes of Social Media Sites.
3. Issues with the privacy policies.
4. Explain the ways to measure the success of a company having social media
5. Benefits of social media users who use social media (5 mks)
6. Explain the benefits of brand building (5 mks)
7. Intention analysis on social media
8. Short notes on: Business use of Social Media and Social Media in Public Sector
9. Summarize ethical issues when mining social media
10. Relate different techniques to secure social media accounts

8. Short notes on: Business use of Social Media and Social Media in Public Sector

Social media has become a powerful tool for businesses, offering various benefits to improve operations, marketing, and customer engagement. Businesses and non-profit organizations use social media in ways similar to the public sector, leveraging it for marketing, customer service, and communication.

Uses of Social Media for Business:

- **Marketing & Advertising**: Businesses use social media to reach potential customers, promote products, and increase brand awareness.
- **Customer Service**: Engaging with customers, addressing inquiries, and resolving issues promptly.
- **Public Relations**: Managing business reputation, responding to feedback, and sharing company news.
- **Talent Recruitment**: Platforms like LinkedIn help attract and recruit skilled employees.
- **Market Research**: Analyzing social media trends, customer feedback, and competitor strategies.
- **Lead Generation**: Connecting with potential customers through targeted content and campaigns.
  Brand Development: Building brand identity and customer loyalty through consistent engagement.
- **Competitive Analysis**: Monitoring discussions about competitors and industry trends.

Benefits of Social Media for Business:

- Increased customer engagement and loyalty.
- Expanded market reach, including international customers.
- Cost-effective advertising and promotion.
  Higher website traffic and improved search engine rankings.
- Exchange of ideas to enhance business operations.

Social Media in Public Sector

Social media is transforming the way public sector organizations communicate, engage, and serve the public. It provides a powerful platform for sharing information, improving transparency, and enhancing public services.

Uses of Social Media in the Public Sector:

- **Improved communication with the public**: Social media platforms allow public sector organizations to directly communicate with the public, providing real-time updates and information about services and initiatives.
- **Increased transparency**: Social media can help increase transparency in government operations by providing a platform for open communication and allowing the public to directly engage with and ask questions of government officials.
- **Improved customer service**: Social media can be used as a customer service tool, allowing public sector organizations to quickly and effectively address issues and concerns raised by the public.
- **Greater reach and engagement:** Social media allows public sector organizations to reach and engage with a wider audience, including those who may not traditionally interact with government.
- **Enhanced community involvement:** Social media can be used to facilitate community involvement in government decision-making and facilitate collaboration between government and community organizations.
- **Increased efficiency and cost savings**: By using social media for tasks such as answering frequently asked questions, public sector organizations can save time and resources that would otherwise be spent on more traditional forms of communication.
- **Enhanced crisis management:** Social media can be used to quickly disseminate important information during a crisis and coordinate responses.

Miscellaneous

1. List and explain the different network layouts(5M)