**1)Discuss the internet users mind set before and after Social Media?**

Answer: Before the rise of social media, internet users typically approached the internet as a tool for finding information and connecting with others through email, forums, and instant messaging. The mindset was generally more focused on seeking out specific information or interacting with others in a more intentional way.

With the advent of social media, internet users began to view the internet as a platform for constant communication and sharing of personal information. Users were encouraged to share their thoughts, feelings, and experiences with others in real-time, and this led to a greater sense of connectedness and community online.

However, the increased emphasis on sharing personal information also led to concerns about privacy and security online. As social media became more integrated into daily life, users became more aware of the potential risks associated with sharing personal information online.

Overall, the rise of social media has led to a shift in the mindset of internet users, from a more focused and intentional approach to the internet to a more open and connected one, but with increased awareness of privacy concerns.

**2)Discuss the advantages of using the social media?**

Ans : Social media has become an integral part of our daily lives, and its advantages are numerous. Here are some of the most significant advantages of using social media:

**Connection:** Social media allows you to connect with people all over the world. It helps you stay in touch with friends and family, connect with people who share your interests, and even make new friends.

**Information:** Social media provides a wealth of information on a variety of topics, from news and current events to entertainment and education. It allows you to stay up-to-date on the latest trends and developments in your industry or field.

**Networking:** Social media provides an excellent platform for networking. It allows you to connect with like-minded professionals, share your work, and find job opportunities.

**Marketing:** Social media is a powerful tool for marketing your products or services. It allows you to reach a vast audience at a fraction of the cost of traditional advertising.

**Community building:** Social media allows you to build and engage with a community of followers around your brand, cause, or interest. It creates a sense of belonging and fosters loyalty.

**Customer support:** Social media allows you to provide excellent customer support by addressing questions and concerns quickly and efficiently. It allows you to build trust and loyalty with your customers.

**Creativity:** Social media provides a platform for creative expression. It allows you to share your art, writing, music, and other creative works with a wide audience and receive feedback and support.

Overall, social media offers numerous advantages, from personal connection to professional networking, marketing, and customer support. However, it is essential to use social media responsibly and be mindful of the potential downsides, such as cyberbullying, misinformation, and addiction.

**3)Discuss the disadvantages of using social media?**

Ans : While social media platforms have many benefits, there are also several disadvantages that users should be aware of. Here are some of the most significant disadvantages of using social media:

**Cyberbullying:** Social media provides a platform for people to harass, bully, and troll others anonymously. This can lead to mental health issues and even suicide in extreme cases.

**Addiction**: Social media addiction is a real problem, and many people find themselves spending hours on social media platforms each day, even when they don't want to.

**False Information**: Social media has been accused of spreading false information, rumors, and conspiracy theories. These can be harmful, especially when they concern health or safety issues.

**Privacy Concerns**: Social media platforms often collect a vast amount of personal data from their users. This data can be used by companies to target users with ads, and it can also be stolen or misused by hackers.

**Distraction:** Social media can be a significant distraction, taking away time that could be spent on more productive activities such as work or spending time with loved ones.

**Social Isolation**: While social media can help people connect with others, it can also lead to social isolation. Some users may become so absorbed in their online life that they neglect their real-world relationships.

**Comparison:** Social media can lead to feelings of inadequacy and low self-esteem as users compare themselves to others who seem to have more successful, happier lives.

Overall, social media platforms have many disadvantages, and it's essential to be aware of them when using these platforms. It's important to use social media responsibly and to balance online interactions with real-world experiences.

**4)Discuss the reasons why million use social media ?**

There are many reasons why millions of people use social media. Here are some of the most common:

**Connecting with friends and family**: Social media allows people to stay connected with friends and family members, no matter where they are in the world. Users can easily share updates, photos, and videos with their loved ones and keep in touch on a regular basis.

**Networking:** Social media platforms provide an opportunity for professionals to connect with others in their industry and build valuable relationships. LinkedIn, for example, is a popular social media platform for business networking.

**Entertainment:** Social media can be a source of entertainment, with users able to watch funny videos, view memes, and explore trending content. Platforms like TikTok and Instagram have gained a huge following for their entertaining content.

**Marketing and advertising**: Social media has become a crucial tool for businesses to reach their target audience and promote their products or services. Advertising on social media can be highly targeted and cost-effective, making it an attractive option for businesses of all sizes.

**News and information**: Social media platforms like Twitter and Facebook are often used to share breaking news and information. Users can also follow news organizations and journalists to stay informed on the latest developments.

**Education and learning:** Social media can also be a valuable resource for learning and education. Platforms like YouTube and Instagram are popular for educational content, with users able to find tutorials, how-to guides, and other instructional videos.

Overall, social media has become an integral part of many people's lives for a variety of reasons, from staying in touch with loved ones to advancing their careers and staying informed on the latest news and trends.

**5)Define social media analytics?**

Social media analytics refers to the process of collecting, analyzing, and interpreting data from various social media platforms to gain insights and extract meaningful information for decision-making and strategic planning. It involves using specialized tools and techniques to track, measure, and evaluate social media performance, user behavior, and trends.

Social media analytics encompasses a wide range of activities, including data collection, data preprocessing, data analysis, and data visualization. It involves monitoring social media channels, such as Facebook, Twitter, Instagram, LinkedIn, and others, to gather data on user interactions, engagement, sentiment, demographics, and other relevant metrics.

The data collected through social media analytics can be used by businesses, organizations, and individuals to understand audience behavior, measure the effectiveness of social media campaigns, track brand sentiment, identify emerging trends, conduct competitor analysis, and make data-driven decisions to optimize their social media strategies.

Social media analytics is a valuable tool for marketing, customer service, brand management, product development, and overall business strategy. It provides insights into how social media is impacting a brand's online presence, customer perception, and business performance, and helps in shaping marketing strategies and tactics accordingly.

**6)Discuss various methodologies used in social media analytics?**

Ans : Social media analytics is the process of analyzing social media data to gain insights and understandings of user behavior, sentiment, trends, and patterns. Here are some of the methodologies used in social media analytics:

**Sentiment Analysis:** Sentiment analysis is a method of analyzing text data to determine the emotional tone and sentiment of the content. It involves categorizing opinions as positive, negative, or neutral.

**Network Analysis**: Network analysis involves analyzing social media networks to understand how users interact with each other. It helps identify influencers, key topics, and user behavior patterns.

**Text Analytics:** Text analytics involves analyzing text data to gain insights into user behavior, sentiment, and opinions. It involves techniques such as keyword analysis, topic modeling, and entity recognition.

**Social Listening**: Social listening involves monitoring social media platforms to identify mentions of a brand, product, or service. It helps businesses understand customer needs, preferences, and feedback.

**Predictive Analytics**: Predictive analytics involves using statistical models to predict future trends and outcomes based on historical data. It helps businesses make data-driven decisions and identify opportunities for growth.

**Image and Video Analysis**: Image and video analysis involves analyzing visual content to understand user behavior, sentiment, and opinions. It involves techniques such as image recognition, object detection, and facial recognition.

**Social Media Monitoring:** Social media monitoring involves tracking social media activity to measure brand performance, identify opportunities for engagement, and improve social media strategy.

Overall, social media analytics is a multidisciplinary field that involves a range of methodologies to gain insights into social media behavior and trends.

**7)Explain the advantages of implementing the social media analytics?**

**8)Discuss the handling of unstructured data?**

**9)Write about the following terms**

**1.tagging unstructured data:**

Tagging of unstructured data refers to the process of adding metadata to unstructured data in order to make it easier to organize, search, and analyze. Unstructured data refers to any data that is not organized in a predefined manner, such as text documents, images, audio files, and videos.

Tagging involves identifying and labeling specific pieces of data within these unstructured sources with relevant keywords, categories, or other descriptive information. This can be done manually or through automated processes like machine learning.

Tagging helps to make unstructured data more searchable, which can be especially helpful for large datasets where finding specific pieces of information can be difficult. It also allows for easier analysis of the data, as tags can be used to group related data together.

**2.tokenization** : Tokenization is the process of breaking a text corpus up into words (most

commonly), phrases, or other meaningful elements, which are then called tokens. The tokens become the basic units for further text processing.

*tokens=nltk.word\_tokenize(verbatim)*

Other techniques are spelling correction, domain knowledge, and grammar

checking

**3.word vector**

is a neural network that is able to generate a dense representation of words. This algorithm, more specifically an unsupervised learning algorithm, try to predict a word based on its neighbors, hence, in this case, the algorithm would predict a follower based on a user.

**4.lemmatization and 5.Stemming**

The main aim of stemming and lemmatization is

to reduce inflectional forms and sometimes derivationally related forms of a word to a common base form. Stemming reduces word forms to so-called stems,whereas lemmatization reduces word forms to linguistically valid lemmas.

Some examples of stemming are cars -> car, men -> man, and went ->go

Such text processing can give added value in some domains, and may

improve the accuracy of practical information extraction tasks

**6.flat file**

A flat file is a type of data file that contains data in a plain text or binary format, where each line of the file represents a record and each field in a record is separated by a delimiter, such as a comma or tab. Flat files are often used to store data that is organized in a simple and straightforward manner, such as lists of names and addresses, inventory data, or financial transactions.

Flat files are easy to create and manipulate using text editors, spreadsheets, or programming languages. They are also portable and can be easily transferred between different systems and applications. However, flat files can become difficult to manage when dealing with large amounts of data or complex data relationships. In these cases, a relational database management system (RDBMS) may be more suitable, as it allows for more efficient data retrieval, querying, and management.

Overall, flat files are a simple and flexible way to store and manage data, but their suitability depends on the specific requirements of the application or system.

**7.no sql database(mongo)**

MongoDB (from humongous) is a cross-platform document-oriented database. Classified as a NoSQL database, MongoDB eschews the traditional table-based relational database structure in favor of JSON-like documents with dynamic schemas (MongoDB calls the format BSON), making the integration of data in certain types of applications easier and faster. Released under a combination of the GNU Affero General Public License and the Apache License, MongoDB is free and open-source software

**8.relational database**

A relational database is a type of database management system (DBMS) that organizes data in a tabular format, with rows and columns, and establishes relationships between the data elements. In a

relational database, data is stored in tables, with each table representing a specific entity or concept, and rows within the table representing individual instances or records of that entity. The columns in the table, also known as fields, define the attributes or properties of the entity, and each row contains values for those attributes.

**10)Discuss unsupervised learning?**

Ans : Unsupervised learning is used when we do not know the outputs. In the case of cars, we only have technical specifications: acceleration, price, engine type. Then we cluster the data points into different groups (clusters) of similar cars. In our case, we will have the clusters with similar price and engine types. Then, we understand similarities and differences between the cars

**Supervised learning:**

Supervised learning assumes that we know what the outputs are of each data point. For example, we learn that a car that costs $80,000, which has an electric engine and acceleration of 0-100 km/h in 3 seconds, is called Tesla; another car, which costs $40,000, has a diesel engine, and acceleration of 0-100 km/h in 9.2 seconds, is called Toyota; and so on. Then, when we look for the name of a car which costs $35,000, has acceleration of 0-100 km/h in 9.8 seconds, and has a diesel engine, it is most probably Toyota and not Tesla.

**11)Discuss about sentiment analysis along with its sub tasks?**

Sentiment analysis is a natural language processing (NLP) technique used to determine the emotional tone of a piece of text. It involves analyzing text data to identify the underlying sentiment or emotional attitude of the writer towards a particular topic, product, service, or entity.

**There are three main sub-tasks involved in sentiment analysis:**

Subjectivity Analysis: This involves identifying whether a piece of text expresses an opinion or a fact. It involves classifying text as either subjective or objective.

Polarity Analysis: This involves identifying the sentiment or emotional tone of a piece of text. It involves classifying text as positive, negative, or neutral.

Aspect-Based Sentiment Analysis: This involves identifying the sentiment of individual aspects or components of a product or service. It involves breaking down a piece of text into its constituent parts and analyzing the sentiment of each part separately.

Sentiment analysis is used in various fields such as market research, social media analysis, customer feedback analysis, and more. It can help businesses and organizations understand the sentiments and opinions of their customers, which can be used to improve their products, services, and customer satisfaction.

**12)Define API’s**

A widely used term, API (Application Programming Interface) is defined as a set of instructions and standards to access a web based software application. But what does it mean in real life? Firstly, APIs allow users to send a request for a particular resource, such as Facebook or Twitter , and receive some data in response. It is worth noting that all API providers fix some limitations on the quantity or type of data which users can obtain. APIs give access data processing resources, such as AlchemyAPI that receives in a request verbatim (textual data) and sends in response all results of the analysis, such as nouns, verbs, entities, and so on. In our case, the APIs are used either to get data from social networks or to execute some complex processing on them. In order to access and manipulate APIs, we have to install the *urllib2* library

*pip3 install urllib2*

In some cases, if you fail to perform the installation. You can also try using the SFRVFTU library, which is compatible with Python 2.x and 3.x

*pip install request*

**13)Discuss various types of API’s?**

Currently, two types of API are available. They are as follows:

**RESTful API**

**Stream API**

RESTful API This is the most common type of API that most social media provides. The information from a REST API is static and is from historical data. The back history of data can vary from platform to platform. Facebook calls its REST API service Graph API. Facebook, Twitter, and Pinterest among others have given access to their APIs to the public through robust data infrastructures. Another real-time version of the RESTful API of Twitter is the Streaming API. The two are inherently similar, but with different usages. We will explore the standard RESTful APIs of the most common platforms and also the Streaming API of Twitter.

REST stands for Representational State Transfer and it relies on the HTTP protocol for data transfer between machines. It has been created to simplify the transfer of data between machines unlike previous web services such as CORBA, RPC, and SOAP. Since the architecture of REST uses the HTTP protocol, it would be fair to assume that the WWW itself is based on RESTful design. Two of the most important uses of RESTful services are

* Get : Procedure to receive data from a distant machine
* Post : Procedure to write data to a distant machine

Almost all the functionalities of a REST API can be used through the preceding two methods

Stream API You need a Stream API when the requirement is to collect data in real time, instead of backdated from the platform. The Stream API of Twitter is widely used to collect real-time data from Twitter. The output is quite similar to that of a REST API apart from the real-time aspect. We'll see examples of the Twitter Stream API and its outputs

**14)Discuss the advantages disadvantages of social media API’s?**

**Advantages of social media APIs Social media APIs have many advantages.**

The main advantages are: Social data: APIs allow you to extract valuable data around Social Media users and content that is used for behavioral analysis and user insights. App development: Thousands of software and applications have been built using Social Media APIs that provide additional services on top of Social Media platforms. Marketing: Social media APIs are useful in automating marketing activities such as social media marketing by posting on platforms. It also helps in enriching marketing data through Social Data acquired about customers.

**Disadvantages of social media APIs Social media APIs have many Disadvantages**

While APIs (Application Programming Interfaces) are incredibly useful for enabling communication and integration between different software systems, they also have some disadvantages. Here are some common disadvantages of APIs:

**Dependence on external systems:** When you use an API to connect your application to an external system or service, you become dependent on the stability, reliability, and availability of that system. If the API provider experiences downtime or changes the API without proper notice, it can disrupt the functioning of your application.

**Versioning and compatibility issues**: APIs can evolve over time, and API providers may release new versions with changes or improvements. This can result in compatibility issues with older versions of the API, requiring you to update your code to work with the latest version. This can be time-consuming and may require additional effort to ensure backward compatibility.

**Security concerns:** APIs can be a potential security risk if not implemented and secured properly. Unauthorized access to APIs, inadequate authentication and authorization mechanisms, or insecure data transmission can result in data breaches or other security vulnerabilities.

**Limited control over functionality:** APIs provide a predefined set of functionalities and may not offer all the features or customization options you need. This can limit your ability to fully customize or tailor the functionality of the API to suit your specific requirements.

**Reliance on third-party providers**: When you use APIs from third-party providers, you are relying on their services and infrastructure. If the third-party provider faces financial or operational challenges, discontinues the API, or goes out of business, it can affect the continuity and stability of your application.

**Performance and latency**: API calls involve communication over the network, which can introduce performance and latency considerations. Slow response times, high network latency, or API rate limits can impact the performance and responsiveness of your application.

**Cost implications**: Some APIs may come with usage limits, pricing tiers, or subscription fees, which can add to the operational costs of your application. Depending on your usage patterns, API costs can become a significant factor in your overall project budget.

**Legal and compliance considerations**: APIs may come with terms of use, licensing agreements, and legal requirements that you need to comply with. Failure to comply with these legal and compliance obligations can result in legal liabilities or regulatory issues.

It's important to thoroughly evaluate the advantages and disadvantages of APIs and carefully consider the specific needs and requirements of your application before incorporating APIs into your software

system. Proper planning, implementation, and ongoing management can help mitigate these disadvantages and ensure successful integration with APIs.

**15)Discuss the step required to put in place a client with Oauth authorization?**

Ans : There are several steps that are required to put in place a client with OAuth authorization

**1. Creating a user/developer account:** First of all, you have to register a user/developer account and provide personal information such as a valid email address, name, surname, country, and in many cases a valid telephone number (the verification process is done by sending you a text message with a code). **2. Creating an application:** Once you create your account, you will have access to a dashboard, which is very often called a developer console. It provides all the functionalities to manage your developer account, create and delete applications, or monitor your quota. In order to obtain access credentials you will have to create your first application via this interface.

**3. Obtaining access tokens**: Then, you generate access tokens for your application and save them in a safe place. They will be used in your code to create an OAuth connection to the API.

**4. Authorizing HTTP requests (optional):** Some APIs require HTTP request authorization, which means that a request has to contain an additional authorization header that provides the server with information about the identity of the application and permission scope.

**5. Setting up permission scopes (optional):** Some APIs have the notion of multilevel permissions. In that case when you generate your API key you need to specify the scope for the key. Scope here refers to a set of allowed actions. Therefore, in cases where an application attempts an action that is out of its scope, it will be refused. This is designed as an additional security layer. Ideally one should use multiple API keys, each with restricted scopes, so that in the scenario where your API key is hijacked, due to the restrictions in its scope the level of potential harm is restricted.

**6. Connecting to the API using obtained access tokens:** When all the preceding steps are configured, you can make requests using your access tokens. Now, the only limitation is the request quota, which depends on each platform

**16)Discuss why we need to use oauth?**

Social media networks APIs aim to provide full interaction with third-party applications allowing all kinds of access within rate limits. Thus, applications can perform actions on behalf of their users and access their data. The main advantage of this protocol is full security and the fact that the connection protocol is standardized. Therefore, there are standard ways of writing code and using request libraries. Moreover, an OAuth connection is the most proper and reliable technique that adheres to the developer policy defined by social network companies. The main advantage for the user is that it gives the highest available quota and very often more API endpoints to collect the data

**17)List the applications of data mining?**

Data mining has a wide range of applications across various industries. Some of the most common applications of data mining include:

**Customer Relationship Management (CRM):** Data mining helps organizations identify customer behavior patterns and preferences, which can be used to improve customer satisfaction and loyalty.

**Marketing:** Data mining can be used to identify potential customers, analyze their behavior, and develop targeted marketing campaigns.

**Fraud Detection:** Data mining can help identify fraudulent activities in financial transactions, insurance claims, and other areas.

**Healthcare:** Data mining can be used to analyze medical data to identify patterns and trends in patient diagnosis and treatment outcomes.

**Retail:** Data mining can help retailers identify buying patterns and preferences of their customers, which can be used to improve sales and customer loyalty.

**Manufacturing:** Data mining can help identify patterns and trends in manufacturing processes, which can be used to improve quality and reduce costs.

**Education:** Data mining can be used to analyze student performance data to identify factors that contribute to academic success and develop targeted interventions.

**Telecommunications:** Data mining can be used to analyze call detail records to identify customer behavior patterns and preferences, which can be used to develop targeted marketing campaigns and improve customer satisfaction.

**Banking:** Data mining can be used to analyze customer transaction data to identify patterns and trends in customer behavior and preferences, which can be used to develop targeted marketing campaigns and improve customer satisfaction.

**Sports:** Data mining can be used to analyze sports performance data to identify factors that contribute to athlete performance and develop targeted training programs.

**18)Discuss the challenges with respect to big data?**

Ans: Big data refers to large volumes of data that are being generated, processed, and analyzed at an unprecedented scale. While big data can provide valuable insights and opportunities, there are several challenges associated with it. Here are some of the major challenges:

**Data quality:** With the huge amount of data generated, ensuring data quality becomes a big challenge. It is essential to ensure that the data is accurate, complete, and consistent, as inaccurate or incomplete data can lead to erroneous results.

**Data security and privacy:** Big data typically contains sensitive and confidential information, and ensuring data security and privacy is a major challenge. Unauthorized access, data breaches, and cyber attacks are a constant threat to big data.

**Data integration:** Big data is often generated from multiple sources, and integrating these data sources can be a major challenge. It is essential to ensure that the data is integrated accurately, and inconsistencies or discrepancies are resolved.

**Scalability**: As the volume of data increases, the systems used to store, process, and analyze the data must be able to scale up to handle the increased workload. This requires a highly scalable infrastructure and the ability to add resources on demand.

**Analysis complexity**: Big data often requires advanced analytics techniques to extract meaningful insights. This requires skilled professionals who can analyze the data, develop models, and interpret the results.

**Cost:** The infrastructure, tools, and resources required to manage big data can be expensive. The cost of storing, processing, and analyzing large volumes of data can be a major challenge, especially for small and medium-sized businesses.

**Legal and ethical issues:** The use of big data raises several legal and ethical issues, including data ownership, privacy, and compliance with regulations such as GDPR and CCPA.

Overall, big data presents several challenges that must be addressed to ensure that organizations can derive value from their data while minimizing risks and costs.

**19)Explain bag of words model?**

The bag of words (BoW) model is a natural language processing technique used for extracting information from textual data. In this model, a text is represented as a bag (multiset) of its words, disregarding grammar and word order but keeping track of word frequency. The model derives its name from the fact that it literally creates a bag of words from the text.

**The BoW model typically involves the following steps:**

**Tokenization:** The text is broken down into individual words or tokens.

Vocabulary creation: A unique vocabulary is created by taking all the unique words from the entire text corpus.

**Encoding:** Each document in the corpus is encoded as a vector by counting the frequency of each word in the vocabulary. The resulting vector represents the document in the high-dimensional space of the vocabulary.

**Similarity calculation**: The similarity between documents can be calculated based on the distance between their respective vectors.

The BoW model is a simple yet effective way to represent textual data, and is often used as a basis for more sophisticated natural language processing techniques such as topic modeling and sentiment analysis. However, it has limitations since it doesn't take into account the order of the words, grammatical structures, or other contextual information that can be important in understanding natural language.

**20)Explain the necessity of simulation?**

Simulation is a powerful tool used in many fields, including engineering, science, finance, and computer science. Its necessity lies in its ability to model complex systems and processes in a way that can be analyzed, tested, and optimized before implementation in the real world.

Here are some of the key reasons why simulation is necessary:

**Risk reduction:** Simulation can help reduce the risk of failure by identifying potential issues and allowing for testing and refinement of a system or process in a virtual environment before it is implemented in the real world.

**Cost savings:** Simulation can also help save costs by reducing the need for physical prototyping and testing, as well as identifying inefficiencies in systems and processes before they are implemented.

**Time savings:** By allowing for testing and optimization in a virtual environment, simulation can save time in the development process, enabling faster and more efficient implementation.

**Improved understanding**: Simulation can provide a deeper understanding of complex systems and processes by allowing for the visualization of data and the ability to manipulate variables and parameters.

**Ethical considerations:** In certain fields, such as healthcare or disaster response, it may be unethical or impractical to test certain scenarios in the real world. Simulation can provide a safe and controlled environment for testing and training in these situations.

In summary, simulation is necessary because it provides a safe, efficient, and cost-effective way to test and optimize complex systems and processes, reducing risk, saving time and money, and improving our understanding of the world around us.