Name: Mehul Sarda
Branch: TT AIML
Roll No. 41

MARCH REMEDIAL

Data Catalogs

A data catalog is a centralized repository that provides a comprehensive inventory of all data assets within an organization. It acts as a directory for locating, managing, and sharing data assets across different teams and departments.

Data catalogs contain metadata about datasets, such as their names, descriptions, formats, owners, and access permissions. By using a data catalog, organizations can improve data governance, reduce data redundancy, enhance data discovery, and facilitate collaboration among teams.

Taxonomies

Taxonomies are hierarchical frameworks used to classify and organize information based on predefined categories. Taxonomies are used in various fields, such as biology, library science, and information management. In the context of data management, taxonomies are used to organize data assets based on their attributes, properties, or relationships.

Taxonomies can improve data discoverability, facilitate data retrieval, and enhance data quality. For example, a taxonomy can be used to categorize different types of products in an e-commerce website.

Multiple hypothesis generator

The multiple hypothesis generator (MHG) is a data analysis technique that generates multiple hypotheses about the underlying structure or patterns in a dataset. The MHG is used in exploratory data analysis to generate hypotheses that can be tested using statistical methods or machine learning algorithms.

The MHG works by iteratively applying statistical or machine learning algorithms to a dataset and generating multiple hypotheses based on the output of these algorithms. By generating multiple hypotheses, the MHG increases the chances of identifying significant patterns in the data.

Distributed data management

Distributed data management refers to the practice of managing data across multiple computing nodes or clusters. In distributed data management, data is stored and processed across multiple machines to improve scalability, performance, and fault tolerance.

Distributed data management is commonly used in big data applications, where the volume of data exceeds the capacity of a single machine. Distributed data management requires specialized tools and technologies, such as distributed databases, distributed file systems, and distributed computing frameworks, such as Apache Hadoop

Parallelism

Parallelism refers to the technique of breaking down a task into smaller sub-tasks that can be executed simultaneously across multiple computing resources. Parallelism is used to improve the performance and efficiency of computing systems, particularly in applications that require high computational power, such as scientific simulations, machine learning, and data analytics.

Parallelism can be achieved through various techniques, such as multi-threading, multi-processing, and distributed computing. Parallelism requires careful design and optimization to avoid issues such as race conditions, deadlocks, and resource contention.

Semi-Supervised learning

Semi-supervised learning is a machine learning technique that combines labeled and unlabeled data to improve the accuracy of a model. In semi-supervised learning, a small portion of the training data is labeled, while the rest is unlabeled. The labeled data is used to train a model, while the unlabeled data is used to improve the model's generalization ability.

Semi-supervised learning is used in situations where labeling data is expensive or time-consuming, or when the labeled data is insufficient for training a high-performing model. Semi-supervised learning has applications in various fields, such as natural language processing, computer vision, and speech recognition.

Name: Aryan Singh
Branch: TT AI&ML
Roll no: 49

MARCH REMEDIAL

RL in finance	**Use cases for RL in finance**
In finance, RL can be used for a wide range of tasks, including algorithmic trading, portfolio optimization, risk management, fraud detection, and more. One example of how RL can be used in finance is for algorithmic trading. An RL agent can be trained to learn from historical price data and make decisions on buying or selling stocks based on the current market conditions. Another use case of RL in finance is for portfolio optimization. An RL agent can be trained to select the optimal mix of assets for a portfolio, based on historical data and other relevant factors such as risk tolerance and investment goals	Algorithmic Trading: RL can be used to train an agent to make buy or sell decisions based on market data. The agent learns from historical data, patterns and can take different actions and evaluate which action results in the highest reward. Risk Management: RL can be used for managing risk in financial institutions. An RL agent can be trained to monitor different risks such as market risk, credit risk, operational risk, and fraud. Pricing: RL can be used for dynamic pricing in financial markets. The agent can learn from historical data, market trends, and other relevant factors to determine the optimal price for financial products such as stocks, bonds, and options.
Advantages of CC over AI **Scalability:** Cloud computing provides the ability to scale up or down resources as needed, which can be beneficial for AI applications that require significant computing power. **Cost-Effective:** Cloud computing can be more cost-effective than building and maintaining an in-house infrastructure for AI applications, as it allows organizations to pay only for the resources they use. **Accessibility:** Cloud computing allows users to access AI applications from anywhere and at any time, as long as they have an internet connection. **Security:** Cloud providers invest heavily in security measures and compliance standards, which can help protect sensitive data and reduce the risk of security breaches.	**What is Data Modelling?** Data modeling is the process of creating a conceptual or logical representation of data and its relationships to other data elements. The goal of data modeling is to create a structured and organized view of data, which can be used to support various business processes and applications. **Types:** conceptual and logical. Conceptual data modeling is a high-level representation of data and its relationships, without taking into account the technical details of how the data will be implemented in a database or system. Logical data modeling, on the other hand, is a more detailed representation of data and its relationships, taking into account the technical aspects of how the data will be stored and processed in a database or system. It often involves the use of a data modeling notation or language, such as the Entity-Relationship (ER) model or Unified Modeling Language (UML).

NLP in Business Problems

- NLP is a set of techniques that extract meaning from text.
- Determine the meaning of a word, phrase, sentence, or document by recognizing the grammatical rules—the predictable patterns within a language.
- NLP applies the same known rules and patterns to make inferences about meaning in a text document.

~ Nidhi Worah, AI&ML 60

Techniques to solve Business Practices in Cognitive Computing

Sentiment Analysis

Analyse customer feedback and reviews.

Chatbots

Understand and respond to customer queries

Text Classification

Classify text into predefined categories

~ Nidhi Worah, AI&ML 60

Sentiment Analysis

- Sentiment Analysis can be used to analyze customer feedback and reviews to understand their sentiment towards a product or service.
- These techniques are used to analyze and extract subjective information from text data, such as customer reviews, social media posts, news articles, comments, etc.
- This helps businesses to deduce insights from the reviews and identify areas of improvement
- Further, the respective teams form strategies to satisfy the customer needs and create a brand name.

Chatbots

- Chatbots are AI programs that can simulate a human conversation.
- They can be integrated into cognitive computing systems to automate customer service and support, sales and marketing, and other business functions
- This helps businesses cater to the needs of the customers efficiently.
- Usually, chatbots are trained with a specific use in mind and the tasks they perform are pre-determined.

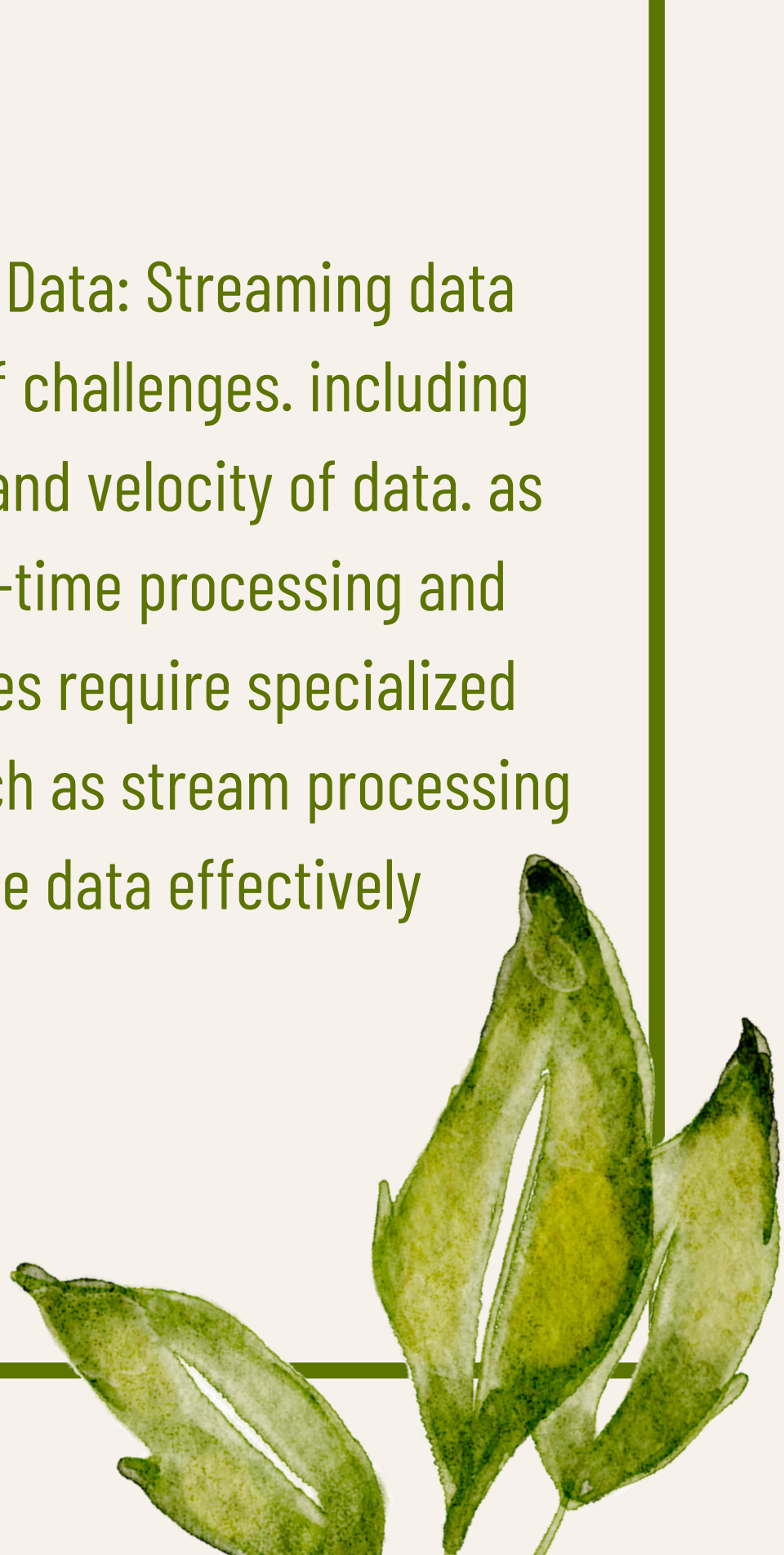
~ Nidhi Worah, AI&ML 60

Text Classification

- Text Classification is a NLP technique which is used to automatically classify text into different predefined categories.
- Ex : We can classify customer feedback into categories such as product quality, customer service, and shipping.
- This can help businesses to filter out unwanted messages and focus on important ones.
- Text classification when integrated with Sentiment Analysis helps a business understand the behaviour of the customers well , which inturn drive sales.

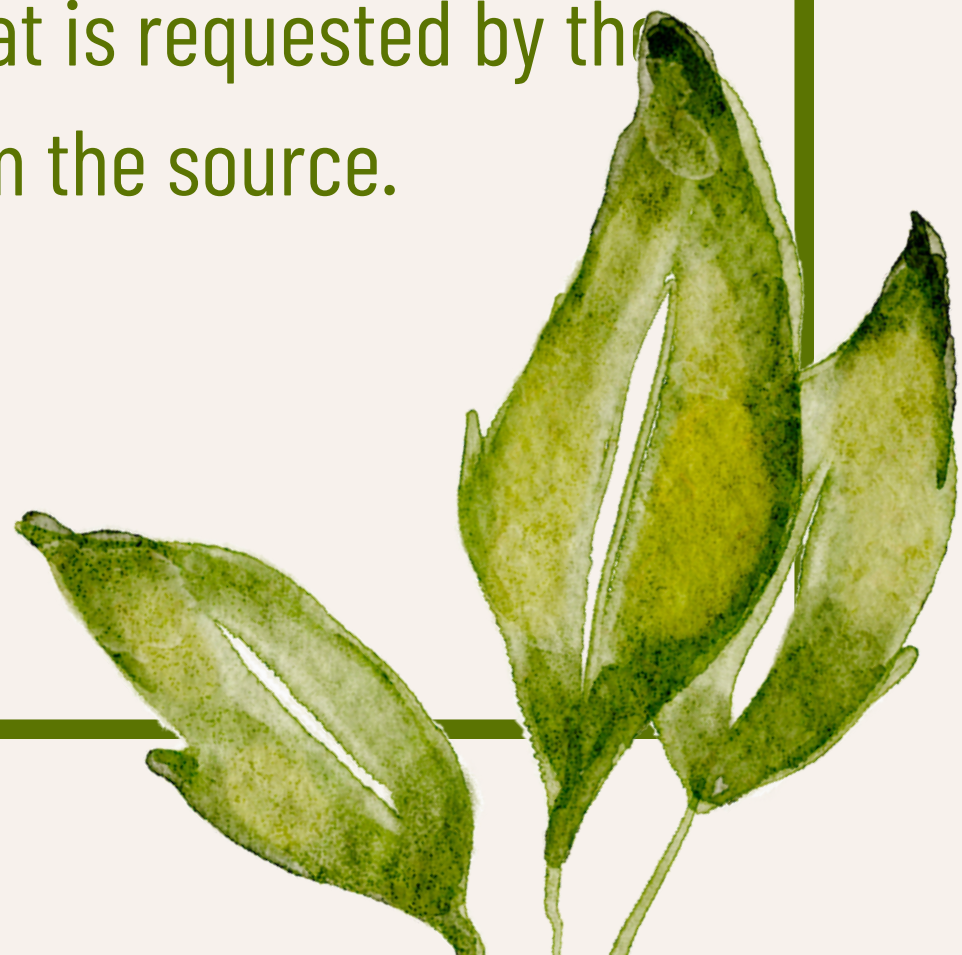
Definition of Streaming Data: Streaming Data refers to data that is continuously generated and sent in real-time from various sources. such as sensors, social media, or other streaming data sources. Streaming data is often used in real-time analytics. machine learning, and other applications that require immediate processing.

Challenges of Streaming Data: Streaming data comes with its own set of challenges. including the high volume. variety, and velocity of data. as well as the need for real-time processing and analysis. These challenges require specialized tools and technologies. such as stream processing engines. to handle the data effectively



Definition of Data in Motion: Data in Motion refers to the continuous flow of data from one point to another in real-time, often over a network. and includes any data that is transmitted, streamed or communicated in real-time.

Types of Data in Motion: There are two types of Data in Motion. which are Push-based and Pull-based data streams. Push-based data streams include data that is automatically pushed from the source to the destination. while pull-based data streams include data that is requested by the destination from the source.



Use Cases for Data in Motion: Data in Motion is used in a variety of industries, including finance, healthcare, retail, and transportation, among others. Some common use cases include fraud detection, real-time inventory management, predictiv maintenance, and personalized marketi

Benefits of Data in Motion: Data in Motion allows businesses to make real-time decisions. improve operational efficiency. and enhance customer experiences. It also enables organizations to capture and analyze data that would be impossible to store and process with traditional methods.