DA Unit-3 complete notes for One Shot video by brevilearning YT compressed

B.tech (Dr. A.P.J. Abdul Kalam Technical University)

Data Analytics

Unit - 3 [One-Shot]

Most important topics:

1. Stream Computing and Sampling in data stream.
2. Decaying window
3. RTAP applications
4. Sentiment analysis and stock predictions [Case-Study]

Stream Computing :

- Also known as stream processing.
- It involves the continuous processing of data streams in real time.
- Unlike batch processing, stream computing processes the data as soon as it arrives.
- This approach is essential for applications that require immediate insights and quick response times.

* Key characteristics of stream computing

1. Real-time Processing : Data is processed in real-time or near real-time, allowing for immediate analysis and action.
2. Scalability : Systems are designed to handle large volumes of data continuously flowing at high velocity.
3. Low Latency : Stream computing systems aim to minimize the delay between data arrival and processing.
4. Event-driven : System react to events or changes in data, triggering specific actions based on predefined rules.

* Technologies used : Apache Kafka, Flink, Storm, etc.

* Applications:

- Fraud detection
- IOT applications
- Social media
- Sentiment analytics
- Telecommunication
- Stock analysis

@brevilearning

⊛ Sampling in data stream:

- Sampling refers to the technique of selecting a subset of data from a continuous stream.
- This helps in handling and analysing large volumes of data when it is impossible to process every single data point.

Types of sampling:

1. Random sampling: Each element in stream gets equal chance of being selected.
2. Systematic sampling: Selecting every n^{th} element from the data stream.
3. Stratified sampling: Dividing stream into strata (groups) and sampling from each stratum.
4. Cluster sampling: Dividing stream into clusters, then randomly selecting entire cluster for sampling.
5. Weighted sampling: Assigning weights to elements, giving some elements a higher chance of being selected.
6. Time-based sampling: Sampling data at fixed time intervals.
7. Event-based sampling: Sampling data based on specific events or triggers.

Like, Share, and Subscribe

Decaying Window Sampling:

- It is a method that gives more importance to recent data while gradually decreasing the significance of older data.
- This approach is particularly useful in senarios where the most recent information is more relevant for making real-time decisions.

Working:

1. Data arrival: As new data points arrive in the stream, they are initially given a full weight of 1.
2. Apply decay: Weights decreases over time using an exponential decay function:

$$\omega(t) = e^{-\lambda t}.$$

3. Update weights: Continuously update weights as time progresses.
4. Weighted Aggregation: Use weights in calculations to ensure recent data has more influence.
5. Data Expiry (Optional): Discards data points with weights below a threshold.

* Applications:

- Real-time analytics
- Financial markets
- Network security
- Sensor data

* Advantages:

- Relevance: Focus on most relevant data (recent)
- Efficiency: Reduces computational burden by down-weighting older data.
- Adaptability: Quickly adapt with changes in data patterns.

RTAP applications:

- Real-time analytics processing involves the immediate analysis of data as it is generated to provide instant insights and support rapid decision-making.

Types of RTAP:

i) Descriptive Analytics: Provides a real-time view of what is happening

ii) Diagnostic Analytics: Identifies why something happened in real-time.

iii) Predictive Analytics: Uses real-time data to forcast future events.

iv) Prescriptive Analytics: Suggests real-time actions based on data insights.

Advantages of RTAP applications:

- Immediate insights: Quick access to actionable data.
- Enhanced decision-making: Faster and more informed decisions.
- Operational efficiency: Steamlined operations and reduced downtime.
- Improved Customer Experience: Personalized and timely interactions.
- Competitive advantages: Staying ahead of market trends and competitors.

Disadvantages of RTAP applications:

- High cost
- More complex
- Security risk due to continuous data flow.

★ Applications of RTAP application:

1. Fraud detection
2. Risk management
3. Customer analysis
4. Sentiment analysis
5. Personalized recommendation
6. Smart traffic management.
7. Smart home devices
8. Marketing
9. Social media
10. IoTs

Case study : Sentiment analysis

♦ Objective: To understand customer's sentiment towards products and services to improve customer satisfaction and brand reputation.

♦ Approach:

1. Data Collection:

- Collect real-time customer reviews from the companies website and product feedbacks.
- Monitor social media platforms for mentions of the companies products and services.

2. Real - Time Processing:

- Use sentiment analysis tools to process the incoming data.
- Apply NLP algorithms to analyze the text and detect sentiments (+ve), (-ve), neutral).

3. Sentiment scoring:

- Assigns sentiment score on each reviews or mentions.
- Categorize these feedbacks based on the sentiment scores to identify overall trends.

4. Actionable insights:

- Identify common themes in negative feedbacks to address product issues or improve services.
- Highlight positive feedbacks in

marketing campaigns to attract more customers.

- Outcomes:

1. Improved Customer Satisfaction:

- Quickly addressed issues raised in negative feedback, leading to better customer service.
- Improve products based on customer's suggestions and complaints.

2. Enhanced marketting Strategies:

- Used positive feedbacks to promote products, improving sales and customer's trust.

⊛ Case Study: Stock predictions

- Objective: To predict stock price movements to make informed investments, and maximize returns.

- Approach:

1. Data collection:

- Grather historical stock price data, trading volumes and market indicators.
- Monitor real-time financial news, social media sentiments, and economic reports.

2. Real-time Processing:

- Uses machine learning algorithms to analyze historical data and identify patterns.
- Utilizes real-time data to continuously update predictions.

3. Prediction models:

- Develop predictive models using techniques such as time series analysis, and neural networks.
- Validate the model using historical data to ensure accuracy.

4. Actionable insights:

- Provides real-time stock prediction to traders.
- Highlight potential investment opportunities and risks based on predicted trends.

• Outcomes:

1. Informed investment decisions:

- Reduced investment risks by identifying potential market downturns early.
- Traders made better investments using accurate, real-time predictions.

2. Increased returns:

- Improved trading strategies led to higher returns.

Thanks for Watching !!