Matillion AI Certification

1. **Which of the following is a common use case for data classification?**

   A. Detecting spam emails

   B. Translating text

   C. Generating automated responses

   D. Summarizing documents

2. What is Retrieval-Augmented Generation (RAG) used for in AI?

A. Data cleaning and preprocessing

B. Hyperparameter tuning

C. Model evaluation

D. Augmenting generative models with external knowledge retrieval

3. What are some emerging AI techniques that could further revolutionize data integration and analytics in the near future?

A. Advanced natural language processing and real-time predictive analytics

B. Automated data archiving and lineage prediction

C. Enhanced spreadsheet functionalities that match data against current market trends

D. Post-processing legacy data management systems to act like modern systems

4. What is the primary benefit of generating automated customer responses?

A. Reducing the need for data visualization

B. Improving customer engagement and satisfaction

C. Enhancing data classification accuracy

D. Simplifying sentiment analysis

5. When generating customer responses, why is it important to analyze the sentiment of the original feedback?

A. To ensure the response is translated correctly

B. To create a structured data format

C. To tailor the response appropriately based on the customer's feelings

D. To classify the feedback into categories

6. What is the primary purpose of data labeling in machine learning?

A. To visualize data

B. To create training data for supervised learning models

C. To analyze sentiment

D. To generate responses

7. In what ways can machine learning algorithms improve predictive analytics within a business data pipeline?

A. By forecasting future trends based on historical data

B. By generating new data that follows trends in the existing data

C. By reducing the amount of historical data required

D. By creating static reports

8. Why is structured data creation important for data visualization?

A. It helps in translating text accurately

B. It aids in sentiment analysis

C. It generates automated responses

D. It helps create the dimensions required for analytics & data visualization.

9. In Matillion , When selecting the "Outputs" property in JSON format, what must be defined?

A. The temperature and top P values

B. The database and schema

C. The key pairs for the JSON output and their context

D. The URL endpoint and API key

10. What is the "Top P" setting used for in the Amazon Bedrock Prompt component in Matillion ?

A. To define the maximum number of tokens to process

B. To specify the primary key column

C. To use nucleus sampling for generating responses

D. To limit the number of rows from the source table

11. Which of the following is NOT a prompt component available in the Data Productivity Cloud in Matillion ?

A. OpenAI Prompt

B. Azure OpenAI Prompt

C. Amazon Bedrock Prompt

D. Google Cloud Prompt

12. In Matillion , The 'Max Tokens' property in prompt component controls the maximum number of rows that can be used with that component in a single pipeline run.

A. True

B. False

13. Which output formats are supported by the Amazon Bedrock Prompt component?

A. XML and CSV

B. TEXT and JSON

C. HTML and PDF

D. DOCX and XLSX

14. TechCo's customer service team receives numerous emails daily, with queries ranging from technical support to billing issues. To streamline responses and improve efficiency, TechCo wants to automate the process of generating personalized responses to these emails. They plan to use the Azure OpenAI Prompt component to process the email content and generate suitable responses based on predefined prompts.

Implementation:

1. Input Data: The source table in their Snowflake data warehouse contains the following columns: Email_ID, Customer_Email, Email_Subject, Email_Body.

2. Prompt Configuration: The team defines a prompt that instructs the model to read the email subject and body, then generate a response addressing the customer's query.

3. Output: The response is stored as a JSON object in a destination table in Snowflake, along with metadata such as the number of tokens used and any error messages. Which input columns should be selected to feed into the Azure OpenAI Prompt component for generating email responses?

A. Email_Subject and Email_Body

B. Email_ID and Customer_Email

C. Customer_Email and Email_Subject

D. Email_ID and Email_Body

15. A company wants to analyze customer reviews to understand the sentiment behind each review. They have a table of customer reviews stored in Snowflake and want to use the Amazon Bedrock Prompt component to classify each review as positive, negative, or neutral. For storing the results of the sentiment analysis, which property should be configured in the Amazon Bedrock Prompt component?

A. Output Format

B. Embedding Provider

C. Data Path

D. Model Name

Matillion AI Certification

16. A company wants to gather insights on potential product feature enhancements based on user feedback. They have a table of feedback entries and wish to use the Amazon Bedrock Prompt component to summarize key enhancement suggestions. Which property should be configured to summarize user feedback into enhancement suggestions?

A. User Context

B. Embedding Model

C. Search Column

D. Pinecone Namespace

17. A retail company wants to automate the process of summarizing product reviews to understand customer feedback more efficiently. 1. Configure the Component: Name: Model: gpt-3.5-turbo API Key: Use the corresponding secret definition for the OpenAI API key. Temperature: Max Tokens: Database: Snowflake Table: ProductReviews Key Column: ReviewID User Context: "Summarize the following product reviews and highlight common positive and negative feedback." Output Format: JSON Outputs: Output: summary Context: "Provide a brief summary of the product review." Output: sentiment Context: "Indicate the overall sentiment of the review (positive, negative, neutral)." 2. Execute the component Sample output: [{ "ReviewID": "1", "summary": "The product quality is excellent, but the delivery was delayed.", "sentiment": "positive" }, { "ReviewID": "2", "summary": "Very poor build quality, not worth the price.", "sentiment": "negative" }, { "ReviewID": "3", "summary": "Satisfied with the purchase, but customer service could be better.", "sentiment": "neutral" }] Which property should be configured to limit the maximum length of the summarization?

A. Temperature

B. Top P

C. Max Tokens

D. N

Matillion AI Certification

18. You work for a software company that receives numerous customer support queries daily. To improve efficiency and accuracy, you decide to automate the response process using the OpenAI Prompt component with Retrieval-Augmented Generation (RAG). Steps: Configure OpenAI Prompt Component: Enable RAG: Run the Pipeline: Analyze Results: Sample Input and Output Data Input Data: [{ "QueryID": "101", "customer_query": "How can I reset my password?", "retrieved_documents": ["To reset your password, go to the login page and click on 'Forgot Password'.", "Enter your registered email address and follow the instructions sent to your email."] }, { "QueryID": "102", "customer_query": "Why is my account locked?", "retrieved_documents": ["Your account can be locked due to multiple failed login attempts.", "Contact customer support to unlock your account."] }, { "QueryID": "103", "customer_query": "How do I update my payment information?", "retrieved_documents": ["To update your payment information, go to your account settings.", "Navigate to the payment section and enter your new payment details."] }] Output Data: [{ "QueryID": "101", "response": "To reset your password, go to the login page and click on 'Forgot Password'. Enter your registered email address and follow the instructions sent to your email." }, { "QueryID": "102", "response": "Your account may be locked due to multiple failed login attempts. Please contact customer support to unlock your account." }, { "QueryID": "103", "response": "To update your payment information, go to your account settings, navigate to the payment section, and enter your new payment details." }] Based on the sample input, which retrieved document snippet would be used to respond to a query about resetting a password?

A. "To reset your password, go to the login page and click on 'Forgot Password'."

B. "To update your payment information, go to your account settings."

C. "Contact customer support to unlock your account."

D. "Your account can be locked due to multiple failed login attempts."

19. What is the purpose of the Pinecone Vector Upsert component?

A. To create data visualizations in Pinecone

B. To convert data into embeddings and store them as vectors in Pinecone

C. To perform real-time data analytics retrieved from the Pinecone API

D. To encrypt data before storage

20. What type of input text does the Cortex Sentiment component analyze?

A. English-language text

B. Any language text

C. Numeric data

D. Code snippets

21. When using the Cortex Summarize component, how does including both source input columns and summary columns help in data analysis?

A. It provides a backup of the original data.

B. It increases the data volume for analysis.

C. It provides context for the summaries.

D. It anonymizes the data

22. How can Large Language Models (LLMs) be utilized for sentiment analysis within a data pipeline?

A. By categorizing text data into predefined sentiment categories

B. By converting text data into numerical values

C. By storing data in a structured format

D. By creating visual dashboards

23. What is the range of the sentiment score returned by the Cortex Sentiment component?

A. -2 to 2

B. -1 to 1

C. 0 to 1

D. 0 to 100

24. What happens if the input text is too short to generate a meaningful summary?

A. The text is deleted.

B. The text is translated.

C. The original text is output.

D. An error message is displayed.

25. What type of input data is NOT suitable for the Cortex Summarize component?

A. Customer reviews

B. Meeting transcripts

C. Research papers

D. Binary files

26. Company wants to monitor its brand reputation by analyzing social media posts mentioning its brand. They decide to use the Cortex Sentiment component to analyze the sentiment of these posts. By generating sentiment scores, they can quickly identify public opinion trends and respond appropriately to maintain and improve their brand image. How can the company use sentiment scores to respond to negative feedback on social media?

A. By ignoring the negative posts.

B. By finding and addressing trends common in negative posts.

C. By deleting negative posts.

D. By increasing the number of posts.

27. An e-commerce company collects product reviews from customers on their website. To understand customer satisfaction and improve product offerings, the company uses the Cortex Sentiment component to analyze the sentiment of these reviews. By generating sentiment scores, the company can identify top-rated products and those needing improvement. Including input columns in the output is mandatory when using the Cortex Sentiment component.

A. True

B. False

28. A large corporation records all of its meetings and transcribes them for future reference. To help employees quickly understand the key points discussed in these meetings, the corporation uses the Cortex Summarize component to generate summaries of the transcripts. This allows employees to stay informed without needing to read through lengthy transcripts. How can employees benefit from the summaries generated by the Cortex Summarize component?

A. By ignoring the detailed transcripts.

B. By quickly understanding the key points discussed in meetings.

C. By reducing the number of meetings.

D. By translating the summaries into different languages.

Matillion AI Certification

29. A multinational company needs to translate internal policy documents from English to multiple languages for their non-English-speaking employees. Using the Cortex Translate component, these documents are efficiently translated into the required languages. Which Snowflake component could be used in conjunction with Cortex Translate to capture the essence of each document?

A. Cortex Summarize

B. Cortex Completions

C. Cortex Sentiment

D. Cortex Translate

30. In Matillion , The Data Productivity Cloud's designer offers Copilot to help build pipelines. How should a user indicate to Copilot which component on the canvas to build from?

A. Copilot does not build from other components in a pipeline. It only creates complete pipelines from a prompt.

B. Specify the name of the component you wish to build from in the prompt.

C. Select the component on the canvas before entering your prompt.

D. Copilot will intuit the best component to build from to fulfill the prompt.

31. Data warehouse credits are consumed as part of the AI note generation process.

A. True

B. False

32. What is the primary purpose of AI Notes in Matillion's Data Productivity Cloud?

A. To annotate data pipeline components using generative AI

B. To perform sentiment analysis on data

C. To add notes into a column of a table

D. To summarize data for reports