Data Science Interview Questions

There is/are 13 Question(s)

Once you click complete, you cannot come back

174897173365895

Question:

How do you determine how many hidden layers to go in your neural network? CNN?

It depends on the number of inputs, the dataset features, size of the image

Question:

How many epochs do you choose to run the algo over your dataset? How do you determine this?

It is usually from 10-100 epochs. It depends on the size of the dataset, if it is a large dataset, one can use early stopping, to bring about better results if met in an epoch

Question:

Explain how you would make a recommendation engine to classify movies by users preferences? Assume you are given the dataset

For classification tasks, using given dataset, the dataset is divided into train and test sets, any classification module can be used like Scikit-learn classification models, or XGBClassifier.
The dataset may contain user data, data containing the movie(like genre and so) and user ratings. So I will use the necessary user data, and movie data, basically rating will be the dependent variable, then the movies with high ratings or genre will be recommended for the user

Question:

Given the following columns, write down the process you would clean/extract/model data to give me an actionable insight? Give me an example recommendat

Order Table
|order_id|customer_email|customer_name|gender|created_at|amount|num_item|tax|total|source|ip|return_customer|order_status

Order Product
|order_item_id|order_id|product_id|product_name|price|quantity|status

First, I can give insights on each table individually, but for more productivity, for it to make sense, merging both tables would be best, I would join the tables on 'order_id' on both tables
For cleaning,
1. I will check for duplicates,
2. check for missing or null values, drop the rows that values can't be imputed for,
3. I would check the datatypes for columns and ensure they are coorect and change the incorrect ones
Further, I will extract features like
1. user details,
2. No. of purchases
3. product features like total quantity sold, price
4. Order features like order size, frequency of order (daytime, evening deeper insight, to keep it basic, normal temporal features)

I can further analyze by
1. predicting purchases
2. create recommendation engine to recommend similar products for users
3. I dentify high value consumers of each product
4. I dentify what months or seasons people buy differnt products at
5. Identify when sales or bonuses can be placed based on temporal trends

Question:

What is a staging table in data science? From the raw data, how do you build out your reporting tables in your data warehouse?

A staging table is a table where transformation or where the data manipulations are performed before applying to the actual data. It is a temporary storage place before moving to reporting tables

Question:

How do you decide how to partition your table? Give us 2 real life scenarios?

When partitioning tables, I consider the size of the data and the strategy for partitioning
When I worked on the AirBnB Market analysis, I analyzed listings and revenue trends over time using tableau, because of the growth of the data, over a long span of time, I chose to partition the data by year, using the date column.
In the bike sales analysis I worked on in my internship, I partition by age ranges, firstly, categorizing the data into Old, Middle Age, Young

Question:

Given the following tables, write the SQL statement to tell me the top 10 items purchased by females between the ages of 20 - 30.
Order Table
|order_id|customer_email|customer_name|gender|created_at|amount|num_item|tax|total|source|ip|return_customer|order_status

Order Product

|order_item_id|order_id|product_id|product_name|price|quantity|status

Product
product_id|product_name|price|quantity|status

Given these tables, an intricate column necessary for this result is missing, which is the age column/field.
Assuming there was a Customer table which contained their demographics, I could join the Order Table with the Customer Table on a similar field customer email.
I would create a CTE first to get the purhcases of females between the ages of 20-30, then I would rank using RANK() on the total quantity genetared, then use LIMIT 10, to get the top ten

```
WITH f_purhases AS (
  SELECT
    op.product_id,
    op.product_name,
    SUM(op.quantity) AS total_quantity
   FROM "Order"
```

Question:

Given the following CSV data, write the python code to clean the data into a format you would put through your modelling?

id,title,description,tags,country,status
1, Need help setting up instant switching between two web hosting companies ,

Question:

Suppose you are given the following task. How would you do it? Explain in detail.

Give a video with audio in English. You need to transcribe the words and display the words below the audio. if a user clicks on a specific word and cha

For example, A fox jumped over the six fences in the farm.
Say we changed six to seven.

So: A fox jumped over the seven fences in the farm.

Question:

Open this link https://imgur.com/a/xbJ8jpv. Read the instructions below.

Pretend this form shows the content describing a product's general features. We want you to generate text about this product without seed data, only th

Question:

Given the following scenario
You are given a sound clip of someone talking outside with construction.
Your task is to split the audio into background noise and the talking voice.
Then you need to translate the talking voice into Spanish and recombine the audio.
How do you solve this problem? Do your best of your knowledge

There are libraries for audio preprocessing, like we have librosa in Python.
I think the audio can be split

Question:

How does caption generation work?
Look at this image
https://tensorflow.org/images/surf.jpg
The caption generated is "a surfer riding on a wave"

Using CNNs, the image is being processed, extract features, like shapes, lines, colors, etc
The machine learning model would analyze the image with this feature and using NLP, it would generate the best description fitting the image

Question:

How would you deploy your model on the server? When do you decide to use a single CPU vs GPU?
What size server would you decide to use?

The server size would depend on the or workload of the service, the amount of request that would be recieved, the size of database.
The model should be optimzed, converted to a compatible format
dependencies should be installed, to work with the server, the choice of server is very important
I would make sure I choose a hosting service or cloud provider that can provide the necessary storage and processing requirements I need

COMPLETE

174897173365895