Scenario based questions

(Data science) Part 4

Mr. Shivan Kumar

Data Scientist & Kaggle Master

- **Qus 1- When would you use random forests Vs SVM and why?**

- **Ans:-** There are a couple of reasons why a random forest is a better choice of the model than a support vector machine:
- Random forests allow you to determine the feature importance. SVM's can't do this.
- Random forests are much quicker and simpler to build than an SVM.
- For multi-class classification problems, SVMs require a one-vs-rest method, which is less scalable and more memory intensive.

Swipe >>>

- **Qus 2- A real estate company wants to predict the sale price of a property. What data sources and machine-learning techniques would be useful for this task?**

- **Ans:-** To predict the sale price of a property, the real estate company can use property data, such as square footage, location, and the number of bedrooms, along with real estate market data, such as housing prices and sales history, to build a predictive model. Feature engineering techniques can be used to extract relevant information from these data sources.

Swipe >>>

- Next, a predictive model, such as regression analysis or a decision tree, can be trained on the data to make the prediction. An evaluation metric, such as mean absolute error or R2 score, can be used to assess the performance of the model.

Swipe >>>

- **Qus 3- How Do You Design an Email Spam Filter?**

- **Ans:-** Building a spam filter involves the following process:

- The email spam filter will be fed with thousands of emails.
- Each of these emails already has a label: ‘spam’ or ‘not spam.’
- The supervised machine learning algorithm will then determine which type of emails are being marked as spam based on spam words like the lottery, free offer, no money, full refund, etc.

Swipe >>>

- The next time an email is about to hit your inbox, the spam filter will use statistical analysis and algorithms like Decision Trees and SVM to determine how likely the email is spam.
- If the likelihood is high, it will label it as spam, and the email won't hit your inbox.
- Based on the accuracy of each model, we will use the algorithm with the highest accuracy after testing all the models.

Swipe >>>

DID YOU FIND THIS POST HELPFUL ?

Follow us on

@engineer_bhaiya_yt

@engineeringwalabhaiya

topmate.io/shivan_kumar1

For Mentorship & Free Text
Queries,
WhatsApp at: +91 6280 963 446

Like, Save, and Share with your friends !!!

Shivan Kumar

Data Scientist & Kaggle Master