

🔗 Google Discussions



Exam Professional Data Engineer All Questions

View all questions & answers for the Professional Data Engineer exam

[Go to Exam](#)

📄 EXAM PROFESSIONAL DATA ENGINEER TOPIC 1 QUESTION 18 DISCUSSION

Actual exam question from Google's Professional Data Engineer

Question #: 18

Topic #: 1

[\[All Professional Data Engineer Questions\]](#)

Business owners at your company have given you a database of bank transactions. Each row contains the user ID, transaction type, transaction location, and transaction amount. They ask you to investigate what type of machine learning can be applied to the data. Which three machine learning applications can you use? (Choose three.)

- A. Supervised learning to determine which transactions are most likely to be fraudulent.
- B. Unsupervised learning to determine which transactions are most likely to be fraudulent.
- C. Clustering to divide the transactions into N categories based on feature similarity.
- D. Supervised learning to predict the location of a transaction.
- E. Reinforcement learning to predict the location of a transaction.
- F. Unsupervised learning to predict the location of a transaction.

[Show Suggested Answer](#)

by [jvg637](#) at March 15, 2020, 12:32 p.m.

Comments

Type your comment...

[Submit](#)

🗄️ 👤 **jvg637** Highly Voted 👍 5 years, 1 month ago

BCD makes more sense to me. Its for sure not unsupervised, since locations are in the data already. Reinforcement also doesn't fit, as there no AI and no interactions with data from the observer.

👍 ↩️ 🚩 upvoted 70 times

🗄️ 👤 **sergio6** 3 years, 8 months ago

D make sense, but i have a doubt: location is a discrete value (no regression), so a multiclass classification model should be applied ... to predict locations?

👍 ↩️ 🚩 upvoted 4 times

🗄️ 👤 **hellofrnds** 3 years, 6 months ago

yes. multiclass classification model should be applied

👍 ↩️ 🚩 upvoted 5 times

🗄️ 👤 **[Removed]** Highly Voted 👍 5 years, 1 month ago

BCD looks more appropriate

👍 ↩️ 🚩 upvoted 8 times

🗄️ 👤 **Parandhaman_Margan** Most Recent 🕒 1 month, 3 weeks ago

Selected Answer: ABC

The three most applicable machine learning applications for analyzing bank transactions are

👍 ↩️ 🚩 upvoted 1 times

🗄️ 👤 **Bulleen** 7 months, 1 week ago

BCD makes sense, but I now agree that BCE is the correct answer.

Say the model predict a location, guessing US or Sweden are both wrong when the answer is Canada. But US is closer, the distance from the correct location can be used to calculate a reward. Through reinforcement learning (E) the model could guess a location with better accuracy than supervised (D).

👍 ↩️ 🚩 upvoted 7 times

🗄️ 👤 **anji007** 7 months, 1 week ago

Ans: B, C and D

i) Fraudulent transaction, is nothing but anomaly detection which falls under Unsupervised.

ii) All transactions can be categorized using type etc - clustering algorithm.

iii) Using location as a label, supervised classification can be developed to predict location.

👍 ↩️ 🚩 upvoted 2 times

🗄️ 👤 **StefanoG** 7 months, 1 week ago

Selected Answer: BCD

As wrote by RP123

B - Not labelled as Fraud or not. So Unsupervised.

C - Clustering can be done based on location, amount etc.

D - Location is already given. So labelled. Hence supervised.

👍 ↩️ 🚩 upvoted 7 times

🗄️ 👤 **ler_mp** 7 months, 1 week ago

Selected Answer: BCD

BCD makes more sense. B, C should not be controversial. For D vs E, in this use case D fits better than usage of reinforcement learning

👍 ↩️ 🚩 upvoted 1 times

🗄️ 👤 **Kyr0** 7 months, 1 week ago

Selected Answer: BCD

BCD makes more sens to me

👍 ↩️ 🚩 upvoted 1 times

🗄️ 👤 **musumusu** 7 months, 1 week ago

Anwer: BCD

Things to understand:

Supervised learning will only predict the column that is labeled. In this case, there is not Fraud or not Fraud column inside which he will train on. So Option A, wrong.

option D: Supervised learning for column (transaction location) is possible as column exist to train on.

Option C: Custerung N-type is possible and also an unsupervised learning to make cluster of similar pattern.

Option B: Its a weaker point here, User should be able to know which clusters are fraud in history. As it doesn't give enough information about past analysis whether user knows potential frauds or not. Ignore this option, if question asked for 2 right options only.

👍 ↩️ 🚩 upvoted 5 times

🗄️ 👤 **iooj** 9 months, 1 week ago

Selected Answer: ABC

Why would you need to predict a location...

   upvoted 2 times

  **Roulle** 10 months ago

Selected Answer: ACD

C and D are good for sure Cliquez pour utiliser cette solution et E, F wrong for sure.

Then, to choose between A and B. Both options indicate that we know which transactions are fraudulent and which are not. Indeed, in order to use unsupervised classification to determine the characteristics of fraudulent transactions, we must already know which ones are fraudulent, either because all transactions in the dataset are fraudulent, or because a variable allows us to identify them. If all transactions were fraudulent, this would probably have been specified in the statement. It is therefore more likely that the "type of transaction" variable can be used to distinguish fraudulent transactions from others.

In this case, we have a target variable to predict, enabling us to build interpretable supervised models to understand the typology of fraudulent transactions. I therefore opt for A, C and D

   upvoted 1 times

  **TVH_Data_Engineer** 1 year, 5 months ago

Options B, E, and F are not as suitable for the given scenario:

B. Unsupervised learning to determine which transactions are most likely to be fraudulent.

Unsupervised learning, while useful for anomaly detection, might not be as effective for fraud detection without labeled data indicating which transactions are fraudulent.

E. Reinforcement learning to predict the location of a transaction.

Reinforcement learning is more suitable for scenarios where an agent learns to make decisions through trial and error, which doesn't seem to align with predicting transaction locations.

F. Unsupervised learning to predict the location of a transaction.

Unsupervised learning typically doesn't involve predicting specific values (like location) without labeled data for training. In summary, A, C, and D are the most appropriate machine learning applications for investigating the provided bank transactions dataset.

   upvoted 3 times

  **rocky48** 1 year, 6 months ago

Selected Answer: BCD

Answer: BCD

   upvoted 1 times

  **Waqasghaloo** 1 year, 7 months ago

Location is already given as attribute so what value is served with predicting location?

   upvoted 2 times

  **youare87** 1 year, 8 months ago

A, B: Data features without the definition of fraudulent, so we can not obtain the answer even if using the unsupervised learning.



C: Kmeans solve this.

D: logistic regression. Just put the location into target.

E: Give the positive reward when the model predicts correct location.

F: Same as C. Use all features but locations, and use similarity to predict new data.



   upvoted 1 times

  **xiaofeng_0226** 1 year, 9 months ago

Selected Answer: BCD

Absolutely

   upvoted 1 times

  **Dip1994** 1 year, 9 months ago

Selected Answer: BCD

makes more sense

   upvoted 1 times

[Load full discussion...](#)



Platform

> [Home](#)

> [Examtopics PRO](#)

> [All Exams](#)

> [Training Courses](#)

