



- Expert Verified, Online, Free.

MENU



Google Discussions



Exam Professional Machine Learning Engineer All Questions

View all questions & answers for the Professional Machine Learning Engineer exam

Go to Exam

EXAM PROFESSIONAL MACHINE LEARNING ENGINEER TOPIC 1 QUESTION 184 DISCUSSI...

Actual exam question from Google's Professional Machine Learning Engineer

Question #: 184

Topic #: 1

[\[All Professional Machine Learning Engineer Questions\]](#)

You work for a retail company. You have a managed tabular dataset in Vertex AI that contains sales data from three different stores. The dataset includes several features, such as store name and sale timestamp. You want to use the data to train a model that makes sales predictions for a new store that will open soon. You need to split the data between the training, validation, and test sets. What approach should you use to split the data?

- A. Use Vertex AI manual split, using the store name feature to assign one store for each set
- B. Use Vertex AI default data split
- C. Use Vertex AI chronological split, and specify the sales timestamp feature as the time variable
- D. Use Vertex AI random split, assigning 70% of the rows to the training set, 10% to the validation set, and 20% to the test set



Show Suggested Answer

by [vale_76_na_xxx](#) at Jan. 8, 2024, 8:19 p.m.

Comments

Type your comment...

Submit

  **fitri001** 6 months, 2 weeks ago



Selected Answer: C

Time-Series Data: Your sales data has timestamps, indicating it's time-series data. A chronological split considers the order of the timestamps, ensuring the model is trained on historical trends.

Predicting for New Store: Since you want to predict sales for a new store, a chronological split is better than a random split (option D) which wouldn't prioritize recent trends.

Vertex AI Functionality: Vertex AI's chronological split functionality is specifically designed for time-series data and leverages the timestamp feature you provide to separate data for training, validation, and testing.

   upvoted 4 times

  **fitri001** 6 months, 2 weeks ago

A. Manual Split by Store: While this might work, it doesn't consider the time element crucial for sales predictions. The new store's performance might not be well-represented by data from a single existing store.

B. Default Split (Random): The default random split in Vertex AI might not prioritize recent data which could be more relevant for predicting sales in the new store.

D. Random Split with Specific Ratios: Similar to the default split, a random approach might not capture the time-series aspect and recent trends that are important for your new store predictions.

   upvoted 1 times

  **guilhermebutzke** 8 months, 2 weeks ago

Selected Answer: C

My answer C:

A: Not Correct: Splitting based on store name wouldn't guarantee temporal separation of data. Furthermore, for this problem is note to assign one store for each set, because the target is for a new store.

B: Not Correct: Randomly choosing data points across different time periods could lead to the model not capturing seasonal trends or temporal patterns effectively.

C: CORRECT: it leverages the chronological nature of the data. Since the dataset contains sales data over time from different stores, using a chronological split ensures that the model is trained on data from earlier time periods and validated/tested on more recent data.

D: Not Correct: Similar to B, a custom random split wouldn't ensure temporal separation and could lead to issues with capturing temporal trends.

   upvoted 2 times

  **shadz10** 9 months, 3 weeks ago

Selected Answer: C

I agree with b1a8fae

   upvoted 1 times

  **BlehMaks** 9 months, 3 weeks ago

Selected Answer: C

<https://cloud.google.com/automl-tables/docs/data-best-practices#time>

   upvoted 1 times

  **b1a8fae** 9 months, 4 weeks ago

Selected Answer: C

Anything different than option C could potentially lead to data leakage imo.

   upvoted 1 times

  **pikachu007** 9 months, 4 weeks ago

Selected Answer: A

By using a manual split based on store names, you can train a model that is more sensitive to the unique characteristics of each store, ultimately leading to better predictions for the new store.

   upvoted 1 times

  **vale_76_na_xxx** 10 months ago

I say C , time-based splitting is always suggest

   upvoted 1 times

[Start Learning for free](#)



Social Media

[Facebook](#) , [Twitter](#)

[YouTube](#) , [Reddit](#)

[Pinterest](#)



We are the biggest and most updated IT certification exam material website.

Using our own resources, we strive to strengthen the IT professionals community for free.



© 2024 ExamTopics

ExamTopics doesn't offer Real Microsoft Exam Questions. ExamTopics doesn't offer Real Amazon Exam Questions. ExamTopics Materials do not contain actual questions and answers from Cisco's Certification Exams.

CFA Institute does not endorse, promote or warrant the accuracy or quality of ExamTopics. CFA® and Chartered Financial Analyst® are registered trademarks owned by CFA Institute.