

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 127 DISCUSSI...

Actual exam question from Google's Professional Machine Learning Engineer

Question #: 127

Topic #: 1

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

While performing exploratory data analysis on a dataset, you find that an important categorical feature has 5% null values. You want to minimize the bias that could result from the missing values. How should you handle the missing values?

- A. Remove the rows with missing values, and upsample your dataset by 5%.
- B. Replace the missing values with the feature's mean.
- C. Replace the missing values with a placeholder category indicating a missing value.
- D. Move the rows with missing values to your validation dataset.

Show Suggested Answer

by hiromi at Dec. 22, 2022, 12:20 a.m.

### Comments

Type your comment...

Submit

fitri001 6 months, 2 weeks ago

Selected Answer: C

Minimizes Bias: Removing rows (A) with missing data can introduce bias if the missingness is not random.expand\_more  
Upsampling the remaining data (A) might not address the underlying cause of missing values.  
Unsuitable for Categorical Features: Replacing with the mean (B) only works for numerical features.  
Transparency and Model Interpretation: A placeholder category (C) explicitly acknowledges the missing data and avoids introducing assumptions during model training. It also improves model interpretability.  
Validation Set Contamination (D): Moving rows with missing values to the validation set (D) contaminates the validation data and hinders its ability to assess model performance on unseen data.  
Using a placeholder category creates a separate category for missing values, allowing the model to handle them explicitly. This approach is particularly suitable for categorical features with a relatively small percentage of missing values (like 5% in this case).

👍 ↩ 🚩 upvoted 4 times

🗄️ 👤 **pinimichele01** 6 months, 1 week ago  
if B nominate mode instead of mean?

👍 ↩ 🚩 upvoted 1 times

🗄️ 👤 **M25** 1 year, 6 months ago

**Selected Answer: C**

<http://webcache.googleusercontent.com/search?q=cache:FzNjYfqNEZ0J:https://towardsdatascience.com/missing-values-dont-drop-them-f01b1d8ff557&hl=de&gl=de&strip=1&vwsrsc=0>

See also #62, #123

👍 ↩ 🚩 upvoted 1 times

🗄️ 👤 **M25** 1 year, 6 months ago

Also, tab "Forecasting":

"For forecasting models, null values are imputed from the surrounding data. (There is no option to leave a null value as null.) If you would prefer to control the way null values are imputed, you can impute them explicitly. The best values to use might depend on your data and your business problem.

Missing rows (for example, no row for a specific date, with a data granularity of daily) are allowed, but Vertex AI does not impute values for the missing data. Because missing rows can decrease model quality, you should avoid missing rows where possible. For example, if a row is missing because sales quantity for that day was zero, add a row for that day and explicitly set sales data to 0."

<https://cloud.google.com/vertex-ai/docs/datasets/data-types-tabular#null-values>

👍 ↩ 🚩 upvoted 1 times

🗄️ 👤 **TNT87** 1 year, 8 months ago

**Selected Answer: C**

C. Replace the missing values with a placeholder category indicating a missing value.

This approach is often referred to as "imputing" missing values, and it is a common technique for dealing with missing data in categorical features. By using a placeholder category, you explicitly indicate that the value is missing, rather than assuming that the missing value is a particular category. This can help to minimize bias in downstream analyses, as it does not introduce any assumptions about the missing data that could bias your results.

👍 ↩ 🚩 upvoted 2 times

🗄️ 👤 **shankalman717** 1 year, 8 months ago

**Selected Answer: C**

When handling missing values in a categorical feature, replacing the missing values with a placeholder category indicating a missing value, as described in option C, is the most appropriate solution in order to minimize bias that could result from the missing values. This approach allows the algorithm to treat missing values as a separate category, avoiding the risk of any assumptions being made about the missing values.

Option A, removing the rows with missing values and upsampling the dataset by 5%, can lead to a loss of valuable data and can also introduce bias into the data. This approach can lead to overrepresentation of certain classes and underrepresentation of others.

Option B, replacing the missing values with the feature's mean, is not appropriate for categorical features as there is no meaningful average value for categorical features.

Option D, moving the rows with missing values to the validation dataset, is not a good solution. This approach may introduce bias into the validation dataset and can lead to overfitting.

👍 ↩ 🚩 upvoted 3 times

🗄️ 👤 **ailiba** 1 year, 8 months ago

I am not really understanding the concept of C. What information should the model learn from that missing value category?

👍 ↩ 🚩 upvoted 1 times

🗄️ 👤 **jdeix** 1 year, 9 months ago



If you want to minimize the bias, why do not you use mean?

👍 ↩ 🚩 upvoted 2 times

  **rayban3981** 1 year, 9 months ago

It is categorical field, you can replace with median or mode not with mean



   upvoted 2 times

  **ares81** 1 year, 10 months ago

**Selected Answer: C**



C, for me.

   upvoted 1 times

  **hargur** 1 year, 10 months ago

C looks correct. We should replace the values with the a placeholder

   upvoted 2 times

  **hiromi** 1 year, 10 months ago

**Selected Answer: C**

C (not sure)

   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.