

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 153 DISCUSSI...

Actual exam question from Google's Professional Machine Learning Engineer

Question #: 153

Topic #: 1

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

You work for a bank and are building a random forest model for fraud detection. You have a dataset that includes transactions, of which 1% are identified as fraudulent. Which data transformation strategy would likely improve the performance of your classifier?

- A. Modify the target variable using the Box-Cox transformation.
- B. Z-normalize all the numeric features.
- C. Oversample the fraudulent transaction 10 times.
- D. Log transform all numeric features.

Show Suggested Answer

by [TNT87](#) at Feb. 9, 2023, 12:48 p.m.

### Comments

Type your comment...



Submit

[fitri001](#) 6 months, 2 weeks ago

**Selected Answer: C**

Oversampling is a common technique to address class imbalance and can significantly improve the performance of the random forest model in fraud detection. It's important to note that oversampling can lead to overfitting, so monitoring the model's performance on unseen data (validation set) is crucial. You might also consider exploring other techniques like undersampling the majority class or using SMOTE (Synthetic Minority Oversampling Technique) for a more balanced approach.

   upvoted 3 times

  **fitri001** 6 months, 2 weeks ago

Class Imbalance: The dataset has a significant class imbalance, with only 1% of transactions being fraudulent (minority class). Random forest models can be biased towards the majority class during training.

Oversampling: Oversampling replicates instances from the minority class (fraudulent transactions) in this case. By increasing the representation of the fraudulent class (10 times in this scenario), the model is exposed to more examples of fraud, improving its ability to learn and detect fraudulent patterns.

   upvoted 1 times

  **pinimichele01** 7 months ago

**Selected Answer: C**

See #60!

   upvoted 1 times


  **M25** 1 year, 5 months ago

**Selected Answer: C**

See #60!

The End. Good luck everyone!!!

   upvoted 2 times

  **Scipione\_** 1 year, 8 months ago

**Selected Answer: C**


The answer is C because it's the only way to improve model performance.

Box-Cox transformation: transform feature values according to normal distribution

Z-normalization: transform feature values according to  $x_{\text{new}} = (x - \mu) / \sigma$  (so  $\{x_{\text{new}}\}$  have mean 0 and std dev 1)

Log transform: just log transformation

Also, the Random Forest algorithm is not a distance-based model but it is a tree-based model, there's no need of normalization process.

   upvoted 4 times

  **TNT87** 1 year, 9 months ago

**Selected Answer: C**

<https://towardsdatascience.com/how-to-build-a-machine-learning-model-to-identify-credit-card-fraud-in-5-steps-a-hands-on-modeling-5140b3bd19f1>

   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.