

- Expert Verified, Online, Free.

■ MENU

C

G 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 36 DISCUSSIO..

Actual exam question from Google's Professional Machine Learning Engineer

Question #: 36

Topic #: 1

[All Professional Machine Learning Engineer Questions]

You are building a model to predict daily temperatures. You split the data randomly and then transformed the training and test datasets. Temperature data for model training is uploaded hourly. During testing, your model performed with 97% accuracy; however, after deploying to production, the model's accuracy dropped to 66%. How can you make your production model more accurate?

- A. Normalize the data for the training, and test datasets as two separate steps.
- B. Split the training and test data based on time rather than a random split to avoid leakage.
- C. Add more data to your test set to ensure that you have a fair distribution and sample for testing.
- D. Apply data transformations before splitting, and cross-validate to make sure that the transformations are applied to both the training and test sets.

Show Suggested Answer

by A maartenalexander at June 22, 2021, 12:46 p.m.

Comments

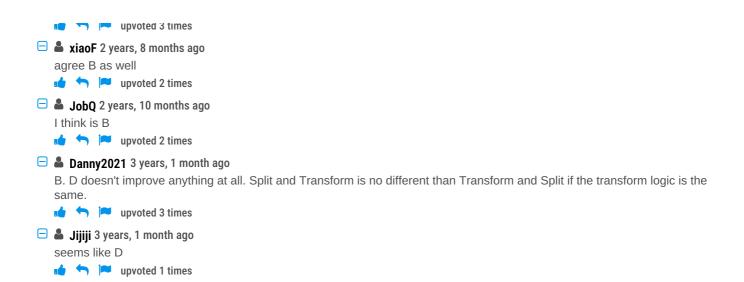
Type your comment...

Submit

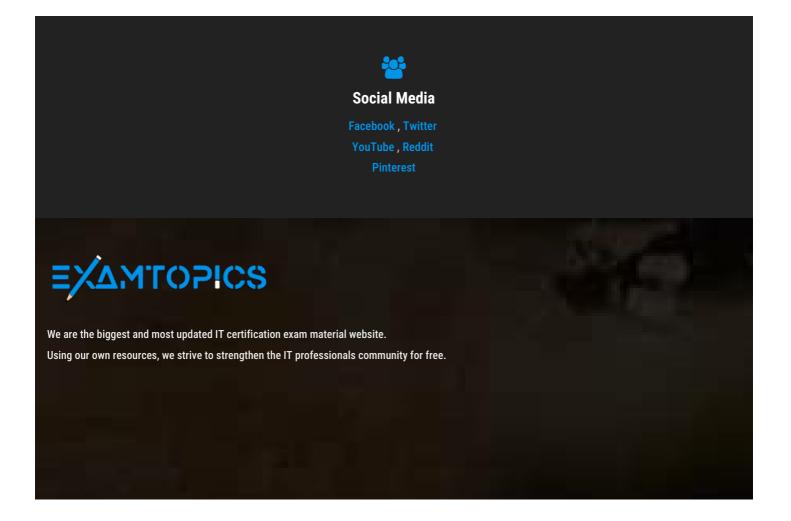
■ maartenalexander Highly Voted 3 years, 4 months ago B. If you do time series prediction, you can't borrow information from the future to predict the future. If you do, you are artificially increasing your accuracy. upvoted 34 times ■ baimus Most Recent ② 1 month, 2 weeks ago Selected Answer: D It's D upvoted 1 times □ baimus 1 month, 2 weeks ago B I mean. Sorry I wrote that comment very early and there is no delete key! upvoted 1 times 😑 🏝 jsalvasoler 2 months, 2 weeks ago Selected Answer: B temporal split is a must in time series forecasting evaluation upvoted 1 times PhilipKoku 4 months, 2 weeks ago **Selected Answer: B** B) Time split to avoid leaking data. upvoted 1 times 🗖 🏜 fragkris 10 months, 3 weeks ago Selected Answer: B Definetely B upvoted 1 times □ Sum_Sum 11 months, 1 week ago **Selected Answer: B** they did not explicitly say forecasting, but splitting by time is the number one rule you learn upvoted 1 times ■ M25 1 year, 5 months ago Selected Answer: B Went with B upvoted 1 times E SergioRubiano 1 year, 7 months ago **Selected Answer: D** D is correct, cross-validate upvoted 2 times ■ Mohamed_Mossad 2 years, 4 months ago **Selected Answer: B** train accuracy 97%, production accuracy 66% ---> time series data ---> random split ---> cause leakage, answer is B upvoted 2 times David_ml 2 years, 5 months ago **Selected Answer: B** You don't split data randomly for time series prediction. upvoted 3 times ■ mmona19 2 years, 6 months ago Selected Answer: B B should be the answer. D is incorrect as normalize before split is going to do data leak https://community.rapidminer.com/discussion/32592/normalising-data-before-data-split-or-after upvoted 2 times 🖃 🚨 giaZ 2 years, 7 months ago Selected Answer: B If you do random split in a time series, your risk that training data will contain information about the target (definition of

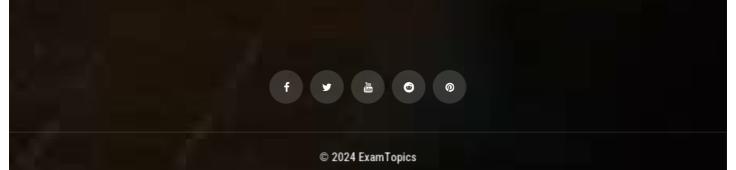
leakage), but similar data won't be available when the model is used for prediction. Leakage causes the model to look

accurate until you start making actual predictions with it.



Start Learning for free





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.