

- 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 76 DISCUSSIO..

Actual exam question from Google's Professional Machine Learning Engineer

Question #: 76

Topic #: 1

[All Professional Machine Learning Engineer Questions]

You are working on a classification problem with time series data. After conducting just a few experiments using random cross-validation, you achieved an Area Under the Receiver Operating Characteristic Curve (AUC ROC) value of 99% on the training data. You haven't explored using any sophisticated algorithms or spent any time on hyperparameter tuning. What should your next step be to identify and fix the problem?

- A. Address the model overfitting by using a less complex algorithm and use k-fold cross-validation.
- B. Address data leakage by applying nested cross-validation during model training.
- C. Address data leakage by removing features highly correlated with the target value.
- D. Address the model overfitting by tuning the hyperparameters to reduce the AUC ROC value.

Show Suggested Answer

by A ares 81 at Dec. 11, 2022, 4:45 p.m.

Comments

Type your comment...

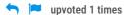
Submit

	pinimichele01 6 months, 2 weeks ago Selected Answer: B
	random cross-validation time series data
	-> B
	♣ gscharly 6 months, 3 weeks ago
	Selected Answer: B
	B with nested cross validation.
	upvoted 2 times
	□ ♣ pinimichele01 6 months, 2 weeks ago
	can you explain me why?
	upvoted 1 times
	Werner123 8 months, 1 week ago
	"99% on training data" -> Data leakage
	"random cross-validation" -> Not suitable for time series, use "nested cross-validation"
	upvoted 3 times
	pmle_nintendo 8 months, 1 week ago
	Selected Answer: D
	Options B and C (Address data leakage by applying nested cross-validation during model training; Address data leakage by removing features highly correlated with the target value) are less relevant in this scenario because the primary concern appears to be overfitting rather than data leakage. Data leakage typically involves inadvertent inclusion of information from the test set in the training process, which may lead to overly optimistic performance metrics. However, there is no indication that data leakage is the cause of the high AUC ROC value in this case.
	upvoted 1 times
	🏝 pico 11 months, 3 weeks ago
	Selected Answer: D Options A and B also address overfitting, but they involve different strategies. Option A suggests using a less complex algorithm and k-fold cross-validation. While this can be effective, it might be premature to change the algorithm without first exploring hyperparameter tuning. Option B suggests addressing data leakage, which is a different issue and may not be the primary cause of overfitting in this scenario.
	humancomputation 1 year, 1 month ago
	Selected Answer: B
	B with nested cross validation.
	upvoted 1 times
	M25 1 year, 6 months ago
	Selected Answer: B
	Went with B
	Selected Answer: B
	Nested cross-validation to reduce data leakage - same as a previous question.
	upvoted 1 times
	♣ Alexarr6 1 year, 8 months ago
	Selected Answer: B
	It's B
_	upvoted 1 times
	Selected Answer: B
	B (same question 48) - https://towardsdatascience.com/time-series-nested-cross-validation-76adba623eb9

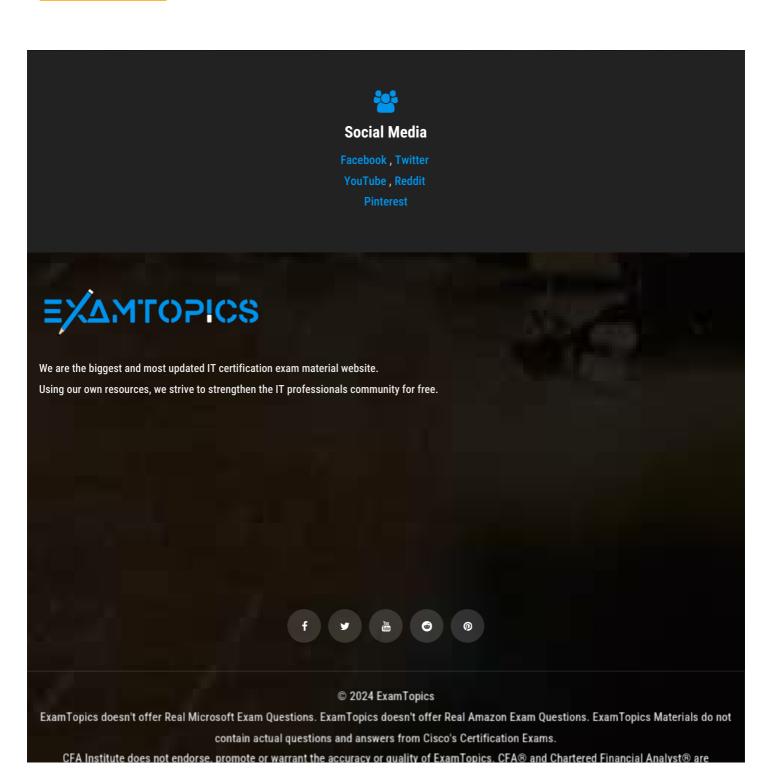
■ ares81 1 year, 11 months ago







Start Learning for free



registered trademarks owned by CFA Institute.