PROFESSIONAL-MACHINE-LEARNING-ENGINEER^{Q&As}

Professional Machine Learning Engineer

Pass Google PROFESSIONAL-MACHINE-LEARNING-ENGINEER Exam with 100% Guarantee

Free Download Real Questions & Answers **PDF** and **VCE** file from:

https://www.passapply.com/professional-machine-learning-engineer.html

100% Passing Guarantee
100% Money Back Assurance

Following Questions and Answers are all new published by Google
Official Exam Center

PROFESSIONAL-MACHINE-LEARNING-ENGINEER PDF Dumps |
PROFESSIONAL-MACHINE-LEARNING-ENGINEER Practice Test |
PROFESSIONAL-MACHINE-LEARNING-ENGINEER Exam Questions

2 / 5

QUESTION 1

Your company manages a video sharing website where users can watch and upload videos. You need to create an ML model to predict which newly uploaded videos will be the most popular so that those videos can be prioritized on your company\\'s website. Which result should you use to determine whether the model is successful?

A. The model predicts videos as popular if the user who uploads them has over 10,000 likes.

B. The model predicts 97.5% of the most popular clickbait videos measured by number of clicks.

C. The model predicts 95% of the most popular videos measured by watch time within 30 days of being uploaded.

D. The Pearson correlation coefficient between the log-transformed number of views after 7 days and 30 days after publication is equal to 0.

Correct Answer: C

QUESTION 2

You work for a retailer that sells clothes to customers around the world. You have been tasked with ensuring that ML models are built in a secure manner. Specifically, you need to protect sensitive customer data that might be used in the models. You have identified four fields containing sensitive data that are being used by your data science team: AGE, IS_EXISTING_CUSTOMER, LATITUDE_LONGITUDE, and SHIRT_SIZE. What should you do with the data before it is made available to the data science team for training purposes?

A. Tokenize all of the fields using hashed dummy values to replace the real values.

B. Use principal component analysis (PCA) to reduce the four sensitive fields to one PCA vector.

C. Coarsen the data by putting AGE into quantiles and rounding LATITUDE_LONGTTUDE into single precision. The other two fields are already as coarse as possible.

D. Remove all sensitive data fields, and ask the data science team to build their models using non-sensitive data.

Correct Answer: A

https://cloud.google.com/blog/products/identity-security/take-charge-of-your-data-how-tokenization-makes-data-usable-without-sacrificing-privacy

QUESTION 3

You are the Director of Data Science at a large company, and your Data Science team has recently begun using the Kubeflow Pipelines SDK to orchestrate their training pipelines. Your team is struggling to integrate their custom Python code into the Kubeflow Pipelines SDK. How should you instruct them to proceed in order to quickly integrate their code with the Kubeflow Pipelines SDK?

A. Use the func_to_container_op function to create custom components from the Python code.

B. Use the predefined components available in the Kubeflow Pipelines SDK to access Dataproc, and run the custom code there.

PROFESSIONAL-MACHINE-LEARNING-ENGINEER PDF Dumps |
PROFESSIONAL-MACHINE-LEARNING-ENGINEER Practice Test |
PROFESSIONAL-MACHINE-LEARNING-ENGINEER Exam Questions

3 / 5

C. Package the custom Python code into Docker containers, and use the load_component_from_file function to import the containers into the pipeline.

D. Deploy the custom Python code to Cloud Functions, and use Kubeflow Pipelines to trigger the Cloud Function.

Correct Answer: A

https://kubeflow-pipelines.readthedocs.io/en/stable/source/kfp.components.html?highlight=func_to_container_op%20#kfp.components.func_to_container_op

QUESTION 4

You work for a public transportation company and need to build a model to estimate delay times for multiple transportation routes. Predictions are served directly to users in an app in real time. Because different seasons and population increases impact the data relevance, you will retrain the model every month. You want to follow Google-recommended best practices. How should you configure the end-to-end architecture of the predictive model?

A. Configure Kubeflow Pipelines to schedule your multi-step workflow from training to deploying your model.

B. Use a model trained and deployed on BigQuery ML, and trigger retraining with the scheduled query feature in BigQuery.

C. Write a Cloud Functions script that launches a training and deploying job on AI Platform that is triggered by Cloud Scheduler.

D. Use Cloud Composer to programmatically schedule a Dataflow job that executes the workflow from training to deploying your model.

Correct Answer: A

https://www.kubeflow.org/docs/components/pipelines/overview/pipelines-overview/) and deploy models

QUESTION 5

You started working on a classification problem with time series data and achieved an area under the receiver operating characteristic curve (AUC ROC) value of 99% for training data after just a few experiments. 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.

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.

Correct Answer: B

https://towardsdatascience.com/time-series-nested-cross-validation-76adba623eb9

PROFESSIONAL-MACHINE-LEARNING-ENGINEER PDF Dumps |
PROFESSIONAL-MACHINE-LEARNING-ENGINEER Practice Test |
PROFESSIONAL-MACHINE-LEARNING-ENGINEER Exam Questions

4 / 5

QUESTION 6

You are experimenting with a built-in distributed XGBoost model in Vertex AI Workbench user-managed notebooks. You use BigQuery to split your data into training and validation sets using the following queries:

CREATE OR REPLACE TABLE `myproject.mydataset.training` AS (SELECT * FROM `myproject.mydataset.mytable` WHERE RAND() A is wrong;

You need to keep doing a good job on older dataset, you can\\'t just ignore it ==> C is wrong;

You know when you are introducing new products, there is no need to wait for a drop in preformaces ==> D is wrong;

B is correct

QUESTION 15

You work on the data science team at a manufacturing company. You are reviewing the company\\'s historical sales data, which has hundreds of millions of records. For your exploratory data analysis, you need to calculate descriptive statistics such as mean, median, and mode; conduct complex statistical tests for hypothesis testing; and plot variations of the features over time. You want to use as much of the sales data as possible in your analyses while minimizing computational resources. What should you do?

A. Visualize the time plots in Google Data Studio. Import the dataset into Vertex AI Workbench user-managed notebooks. Use this data to calculate the descriptive statistics and run the statistical analyses.

B. Spin up a Vertex AI Workbench user-managed notebooks instance and import the dataset. Use this data to create statistical and visual analyses.

C. Use BigQuery to calculate the descriptive statistics. Use Vertex AI Workbench user-managed notebooks to visualize the time plots and run the statistical analyses.

D. Use BigQuery to calculate the descriptive statistics, and use Google Data Studio to visualize the time plots. Use Vertex AI Workbench user-managed notebooks to run the statistical analyses.

Correct Answer: C

https://cloud.google.com/architecture/data-science-with-r-on-gcp-eda#ai_platform_notebooks

https://cloud.google.com/vertex-ai-workbench#section-5

PROFESSIONAL-MACHIN
E-LEARNING-ENGINEER
PDF Dumps

PROFESSIONAL-MACHIN
E-LEARNING-ENGINEER
Practice Test

PROFESSIONAL-MACHIN
E-LEARNING-ENGINEER
Exam Questions