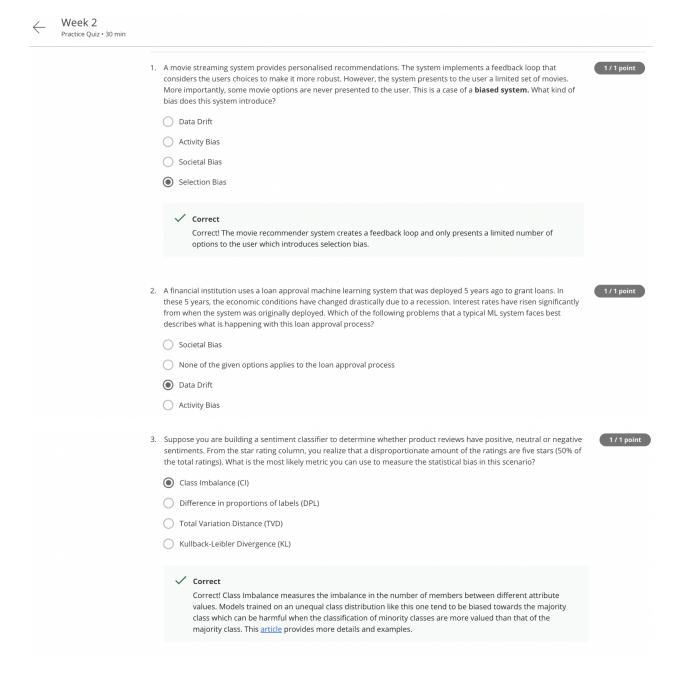
Practical Data Science Specialization



Hong Hanh

Practical Data Science Specialization



4. As a data scientist who works in a sales company, you are asked to predict the sales for December 2020. Your dataset contains daily sales data from January 2020 to September 2020. You successfully train your model on this data, but the actual sales measured in December are completely different from what your model predicted. These types of events are called data drift. After some research, you realize that this change occurred because there is always a sharp increase in sales due to the holiday season. What specific kind of data drift best describes this change in sales? Covariate shift	point
 Prior probability shift Concept jump None of the above 	
 Correct! A prior probability shift occurs when there is a shift in the target variable. 	
 5. Properly measuring statistical data bias is key to build fair and successful ML models. Select the most appropriate metric to detect an imbalance of positive outcomes between different facet values. Total Variation Distance (TVD) 	point
 Kullback-Leibler Divergence (KL) Class Imbalance(Cl) Difference in proportions of labels (DPL) 	
statistical bias in your dataset due to the product category. Which of the following tools from the AWS toolkit best provides the ability to detect statistical bias in the dataset? AWS Glue Amazon Athena Amazon SageMaker Clarify	point
 ✓ Correct Correct! Amazon SageMaker Clarify can be used to identify statistical bias during data preparation without having to write your own code. 	
 7. Feature importance refers to a metric that assigns a score to input features based on how useful they are at predicting a target variable. True or false: applying feature importance always improves model performance. True False 	point