

## Test Questions AIP-210

When you submit this form, it will not automatically collect your details like name and email address unless you provide it yourself.

1

You and your team need to process large datasets of images as fast as possible for a machine learning task. The project will also use a modular framework with extensible code and an active developer community. Which of the following would BEST meet your needs? (1 Point)

- ☐ Caffe
- ☐ Microsoft Cognitive Services
- ☐ Keras
- ☐ TensorBoard

2

Which of the following principles supports building an ML system with a Privacy by Design methodology? (1 Point)

- ☐ Understanding, documenting, and displaying data lineage.
- ☐ Utilizing quasi-identifiers and non-unique identifiers, alone or in combination.
- ☐ Collecting and processing the largest amount of data possible.

- ☐ Avoiding mechanisms to explain and justify automated decisions.

3

A data scientist is tasked to extract business intelligence from primary data captured from the public. Which of the following is the most important aspect that the scientist cannot forget to include? (1 Point)

- ☐ Cybersecurity
- ☐ Data security
- ☐ Cyberprotection
- ☐ Data privacy

4

For a particular classification problem, you are tasked with determining the best algorithm among SVM, random forest, K-nearest neighbors, and a deep neural network. Each of the algorithms has similar accuracy on your data. The stakeholders indicate that they need a model that can convey each feature's relative contribution to the model's accuracy. Which is the best algorithm for this use case? (1 Point)

- ☐ K-nearest neighbors
- ☐ Deep neural network
- ☐ Random forest
- ☐ SVM

5

A dataset can contain a range of values that depict a certain characteristic, such as grades on tests in a class during the semester. A specific student has so far received the following grades: 76, 81, 78, 87, 75, and 72. There is one final test in the semester. What minimum grade would the student need to achieve on the last test to get an 80% average? (1 Point)

☐ 91☐ 89☐ 82☐ 94

6

Which three security measures could be applied in different ML workflow stages to defend them against malicious activities? (Select three.) (1 Point)

Please select 3 options.

☐ Launch ML Instances In a virtual private cloud (VPC).☐ Monitor model degradation.☐ Use max privilege to control access to ML artifacts.☐ Use Secrets Manager to protect credentials.☐ Disable logging for model access.☐ Use data encryption.

7

A healthcare company experiences a cyberattack, where the hackers were able to reverse-engineer a dataset to break confidentiality. Which of the following is TRUE regarding the dataset parameters? (1 Point)

- ☐ The model is underfitted and trained on a low quantity of patient records.
- ☐ The model is overfitted and trained on a high quantity of patient records.
- ☐ The model is overfitted and trained on a low quantity of patient records.
- ☐ The model is underfitted and trained on a high quantity of patient records.

8

When working with textual data and trying to classify text into different languages, which approach to representing features makes the most sense? (1 Point)

- ☐ Clustering similar words and representing words by group membership
- ☐ Bag of words model with TF-IDF
- ☐ Bag of bigrams (2 letter pairs)
- ☐ Word2Vec algorithm

9

Which of the following options is a correct approach for scheduling model retraining in a weather prediction application? (1 Point)

- ☐ Once a month

- ☐ When the input format changes
- ☐ When the input volume changes
- ☐ As new resources become available

10

Which of the following tools would you use to create a natural language processing application? (1 Point)

- ☐ DeepDream
- ☐ NLTK
- ☐ AWS DeepRacer
- ☐ Azure Search

11

A classifier has been implemented to predict whether or not someone has a specific type of disease. Considering that only 1% of the population in the dataset has this disease, which measures will work the BEST to evaluate this model? (1 Point)

- ☐ Mean squared error
- ☐ Precision and accuracy
- ☐ Recall and explained variance
- ☐ Precision and recall

12

Which of the following describes a typical use case of video tracking? (1 Point)

- ☐ Traffic monitoring
- ☐ Video composition
- ☐ Augmented dreaming
- ☐ Medical diagnosis

13

You are developing a prediction model. Your team indicates they need an algorithm that is fast and requires low memory and low processing power. Assuming the following algorithms have similar accuracy on your data, which is most likely to be an ideal choice for the job? (1 Point)

- ☐ Random forest
- ☐ Deep learning neural network
- ☐ Ridge regression
- ☐ Support-vector machine

14

For each of the last 10 years, your team has been collecting data from a group of subjects, including their age and numerous biomarkers collected from blood samples. You are tasked with creating a prediction model of age using the biomarkers as input. You start by performing a linear regression using all of the data over the 10-year period, with age as the dependent variable and the

biomarkers as predictors. Which assumption of linear regression is being violated? (1 Point)

- ☐ Equality of variance (Homoscedastidty)
- ☐ Normality
- ☐ Independence
- ☐ Linearity

15

When should you use semi-supervised learning? (Select two.) (1 Point)

Please select 2 options.

- ☐ Labeling data is challenging and expensive.
- ☐ A small set of labeled data is biased toward one class.
- ☐ A small set of labeled data is available but not representative of the entire distribution.
- ☐ There is a large amount of unlabeled data to be used for predictions.
- ☐ There is a large amount of labeled data to be used for predictions.

16

Which of the following can benefit from deploying a deep learning model as an embedded model on edge devices? (1 Point)

- ☐ A more complex model
- ☐ Guaranteed availability of enough space

- ☐ Reduction in latency
- ☐ Increase in data bandwidth consumption

17

Which of the following is the definition of accuracy? (1 Point)

- ☐  $(\text{True Positives} + \text{False Positives}) / \text{Total Predictions}$
- ☐  $\text{True Positives} / (\text{True Positives} + \text{False Positives})$
- ☐  $(\text{True Positives} + \text{True Negatives}) / \text{Total Predictions}$
- ☐  $\text{True Positives} / (\text{True Positives} + \text{False Negatives})$

18

Personal data should not be disclosed, made available, or otherwise used for purposes other than specified with which of the following exceptions? (Select two.) (1 Point)

Please select 2 options.

- ☐ If it was with consent of the person it is collected from.
- ☐ If it was collected accidentally.
- ☐ If the data is only collected once.
- ☐ If it was requested by the authority of law.
- ☐ If it is for a good cause.



19

Which of the following sentences is TRUE about the definition of cloud models for machine learning pipelines? (1 Point)

- ☐ Software as a Service (SaaS) can provide AI practitioner data science services such as Jupyter notebooks.
- ☐ Platform as a Service (PaaS) can provide some services within an application such as payment applications to create efficient results.
- ☐ Data as a Service (DaaS) can host the databases providing backups, clustering, and high availability.
- ☐ Infrastructure as a Service (IaaS) can provide CPU, memory, disk, network and GPU.

20

In a self-driving car company, ML engineers want to develop a model for dynamic pathing. Which of following approaches would be optimal for this task? (1 Point)

- ☐ Reinforcement learning
- ☐ Dijkstra Algorithm
- ☐ Supervised Learning.
- ☐ Unsupervised Learning

21

R-squared is a statistical measure that: (1 Point)

- ☐ Expresses the extent to which two variables are linearly related.
- ☐ Is the proportion of the variance for a dependent variable that's explained by independent variables.
- ☐ Represents the extent to which two random variables vary together.
- ☐ Combines precision and recall of a classifier into a single metric by taking their harmonic mean.

22

Which of the following equations best represent an L1 norm? (1 Point)

- ☐  $|x| - |y|$
- ☐  $|x|^2 + |y|^2$
- ☐  $|x| + |y|$
- ☐  $|x| + |y|^2$

23

Which of the following statements are true regarding highly interpretable models? (Select two.) (1 Point)

Please select 2 options.

- ☐ They are usually easier to explain to business stakeholders.
- ☐ They are usually very good at solving non-linear problems.
- ☐ They are usually binary classifiers.
- ☐ They are usually referred to as "black box" models.

- ☐ They usually compromise on model accuracy for the sake of interpretability.

24

Which two of the following decrease technical debt in ML systems? (Select two.)  
(1 Point)

Please select 2 options.

- ☐ Boundary erosion
- ☐ Design anti-patterns
- ☐ Documentation readability
- ☐ Model complexity
- ☐ Refactoring

25

Which of the following describes a neural network without an activation function? (1 Point)

- ☐ A form of a linear regression
- ☐ A radial basis function kernel
- ☐ An unsupervised learning technique
- ☐ A form of a quantile regression

26

The following confusion matrix is produced when a classifier is used to predict labels on a test dataset. How precise is the classifier?

(1 Point)

- ☐  $37/(37 + 8)$
- ☐  $48/(48 + 37)$
- ☐  $(48 + 37)/100$
- ☐  $37/(37 + 7)$

27

Given a feature set with rows that contain missing continuous values, and assuming the data is normally distributed, what is the best way to fill in these missing features? (1 Point)

- ☐ Delete entire columns that contain any missing features.
- ☐ Fill in missing features with the average of observed values for that feature in the entire dataset.
- ☐ Fill in missing features with random values for that feature in the training set.
- ☐ Delete entire rows that contain any missing features.

28

In addition to understanding model performance, what does continuous monitoring of bias and variance help ML engineers to do? (1 Point)

- ☐ Respond to hidden attacks

- ☐ Recover from hidden attacks
- ☐ Prevent hidden attacks
- ☐ Detect hidden attacks

29

A company is developing a merchandise sales application. The product team uses training data to teach the AI model predicting sales, and discovers emergent bias. What caused the biased results? (1 Point)

- ☐ The AI model was trained in winter and applied in summer.
- ☐ The application was migrated from on-premise to a public cloud.
- ☐ The training data used was inaccurate.
- ☐ The team set flawed expectations when training the model.

30

You train a neural network model with two layers, each layer having four nodes, and realize that the model is underfit. Which of the actions below will NOT work to fix this underfitting? (1 Point)

- ☐ Increase the complexity of the model
- ☐ Add features to training data
- ☐ Train the model for more epochs
- ☐ Get more training data

31

Which of the following is NOT an activation function? (1 Point)

- ☐ ReLU
- ☐ Sigmoid
- ☐ Additive
- ☐ Hyperbolic Tangent (tanH)

32

Which of the following items should be included in a handover to the end user to enable them to use and run a trained model on their own system? (Select three.) (1 Point)

Please select 3 options.

- ☐ Information on the folder structure in your local machine
- ☐ README document
- ☐ Intermediate data files
- ☐ Sample input and output data files
- ☐ Link to a GitHub repository of the codebase

33

Which of the following pieces of AI technology provides the ability to create fake videos? (1 Point)

- ☐ Generative adversarial networks (GAN)
- ☐ Long short-term memory (LSTM) networks
- ☐ Support-vector machines (SVM)
- ☐ Recurrent neural networks (RNN)

34

Which database is designed to better anticipate and avoid risks of AI systems causing safety, fairness, or other ethical problems? (1 Point)

- ☐ Configuration Management
- ☐ Code Repository
- ☐ Incident
- ☐ Asset

35

What is the open framework designed to help detect, respond to, and remediate threats in ML systems? (1 Point)

- ☐ Adversarial ML Threat Matrix
- ☐ OWASP Threat and Safeguard Matrix
- ☐ Threat Susceptibility Matrix
- ☐ MITRE ATT&CK® Matrix

36

Which two techniques are used to build personas in the ML development lifecycle? (Select two.) (1 Point)

Please select 2 options.

- ☐ Population variance
- ☐ Population estimates
- ☐ Population regression
- ☐ Population triage
- ☐ Population resampling

37

Which of the following text vectorization methods is appropriate and correctly defined for an English-to-Spanish translation machine? (1 Point)

- ☐ Using TF-IDF because in translation machines, we need to consider the order of the words.
- ☐ Using Word2vec because in translation machines, we need to consider the order of the words.
- ☐ Using TF-IDF because in translation machines, we do not care about the order of the words.
- ☐ Using Word2vec because in translation machines, we do not care about the order of the words.

38

Word Embedding describes a task in natural language processing (NLP) where:



(1 Point)

- ☐ Words are featurized by taking a matrix of bigram counts.
- ☐ Words are grouped together into clusters and then represented by word cluster membership.
- ☐ Words are featurized by taking a histogram of letter counts.
- ☐ Words are converted into numerical vectors.

39

You are building a prediction model to develop a tool that can diagnose a particular disease so that individuals with the disease can receive treatment. The treatment is cheap and has no side effects. Patients with the disease who don't receive treatment have a high risk of mortality. It is of primary importance that your diagnostic tool has which of the following? (1 Point)

- ☐ High negative predictive value
- ☐ High positive predictive value
- ☐ Low false negative rate
- ☐ Low false positive rate

40

Which of the following algorithms is an example of unsupervised learning? (1 Point)

- ☐ Neural networks
- ☐ Principal components analysis

- ☐ Random forest
- ☐ Ridge regression

41

What is Word2vec? (1 Point)

- ☐ A word embedding method that finds characteristics of words in a very large number of documents.
- ☐ A word embedding method that builds a one-hot encoded matrix from samples and the terms that appear in them.
- ☐ A bag of words.
- ☐ A matrix of how frequently words appear in a group of documents.

42

We are using the k-nearest neighbors algorithm to classify the new data points. The features are on different scales. Which method can help us to solve this problem? (1 Point)

- ☐ Square-root transformation
- ☐ Log transformation
- ☐ Standardization
- ☐ Normalization

43

An HR solutions firm is developing software for staffing agencies that uses machine learning. The team uses training data to teach the algorithm and discovers that it generates lower employability scores for women. Also, it predicts that women, especially with children, are less likely to get a high-paying job. Which type of bias has been discovered? (1 Point)

- ☐ Automation
- ☐ Emergent
- ☐ Preexisting
- ☐ Technical

44

Which two encodes can be used to transform categories data into numerical features? (Select two.) (1 Point)

Please select 2 options.

- ☐ Count Encoder
- ☐ Median Encoder
- ☐ Mean Encoder
- ☐ Log Encoder
- ☐ One-Hot Encoder

45

Which of the following is the primary purpose of hyperparameter optimization?  
(1 Point)

- ☐ Increases recall over precision
- ☐ Makes models easier to explain to business stakeholders
- ☐ Improves model interpretability
- ☐ Controls the learning process of a given algorithm

46

In which of the following scenarios is lasso regression preferable over ridge regression? (1 Point)

- ☐ The number of features is much larger than the sample size.
- ☐ There is high collinearity among some of the features associated with the dependent variable.
- ☐ There are many features with no association with the dependent variable.
- ☐ The sample size is much larger than the number of features.

47

Which of the following is the correct definition of the quality criteria that describes completeness? (1 Point)

- ☐ The degree to which the measures conform to defined business rules or constraints.

- ☐ The degree to which a set of measures are specified using the same units of measure in all systems.
- ☐ The degree to which all required measures are known.
- ☐ The degree to which a set of measures are equivalent across systems.

48

You have a dataset with many features that you are using to classify a dependent variable. Because the sample size is small, you are worried about overfitting. Which algorithm is ideal to prevent overfitting? (1 Point)

- ☐ XGBoost
- ☐ Logistic regression
- ☐ Random forest
- ☐ Decision tree

49

Which of the following approaches is best if a limited portion of your training data is labeled? (1 Point)

- ☐ Reinforcement learning
- ☐ Semi-supervised learning
- ☐ Probabilistic clustering
- ☐ Dimensionality reduction

50

Which of the following can take a question in natural language and return a precise answer to the question? (1 Point)

- ☐ Spark ML
- ☐ Databricks
- ☐ Pandas
- ☐ IBM Watson

51

Which of the following is TRUE about SVM models? (1 Point)

- ☐ They can be used only for classification.
- ☐ They use the sigmoid function to classify the data points.
- ☐ They can be used only for regression.
- ☐ They can take the feature space into higher dimensions to solve the problem.

52

Which of the following sentences is true about model evaluation and model validation in ML pipelines? (1 Point)

- ☐ Model evaluation and validation are the same.
- ☐ Model validation occurs before model evaluation.

- ☐ Model validation is defined as a set of tasks to confirm the model performs as expected.
- ☐ Model evaluation is defined as an external component.

53

Which of the following methods can be used to rebalance a dataset using the rebalance design pattern? (1 Point)

- ☐ Bagging
- ☐ Boosting
- ☐ Stacking
- ☐ Weighted class

54

Why do data skews happen in the ML pipeline? (1 Point)

- ☐ There is insufficient training data for evaluation.
- ☐ There is a mismatch between live output data and offline data.
- ☐ There Is a mismatch between live input data and offline data.
- ☐ Test and evaluation data are designed incorrectly.

55

Which of the following models are text vectorization methods? (Select two.) (1 Point)

Please select 2 options.

- ☐ Lemmatization
- ☐ Skip-gram
- ☐ PCA
- ☐ TF-IDF
- ☐ t-SNE
- ☐ Tokenization

56

Your dependent variable Y is a count, ranging from 0 to infinity. Because Y is approximately log- normally distributed, you decide to log-transform the data prior to performing a linear regression. What should you do before log-transforming Y? (1 Point)

- ☐ Subtract the mean of Y from all the Y values.
- ☐ Divide all the Y values by the standard deviation of Y.
- ☐ Add 1 to all of the Y values.
- ☐ Explore the data for outliers.

57

Which of the following describes a benefit of machine learning for solving business problems? (1 Point)

- ☐ Improving the quality of original data



- ☐ Increasing the speed of analysis
- ☐ Increasing the quantity of original data
- ☐ Improving the constraint of the problem

58

An AI practitioner incorporates risk considerations into a deployment plan and decides to log and store historical predictions for potential, future access requests. Which ethical principle is this an example of? (1 Point)

- ☐ Safety
- ☐ Privacy
- ☐ Fairness
- ☐ Transparency

59

An AI system recommends New Year's resolutions. It has an ML pipeline without monitoring components. What retraining strategy would be BEST for this pipeline? (1 Point)

- ☐ Periodically before New Year's Day and after New Year's Day
- ☐ Periodically every year
- ☐ When data drift is detected
- ☐ When concept drift is detected

60

Which two of the following criteria are essential for machine learning models to achieve before deployment? (Select two.) (1 Point)

Please select 2 options.

- ☐ Data size
- ☐ Portability
- ☐ Scalability
- ☐ Explainability
- ☐ Complexity

61

Which two of the following statements about the beta value in an A/B test are accurate? (Select two.) (1 Point)

Please select 2 options.

- ☐ The Beta value is the rate of type I errors for the test.
- ☐ The Beta value is the rate of type II errors for the test.
- ☐ The Beta in an Alpha/Beta test represents one of the two variants of the A/B test.
- ☐ The statistical power of a test is the inverse of the Beta value, or  $1 - \text{Beta}$ .

62

A change in the relationship between the target variable and input features is:

(1 Point)

- ☐ concept drift.
- ☐ model decay.
- ☐ covariate shift.
- ☐ data drift.

63

When should the model be retrained in the ML pipeline? (1 Point)

- ☐ A new monitoring component is added.
- ☐ Concept drift is detected in the pipeline.
- ☐ Some outliers are detected in live data.
- ☐ More data become available for the training phase.

64

Which of the following regressions will help when there is the existence of near-linear relationships among the independent variables (collinearity)? (1 Point)

- ☐ Linear regression
- ☐ Ridge regression
- ☐ Clustering

☐ Polynomial regression

65

What is the primary benefit of the Federated Learning approach to machine learning? (1 Point)

- ☐ It does not require a labeled dataset to solve supervised learning problems.
- ☐ It protects the privacy of the user's data while providing well-trained models.
- ☐ It uses large, centralized data stores to train complex machine learning models.
- ☐ It requires less computation to train the same model using a traditional approach.

66

Which of the following unsupervised learning models can a bank use for fraud detection? (1 Point)

- ☐ DB5CAN
- ☐ Hierarchical clustering
- ☐ Anomaly detection
- ☐ k-means

67

Which of the following are true about the transform-design pattern for a machine learning pipeline? (Select three.) It aims to separate inputs from features. (1 Point)

Please select 3 options.

- ☐ It represents steps in the pipeline with a directed acyclic graph (DAG).
- ☐ It ensures reproducibility.
- ☐ It encapsulates the processing steps of ML pipelines.
- ☐ It seeks to isolate individual steps of ML pipelines.
- ☐ It transforms the output data after production.

68

Which of the following scenarios is an example of entanglement in ML pipelines? (1 Point)

- ☐ Change the way output is visualized in the monitoring step.
- ☐ Add a new pipeline for retraining the model in the model training step.
- ☐ Change in normalization function in the feature engineering step.
- ☐ Add a new method for drift detection in the model evaluation step.

69

Which of the following is NOT a valid cross-validation method? (1 Point)

- ☐ Bootstrapping
- ☐ K-fold
- ☐ Stratification

☐ Leave-one-out

70

Which of the following is a privacy-focused law that an AI practitioner should adhere to while designing and adapting an AI system that utilizes personal data? (1 Point)

- ☐ PCIDSS
- ☐ Sarbanes Oxley (SOX)
- ☐ General Data Protection Regulation (GDPR)
- ☐ ISO/IEC 27001

71

The graph is an elbow plot showing the inertia or within-cluster sum of squares on the y-axis and number of clusters (also called K) on the x-axis, denoting the change in inertia as the clusters change using k-means algorithm. What would be an optimal value of K to ensure a good number of clusters? (1 Point)

- ☐ 5
- ☐ 3
- ☐ 9
- ☐ 2

72

Which of the following is a common negative side effect of not using

regularization? (1 Point)

- ☐ Overfitting
- ☐ Slow convergence time
- ☐ Higher compute resources
- ☐ Low test accuracy

73

Which SQL query provides the Directors' Firstname, Lastname, the name of their departments, and the average employee's salary? (1 Point)

- ☐ SELECT m.Firstname, m.Lastname, [d.Name](#), AVG(e.Salary) as Dept\_avg\_Salary FROM Employee\_Table as e RIGHT JOIN Departmenttable as d on e.Dept = [d.Name](#) INNER JOIN Directorable as m on [d.ID](#) = m.DeptJD GROUP BY [d.Name](#)
- ☐ SELECT m.Firstname, m.Lastname, [d.Name](#), AVG(e.Salary) as Dept\_avg\_Salary FROM Employee\_Table as e RIGHT JOIN Department\_Table as d on e.Dept = [d.Name](#) INNER JOIN Directorable as m on [d.ID](#) = m.DeptID GROUP BY m.Firstname, m.Lastname, [d.Name](#)
- ☐ SELECT m.Firstname, m.Lastname, [d.Name](#), AVG(e.Salary) as Dept\_avg\_Salary FROM Employee\_Table as e RIGHT JOIN Department\_Table as d on e.Dept = [d.Name](#) INNER JOIN Directorable as m on [d.ID](#) = m.DeptJD GROUP BY e.Salary
- ☐ SELECT m.Firstname, m.Lastname, [d.Name](#), AVG(e.Saiary) as Dept\_avg\_Saiary FROM Employee\_Table as e LEFT JOIN Department\_Table as d on e.Dept = [d.Name](#) LEFT JOIN Directorable as m on [d.ID](#) = m.DeptJD GROUP BY m.Firstname, m.Lastname, [d.Name](#)

74

An organization sells house security cameras and has asked their data scientists to implement a model to detect human feces, as distinguished from animals, so they can alert th customers only when a human gets close to their house.

Which of the following algorithms is an appropriate option with a correct reason? (1 Point)

- ☐ Logistic regression, because this is a classification problem and our data is linearly separable.
- ☐ A decision tree algorithm, because the problem is a classification problem with a small number of features.
- ☐ Neural network model, because this is a classification problem with a large number of features.
- ☐ k-means, because this is a clustering problem with a small number of features.

75

You create a prediction model with 96% accuracy. While the model's true positive rate (TPR) is performing well at 99%, the true negative rate (TNR) is only 50%. Your supervisor tells you that the TNR needs to be higher, even if it decreases the TPR. Upon further inspection, you notice that the vast majority of your data is truly positive. What method could help address your issue? (1 Point)

- ☐ Normalization
- ☐ Principal components analysis
- ☐ Oversampling
- ☐ Quality filtering

76

Normalization is the transformation of features: (1 Point)

- ☐ By subtracting from the mean and dividing by the standard deviation.



- ☐ Into the normal distribution.
- ☐ So that they are on a similar scale.
- ☐ To different scales from each other.

77

You have a dataset with thousands of features, all of which are categorical. Using these features as predictors, you are tasked with creating a prediction model to accurately predict the value of a continuous dependent variable. Which of the following would be appropriate algorithms to use? (Select two.)  
(1 Point)

Please select 2 options.

- ☐ Lasso regression
- ☐ K-means
- ☐ K-nearest neighbors
- ☐ Logistic regression
- ☐ Ridge regression

78

Which of the following tests should be performed at the production level before deploying a newly retrained model? (1 Point)

- ☐ Performance test
- ☐ A/B test
- ☐ Option 2

- ☐ Security test
- ☐ Unit test

79

A product manager is designing an Artificial Intelligence (AI) solution and wants to do so responsibly, evaluating both positive and negative outcomes. The team creates a shared taxonomy of potential negative impacts and conducts an assessment along vectors such as severity, impact, frequency, and likelihood. Which modeling technique does this team use? (1 Point)

- ☐ Threat
- ☐ Process
- ☐ Harms
- ☐ Business

80

In general, models that perform their tasks: (1 Point)

- ☐ More accurately are less robust against adversarial attacks.
- ☐ Less accurately are less robust against adversarial attacks.
- ☐ Less accurately are neither more nor less robust against adversarial attacks.
- ☐ More accurately are neither more nor less robust against adversarial attacks.

81

Which type of regression represents the following formula:  $y = c + b \cdot x$ , where  $y$  = estimated dependent variable score,  $c$  = constant,  $b$  = regression coefficient, and  $x$  = score on the independent variable? (1 Point)

- ☐ *Linear regression*
- ☐ *Ridge regression*
- ☐ *Polynomial regression*
- ☐ *Lasso regression*

82

Which of the following metrics is being captured when performing principal component analysis? (1 Point)

- ☐ Kurtosis
- ☐ Skewness
- ☐ Missingness
- ☐ Variance

83

Workflow design patterns for the machine learning pipelines: (1 Point)

- ☐ Separate inputs from features.
- ☐ Represent a pipeline with directed acyclic graph (DAG).

- ☐ Seek to simplify the management of machine learning features.
- ☐ Aim to explain how the machine learning model works.

84

A big data architect needs to be cautious about personally identifiable information (PII) that may be captured with their new IoT system. What is the final stage of the Data Management Life Cycle, which the architect must complete in order to implement data privacy and security appropriately?

(1 Point)

- ☐ Destroy
- ☐ Option 2
- ☐ De-Duplicate
- ☐ Detain
- ☐ Duplicate

85

You are implementing a support-vector machine on your data, and a colleague suggests you use a polynomial kernel. In what situation might this help improve the prediction of your model? (1 Point)

- ☐ When it is necessary to save computational time.
- ☐ When there is high correlation among the features.
- ☐ When the distribution of the dependent variable is Gaussian.

- ☐ When the categories of the dependent variable are not linearly separable.

86

Which of the following best describes distributed artificial intelligence? (1 Point)

- ☐ relies on a distributed system that performs robust computations across a network of unreliable nodes.
- ☐ It does not require hyperparameter tuning because the distributed nature accounts for the bias.
- ☐ It intelligently pre-distributes the weight of starting a neural network.
- ☐ It uses a centralized system to speak to decentralized nodes.

87

Your dependent variable data is a proportion. The observed range of your data is 0.01 to 0.99. The instrument used to generate the dependent variable data is known to generate low quality data for values close to 0 and close to 1. A colleague suggests performing a logit-transformation on the data prior to performing a linear regression. Which of the following is a concern with this approach? (1 Point)

Definition of logit-transformation: If  $p$  is the proportion:  $\text{logit}(p) = \log(p/(1-p))$

- ☐ Values near 0.5 before logit-transformation will be near 0 after.
- ☐ Noisy data could become more influential in your model.
- ☐ After logit-transformation, the data may violate the assumption of independence.
- ☐ The model will be more likely to violate the assumption of normality.

88

Which of the following occurs when a data segment is collected in such a way that some members of the intended statistical population are less likely to be included than others? (1 Point)

- ☐ Algorithmic bias
- ☐ Stereotype bias
- ☐ Systematic value distortion
- ☐ Sampling bias

89

A market research team has ratings from patients who have a chronic disease, on several functional, physical, emotional, and professional needs that stay unmet with the current therapy. The dataset also captures ratings on how the disease affects their day-to-day activities. A pharmaceutical company is introducing a new therapy to cure the disease and would like to design their marketing campaign such that different groups of patients are targeted with different ads. These groups should ideally consist of patients with similar unmet needs. Which of the following algorithms should the market research team use to obtain these groups of patients? (1 Point)

- ☐ Naive-Bayes
- ☐ k-means clustering
- ☐ k-nearest neighbors
- ☐ Logistic regression

90

Which of the following is a type 1 error in statistical hypothesis testing? (1 Point)

- ☐ The null hypothesis is false and is rejected.
- ☐ The null hypothesis is true, but is rejected.
- ☐ The null hypothesis is false, but fails to be rejected.
- ☐ The null hypothesis is true and fails to be rejected.

Never give out your password. [Report abuse](#)



This content is created by the owner of the form. The data you submit will be sent to the form owner. Microsoft is not responsible for the privacy or security practices of its customers, including those of this form owner. Never give out your password.

**Microsoft Forms** | AI-Powered surveys, quizzes and polls [Create my own form](#)

The owner of this form has not provided a privacy statement as to how they will use your response data. Do not provide personal or sensitive information. | [Terms of use](#)