

Prodotto concepito per la preparazione all' esame di certificazione
Information Technology Specialist in Artificial Intelligence

Versione 2.0 del 03/02/2025

Info simulatore:

Il presente simulatore comprende 90 domande

info esame reale:

Numero esame: **Information Technology in Artificial Intelligence**

Durata: **50 Minuti**

Domande: **50**

Punteggio minimo: **700**

Importante:

per poter sostenere l'esame reale è necessario registrarsi su questo sito:

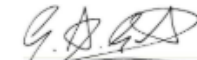
<https://app.certiport.com/portal/>



Mario Rossi

has successfully completed the certification requirements for

Artificial Intelligence


Dr. Gary A. Gates
Managing Director
Pearson VUE

May 25, 2023
Date Awarded

XXXX-XXXX
verify.certiport.com



Il simulatore, e quindi le relative domande e risposte, hanno l'unico obiettivo di fornire allo studente una preparazione che si avvicini quanto più possibile al contenuto reale dell'esame di certificazione. Si precisa inoltre che l'esame ufficiale è coperto da copyright e, quindi, non è né corretto né utile, didatticamente, considerare il simulatore come una copia dell'esame di certificazione.

1

A car dealership wants to use AI to predict the number of customers to expect on any given day, based on previous sales data. The dealership plans to use this data to make plans weekly.

What is the most important metric of success for the AI model from the dealership's perspective?

- ☒ Generated predictions help the dealership manage inventory and staffing to increase profits
- ☐ The model runs efficiently and in real time.
- ☐ The model can be updated as often as weekly.
- ☐ Generated predictions vary as little as possible from day to day.

2

An insurance company determines whether clients are eligible for discounts based on their answers to 10 Yes/No questions. The company CEO wants to develop a program that will determine a person's eligibility instead of having the insurance agents ask the questions. The program must be as reliable and efficient as possible. The CEO asks you what type of AI you recommend for this situation.

How should you respond to the CEO?

- ☐ A clustering algorithm that clusters people with similar answers into discount groups would be the best solution.
- ☐ A neural network that learns to predict a person's eligibility for insurance discounts based on their answers to the questions would be the best solution.
- ☒ AI is not the most efficient solution in this case because the problem is well handled with a small set of decision statements.
- ☐ Any type of machine learning algorithm will be able to efficiently learn the correct discount categories given sufficient training data that is representative of all discount classes.

3

Which artificial intelligence algorithm or application uses unlabeled data?

- ☒ Clustering
- ☐ Classification
- ☐ Object detection
- ☐ Machine translation

4

A bank wants to detect fraudulent transactions. It has the ability to collect data, such as the amount and item purchased, for each transaction.

Which type of machine learning problem is this?

- ☐ Supervised learning, regression
- ☐ Supervised learning, classification
- ☒ Unsupervised learning, clustering
- ☐ Reinforcement learning, game playing

5

A company offers a wide range of products to users. The company wants to organize the products into different groups so they can develop marketing strategies for each group of products.

Which type of machine learning problem is this?

- ☐ Supervised learning, clustering
- ☒ Unsupervised learning, clustering
- ☐ Unsupervised learning, classification
- ☐ Supervised learning, regression

6

You are working with a team on a product that can act as an opponent in a video game. The team is using an algorithm based on rewards: if it does well, the quality of the outcome is improved.

Which type of machine learning is this?

- ☒ Reinforcement learning
- ☐ Supervised learning
- ☐ Semi-supervised learning
- ☐ Unsupervised learning

7

Why should you include subject-matter experts who do not have technical expertise on your AI project team?

- ☒ To feed domain-specific data into the algorithm so the AI system is efficient and unbiased
- ☐ To evaluate tools used to integrate AI into existing applications
- ☐ To organize and cleanse the structured and unstructured data in a custom AI solution
- ☐ To train them in applied AI and data engineering skills

8

You've been hired as a project manager by an animal hospital. The hospital wants to use machine learning to predict which characteristics of tumors in cats and dogs, that can be identified on X-rays, lead to the most successful risk classification for different patients. You will have access to the medical records (including imaging) for all the animals that were treated for tumors at the hospital in the last 10 years. Your first job is to assemble the team that will work on this project.

- ☐ The owner of a local pet boarding business who specializes in medically fragile pets
- ☒ An AI developer who has experience with image-classification tasks.
- ☐ A veterinary technician who often assists in surgeries and cares for pets in the operating room and during their chemotherapy appointments.
- ☐ A pharmaceutical sales rep who provides chemotherapy drugs to the hospital.
- ☒ An expert veterinarian radiologist who excels at recognizing tumors in imaging scans
- ☐ The hospital CFO, who is responsible for the purchase and maintenance of imaging machinery.

What is a vulnerability that a malicious actor can exploit to attack or undermine an AI system?

- ☐ Unsupervised learning
- ☒ Data extraction
- ☐ Data encryption
- ☐ Overfitting

10

How does a decision tree train a classification model?

- ☐ It finds a linear relationship between independent and dependent variables to split the dataset into two or more groups with similar features.
- ☒ It uses the features to split the dataset into two or more groups with similar features.
- ☐ It calculates the similarities between data points based on their features to split them into two or more groups with similar features.
- ☐ It finds one or more lines to split the dataset into two or more groups with similar features

11

Several companies are contributing to and monitoring an industry-specific database of AI-related safety incidents. The companies hope that by sharing knowledge, they can improve the safety of AI systems deployed in the industry.

Which project team member is responsible for monitoring incidents in the AI incident database?

- ☐ UI Designer, improving the look and feel of product software.
- ☒ Product Manager, evaluating the addition of a new feature her team is developing.
- ☐ Developer, producing efficient and elegant code based on requirements.
- ☐ Project Manager, measuring cost and schedule overruns.

12

A school counselor wants to understand the main criteria for admission to highly selective universities. What kind of data does the counselor need to train the most effective machine learning model?

- ☐ Past years' data from the city in which the school operates for as far back as possible.
- ☐ Past years data from the counselor's school for as far back as possible
- ☒ Past years' data from schools in as many diverse locations as possible
- ☐ Data from the past decade for the school.

13

Your team is building a farming robot. The robot will use a vision system to recognize weeds. Then it will kill the weeds without harming the crops around them.

You have been asked to develop a training dataset for the project.

Which images should the training set include, and why?

- ☐ Images of items found around a farm including tools, livestock, and crops, so the robot can recognize all items it might encounter on a working farm
- ☐ A wide variety of images, so the robot will be able to perform general object recognition in different environments
- ☐ Only images of weeds, so the system can build a robust recognition model that is tuned to recognize weeds
- ☒ Images of weeds and crop plants, so the system can learn to distinguish between the two categories of plants it is likely to see

14

You are choosing between various datasets for a model with a specified goal.

Which two factors are likely to be important elements in making a choice of dataset to meet this goal?

(Choose 2)

- ☐ Whether the dataset was released by an academic or commercial entity.
- ☐ The file format of dataset elements such as images.
- ☐ Whether the dataset was located by using a dataset search engine.
- ☒ Missing elements or the presence of corrupt elements
- ☒ Ensuring that the dataset is sufficient for the purpose.

15

You are trying to build a model that predicts whether incoming email messages are spam. The following table displays a representative sample of the dataset.

What issue does the dataset present that you will need to address?

Email Source Type	Sender Name	is_Spam
organization	Superstore	No
personal	John Doe	No
personal	Max Front	No
personal	Taylor Tires	No
organization	Grocery Store	No
organization	Beauty Company	No
personal	Mark Miller	Yes
personal	Anna Brown	No

- ☒ Imbalanced data
- ☐ Outliers
- ☐ Missing data
- ☐ Corrupted data

16

A company developed a smartphone app that used GPS and accelerometer data to identify and report potholes in a city's streets. Their intent was to help allocate resources to patch thousands of potholes. However, many of the city's elderly and lower-income residents did not have smartphones, resulting in datasets biased in favor of more affluent neighborhoods

Which type of bias is this an example of?

- ☐ Human cognitive bias
- ☐ Reinforcement bias
- ☒ Data collection bias
- ☐ Automation bias

17

Move each description from the list on the left to the corresponding data storage type on the right.

Data Storage Types

Cloud Storage

Costs vary by the amount of space needed at given-time

There are no extra costs for maintenance and upgrades

Local Server

Involves the cost of installing and maintaining hardware

It is accessible even without an internet connection

18

You intend to encode visual information into a form that a machine vision system can interpret. However, memory requirements and the size of model are constraining factors. Assuming identical resolution, which form of image data will require the smallest memory footprint?

- ☐ 256 color
- ☒ Monochrome
- ☐ Grayscale
- ☐ High dynamic range

19

Move each AI problem from the list on the left to the most appropriate data type on the right.

AI Problems

Data Types

Stock price prediction

Time series data

*Voting patterns across
different areas*

Tabular data

Email sentiment analysis

Text data

Object recognition

Image data

20

You are building an AI to predict whether a loan applicant should be approved or denied for a loan. The data has many features because your company knows a lot about its customers. You want to start with only the most-valuable features to give the AI its best chance for success.

How should you do this?

- ☐ Keep only the features that describe personal things about the customer because those are the most distinctive
- ☐ Ask the customer what they think would be the best things about them to consider when evaluating the application.
- ☒ Talk with the people who manually approve the loans and ask them what factors to consider.
- ☐ Keep only the features that are numbers (like the customer's age) because AIs only work with numbers.

21

Data scientist will use data represented by the table to predict how much money costumers will invest through a bank in next year.

Which two columns should the data scientists clean during the feature-selection step, and why? (Choose 2.)

Note: You will receive partial credit for each correct selection.

customer_id	age	city	degree	birth_year	salary
1	25	New York City	Master's	1996	90,000
2	32	New York City	Bachelor's	1989	120,000
3	21	New York City	Master's	2000	110,000
4	29	New York City	Doctorate	1992	320,000
5	24	New York City	Bachelor's	1997	95,000

- ☒ age, because this column can be generated using birth_year.
- ☒ city, because this column contains a single value
- ☐ city, because the place of residence does not affect the amount of money invested
- ☐ degree, because this column contains categorical values

22

A hospital is building an AI solution to predict whether a patient is likely to cancel an appointment. The data includes the patient's name, medical conditions, age, date and time of the last appointment, whether they missed their last appointment, and the date and time of the upcoming appointment. The label is Yes (they will miss the appointment) or No (they will not miss the appointment).

Which two operations are likely to occur during feature engineering? (Choose 2.)

- ☐ Hyperparameter tuning the AI algorithm.
- ☒ Encoding Yes and No to numerical values
- ☐ Evaluating whether the patient's privacy requirements will be met during the development of the AI
- ☐ Selecting the optimal classification algorithms for experimentation on this dataset
- ☒ Encoding the medical conditions to numerical values

23

You are trying to create a model to estimate the land area covered by ice in the Arctic Circle. You have 100,000 labeled satellite images of the Arctic Circle. You can use the images to train a machine-learning model to distinguish between ice, water, and land in order to create your land area estimates. How should you allocate images to the training and testing datasets?

- ☐ Select approximately 50% of the images for training and use the rest for testing.
- ☐ Train and test the model on 100% of the labeled images.
- ☒ Select approximately 80% of the images for training and use the rest for testing.
- ☐ Use all the data to train the model and use a subset of the data for testing.

You need to create an easily explainable guidance system for medical technicians.
Which form of machine-learning model should you use?

- ☐ Neural network
- ☐ K-means clustering
- ☐ Support-vector machine
- ☒ Decision tree

25

Which two problems can be solved using a Decision Tree algorithm? (Choose 2.)

- ☐ Text generation
- ☒ Classification
- ☒ Regression
- ☐ Clustering
- ☐ Reinforcement learning

26

Move each dataset split from the list on the left to the AI modeling step in which it is used on the right.

Dataset Splits

AI Modelling Steps

Training dataset

Training and modelling

Test dataset

*Evaluating the model on new
instances*

Validation dataset

Training the hyperparameters

How does a decision tree train a classification model?

- ☒ It uses the features to split the dataset into two or more groups with similar features
- ☐ It calculates the similarities between data points based on their features to split them into two or more groups with similar features
- ☐ It finds a linear relationship between independent and dependent variables to split the dataset into two or more groups with similar features
- ☐ It finds one or more lines to split the dataset into two or more groups with similar features

The chart shown depicts the percentage of a population, by gender and by year, that has a bank account. Which type of visualization is this?



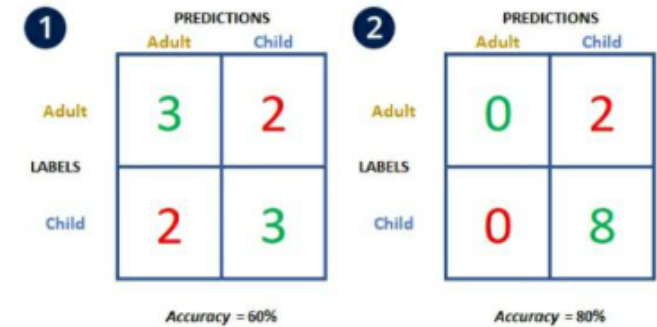
- ☐ Change over time
- ☒ Intersectional
- ☐ Factor
- ☐ Contrast

29

You build two AIs to classify adults and children. The confusion matrix for each AI is shown

Which AI is better, and why?

- ☐ AI 1 because its accuracy is 60%.
- ☐ AI 2 because its accuracy is 80%
- ☐ AI 2 because it has the highest number of correct child predictions.
- ☒ AI 1 because it has an equal number of correct adult and child predictions



Which metric is commonly used to measure the quality of regression AIs?

- ☐ Precision
- ☐ Accuracy
- ☐ Recall
- ☒ Root Mean Squared Error (RMSE)

You need to evaluate the performance of a model that trained on transactional data for fraud detection. Which metric should you use?

- ☒ Confusion matrix
- ☐ Accuracy
- ☐ True positive rate
- ☐ BLEU

What is the difference between accuracy and precision?

- ☒ Accuracy refers to how close measurements are to the "true" value, while precision refers to how close measurements are to each other
- ☐ Accuracy is the ratio of correctly predicted positive observations to the total predicted positive observations, while precision is the ratio of correctly predicted positive observations to the total predicted positive observations.
- ☐ Accuracy is the ratio of correctly predicted observations to the total observations, while precision is the ratio of correctly predicted positive observations to all the observations in the actual class.
- ☐ Accuracy is the weighted average of precision and recall, while precision is the ratio of correctly predicted positive observations to the total predicted positive observations

What can be done to avoid overfitting in decision trees?

- ☐ Ensure the dataset is well skewed
- ☐ Determine the minimum number of leaves
- ☒ Specify a limit for the depth of the tree
- ☐ Remove some of the observations at random from the training dataset.

A school is using an AI model to help with admissions and is training it on 10 years of past admissions data. In order to have need-blind admissions, income data for applicants is excluded from the input. The school noticed that the results, while useful, are still biased toward higher income students. What is a likely explanation?

- ☒ Other features (e.g., postal code, test scores) are correlated with income level.
- ☐ The feature set chosen is too small, leading to overfitting.
- ☐ There is likely an error in the code for the AI model.
- ☐ The training dataset is too small to achieve accurate results.

35

An AI team has implemented a model that looks like it is overfit.

Which three approaches can the team take to solve this problem? (Choose 3.)

- ☒ Tune the hyperparameters
- ☒ Collect more data
- ☐ Use nonlinear models instead
- ☐ Implement a more complicated model.
- ☒ Make sure the data are not unbalanced

Which artifacts can help AI project teams manage regulatory compliance?

- ☒ Data-sharing agreement, incident log, risk register
- ☐ Data model, debugging log, change request
- ☐ Data strategy, issue log, project plan
- ☐ Data quality scorecard, security log, storyboard

You create and test a model. Before deployment, you decide to consult clients for their impressions. Which two tasks are essential in this process? (Choose 2.)

- ☒ Benchmark and communicate potential risks based on results.
- ☐ Compare your results with competitors' results.
- ☐ Increase the scale of the model used
- ☐ Search for improved technologies
- ☒ Meet with clients to evaluate the solution.

Your company builds facial recognition software that will be deployed as part of the security system in an office building. Your job is to create the training plan to help the building security guards understand how to use the software.

What should your training plan contain?

- ☐ A testing method so that the security guards can test the AI for accuracy
- ☒ A review of AI explainability so that the security guards can understand why the AI does what it does
- ☐ An introduction to the latest neural networks used in your software
- ☐ A process for handling complaints that the AI misidentified a person.

A bank builds an AI that predicts whether a customer is likely to default on a loan. The bank intends to use the AI predictions to approve or deny loan applications. Some customers may dispute the loan decisions. The bank wants to make sure that it keeps appropriate records to explain to customers how decisions are made.

Which information should the bank use to mitigate future disputes?

- ☐ The engineer who wrote the prediction code
- ☒ The dataset that was used to train the AI
- ☐ The number of training experiments that were run to select the best algorithm
- ☐ The number of predictions the AI has made in production

Which term is commonly used to describe AI predictions?

- ☐ K-fold cross-validation
- ☐ Hyperparameter tuning
- ☐ Confusion matrix
- ☒ Inference

An organization relies on its continuous pipeline to keep its ever-increasing number of machine-learned models up to date.

Which two types of algorithms are deployed in production AI operations? (Choose 2.)

- ☐ Incident ticketing
- ☐ Feature repository
- ☐ Immutable infrastructure patterns
- ☒ AI health monitoring
- ☒ Inference scaling

Your company's goal is to run models that are highly relevant and that perform optimally at any point in time. You must determine how often the model should be retrained and are considering three retraining strategies: periodic, performance-based, or based on data changes.

What important factor, besides availability of feedback and volume of data, should you use in your assessment?

- ☐ Resource cost
- ☐ Ease of implementation
- ☒ Visibility to performance of the model in production
- ☐ Periodicity

43

You work for an insurance company. Your job is to help the data scientists and engineers prepare to put a new model into production. This model will be used to assess whether to approve client claims. You need to ensure that your company follows the best practices for the industry and location. Which two items should you pay attention to? (Choose 2.)

- ☒ Whether the process of building and evaluating the AI has been well documented
- ☐ Whether only public data was used to develop the AI model
- ☒ Whether there are any laws to be followed with respect to validating the model and getting it approved
- ☐ Whether the engineers built the AI in the cloud where there is a lot of computing power to be used
- ☐ Whether the most recent AI technology on the market has been used.

You have put into production an AI that recommends products to website users. The AI has been in production for over a year and a new model update was added last week based on a new algorithm that the data scientists have discovered.

How can you assess whether this new algorithm is working better for your business than the old one?

- ☐ Review the data scientists' experiments and make sure that they have measured both algorithms on the same test cases.
- ☒ Over a period of three months, compare the sales of recommended products for the new algorithm relative to the old.
- ☐ Compare the sales on the busiest shopping day of the year when the old algorithm was used to the sales of the first day of the new algorithm.
- ☐ Review the code of both algorithms and see if one has more bugs than the other.

45

Your new AI system has been test piloted, and you can now start gathering metrics in the field. The results appear to be as you expected, and you hope to fully deploy the system soon.

What is an important and appropriate ethical step to take at this stage?

- ☐ Review new technologies to determine whether the system can be upgraded to have new capabilities.
- ☐ Accept pre-orders for the new system from early adopter customers, based on the pilot metrics.
- ☒ Determine whether the technology might disproportionately affect any group
- ☐ Reserve an appropriate number of server instances for imminent intense usage.

46

You develop an AI that recommends songs to users. Users can provide feedback about the recommendations by selecting a "thumbs up" or "thumbs down" button.

Which three benefits should you expect to gain by receiving this user feedback? (Choose 3.)

- ☐ Reduced model size.
- ☒ Effective A/B testing of different versions of the recommendation algorithm.
- ☒ Opportunities to profile users more easily.
- ☐ Insights on the latest technology developments
- ☒ Insights on user satisfaction with the system.

A grocery store website uses AI to recommend products for customers to buy. The AI has been in use for several months.

Which factor should the store consider when deciding whether to keep or discontinue the AI?

- ☐ Whether the AI used an explainable algorithm.
- ☐ Whether the original accuracy of the AI was above 85%.
- ☐ Whether the AI used a supervised or unsupervised algorithm.
- ☒ Whether customer purchases increased based on the AI's recommendations.

You have the following dataset.

You need to identify the encoding method of the categorical degree feature of the dataset used in each of the following tables.

person_id	age	degree	salary
1	25	Master's	120000
2	32	Bachelor's	200000
3	21	Bachelor's	60000
4	29	Doctorate	250000
5	24	Master's	90000

Table A

person_id	degree
1	2
2	1
3	1
4	3
5	2

Table B

person_id	degree_1	degree_2	degree_3
1	0	1	0
2	1	0	0
3	1	0	0
4	0	0	1
5	0	1	0

Table C

person_id	degree_1	degree_2
1	1	0
2	0	1
3	0	1
4	1	1
5	1	0

Which table uses one-hot encoding?

Table B

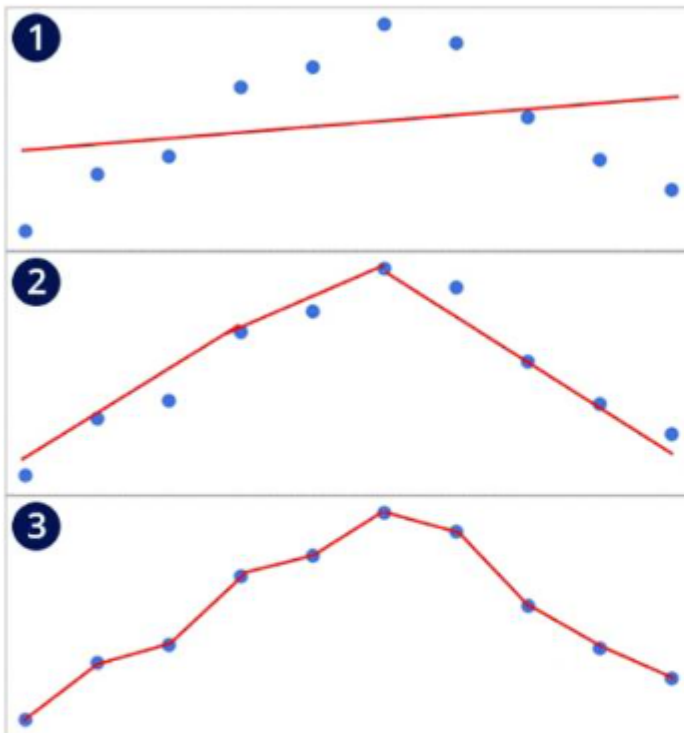
Which table uses dummy encoding?

Table C

Which table uses label encoding?

Table A

The three models shown are based on the same dataset.
Review the models and answer the following questions



Which model is optimal?

Model 2

Which model is overfit?

Model 3

Which model is underfit?

Model 1

50

A company wants to streamline several tasks by using automation. Which three tasks are appropriate for an AI solution? (Choose 3.)

Note: You will receive partial credit for each correct

- ☒ Understanding documents
- ☐ Logging into applications
- ☒ Discovering tasks and processes to automate
- ☒ Copying and pasting data

Which two projects would benefit from the use of AI modeling? (Choose 2.)

- ☐ Determining how many recipients of promotional emails check a company's website
- ☐ Checking whether a user's password matches the password set by the account owner
- ☒ Determining the opinion of people on a subject in social media from their comments, hashtags, and messages
- ☒ Selecting bank customers to receive credit card offers based on their behavior on the bank's website

A museum has collected 15 years of detailed visitor information, including timestamped entry and exit information. The museum director wants you to develop an AI model that uses this data to determine the number of museum visitors at a specific time on a specific date in the past.

What advice should you give the director?

- ☒ This application does not need AI because you can directly query the data to find the result
- ☐ This application is best suited for a regression model because the output is numerical
- ☐ This application is best suited for supervised learning because the input data is labeled
- ☐ This application is best suited for an unsupervised learning model because the data is unlabeled

A company wants to improve customer satisfaction by knowing its customers better. As a first step, the company wants to group customers based on similarities in their profiles and shopping histories.

Which type of machine learning problem is this?

- ☐ Reinforcement learning
- ☐ Supervised learning
- ☒ Unsupervised learning
- ☐ Meta learning

Which problem can be addressed through the use of regression?

- ☐ Detecting whether a product is faulty on an assembly line by analyzing a picture of the product
- ☐ Providing a user with a rank ordered list of recommendations of movies to watch
- ☒ Deciding the asking price for a house based on the characteristics of the house and its location.
- ☐ Determining whether someone has a disease from an MRI scan of their brain

Which AI problem is a classification problem?

- ☒ Determining whether an email sent to customer support has positive or negative sentiment
- ☐ Proposing a target price for selling a house based on similar homes nearby
- ☐ Determining the best strategy to win in a game of chess
- ☐ Forecasting the price of a stock when the market opens tomorrow

You are tasked with using AI to predict annual fish populations in the local river, based on environmental and climate features such as temperature, date of first snowfall, and amount of precipitation

Why should you use regression modeling?

- ☐ The input can be represented using feature vectors
- ☐ The output is a discrete label
- ☐ The input is a set of quantitative values
- ☒ The output is a quantitative value.

When should you purchase an AI solution instead of developing it yourself?

- ☒ You verified that available solutions are proven to be effective for your organization's purpose
- ☐ Your long-term plan includes building a world-class technical team that can quickly shift technology strategies as your business needs change and industry evolves
- ☐ Your priority is gaining efficiency from a program specifically designed with your needs in mind
- ☐ You want to garner a competitive advantage relative to your competitors

A bank plans to implement an AI that will decide which customers qualify for a special low-interest mortgage loan. You are in charge of planning how to bring the AI to production and make it customer-ready.

You need to be able to demonstrate that the AI system implementation is transparent.

Which activities will be critical?

- ☐ Ensuring that the bank issues a press release stating that AI is being used for this purpose
- ☐ Publishing all the source code to a shared repository
- ☒ Tracking all models used by the AI in production
- ☐ Ensuring that the project manager knows the details of what every engineer is working on

Which term refers to the sharing of data among private companies, research institutions, and government agencies to help solve public problems?

- ☐ Open data
- ☒ Data collaboratives
- ☐ Public data
- ☐ Data exchange

A company wants to check consumers' opinions about a new product and perform a sentiment analysis on a social media platform. The potential number of consumers is quite large.

How can the company collect the data in a structured manner?

- ☐ Access the physical storage of the social media platform's data
- ☐ Ask customers to provide their usernames, then crawl their timeline
- ☐ Ask customers to provide their usernames, then copy their timeline manually
- ☒ Use the social media platform's API

A large multinational corporation wants to use AI to screen resumes from job applicants for those who are most likely to perform well. They have given you ten years of employee performance reviews to use as training data.

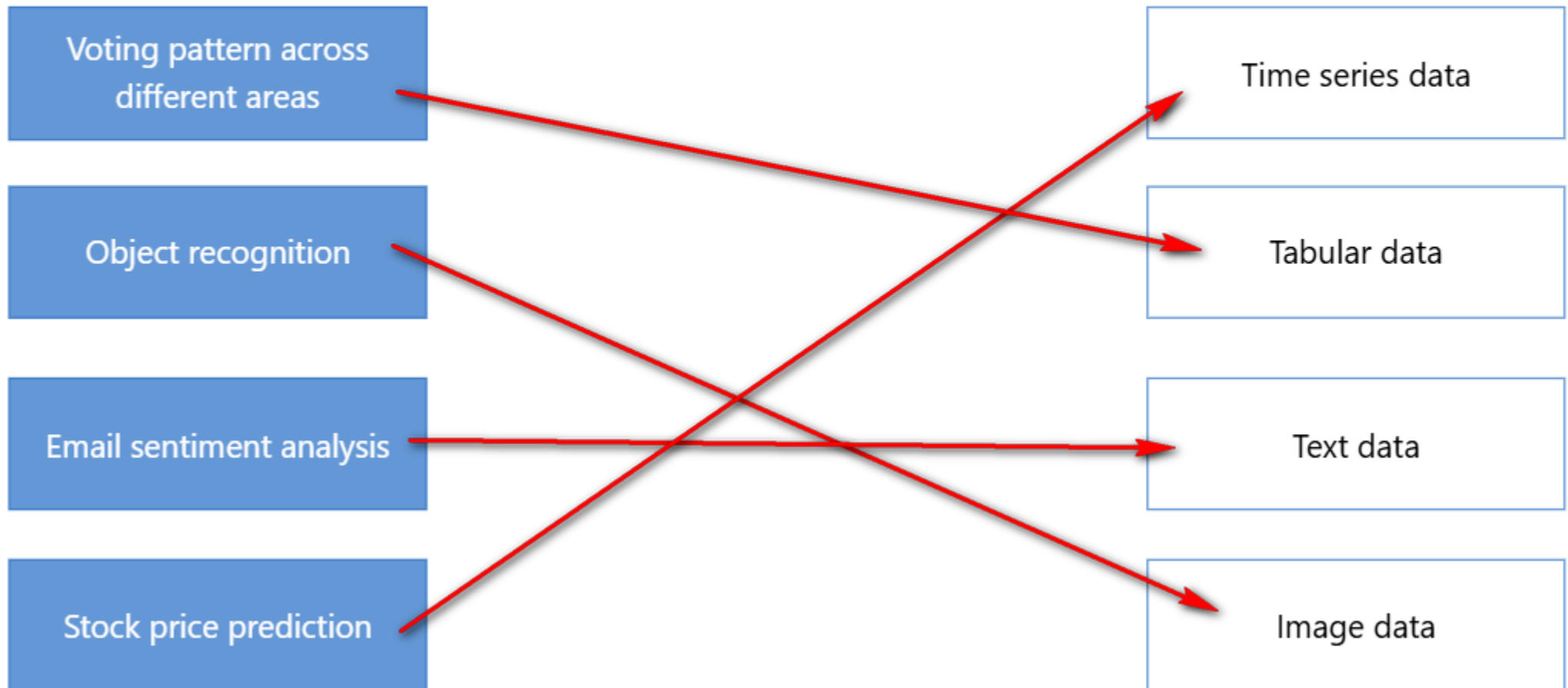
What concern should you raise about this data?

- ☐ The dataset may have missing information that will impact the performance of the model
- ☐ Useful features will be impossible to extract from employee performance data
- ☒ Past biases in hiring will be perpetuated using this dataset
- ☐ This dataset is unlikely to contain enough data for effective training

Move each AI problem from the list on the left to the most appropriate data type on the right

AI Problems

Data types



You need to make a text dataset understandable to a machine learning model.
Which technique should you use?

- ☒ Assign a unique numerical value to each unique word in the dataset
- ☐ Assign a unique categorical value to each unique word in the dataset
- ☐ Assign multiple categorical values to each unique word in the dataset
- ☐ Assign multiple numerical values to each unique word in the dataset

You are building a system to help a gym decide which customers are most likely to cancel their memberships. You have over 100 features in the customer data.

How can you effectively select good features for your AI?

- ☐ Use as many features as possible that relate to the customers' personal characteristics, such as gender and race
- ☐ Randomly select a group of 10 features to use
- ☒ Speak with customer representatives and find out what previous customers have said about cancellations
- ☐ Choose only numerical features

Which type of feature should be removed in the feature-selection step in order to increase accuracy?

- ☐ One that is highly correlated with the target
- ☐ One that has multiple values
- ☐ One that is highly correlated with the target
- ☒ One that has no correlation with the target

66

You are building a regression AI to predict the price of a vehicle. You will be using one-hot encoding in your feature engineering.

To which type of feature will you need to apply one-hot encoding?

- ☐ The number of doors in the vehicle
- ☒ The type of vehicle (truck, sedan, or van)
- ☐ The amount of time to go from 0 to 100 kilometers per hour
- ☐ The fuel efficiency of the car on regular streets (not freeways/highways)

Which technique can help to reduce an AI model's input dataset's dimensions?

- ☐ Undersampling
- ☐ Removing features with missing values
- ☐ One-hot encoding
- ☒ Principal Component Analysis

You are building an AI model.

Which approach should you take for the initial dataset?

- ☐ 80% testing data, 100% training data, keeping the testing data constant
- ☐ 20% testing data, 100% training data, keeping the testing data constant
- ☐ 80% testing data, 20% training data, keeping the testing data constant
- ☒ 20% testing data, 80% training data, keeping the testing data constant

A company makes loan-approval decisions. A customer is disputing the conclusion of the AI. Which type of documentation is required for auditability?

- ☐ Detailed properties of data elements, i.e., data type, size, nullability, optionality, and indexes
- ☐ Diagram describing data flow, inputs and outputs, storage points, and sub-processes
- ☐ Database schema representing how the data is organized and relations among them are associated
- ☒ Methods of data collection and consent, chain of responsibility, and quality and accuracy assessments

Which three algorithms can be used to solve a classification problem? (Choose 3.)

- ☒ Logistic regression
- ☐ Principal Component Analysis
- ☐ K-means
- ☒ Support Vector Machine (SVM)
- ☐ Linear regression
- ☒ Decision tree

What is the difference between linear and logistic regression?

- ☐ Using a given set of independent variables, linear regression predicts the categorical dependent variable. Logistic regression predicts the continuous dependent variable
- ☐ Linear regression assumes that the relationship between the independent and dependent variables is linear. In logistic regression the assumption is that the relationship is nonlinear
- ☒ Linear regression is used for regression problems. Logistic regression is used for classification problems
- ☐ In linear regression, we predict the values of categorical variables. In logistic regression we predict the values of continuous variables

You want to tell a data story that investigates the relationship between access to bank account balance information and having enough money to cover basic daily necessities.

Which requirements apply to your dataset?

- ☐ Layered data with minor variation
- ☐ Data for one customer over the past ten years
- ☐ Granular level of data
- ☒ One dataset with lots of variables, or more than one similar dataset

What does underfitting mean for a model?

- ☐ There is high variance and low bias
- ☒ There is low variance and high bias
- ☐ There is too little data in the test set
- ☐ There is high variance and low bias

You trained a model to predict customer churn for an online bank. However, evaluations show that the model has a poor performance.

Which two methods can improve the performance of the model? (Choose 2.)

Note: You will receive partial credit for each correct selection

- ☒ Fine-tune the hyperparameters.
- ☐ Use a faster programming language
- ☐ Select features that vary less
- ☒ Revise selected features
- ☐ Minimize the initial dataset

You are building an AI model that can recognize different species of birds from pictures taken by amateur birders. You've just trained your model and tested by making predictions for a set of images previously unseen by the model.

In order to compute the precision of your model, what calculation do you need to do?

- ☐ Compute the ratio of correct predictions to the incorrect predictions made
- ☐ Subtract the number of incorrect predictions from the number of correct predictions
- ☐ Divide the number of correct predictions by the size of your training set
- ☒ Compute the ratio of correct predictions to total predictions made

An AI team in a banking organization has implemented a model to predict whether a transaction is fraudulent. They get 99.7% accuracy for the training set, but 72.45% accuracy for the test set.

What explanation fits the information provided?

- ☐ The model is performing well enough on both training and test sets, and it is not a problem that the test set is resulting in lower accuracy.
- ☐ The model is low bias because it is not performing well on the test set
- ☒ The model is overfit because it is not performing well on test set
- ☐ The model is underfit because it is not performing well on test set

Regulators and users increasingly demand transparency with regard to how AI models make decisions. Which two types of documentation help developers provide assurance of transparency for their models? (Choose 2.)

Note: You will receive partial credit for each correct selection

- ☒ Explainable features with all hyperparameter configurations and settings.
- ☒ Method and means by which models will be versioned
- ☐ Specific filtering approaches, methods, or criteria for data inclusion or exclusion defined for all externally sourced data
- ☐ Specific filtering approaches, methods, or criteria for data inclusion or exclusion defined for all externally sourced data
- ☐ Identification of a minimum of 80% of datasets used for training the model

You are the project manager for a team that is developing an AI model. The AI will recommend products to customers. You are required to get internal stakeholder approval before you can launch the AI to customers. This is one of one hundred AI projects in your company. Who is likely to be a internal stakeholder?

- ☐ An engineer who built a similar AI in open source
- ☐ A Human Resources specialist from your company
- ☐ Customers whose data was used to train the model
- ☒ The production machine-learning engineering team

A bank has built an AI to predict which customers are at risk of closing their bank accounts and moving to another bank. Before the bank puts this AI into use, it needs to train the bank employees who will use the AI on how best to use it.

What should that training include?

- ☐ The algorithm that was used in the AI's prediction and what metrics were used to evaluate the AI
- ☒ The types of predictions that the AI is able to make and the information it uses to make those predictions
- ☐ What programming language the AI prediction pipeline was built on
- ☐ How much historical data was used to train the AI model and how the model was tuned

Your company builds an AI technology that detects whether lettuce is diseased. The AI uses a very advanced neural network to analyze photos of the lettuce leaves. The AI worked very well in the experiments.

You need to determine scenarios in which the production AI might wrongly detect the presence of disease.

Which two scenarios might cause incorrect diagnosis of disease? (Choose 2.)

Note: You will receive partial credit for each correct answer

- ☐ Lettuce farmers might distrust the results of the AI and refuse to use it.
- ☒ Lettuce might look different in the field than in the training photos
- ☒ Users might try to use the AI to predict diseases in other types of leafy green vegetables
- ☐ Researchers might come up with a better way to build neural networks for images like the training photos
- ☐ A competitor might come up with a way to detect the same diseases and claim that their solution is better than yours

81

A turbine manufacturer has designed an AI to predict whether a turbine is likely to catch fire. The time from a turbine displaying any indication of failure to the turbine catching fire is less than 100 milliseconds. The AI must make a prediction within 15 milliseconds so that the turbine can be turned off before it catches fire. Each turbine design has a custom AI model that is retrained once a year.<

Which two deployment approaches would be responsive? (Choose 2.)

Note: You will receive partial credit for each correct selection

- ☒ Run the prediction program on a computer next to the turbine and have the turbine contact the computer
- ☐ Gather up all of the turbines' data and do a health check over the internet once a day
- ☐ Run a prediction program over the internet and have the turbine get its predictions via Wi-Fi
- ☒ Integrate the model directly into the turbine's software
- ☐ Gather up all of the turbines' data and do a health check in the factory once a day

Your company builds an AI to recommend products on a website. You are in charge of creating the production pipeline and integrating it with the website.

What should be part of your considerations?

- ☐ Whether the latest AI algorithms were used to make the recommendation
- ☒ Whether the predictions will be fast enough if the number of website customers increases tenfold
- ☐ Whether the predictions are good enough and customers buy the recommended products.
- ☐ Whether enough customer data was used to train the AI algorithm

You have been tracking the metrics of your AI system and everything seems to be working as expected. Suddenly, you start getting very unexpected results that significantly differ from earlier ones. Which two are plausible causes to check for first? (Choose 2.)

- ☒ A software upgrade on users' devices
- ☐ Users running out of space on their devices
- ☐ The sensors on user's devices getting dirty and degraded
- ☐ Hardware malfunctions
- ☒ A change in user behavior or the types of users accessing a system.

You inherit responsibility for a machine vision system. While analyzing it for biases, you find that it does not respond effectively to people with darker skin.

What are two appropriate safety and ethics steps to take in this instance? (Choose 2.)

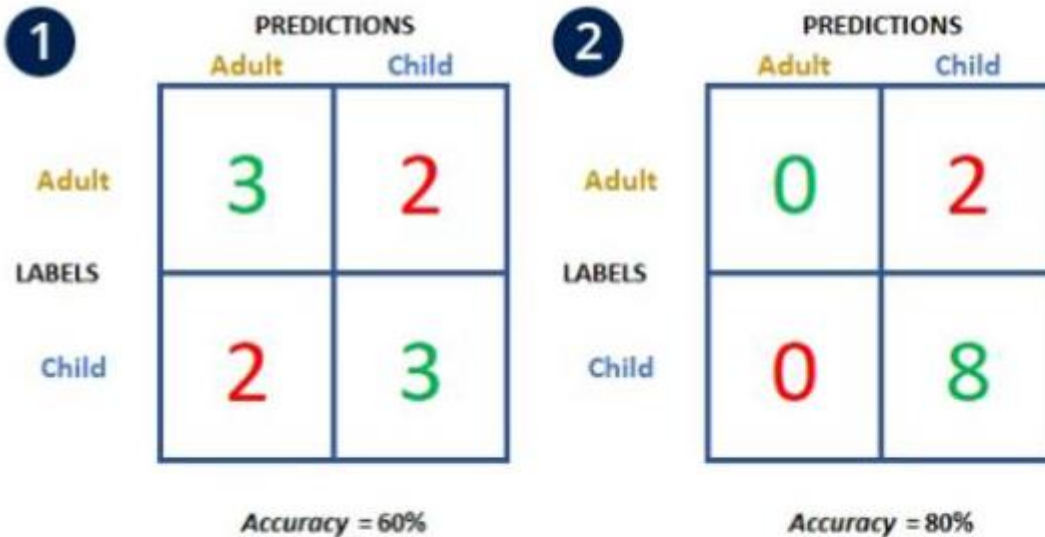
- ☐ Fine tune marketing for the system toward locations where a smaller percentage of people tend to have darker skin
- ☐ Upgrade the version of the software used to implement the algorithm
- ☒ Seek to add new training data of classification methods to improve performance
- ☐ Degrade the performance for all skin types to match performance upon darker skin
- ☒ Take the system offline as soon as possible

An online store uses an AI system that recommends to online shoppers products to add to their orders. Which two measures of customer behavior can you use to evaluate efficacy of the AI ? (Choose 2.)

Note: You will receive partial credit for each correct selection

- ☒ Did the customer purchase one of the recommended products?
- ☐ Does the AI report better accuracy metrics after customers started using it?
- ☒ Did the customer follow a link to one of the recommended products?
- ☐ Does the AI successfully make predictions for every customer?
- ☐ Did more customers visit the website after the AI was put into production?

You build two AIs to classify adults and children. The confusion matrix for each AI is shown.



Which AI is better, and why?

- ☐ AI 1 because its accuracy is 60%.
- ☐ AI 2 because its accuracy is 80%
- ☒ AI 1 because it has an equal number of correct adult and child predictions.
- ☐ AI 2 because it has the highest number of correct child predictions.

Which application is implemented without the use of AI?

- ☐ Email spam filter
- ☐ Email autocomplete
- ☐ Predictive browser search bar
- ☒ Messaging platform word-to-emoji converter

What is an issue that an adversary or malicious actor can exploit to attack or undermine an AI system?

- ☐ Unsupervised learning
- ☒ Data extraction
- ☐ Overfitting
- ☐ Data skew

You put into production an AI application that is trained to distinguish between cats, dogs, and rabbits. Which scenario is an example of model drift in production?

- ☐ You later add a set of images of monkeys, and retrain the AI application to detect all four types of animals
- ☐ You retrain the AI with more images of cats, dogs, and rabbits in different lighting conditions
- ☒ The model is frequently asked to distinguish between tigers and monkeys
- ☐ The model is only used to distinguish between cats and rabbits.

Your organization is developing an AI system for a specific purpose. However, the same system could easily be used for other purposes also Why might a system having multiple potential uses lead to an ethical concern?

- ☐ It could open up new market opportunities
- ☒ It might be used for an unauthorized purpose
- ☐ Another organization might want to buy the AI system.
- ☐ It might be more likely to break down.