DataScience Specialization Test Series - Test 3 - Batch A & B - 14th October 2023

Total points 44/60

The respondent's email (**rajendraku0000@gmail.com**) was recorded on submission of this form.

0 of 0 points

Name *

rajendraku0000@gmail.com

DataScience Specialization Test Series - Test 3 - Batch A & B - 14th October 2023 **44 of 60 points**

✗ What is the primary purpose of a validation set in the machine learning workflow? 0/1

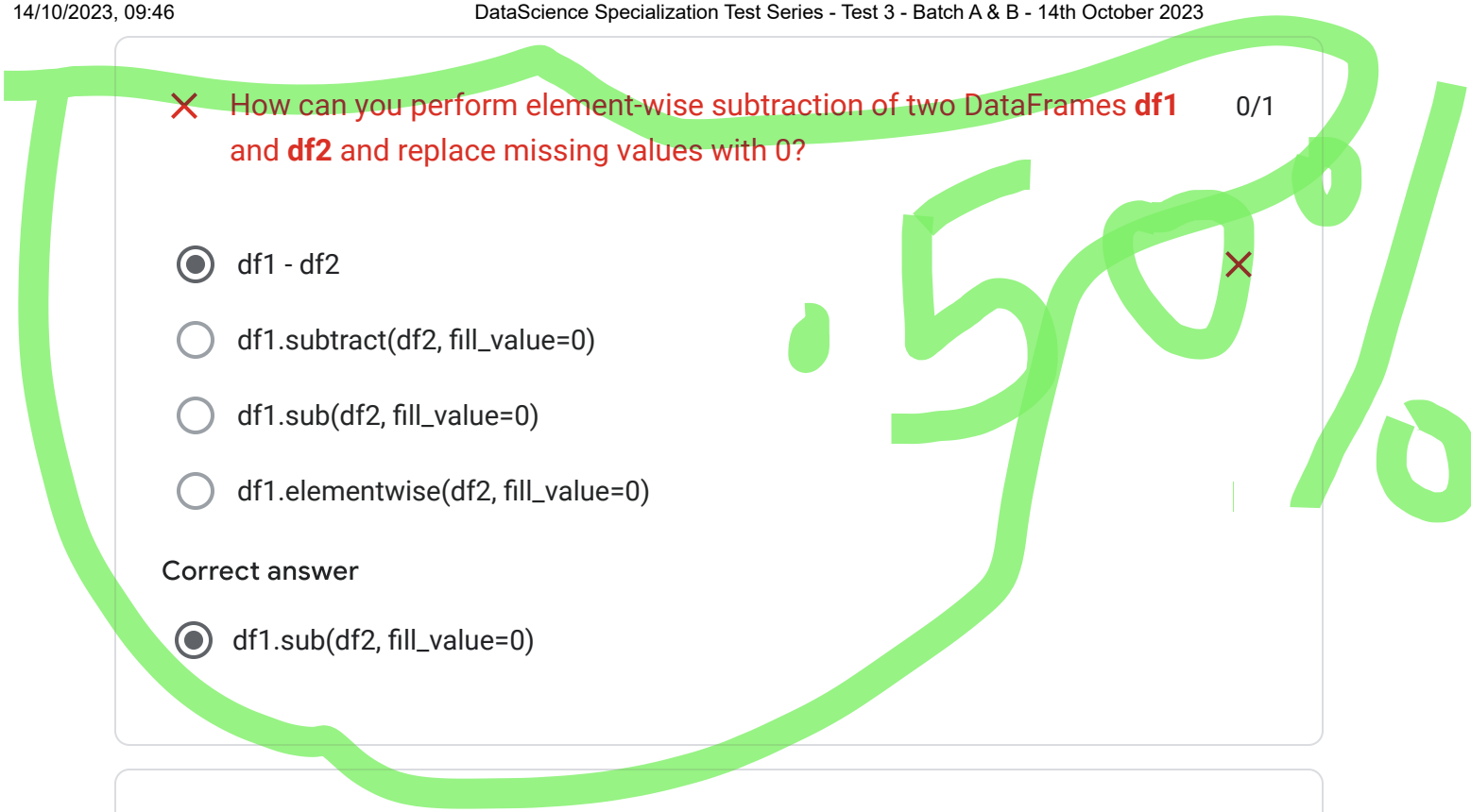
- ○ To train the model
- ○ To fine-tune hyperparameters
- ◉ To evaluate the model's performance on unseen data ✗
- ○ To test the model's generalization

Correct answer

- ◉ To fine-tune hyperparameters

✓ What does the **drop_duplicates()** function in Pandas do? 1/1

- ○ Drops rows with missing values from a DataFrame.
- ◉ Drops duplicate rows from a DataFrame. ✓
- ○ Drops columns with missing values from a DataFrame.
- ○ Drops columns with duplicate values from a DataFrame.



✗ How can you perform element-wise subtraction of two DataFrames **df1** and **df2** and replace missing values with 0? 0/1

- (●) df1 - df2 ✗
- ( ) df1.subtract(df2, fill_value=0)
- ( ) df1.sub(df2, fill_value=0)
- ( ) df1.elementwise(df2, fill_value=0)

Correct answer

- (●) df1.sub(df2, fill_value=0)

✓ How can you replace all occurrences of a specific value (e.g., 'old_value') in a DataFrame **df** with a new value (e.g., 'new_value')? 1/1

- (●) df.replace('old_value', 'new_value') ✓
- ( ) df.replace('old_value', 'new_value', inplace=True)
- ( ) df.replace_values({'old_value': 'new_value'})
- ( ) df.replace_values('old_value', 'new_value')

✓ How can you rename the column 'old_name' to 'new_name' in a DataFrame **df**? 1/1

- ○ df.rename_column('old_name', 'new_name')
- ◉ df.rename(columns={'old_name': 'new_name'}) ✓
- ○ df.column_rename('old_name', 'new_name')
- ○ df.rename_column({'old_name': 'new_name'})

✓ In the context of deep learning, what does the term "backpropagation" refer to? 1/1

- ◉ The process of training a model by updating its parameters ✓
- ○ The process of preprocessing data before training a model
- ○ The process of evaluating a model's performance on a validation set
- ○ The process of selecting hyperparameters for a model

✓ What is the primary purpose of data augmentation in machine learning? 1/1

- ○ To reduce model complexity
- ○ To add noise to the data
- ○ To increase the dimensionality of the data
- ◉ To generate additional training data by applying transformations ✓

✓ In deep learning, what is the role of the activation function in a neural network? 1/1

- ○ To initialize the network weights
- ○ To determine the learning rate of the network
- ◉ To introduce non-linearity to the model ✓
- ○ To define the network architecture

✓ What does the **to_datetime()** function in Pandas do? 1/1

- ◉ Converts a Series to a datetime data type. ✓
- ○ Converts a Series to a timedelta data type.
- ○ Converts a Series to a string representation of the datetime.
- ○ Converts a datetime to a string representation.

✗ What is the primary difference between bagging and boosting in ensemble learning? 0/1

- ○ Bagging combines multiple weak learners into a strong learner, while boosting uses a single strong learner.
- ◉ Bagging focuses on reducing model variance, while boosting focuses on reducing model bias. ✗
- ○ Bagging uses only decision trees, while boosting uses various machine learning algorithms.
- ○ Bagging is a form of data augmentation, while boosting is a form of feature engineering

Correct answer

- ◉ Bagging combines multiple weak learners into a strong learner, while boosting uses a single strong learner.

✓ Which deep learning architecture is used for generating new data samples, such as images or text? 1/1

- ○ CNN
- ○ RNN
- ◉ GAN ✓
- ○ Autoencoder

✗ Which technique is used to combat the problem of class imbalance in deep learning? 0/1

- ○ Data augmentation
- ◉ Batch normalization ✗
- ○ Regularization
- ○ Gradient clipping

Correct answer

- ◉ Data augmentation

✓ What is a hyperparameter in the context of deep learning? 1/1

- ○ A parameter that is learned during training
- ○ A weight value in a neural network
- ◉ A configuration setting that is set before training and affects the model's performance ✓
- ○ A bias term in a neural network

✓ What is the primary goal of feature selection in machine learning? 1/1

- ○ To increase model complexity
- ○ To add noise to the data
- ○ To create new features from existing data
- ◉ To choose the most relevant features for a model ✓

✗ How can you handle missing values in a DataFrame **df** by filling them with the mean value of the column? 0/1

- ○ df.fillna(df.mean())
- ○ df.fillna(df.mean(), axis=0)
- ○ df.fillna(df.mean(), axis=1)
- ◉ df.fillna(df.mean(), inplace=True)

Correct answer

- ◉ df.fillna(df.mean())

✓ What is the main objective of semi-supervised learning in machine learning? 1/1

- ○ To reduce the dimensionality of data
- ○ To address class imbalance in the training data
- ◉ To improve model performance by using both labeled and unlabeled data ✓
- ○ To build ensemble models

✓ How can you calculate the element-wise product of two DataFrames **df1** and **df2**? 1/1

- (●) df1 * df2 ✓
- ( ) df1.mul(df2)
- ( ) df1.product(df2)
- ( ) df1.elementwise(df2)

✓ In machine learning, what is the primary goal of data preprocessing? 1/1

- ( ) To build the model
- ( ) To evaluate the model's performance
- (●) To transform and clean the data to make it suitable for modeling ✓
- ( ) To validate the model's predictions

✗ Which machine learning technique is suitable for solving problems involving a sequence of decisions, such as game playing or robot control? 0/1

- ( ) Reinforcement Learning
- (●) Unsupervised Learning ✗
- ( ) K-Means Clustering
- ( ) Naive Bayes

Correct answer

- (●) Reinforcement Learning

✗ Which machine learning algorithm is best suited for outlier detection in a dataset? 0/1

- ○ Support Vector Machine (SVM)
- ○ Random Forest
- ○ Principal Component Analysis (PCA)
- ◉ k-Nearest Neighbors (KNN) ✗

Correct answer

- ◉ Support Vector Machine (SVM)

✓ How can you convert a DataFrame **df** to a CSV file named 'output.csv'? 1/1

- ○ df.write_csv('output.csv')
- ○ df.save_csv('output.csv')
- ◉ df.to_csv('output.csv') ✓
- ○ df.export_csv('output.csv')

✓ What is the primary goal of outlier detection in machine learning? 1/1

- ○ To identify patterns and clusters in data
- ○ To uncover hidden associations in data
- ◉ To detect and remove data points that deviate significantly from the norm ✓
- ○ To classify data into predefined categories

✓ Which machine learning algorithm is commonly used for recommendation systems, such as those used by Netflix or Amazon to suggest movies or products? 1/1

- ○ Naive Bayes
- ○ K-Means Clustering
- ◉ Collaborative Filtering ✓
- ○ Principal Component Analysis (PCA)

✗ Which machine learning technique is often used for imbalanced classification tasks, where one class has significantly fewer examples than the other? 0/1

- ◉ K-Means Clustering ✗
- ○ Random Forest
- ○ Logistic Regression
- ○ Synthetic Minority Over-sampling Technique (SMOTE)

Correct answer

- ◉ Synthetic Minority Over-sampling Technique (SMOTE)

✗ Which type of neural network architecture is used for sequence-to-sequence tasks such as machine translation? 0/1

- ○ Convolutional Neural Network (CNN)
- ○ Recurrent Neural Network (RNN)
- ◉ Long Short-Term Memory (LSTM) ✗
- ○ Autoencoder

Correct answer

- ◉ Recurrent Neural Network (RNN)

✓ What is the primary purpose of the backpropagation algorithm in deep learning? 1/1

- ○ To compute the forward pass of a neural network
- ○ To initialize the weights of the network
- ◉ To update the model's weights based on the computed gradients ✓
- ○ To determine the optimal learning rate

✗ How can you select rows from a DataFrame **df** where the values in column 'column_name' are not null? 0/1

- ○ df[df.column_name.notnull()]
- ○ df[df.column_name != null]
- ○ df[df.column_name.notna()]
- ◉ All of the above ✗

Correct answer

- ◉ df[df.column_name.notna()]

✓ In deep learning, what is the purpose of a loss function? 1/1

- ○ To regularize the model
- ○ To initialize the weights of the neural network
- ◉ To measure the error between the predicted and actual values ✓
- ○ To prevent overfitting

✗ What does the **resample()** function in Pandas do? 0/1

- ○ Aggregates data over a specific time period.
- ○ Fills missing values in a DataFrame.
- ◉ Reindexes a DataFrame with a new time frequency. ✗
- ○ None of the above.

Correct answer

- ◉ Aggregates data over a specific time period.

✓ How can you extract the unique values from a specific column 'column_name' in a DataFrame **df**? 1/1

- ○ df.unique('column_name')
- ◉ df['column_name'].unique() ✓
- ○ df.get_unique('column_name')
- ○ df.column_name.unique()

✓ In machine learning, what does the term "ensemble method" refer to? 1/1

- ○ A method for preprocessing data before training
- ○ A method for dimensionality reduction
- ◉ A technique that combines the predictions of multiple models to improve overall performance ✓
- ○ A type of feature extraction technique

✓ What is the primary goal of hyperparameter tuning in machine learning? 1/1

- ○ To eliminate bias in the model
- ○ To minimize overfitting
- ◉ To identify the best set of model parameters and settings for optimal performance ✓
- ○ To reduce the dimensionality of the data

✗ Which technique is used for reducing the dimensionality of data in deep learning? 0/1

- ○ Principal Component Analysis (PCA)
- ○ Batch normalization
- ◉ Max-pooling ✗
- ○ One-hot encoding

Correct answer

- ◉ Principal Component Analysis (PCA)

✗ Which machine learning algorithm is often used for anomaly detection in time series data, such as identifying unusual patterns in network traffic or fraud detection? 0/1

- (●) Decision Trees ✗
- ( ) Support Vector Machine (SVM)
- ( ) Hidden Markov Models (HMM)
- ( ) Logistic Regression

Correct answer

- (●) Hidden Markov Models (HMM)

✓ Which evaluation metric is commonly used for regression problems to assess the goodness of fit of a model's predictions to the actual values? 1/1

- ( ) F1 Score
- (●) Mean Absolute Error (MAE) ✓
- ( ) Precision
- ( ) Recall

✓ What is the primary difference between supervised learning and unsupervised learning in deep learning? 1/1

- (●) Supervised learning uses labeled data for training, while unsupervised learning uses unlabeled data. ✓
- ( ) Supervised learning is used for regression tasks, while unsupervised learning is used for classification tasks.
- ( ) Supervised learning has a higher degree of model interpretability compared to unsupervised learning.
- ( ) Unsupervised learning uses reinforcement learning algorithms for training

✓ Which technique can be used to address the curse of dimensionality in high-dimensional datasets? 1/1

- ( ) Adding more features
- ( ) Feature scaling
- (●) Dimensionality reduction ✓
- ( ) Imputation

✓ How can you create a new column 'new_column' in a DataFrame **df** that contains the sum of two existing columns 'col1' and 'col2'? 1/1

- (●) df['new_column'] = df['col1'] + df['col2'] ✓
- ( ) df.add_column('new_column', df['col1'] + df['col2'])
- ( ) df.new_column = df.col1 + df.col2
- ( ) df.set_column('new_column', df.col1 + df.col2)

✓ Which machine learning algorithm is commonly used for sentiment analysis and text classification tasks, particularly when dealing with text data? 1/1

- ○ K-Means Clustering
- ○ Decision Trees
- ◉ Naive Bayes ✓
- ○ Random Forest

✓ Which machine learning technique is used for solving sequential decision-making problems with a Markovian property, such as autonomous driving or game playing? 1/1

- ◉ Reinforcement Learning ✓
- ○ Unsupervised Learning
- ○ Principal Component Analysis (PCA)
- ○ Genetic Algorithms

✓ What is the primary objective of Natural Language Processing (NLP) in machine learning? 1/1

- ○ To identify patterns in numerical data
- ◉ To understand and process human language data ✓
- ○ To perform clustering of text documents
- ○ To predict stock market prices

✓ How can you create a new column 'new_column' in a DataFrame **df** that contains the row-wise mean of all columns? 1/1

- ( ) df.new_column = df.mean(axis=1)
- (•) df['new_column'] = df.mean(axis=1) ✓
- ( ) df.insert('new_column', df.mean(axis=1))
- ( ) df['new_column'] = df.sum(axis=1) / len(df.columns)

✓ How can you calculate the percentage change of a DataFrame **df** along the columns? 1/1

- (•) df.pct_change() ✓
- ( ) df.calculate_change()
- ( ) df.percent_change()
- ( ) df.change()

✓ What is the primary purpose of the Expectation-Maximization (EM) algorithm in machine learning? 1/1

- (•) To estimate the parameters of a statistical model when some data is missing or hidden ✓
- ( ) To maximize the accuracy of a model on the training data
- ( ) To optimize the hyperparameters of a neural network
- ( ) To find the global minimum of a non-convex loss function

✗ What does the **merge_asof()** function in Pandas do? 0/1

- ○ Merges two DataFrames based on the closest match to a key column.
- ◉ Merges two DataFrames based on a common column. ✗
- ○ Merges two DataFrames based on a specified time interval.
- ○ None of the above.

Correct answer

- ◉ Merges two DataFrames based on the closest match to a key column.

✓ What is the purpose of transfer learning in deep learning? 1/1

- ○ Learning to transfer data between different devices
- ◉ Using pre-trained models to improve performance on a new task ✓
- ○ Transferring data between the training and validation sets
- ○ Transferring gradients between layers in a neural network

✓ What is the purpose of Principal Component Analysis (PCA) in dimensionality reduction? 1/1

- ○ To remove noisy data points from the dataset
- ◉ To eliminate irrelevant features and reduce data dimensionality ✓
- ○ To increase the dimensionality of the data
- ○ To separate data into distinct clusters

✓ How can you create a pivot table with multiple aggregation functions for a DataFrame **df**? 1/1

- (●) df.pivot_table(values='value_column', index='index_column', columns='columns_column', aggfunc=['sum', 'mean']) ✓
- ( ) df.pivot_table(values='value_column', index='index_column', columns='columns_column', aggfunc='sum', aggfunc='mean')
- ( ) df.pivot_table(values='value_column', index='index_column', columns='columns_column', aggfunc='sum, mean')
- ( ) df.pivot_table(values='value_column', index='index_column', columns='columns_column', aggfunc='sum_mean')

✗ How can you calculate the z-scores of a specific column 'column_name' in a DataFrame **df**? 0/1

- (●) df['column_name'].zscore() ✗
- ( ) (df['column_name'] - df['column_name'].mean()) / df['column_name'].std()
- ( ) df['column_name'].transform(lambda x: (x - x.mean()) / x.std())
- ( ) All of the above

Correct answer

- (●) All of the above

✓ In machine learning, what is the main purpose of transfer learning? 1/1

- ○ To transfer data between different machines
- ○ To transfer models between different machine learning frameworks
- ◉ To transfer knowledge from one task or domain to another, typically by fine-tuning a pre-trained model ✓
- ○ To transfer labeled data to an unlabeled dataset

✓ What does the term "bias-variance trade-off" refer to in machine learning? 1/1

- ○ The trade-off between precision and recall in classification models
- ○ The trade-off between model complexity and model accuracy
- ○ The balance between the mean squared error and the number of features
- ◉ The trade-off between underfitting and overfitting in a model ✓

✓ Which type of machine learning task involves grouping similar data points into clusters based on their features, without predefined class labels? 1/1

- ○ Regression
- ○ Classification
- ◉ Clustering ✓
- ○ Anomaly Detection

✓ What is the vanishing gradient problem in deep learning? 1/1

- ○ It occurs when gradients are too large and lead to divergence.
- ○ It is a problem related to overfitting.
- ◉ It happens when gradients in a neural network become very small during training. ✓
- ○ It is a term used in generative adversarial networks (GANs).

✓ How can you calculate the cumulative sum of a DataFrame **df** along the columns? 1/1

- ○ df.cumulative_sum(axis=0)
- ◉ df.cumsum(axis=1) ✓
- ○ df.cumulative(axis=1)
- ○ df.cumulative_sum(axis=1)

✓ In the context of deep learning, what is a "vanishing gradient" problem, and how is it typically addressed? 1/1

- It is a problem where gradients become extremely large, leading to numerical instability. It is addressed by using gradient clipping.
- It is a problem where gradients become very small, making it difficult to update model parameters in deep networks. It is addressed by using activation functions that mitigate this issue, such as ReLU. ✓
- . It is a problem where gradients converge to a local minimum, leading to suboptimal solutions. It is addressed by increasing the learning rate.
- It is a problem that occurs when there are too many layers in a neural network, leading to slow training times. It is addressed by using shallow networks.

✓ How can you sort a DataFrame **df** based on values in the column 'column_name' in ascending order? 1/1

- df.sort('column_name')
- df.sort_values('column_name', ascending=True) ✓
- df.sort_by('column_name')
- df.order_by('column_name')

✗ What does the **query()** function in Pandas do? 0/1

- ○ Selects rows from a DataFrame based on a query string.
- ○ Performs SQL-like operations on a DataFrame.
- ○ Calculates the mean value of each group.
- ◉ None of the above. ✗

Correct answer

- ◉ Selects rows from a DataFrame based on a query string.

✓ What is the primary purpose of a Confusion Matrix in machine learning? 1/1

- ○ To visualize the decision boundary of a classification model
- ◉ To summarize the performance of a classification model, showing the number of true positives, true negatives, false positives, and false negatives ✓
- ○ To reduce overfitting in a neural network
- ○ To optimize hyperparameters in an ensemble learning model

✓ Which of the following is an example of a non-parametric machine learning algorithm? Name *1/1

- ○ Linear Regression
- ○ K-Means Clustering
- ○ Support Vector Machine (SVM)
- ◉ k-Nearest Neighbors (KNN) ✓

✓ In ensemble learning, what is the purpose of stacking (stacked generalization)? 1/1

- ○ To create a stack of physical hardware for distributed machine learning
- ○ To stack multiple layers of neurons in a neural network
- ◉ To combine predictions from multiple base models using a meta-model ✓
- ○ To perform feature engineering by stacking features from different sources