Grand Assessment - Applied Data Science

Please enter correct Name, Email Address, NM ID and Anna University Registration Number. Wrong information will lead to zero score.

Points: 28/30   Time: 27:58

1. Name *

    Melvin Jones E

2. Email Address **(please use your IBM linked email address only)** *

    21csmelvinjonese@jeppiaarinstitute.org

3. NM ID *

    au210621104031

4. Anna University Registration Number *

    210621104031

5. College Code *

   2106

   ✓ **Correct**   1/1 Points

6. What is the purpose of feature scaling in data preprocessing?
   *

   ○ To remove outliers from the data

   ○ To convert categorical features into numerical format

   ● To normalize the data to a similar scale

   ○ To handle missing values in the dataset

   ✓ **Correct**   1/1 Points

7. Which evaluation metric is used to assess the performance of a regression model by measuring the average difference between predicted and actual values? *

   ○ F1 Score

   ○ R-squared (R2) Score

   ● Mean Absolute Error (MAE)

   ○ Precision

   ✓ **Correct**   1/1 Points

8. Which technique is used to handle the class imbalance problem by generating synthetic samples for the minority class? *

   ○ Ensemble Learning

○ SMOTE (Synthetic Minority Over-sampling Technique)

○ Ridge Regression

○ Recursive Feature Elimination (RFE)

✓ **Correct**   1/1 Points

9. Which machine learning algorithm is used for anomaly detection and novelty detection? *

○ Naive Bayes

○ Random Forest

○ Support Vector Machines (SVM)

◉ Isolation Forest

✓ **Correct**   1/1 Points

10. What is the process of cleaning, transforming, and enriching raw data into a usable form for analysis? *

◉ Data Wrangling

○ Data Visualization

○ Data Mining

○ Data Analysis

✓ **Correct**   1/1 Points

11. Which Python library is commonly used for data manipulation and analysis *

○ TensorFlow

○ Keras

◉ Pandas

○ Scikit-learn

✓ **Correct**   1/1 Points

12. What type of machine learning algorithm is used for predicting numeric values, such as house prices? *

○ Classification

○ Clustering

◉ Regression

○ Reinforcement Learning

✓ **Correct**   1/1 Points

13. What is the main advantage of using cloud services for AI and ML applications? *

○ Lower computational power

◉ Reduced cost and scalability

○ Restricted access to AI models

○ Limited storage capabilities

✓ **Correct**   1/1 Points

14. Which statistical measure gives an idea of how much the values in a dataset vary from the mean? *

○ Mean Absolute Deviation (MAD)

◉ Standard Deviation

◯ Variance

◯ Median Absolute Deviation (MAD)

✓ **Correct**   1/1 Points

15. What type of data preprocessing technique is used to convert text data into numerical format for machine
learning algorithms?
*

◯ Data Imputation

◯ Feature Scaling

◯ Feature Engineering

◉ Text Encoding

✓ **Correct**   1/1 Points

16. Which data type is used for categorical variables that have no intrinsic ordering?
*

◯ Integer

◉ String

◯ Float

◯ Boolean

✓ **Correct**   1/1 Points

17. In a confusion matrix, which term represents the number of correctly predicted
positive instances? *

○ True Positive (TP)

○ False Positive (FP)

○ True Negative (TN)

○ False Negative (FN)

✓ **Correct**   1/1 Points

18. Which supervised learning algorithm is used for making predictions with discrete or categorical target variables? *

○ Linear Regression

○ Decision Trees

◉ Logistic Regression

○ K-means Clustering

✓ **Correct**   1/1 Points

19. Which Python library provides tools for data manipulation and analysis, as well as mathematical functions and arrays? *

○ Scikit-learn

◉ NumPy

○ Pandas

○ Matplotlib

✓ **Correct**   1/1 Points

20. In machine learning, what term is used for the dataset used to test the model's performance after training on

the training dataset? *

  ◯   Validation set

  ◉   Test set

  ◯   Training set

  ◯   Unlabeled set

✓ **Correct**   1/1 Points

21. Which technique is used for reducing the dimensionality of data while preserving its variance?
*

  ◉   Principal Component Analysis (PCA)

  ◯   K-means Clustering

  ◯   Decision Trees

  ◯   Ridge Regression

✓ **Correct**   1/1 Points

22. What is the term used for machine learning algorithms that learn from labeled data to make predictions or
decisions? *

  ◯   Unsupervised Learning

  ◉   Supervised Learning

  ◯   Reinforcement Learning

  ◯   Semi-Supervised Learning

✕ **Incorrect**   0/1 Points

23. Which cloud platform is known for providing AI services like natural language processing and computer vision? *

◯   Google Cloud Platform

🔘   b) Microsoft Azure

◯   c) IBM Cloud

◯   d) Amazon Web Services (AWS)

✓ **Correct**   1/1 Points

24. Which algorithm is commonly used for clustering similar data points together? *

◯   K-nearest Neighbors (KNN)

◯   Decision Trees

🔘   K-means

◯   Support Vector Machines (SVM)

✓ **Correct**   1/1 Points

25. Which technique is used to replace missing values in a dataset with appropriate values? *

◯   Outlier Detection

◯   Feature Scaling

🔘   Data Imputation

◯   Feature Engineering

✓ **Correct**   1/1 Points

26. Which technique is used to find the best set of hyperparameters for a machine learning model? *

   ○ Gradient Descent

   ○ Feature Engineering

   ◉ Hyperparameter Optimization

   ○ K-Nearest Neighbors

✓ **Correct**   1/1 Points

27. Which module in data science focuses on finding patterns and relationships in data without using labeled examples? *

   ◉ Unsupervised Learning

   ○ Supervised Learning - Classification

   ○ Model Evaluation Metrics

   ○ Hyper-parameter Optimization

✓ **Correct**   1/1 Points

28. Which statement best describes the term "overfitting" in machine learning? *

   ◉ The model performs well on the training data but poorly on unseen data.

   ○ The model performs equally well on both the training and test data.

   ○ The model cannot capture complex patterns in the data.

○ The model is under-trained and lacks accuracy.

✓ **Correct**   1/1 Points

29. Which programming language is widely used for data science and machine learning? *

○ Java

○ C++

◉ Python

○ Ruby

✓ **Correct**   1/1 Points

30. Which method can be used to handle imbalanced datasets in classification problems?
*

○ Randomly removing samples from the majority class

○ Using accuracy as the evaluation metric

◉ Oversampling the minority class

○ Ignoring the class imbalance and training the model as usual

✓ **Correct**   1/1 Points

31. What is the purpose of cross-validation in machine learning? *

◉ To evaluate the model's performance on unseen data

○ To compare different machine learning algorithms

○ To handle missing values in the dataset

○ To increase the model's complexity

✓ **Correct**   1/1 Points

32. Which data preprocessing technique is used to transform categorical variables into numerical form by assigning each category a unique integer?

*

○ Data Normalization

○ Feature Scaling

○ One-Hot Encoding

◉ Label Encoding

✓ **Correct**   1/1 Points

33. Which evaluation metric is commonly used for classification problems and represents the ratio of correctly predicted instances to the total instances? *

○ Mean Squared Error (MSE)

○ Area Under the Curve (AUC)

○ F1 Score

◉ Accuracy

✓ **Correct**   1/1 Points

34. Which unsupervised learning algorithm is used to find the optimal number of clusters in a dataset? *

○ K-means Clustering

○ Hierarchical Clustering

○ Density-Based Spatial Clustering of Applications with Noise (DBSCAN)

⦿ Elbow Method

✕ **Incorrect**   0/1 Points

35. Which evaluation metric is commonly used for imbalanced datasets and represents the ability of a model to identify positive instances correctly? *

○ F1 Score

○ Accuracy

⦿ Precision

○ Area Under the Curve (AUC)

Keep the information with you by saving your response.

Save my response

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.

Powered by Microsoft Forms |
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