1. **A feature F1 can take certain value: A, B, C, D, E, & F and represents grade of students from a college.**

**Which of the following statement is true in following case?**

A) Feature F1 is an example of nominal variable.

**B) Feature F1 is an example of ordinal variable.**

C) It doesn’t belong to any of the above category.

D) Both of these

1. **Which of the following is an example of a deterministic algorithm?**

A) **PCA**

B) K-Means

C) None of the above

1. **A Pearson correlation between two variables is zero but, still their values can still be related to each other.**

A) TRUE

B) FALSE

1. **Suppose a new cancer treatment has been discovered, claiming to increase the one year survival rate for pancreatic cancer to 40%. In other words, the probability that a patient suffering from pancreatic cancer would survive for at least one year after receiving this treatment is 40%.**

**Suppose a hospital is planning to use this treatment for its pancreatic cancer patients.**

**Pancreatic Cancer Hospital**

The hospital has a total of 10 patients suffering from pancreatic cancer. What is the probability that exactly 4 of these patients would survive the first year after receiving this treatment?

A) 2.2%

B) 12.6 %

C) 25.1 %

D) 21.6%

1. **Suppose a new cancer treatment has been discovered, claiming to increase the one year survival rate for pancreatic cancer to 40%. In other words, the probability that a patient suffering from pancreatic cancer would survive for at least one year after receiving this treatment is 40%.**

**Suppose a  hospital is planning to use this treatment for its pancreatic cancer patients.**

**Pancreatic Cancer Hospital**

What is the probability that the number of patients that survive the first year after receiving the treatment would not be more than 2?

A) 14.9%

B) 16.7%

C) 22.1%

D) 15.7%

1. **A Singapore-based startup Healin launched an app called JustShakeIt that enables a user to send an emergency alert to emergency contacts and/or caregivers simply by shaking the phone with one hand. The program uses a machine learning algorithm to distinguish between actual emergency shakes and everyday jostling, using data with labels to distinguish between everyday jostling and emergency shaking.**

**What kind of problem is this?**

A) Supervised learning - Regression

**B) Supervised learning - Classification**

C) Unsupervised learning - Clustering

D) Can’t say with this information

1. **Which of the following is a tool of Machine Learning Library?**
2. Persistence
3. Utilities like linear algebra, statistics
4. Pipelines
5. All of the above
6. **Is MLlib deprecated?**
7. Yes
8. No
9. **For Regression problem which algorithm is not the solution**
10. Logistic Regression
11. Ridge Regression
12. **Decision Trees**
13. Gradient-Boosted Trees
14. **Which of the following is true for Spark MLlib?**
15. Provides an execution platform for all the Spark applications
16. **It is the scalable machine learning library which delivers efficiencies**
17. enables powerful interactive and data analytics application across live streaming data
18. None of the above
19. **For Multiclass classification problem which algorithm is not the solution?**
20. Naive Bayes
21. Random Forests
22. Logistic Regression
23. Decision Trees
24. **The primary Machine Learning API for Spark is now the \_\_\_\_\_ based API**
25. DataFrame
26. Dataset
27. RDD
28. All of the above
29. **Which of the following algorithm is not present in MLlib?**
30. Streaming Linear Regression
31. Streaming KMeans
32. Tanimoto distance
33. None of the above
34. **We have real-time Twitter feed. We need to build an application that is near real-time and classifies the Twitter feeds based on relevant and not relevant, where "relevant means that it contains the words "FIFA," Women's" and "World Cup." Which of the following Apache Spark libraries we could not use in the application?**
35. Spark SQL
36. Spark Streaming
37. Spark MLib
38. Spark GraphX
39. **Which flow below correctly describes a supervised learning window?**
40. Load sample data 2. Extract features 3. Use algorithm to group data based on common features 4. Use validated algorithm with new data
41. Load sample data 2. Extract features 3. Split data into training and test sets 4. Build, test, and tune the model 5. Use validated model with new data
42. Load sample data 2. Extract features 3. Split the data into training and test sets 4. Use algorithm to discover model features 5. Use model to classify new data features
43. None of Above
44. **You have a dataset of in-game purchases from mobile game users. You want to group these users for upsell. Which of the Spark machine learning algorithms could you use?**
45. Random decision forest
46. Linear regression
47. **K-means**
48. **You have a dataset of houses sold. You want to predict a house's price based on the size of the house. Which Spark Machine learning algorithm could you use?**
49. Clustering
50. ALS
51. **Linear regression**
52. **Which of these is an object used for classification in Spark**
53. GaussianMixture
54. RandomForestModel
55. KMeansModel
56. PowerIterationCluster
57. **Which of the following is not a component of Spark Ecosystem**
58. Sqoop
59. GraphX
60. MLlib
61. BlinkDB
62. **Can you combine the libraries of Apache Spark into the same Application, for example, MLlib, GraphX, SQL and DataFrames etc.**
63. Yes
64. No