INSEM EXAM - Honor Course (Data Science and Visualization)

* Required

INSEM Paper

..... involves extraction of knowledge from large volumes of different types of data. *

- ( ) Data Analysis
- ( ) Data Science
- ( ) Descriptive Analytics
- (•) None of the mentioned

..... which is original data source. *

- (•) Raw data
- ( ) Pre-processed data
- ( ) Raw data is the data obtained after processing steps
- ( ) None of the mentioned

Data Scientist performs.....? *

- ○ Defining the question
- ○ Creating reproducible code
- ○ Challenge results
- ◉ All of the mentioned

Which of following is not characteristic of Big Data? *

- ◉ Handling Massive amount of data
- ○ Processing Different Data types and Structure.
- ○ Handling by traditional tools.
- ○ Analyze the data

Which from the below relates correct to Velocity in Big Data? *

- ○ The amount of data handling by processing & analysing.
- ◉ The fastest rate at which data is received or generated.
- ○ Data types requiring additional pre-processing to derive meaning and support metadata. Structured, Semi Structured, Unstructured Data.
- ○ Accurate or truthful a data set may be, the quality of the data, trustworthy the data source, type, and processing of it is.

The characteristics of big data are often referred to as..... *

- ○ volume, velocity, and variability.
- ○ volume, veracity, and variety.
- ○ volume, volatility, and variability.
- ◉ volume, velocity, and variety

Which of the following is not a key skill for data science? *

- ◉ None of the above
- ○ Machine Learning
- ○ Data Visualization
- ○ Computing, and Algorithm building

In data science and big data the main categories of data are.... *

- ○ Structured and Unstructured
- ○ Graph-based
- ○ Audio, video, and images
- ◉ All of the above

.....data is data that depends on a data model and resides in a fixed field within a record. *

- [x] Structured
- [ ] Unstructured
- [ ] Graph-based
- [ ] Audio, video, and images

The data science process isn't linear it can be divided into steps..... *

- [x] Setting the research goal, Gathering data, Data preparation, Data exploration, Modeling, Presentation and automation.
- [ ] Setting the research goal, Preparing the model, Data isolation, Data exploration, Modeling, Presentation and automation
- [ ] Setting the research goal, Gathering data, Data preparation, Data visualization, Modeling, Presentation and automation.
- [ ] Setting the research goal, Gathering data, Data visualization, Data exploration, Modeling, Presentation and automation.

The goal of this .....step is to gain a deep understanding of the data. *

- [x] Data Exploration
- [ ] Data Preparation
- [ ] Model Building
- [ ] Gathering data

Which one is not a format of a data can be stored in official data repositories..... *

- databases
- data parts
- data warehouses
- data lakes

The Different ways of Combining Data are.... *

- Data enrichment
- Joining tables
- Appending tables
- All the above

.....is a field of study that gives computers the ability to learn without being explicitly programmed. *

- Machine learning
- Data Science
- Data Analytics
- Self Intelligence

.....and.....are of primary importance to a data scientist, to achieve these goals, a data scientist uses is machine learning. *

- (●) Regression, classification
- ( ) Data preparation and classification
- ( ) Cleaning, Regression
- ( ) Regression, data hiding

The modeling phase consists of ..... *

- ( ) Feature engineering and model selection
- ( ) Training the model
- ( ) Model validation and selection
- ( ) Applying the trained model to unseen data
- (●) All of the above

Unsupervised learning techniques don't rely on labeled data *

- (●) True
- ( ) False

Which of the following is not a application for data science? *

- [ ] Recommendation Systems
- [ ] Image & Speech Recognition
- [ ] Online Price Comparison
- [x] Privacy Checker

Mean of a random variable is the ..... *

- [x] Midpoint of its distribution.
- [ ] Skewness of the distribution
- [ ] deviation in the distribution
- [ ] Bayesian inference.

The median of 7, 6, 4, 8, 2, 5, 11 is..... *

- [x] 6
- [ ] 12
- [ ] 11
- [ ] 4

..... refers to measures of how spread out our data is. *

- [ ] Dispersion
- [ ] Randomness
- [x] Standard deviation
- [ ] Mode

.....is a variable whose possible values have an associated probability distribution. *

- [ ] Discrete Variable
- [x] Random variable
- [ ] Standard deviation
- [ ] Discrete random variables

.....is the classic bell curve–shaped distribution and is completely determined by two parameters: its mean μ and its standard deviation *

- [x] Binomial distribution
- [ ] Continuous distribution
- [ ] Bernoulli Distribution
- [ ] Normal distribution

Regression is one of the type of clustering? *

- ○ True
- ◉ False

Extremely sensitive to the outliers is.....? *

- ◉ K-means clustering algorithm.
- ○ K-medians clustering algorithm
- ○ K-modes clustering algorithm
- ○ K-medoids clustering algorithm

In which of the following cases will K-Means clustering fail to give good results? 1. Data points with outliers 2. Data points with different densities 3. Data points with round shapes 4. Data points with non-convex shapes *

- ○ 1 and 2
- ○ 2 and 3
- ○ 2 and 4
- ◉ 1, 2 and 4
- ○ 1, 2, 3 and 4

Iterations count in apriori _____ *

- ○ increases with the size of the data
- ◉ increases with the size of the maximum frequent set
- ○ decreases with the increase in size of the data
- ○ decreases with increase in size of the maximum frequent set

Which of the following are interestingness measures for association rules? *

- ○ recall
- ○ lift
- ○ accuracy
- ◉ interestingness measures

..... refers to a specific implementation of association rules. *

- ◉ Market basket analysis
- ○ Churn Prediction
- ○ Image Classification
- ○ Spam filtering

Which is an example of regression ...

- [x] Predict residential home prices.
- [ ] Analyze effect of proposed radiation treatment
- [ ] Document analysis
- [x] Spam Filtering

In linear regression modeling, the outcome variable is .... *

- ( ) Continuous
- ( ) Categorical
- ( ) None of the above
- (•) Both a and b

.....is one of the most popular data mining approach for finding frequent item sets from a transaction dataset and derive association rules. *

- ( ) Binary
- ( ) Greedy
- ( ) Randomized
- (•) Apriori

.....measures the difference in the probability of X and Y appearing together compared to statistical independence. *

- ○ Leverage
- ○ Lift
- ○ Confidence
- ◉ Regression

Which method use data with label for training..... *

- ◉ Supervised learning
- ○ Unsupervised learning
- ○ Reinforcement learning
- ○ None of these

..... method is supervised method. *

- ○ Classification
- ○ Clustering
- ◉ Both a and b
- ○ None of the above

Where should we apply Bayes rule..... *

- ○ Solving queries
- ○ Increasing complexity
- ○ Decreasing complexity
- ◉ Answering probabilistic query

..... types of discrete probability distributions *

- ○ Poisson
- ○ Bernoulli
- ○ Binomial
- ◉ All of the above

Find the mode in given example. 2,5,6,2,2,2,5,6,2 *

- ◉ 2
- ○ 4
- ○ 6
- ○ 5

..... in which correlations can be misleading when confounding variables are ignored. *

- ○ Simpsons Paradox
- ○ Correlational Caveats
- ○ Bayes's Theorem
- ◉ Confounding and collinearity

Which of the following statements about Naive Bayes is correct? *

- ○ Attributes are not equally important.
- ◉ Attributes are statistically dependent of one another given the class value.
- ○ Attributes are statistically independent of one another given the class value.
- ○ Attributes can't be nominal or numeric.

Back Submit