Connect, Message, Like, Follow & Share, 100% Free Counselling ➔ Thank You

Data Science Part-1

Interview Questions & Answers

1. What are the different types of Discrete ProbabilityDistributions?

Binomial, Poisson, Negative Binomial, Geometric, Hypergeometric arethe examples of Discrete Probability Distributions.

2. What are the different types of Continuous ProbabilityDistribution?

Normal, Exponential, t, f, Chi-square, Uniform, Weibull are few of the examples of Continuous Probability Distributions.

3. What do you mean by Binomial Distribution?

Binomial Distribution can be simply thought of as the probability of Success of Failure outcome in an experiment that is conducted multiple times. Examples: Head and Tail outcomes after tossing a coin, Pass or Fail after appearing for an examination.

4. What is a Poisson Distribution?

Poisson Distribution gives the probability of a number of events happening in a fixed interval or space. Number of customers visiting a restaurant every day.

5. What is logistic regression? Or State an example when you have used logistic regression recently?

Logistic Regression often referred as logit model is a technique to predictthe binary outcome from a linear combination of predictor variables. For example, if you want to predict whether a particular political leader will

win the election or not. In this case, the outcome of prediction is binary i.e. 0 or 1 (Win/Lose). The predictor variables here would be the amountof money spent for election campaigning of a particular candidate, the amount of time spent in campaigning, etc.

6. What are Recommender Systems?

A subclass of information filtering systems that are meant to predict the preferences or ratings that a user would give to a product. Recommender systems are widely used in movies, news, research articles, products, social tags, music, etc.

7. Why data cleaning plays a vital role in analysis?

Cleaning data from multiple sources to transform it into a format that data analysts or data scientists can work with is a cumbersome process because – as the number of data sources increases, the time take to clean the data increases exponentially due to the number of sources and the volume of data generated in these sources. It might take up to 80% of the time for just cleaning data making it a critical part of analysis task.

8. Differentiate between univariate, bivariate andmultivariate analysis

These are descriptive statistical analysis techniques which can be differentiated based on the number of variables involved at a given pointof time. For example, the pie charts of sales based on territory involve only one variable and can be referred to as univariate analysis.

If the analysis attempts to understand the difference between 2 variablesat time as in a scatterplot, then it is referred to as bivariate analysis. For example, analyzing the volume of sale and a spending can be consideredas an example of bivariate analysis.

Analysis that deals with the study of more than two variables to understand the effect of variables on the responses is referred to as multivariate analysis."

9. What is Machine Learning?

Machine learning is the science of getting computers to act without being explicitly programmed. Machine learning has given us self-driving cars, practical speech recognition, effective web search, and a vastly improved understanding of the human genome. It is so widespread that unknowingly we use it many a times in our daily life.

10. Why the probability associated with a single value of a continuous random variable is considered to be zero?

A continuous random variable takes an infinite number of values. As the number of values assumed by the random variable isinfinite, the probability of observing a single value is zero.