Data Science Interview Questions

1. What is Data Science? List the difference between supervised and unsupervised learning.
2. What is selection bias?
3. What is a bias-variance trade-off?
4. In your choice of language, write a program that prints the numbers ranging from 1 to 50.
5. What is a confusion matrix?
6. Describe Markov chains?
7. What do you understand by true and false positive rates?
8. Why R is used in Data Visualization?
9. What is the ROC curve?
10. What are dimensionality reduction and its benefits?
11. How do you find RMSE and MSE in a linear regression model?
12. How do deal with unbalanced binary classification?
13. What is the difference between a box plot and a histogram?
14. What does NLP stand for?
15. Walkthrough the probability fundamental?
16. Describe different regularization methods, such aa L1 and L2 regularization?
17. How should you maintain a deployed model?
18. Write the equation and calculate the precision and recall rate?
19. Why do we use the summary function?
20. How will you measure the Euclidean distance between the two arrays in NumPy?
21. What is the difference between an error and residual error?
22. Difference between Normalization and Standardization?
23. Wat is the difference between ‘long’ and ‘wide’ format data?
24. What do you understand by the term Normal Distribution?
25. What are correlation and covariance in statistics?
26. What is the difference between Point Estimates and Confidence Intervals?
27. What is the goal of A/B testing?
28. What is the p-value?
29. In any 15 minutes interval, there is a 20% probability that you will see at least one shooting star. What is the probability that you see at least one shooting star in the period of an hour?
30. How can you generate a random number between 1-7 with only a die?
31. A certain couple tells you that they have two children, at least one of which is a girl. What is the probability that they have two girls?
32. A jar has 1000 coins, of which 999 are fair and 1 is double-headed. Pick a coin at random, and toss it 100 times. Given that you see 10 heads, what is the probability that the next toss of that coin is also a head?
33. What do you understand by statistical power of sensitivity and how do you calculate it?
34. What is Re-Sampling done?
35. What is the difference between overfitting and underfitting?
36. How to combat overfitting and underfitting?
37. What is regularisation? Why is it useful?
38. What is the law of large numbers?
39. What are confounding variables?
40. What are the types of biases that can occur during sampling?
41. What is Survivorship Bias?
42. What is selection Bias?
43. Explain how a ROC curve works?
44. What is TF/IDF vectorization?
45. Why do we generally use se SoftMax non-linearity function as the last operation in networks?
46. Python or R which one would you prefer for text analytics?
47. How does data cleaning play a Vitol role in the analysis?
48. Differentiate between univariate, bivariate and multivariate analysis?
49. Explain Star Schema?
50. What is cluster sampling?