100 Data Science Machine Learning Interview Questions

1. What is Normal distribution? What if data is skewed, how it impacts machine learning.

2. What is Feature Creation? Where it can be used?

3. How K-means clustering work on categorical data.

4. What is decision boundary? Can we categories data having spherical decision boundary using Logistic regression?

5. What is polynomial Regression?

6. What are the drawbacks of Random forest?

7. How you decide which models/algorithm you need to use?

8. How do you treat balanced and imbalanced data?

9. Explain Confusion Matrix? How do you read a 4X4 confusion matrix?

10. Explain Linear regression with formula. Also explain about the data sets best for linear regression to be applied.

11. How to plot a data set with multiple independent variable. Also explain multiple regression model with mathematical calculation.

12. Can linear regression be applied for multi class classification?

13. Explain real time scenario where polynomial regression can be applied. Also explain how degree of polynomial can be adjusted.

14. Explain about the python libraries used for linear regression, multiple regression and polynomial regression.

15. What is support vector regressor Explain some real time cases where support vector regressor can be applied?

16. Explain steps to be followed in data cleaning and noise removal.

17. Explain about feature selection and feature extraction in data prepossessing step. Also explain methods applied for feature selection and extraction.

18. Explain the cases when dimensionality reduction is required and when new features should be derived.

19. Explain PCA and ICA.

20. How to evaluate regression model and classification model.

21. Explain about AME, RMSE, MSE, R2 Errors. What is the best value for these errors?

22. How do you treat Null Values and Outlier?

23. How do you decide when to impute mean/median or mode while treating Null values or Outlier?

24. Difference between List and tuples? Which one is Faster?

25. What does P value Signifies? When P > 0.05, P < 0.05 and P = 0.05?

26. What is co-linearity and co-variance?

27. What is Machine Learning and its type?

28. What is Supervised Learning?

29. What is Unsupervised Learning?

30. What are the Various Classification Algorithms?

31. What is Logistic regression? State an example when you have used logistic regression recently.

32. What are Recommender Systems?

33. What is Linear Regression?

34. What is Collaborative filtering?

35. How can be outlier values be treated?

36. What are various steps involved in analytic project?

37. During Analysis, how do you treat missing values?

38. Explain about classification and confusion metrics.

39. How to read a 4X4 confusion metrics.

40. Explain about ensemble learning.

41. Explain Bagging and boosting.

42. Explain about classification boundary for classification algorithms.

43. How Decision tree works. How to calculate entropy and information gain. How root node is selected to build a decision tree.

44. How to decide best branches for pruning in decision tree.

45. What is Selection Bias?

46. Difference between "long" and "wide" format data?

47. What do you understand by the term Normal Distribution?

48. What is the goal of A/B Testing?

49. What do you understand by statistical power of sensitivity and how do you calculate it?

50. Difference between overfitting and underfitting?

51. Differentiate between univariate, bivariate and multivariate analysis?

52. What is Cluster Sampling?

53. What is Systematic Sampling?

54. What is Eigenvalue and Eigenvector?

55. Can you cite some Examples where a false positive is important than a false negative?

56. Can you cite some Examples where a false negative is important than a false positive?

57. Can you cite some Examples where both false positive and false negatives both are equally important?

58. Can you explain the difference between a Test Set and a Validation Set?

59. Explain Cross Validation?

60. How will you define the number of clusters in a clustering algorithm?

61. What are feature vectors?

62. Explain the steps in making a decision Tree .

63. Do gradient descent methods at all time converge to a similar point?

64. Drawbacks of Linear model?

65. What is the Law of Large Numbers?

66. What are Confounding Variables?

67. Why is resampling done?

68. How regularly must an algorithm must be updated?

69. How do you work towards a random forest?

70. What is the life cycle of a Data Science Project?

71. Explain Curse of Dimensionality?

72. How to test if regression & classification models are working fine?

73. When will you use classification over regression?

74. What is 'Training Set' and 'Test Set' in a machine Learning model?

75. How much data will you allocate for your training, validation and test sets?

76. How do you handle missing or corrupted data in a dataset?

77. How can you choose a classifier based on training set size?

78. Explain Confusion Matrix with respect to Machine Learning algorithms?

79. Difference between inductive Machine Learning and deductive Machine Learning?

80. Explain K nearest Neighbor algorithm.

81. Compare K- Means and KNN algorithms.

82. What is Vanishing Gradient and Exploding gradient.

83. What is backpropagation in Artificial Neural Network.

84. Folding and Unfolding in terms of RNN.

85. How LSTM store data?

86. What is the significance of Dropout?

87. What is RNN? How it works? How it stores weights?

88. What is specificity and sensitivity?

89. How to control false positive and false negative rates.

90. How does convolution neural networks work.

91. What is loss function or Cost function?

92. Explain hyperparameter tuning or Model tuning?

93. What Is a Multi-layer Perceptron(MLP)?

94. What Is the Role of Activation Functions in a Neural Network?

95. Explain Data Normalization.

96. What are the unsupervised learning algorithms in Deep learning?

97. What are the main benefits of Mini-batch Gradient Descent?

98. What are the different layers of Autoencoders? Explain briefly.

99. What do you understand by Autoencoder?

100. What do you understand by Deep Autoencoders?