

# Machine Learning Interview Questions

## Machine Learning Questions:

1. What are the activation functions?
2. Do you know about RNN?
3. Difference between bias and variance
4. Do you know about bagging and boosting, in which machine learning model we use them
5. What are word embeddings and how are they used in NLP
6. What is sentiment analysis and how is it performed in NLP
7. Do you know about multi regression
8. **Do you know about ensemble learning?**
9. How to evaluate regression models? Like which type of accuracy matrix you use in regression models?
10. Do you know about cross-validation?
11. **Do you know stacking in machine learning?**
12. **Do you know how a decision tree splits?**
13. What is PCA? Why do we use? What alternatives are there?
14. **What is back propagation?**
15. can you tell me about the different types of layers used in neural network?
16. **What is pruning in decision trees?**
17. **Explain the assumptions of linear regression.**
18. **How do decision trees differ from random forests? In what scenarios would you choose one over the other?**
19. **What is A/B testing?**
20. What to do if your time series data is not stationary?
21. **Machine Learning Interview Questions**
22. What is Regression?
23. What are the Types of regression?
24. Explain the assumptions of Linear regression.
25. How to evaluate regression models?
26. Explain different types of model Testing methods.
27. Explain Benefits of cross validation?
28. What is overfitting?
29. What are different ways to fix overfitting?
30. How to handle imbalanced data in a dataset?
31. What to do if your data is skewed?
32. What if your data has outliers, how to identify and different ways to fix?
33. What is difference between lasso and ridge regression?
34. What is cross entropy?
35. Difference between precision and recall?
36. What is a confusion matrix?
37. What is AUC?
38. What is bias variance trade-off?
39. What is correlation? How is it related to covariance?
40. What are activation function? Explain different Types? Where and why it is used? Sigmoid vs tan h vs relu?
41. What is classification?

42. How to evaluate classification models?
43. What are ensemble models?
44. What is clustering?
45. What are types of clustering?
46. How to decide the K in k means clustering? Can we automate it? If yes how?
47. What is hypothesis testing?
48. What is A/B testing?
49. How does time series work? Explain how to approach a time series problem step by step.
50. How do you treat seasonality in time series?
51. What to do if your time series data is not stationary?
52. Difference between ARIMA and SARIMA?
53. What is auto correlation?
54. Can we use deep learning for time series forecasting? If yes what models to use? How do they work?
55. What is null hypothesis?

## **NLP Questions:**

56. What is word embedding in NLP? Types of embedding? Best one to use ?
57. What is N-grams?
58. What is bag of words model?
59. What is named entity recognition? Is it a classification or regression model?
60. How to approach a named entity recognition problem?
61. How to handle large data size in case of a NER?
62. What is transformer architecture?
63. How does BERT works? explain Architecture?
64. What is attention? Why is it important?
65. How to create a model similar to BERT manually?
66. What is LLM? How are they designed?
67. In which cases LLMs are disadvantageous?
68. What are benefits of LLM?
69. How does ChatGPT (Any Generative AI chatbot) remember previous text messages?
70. Where should you use open source models like falcon, llama over ChatGPT?
71. How does LLM like chatgpt work in the background?
72. Can we create a Private LLM like ChatGPT that can Work only on Prem devices? If yes how?
73. How to Fine Tune LLM?
74. How does ChatGPT generate text response? Can it generate without any manual intervention? If yes how?
75. Let's say you have created a chatbot. How would you test it? Can you test it without launching to the customers?
76. How to design conversations in a chatbot?
77. How does dialogflow or rasa work in designing chatbots?
78. How to evaluate a chatbot?
79. What metrics to use to monitor a chatbot performance once it's live?
80. How the chatbot knows what the user is saying and make a response?
81. Let's say you need to generate caption from an input image? How would you design? What steps and models to use? How image embedding works? How an image is converted to generate text input?

82. What is topic modelling? What steps and models to use? If given a text data corpus of conversations, how would you generate topic from a unlabeled unsorted data corpus?
83. What is sentiment analysis? How to approach? Models to use? How to evaluate?
84. How can we create prompt templates and use them in LLMs?
85. How vector storage/DBs are used to store information in LLMs?

## **Other ML Questions**

86. What is bootstrapping?
87. What is conditional probability?
88. What is word2vec?
89. What is seq2seq and encoder decoder Neural Networks?
90. I haven't been asked these following questions yet. I am adding them to cover other important topics.
91. What is RNN? Where is it used? Drawbacks of RNN?
92. What is vanishing gradient problem and where does it occurs? How to fix?
93. What is LSTM? How does it works? Explain LSTM architecture?
94. What is bagging and boosting? Explain Xgboost?
95. Why do we need an optimizer? Explain Adam optimizer?
96. What is LDA in ML?
97. What is PCA? Why do we use? What alternatives are there?
98. What is pruning in decision trees?
99. What is back propagation?
100. What is gradient decent?