

## 50 Essential Machine Learning Interview Questions

1. What are the different types of machine learning?
2. What's the trade-off between bias and variance?
3. What is over fitting, and how can you avoid it?
4. How is KNN different from k-means clustering?
5. What is the difference between supervised and unsupervised machine learning?
6. Explain how a roc curve works
7. What is 'training set' and 'test set' in a machine learning model?
8. How much data will you allocate for your training, validation, and test sets?
9. How do you handle missing or corrupted data in a dataset?
10. How can you choose a classifier based on a training set data size?
11. Explain the confusion matrix with respect to machine learning algorithms.
12. What is the difference between classification and regression?
13. How to ensure that your model is not over fitting?
14. List the main advantage of naive bayes?
15. Explain ensemble learning.
16. Explain dimension reduction in machine learning.
17. What should you do when your model is suffering from low bias and high variance?
18. Explain differences between random forest and gradient boosting algorithm.
19. What is 'naive' in the naive bayes classifier?
20. How is amazon able to recommend other things to buy? How does the recommendation engine work?
21. When will you use classification over regression?
22. Considering a long list of machine learning algorithms, given a data set, how do you decide which one to use?
23. What is bias and variance in a machine learning model?
24. Define precision and recall.
25. What is pruning in decision trees, and how is it done?

26. Briefly explain logistic regression.
27. What are some methods of reducing dimensionality?
28. What are 3 ways of reducing dimensionality?
29. Explain principle component analysis (pca).
30. What are collinearity and multicollinearity?
31. Explain the difference between l1 and l2 regularization.
32. What is the roc curve and what is auc (a.k.a. AUROC)?
33. What's your favourite algorithm, and can you explain it to me in less than a minute?
34. What's the difference between type i and type ii error?
35. What's the difference between probability and likelihood?
36. What cross-validation technique would you use on a time series dataset?
37. Which is more important to you– model accuracy, or model performance?
38. What's the f1 score? How would you use it?
39. How would you handle an imbalanced dataset?
40. What do you understand by eigenvectors and eigenvalues?
41. How would you implement a recommendation system for our company's users?
42. How can we use your machine learning skills to generate revenue?
43. Explain false negative, false positive, true negative and true positive with a simple example.
44. What is the difference between inductive and deductive learning?
45. What is the difference between entropy and information gain?
46. What is bagging and boosting in machine learning?
47. How would you screen for outliers and what should you do if you find one?
48. Difference between linear regression and logistics regression.
49. Explain your favourite machine learning algorithm in depth.
50. How to select root node in decision tree?