Question Bank – ML Interviews

1. How regression is done in Decision Trees?
2. Explain the gini impurity formula through examples?
3. How Random Forest achieves Randomness? How does it resolve overfitting?
4. Precision vs Recall?
5. Which challenge does ROC AUC solve?
6. Explain situations in ML when high/low std dev is preferred?
7. What challenge does correlation vs causation pose in ML?
8. If you have 10 models (with no training data, no loss metric value). You have access to a test dataset. How would you find out the best model among the 10?
9. What are different methods to resolve the overfitting problem?
10. What are different methods to resolve the imbalanced dataset problem?
11. Why model retraining is required? Explain it through feature and model drift?
12. What is model lift? In what scenarios it is used in ML?
13. Bagging vs Boosting?
14. Among Bagging vs Boosting, whose base learners will have high variance?
15. What is curse of dimensionality? How to resolve it?
16. Fit vs Transform vs Fit_Transform?
17. What are demerits of PCA? Give real life project examples?
18. Explain Degree of freedom in layman's terms?
19. Explain Bessel's correction?
20. Can correlation be applied on non-linear relationships?
21. Explain the formula for VIF?
22. What is the range of R-squared?
23. Explain through examples why multicollinearity is bad for an ML model?
24. In hypothesis testing, what factors should be used to assume the value of alpha?
25. When should we use Standard Scalar instead of Minmax Scalar?
26. Why scaling is important in ML? Explain it through examples.
27. What are different types of cross validation techniques?
28. What are different types of hyperparameter tuning techniques?
29. What are different types of label encodings? When to use which? Explain with examples.
30. Why is it important to check the value of loss metric on the training dataset itself before on testing dataset?
31. What is the significance of Random State hyperparameter?
32. Supervised ML vs Unsupervised ML?
33. Can we do oversampling/undersampling/smote after doing the train test split?
34. What are type-I and type-II errors? Explain in layman's terms.
35. Chisquare vs Anova vs T-test?
36. What are the assumptions for conducting Chisquare, Anova, and T-test?
37. How Chisquare, Anova, and T-test can be used in feature selection in ML?
38. What is a p-value? How is it used in hypothesis testing?

39. Why do we say that we "fail to reject" but not "accept" the null hypothesis?
40. L1 vs L2 regularization?
41. What is the significance of pruning in Decision Trees? How is it achieved using hyperparameters?
42. Why bootstrapping is important in Random Forests?
43. What is the advantage/disadvantage of choosing a high value for 'p' (the number of attributes chosen at each node) in Random Forest?
44. When there are more than 2 classes in the target variable, how to calculate recall, precision and F1 score?
45. What is TPR and FPR? Can you calculate TPR and FPR when there are more than 2 classes in the target variable?
46. What is the formula for loss function in Logistic Regression? Explain in layman's terms.
47. Explain Alpha, Beta, and Log Odds in Logistic Regression?
48. Explain Odds and Log-Odds using probability?
49. Explain the loss function in K-means clustering?
50. Explain the Elbow method in K-means clustering?
51. In what scenarios A/B Testing is used in ML?
52. What is the significance of Learning Rate in boosting?
53. Why do we take smaller values of the learning rate?
54. Explain the process Gradient Descent in simple words?
55. What are some limitations of AutoML?
56. What is the difference between prediction and forecast?
57. How to explain feature influence/importance in the final model of different ML algorithms?
58. Name some methods other than Mean, Median, Mode to impute null values in a dataset?
59. What are the challenges you will face if you don’t handle outliers in the dataset?
60. Explain Bayes Theorem in probability? Where is it used in ML?
61. During model training, why do we minimize a loss function but not a loss metric?