ML FOR DS IMPORTANT QUESTIONS

1. What is Hashing? How do you resolve collision in hashing.
2. Define the Longest Common Subsequence (LCS) problem with an example
3. How does genomic sequencing work?
4. Provide examples of predictive modeling or analysis techniques used in personal genomics.
5. Differentiate machine learning with data science
6. Define Randomization? List the applications of Randomization
7. Illustrate BFS and DFS with an example
8. Describe the difference between objective function and constraints in linear programming.
9. List the applications of Stable Marriages Problem
10. What are some common techniques used for model selection?
11. Define probabilistic modeling in the context of machine learning.
12. Describe the properties and advantages of binary search trees (BSTs) in comparison to other tree data structures
13. Explain the concept of the divide and conquer algorithm. What are the advantages of using divide and conquer algorithms, specifically in predictive science?
14. Explain
    i. Control abstraction for dynamic programming
    ii. Principle of optimality
    iii. Example of dynamic programming
15. Using the Gale-Shapley algorithm, solve the stable marriages problem for the following set of preferences:

Men's preferences:

Alice	3	2	1
Bob	2	1	3
Charlie	1	3	2

Women's preferences:

Amy	2	3	1
Beth	3	1	2
Carol	1	2	3

Provide the stable pairings obtained and explain the algorithm's steps involved in the process.

16. Compare and contrast Dijkstra's algorithm and the A* algorithm in the context of finding the shortest path on a map.
17. Discuss NP-hard problems and explain their relevance

18. What is linear classification? Explain the concept of decision boundaries and how they relate to linear classifiers.
19. Explain the difference between bagging and boosting in ensemble learning.
20. Explain the concept of regularization in machine learning and its relationship to statistical inference. How does regularization help prevent overfitting?
21. Compare and contrast cross-validation and holdout method in terms of their advantages and disadvantages.
22. Evaluate the strengths and limitations of using machine learning for predicting preterm birth compared to traditional statistical methods

23. Define topic modeling and discuss its applications in natural language processing (NLP)
24. What is an algorithm? What are the tools to analyze algorithms
25. Define DFS with example.
26. Define Principle of Optimality
27. What is linear programming. What are the key components of a linear programming problem?
28. List out NP-hard and Np-complete problems
29. What is cross-validation, and why is it used in machine learning?
30. Provide an example of a real-world application where probabilistic modeling is used.
31. What is a randomized algorithm? Explain randomized quicksort algorithm with example.
32. Discuss the advantages of using divide and conquer algorithms, specifically in predictive science?
33. Explain about dictionary and hashing? What are the collision resolution techniques in hashing.
34. Apply Gale-Shapley algorithm to solve the Stable Marriages Problem? Write its algorithm and provide a step-by-step example.
35. What is personal genomics and discuss how data science is useful in personal genomics
36. Differentiate between NP-complete and NP-hard problems.
37. What is model selection,and why is it important in machine learning? Describe at least two common techniques for model selection.
38. Explain the difference between bagging and boosting in ensemble learning.
39. What is probabilistic modeling in the context of machine learning? How does it differ from deterministic modeling?
40. Compare and contrast machine learning and statistics. How do they differ in their goals, methodologies, and applications?
41. Provide an example of a problem that can be efficiently solved using dynamic programming, explaining the steps involved in the approach.
42. Describe the properties and advantages of binary search trees (BSTs) in comparison to other tree data structures