Dr. D. Y. Patil Educational Federation's
Dr. D. Y. PATIL COLLEGE OF ENGINEERING & INNOVATION
Survey No. 27/A/1/2C, Varale Campus, Near Talegaon Railway Station,
Tal. Maval, Dist. Pune 410 507, Ph: 020 48522561, 565,566
Web Site: www.dypcoei.edu.in, Email: principal.dypcoei@dypatilef.com

Dr. D. Y. Patil	Dr. Sushant Patil	Dr. Suresh Mali
Founder, Dr. D. Y. Patil Group	**President, DYPEF**	**Principal, DYPCOEI**

DSBDA Lab Oral Questions

1. Explain Data Frame with Suitable example.

2. What is the limitation of the label encoding method?

3. What is the need of data normalization?

4. What are the different Techniques for Handling the Missing Data?

5. Explain the methods to detect the outlier.

6. Explain data transformation methods.

7. Write the algorithm to display the statistics of Null values present in the dataset.

8. Write an algorithm to replace the outlier value with the mean of the variable.

9. Explain Measures of Central Tendency with examples.

10. What are the different types of variables? Explain with examples.

11. Which method is used to statistic the data frame? write the code.

12. How are Hadoop and Big Data related?

13. Explain the core components of Hadoop

14. Explain the features of Hadoop.

15. How is HDFS different from traditional NFS?

16. What is Regression?

17. Explain the Difference between Regression and classification.

18. What is Classification?

19. What is Naïve Base?

20. What is Decision Tree?

DYPCOEI-

21. What is TF & IDF write down its formula.

22. What is NumPY?

23. What is Scikit Learn?

24. What do you mean by Seaborn Library?

25. Explain Logistic Regression.

26. Explain the Confusion Matrix with Respect to Machine Learning Algorithms.

27. What Are False Positive and False Negative and How Are They Significant?

28. What Are the Differences Between Machine Learning and Deep Learning?

29. What is Supervised Learning?

30. What is Unsupervised Learning?

31. What is the 'training Set' and 'test Set' in a Machine Learning Model? How Much Data Will You Allocate for Your Training, Validation, and Test Sets?

32. How Do You Handle Missing or Corrupted Data in a Dataset?

33. How Do You Design an Email Spam Filter?

34. Explain the K Nearest Neighbour Algorithm.

35. What you did in your Mini Project explain it.