BITS Pilani
Pilani | Dubai | Goa | Hyderabad
Work Integrated Learning Programmes

Title of the Work

Group Number - Student Name, Student Name, Student Name,

Introduction to Data Science

M.Tech Data Science and Engineering

1. Business Understanding

<Understanding the business problem>
< What is the problem that you are trying to solve?>

2. Ideas

<Write some ideas here to solve the problem. >
< How can you use data to answer the questions? >

3. Selected Idea

<Write the selected idea.>
< What data do you need to answer the question?>

4. Data Acquisition

<How many files?>
<Read the data>
<Size of dataset>
<Numerical or character data>
<Balanced or imbalanced dataset >

5. Understanding

<How many features?>
< In what way can the data be visualized to get to the answer that is required?>

6. Wrangling 1

<Missing data, which imputation technique>
<discretization, which features and method>
<reduction, sampling/PCA? >
<Normalization?>
<Use table, plots to show >

7. Wrangling 2

<Missing data, which imputation technique>
<discretization, which features and method>
<reduction, sampling/PCA? >
<Normalization?>
<Use table, plots to show >

8. Feature Selection

<Two Feature engineering technique used should be mentioned here clearly and crisply>

9. Results of FE technique 1

Mention only those that are applicable in your approach.
<New features created?>
<Feature ranking technique?>
<Use horizontal bar chart to show the top 5/10 features found>

10. Results of FE technique 2

Mention only those that are applicable in your approach.
<New features created?>
<Feature ranking technique?>
<Use horizontal bar chart to show the top 5/10 features found>

11. Modeling

<Use two ML techniques to implement the idea should be mentioned here clearly and crisply. >

12. Results of ML technique 1

<Classification report>
<Confusion matrix>
<Silhouette coefficient / F1 score>
<Plot of decision boundary or clusters formed>

13. Results of ML technique 2

<Classification report>
<Confusion matrix>
<Silhouette coefficient / F1 score>
<Plot of decision boundary or clusters formed>

14. Comparison

<Use a table / graphs to compare the 4 results>
<FE1+ ML1 = result 1>
<FE2+ ML1 = result 2>
<FE1+ ML2 = result 3>
<FE2+ ML2 = result 4>

15. Conclusion

Here model means the optimal FE + ML combination. <Which model is better>
< Does the model used really answer the initial question or does it need to be adjusted?>
<Can you put the model into practice?>

16. Recommendation / Suggestion

<Recommendation / suggestion to solve the business problem.>
<Can you get constructive feedback into answering the question?>

NB: Using the principles of Data Science, Preprocessing, Feature engineering, data visualization and storytelling.

Single page