2b. Selected Idea

2 b <Write the selected idea.>

2 b< What data do you need to answer the question?>

3. Data Acquisition

3 a <How many files?>

3 b <Read the data>

3 c <Size of dataset>

3 d <Numerical or character data>

3 e <Balanced or imbalanced dataset >

1. Business Understanding

1 a i <Understanding the business problem>

1 a ii < What is the problem that you are trying to solve?>

2a. Ideas

2 a <Write some ideas here to solve the problem. >

2 a < How can you use data to answer the questions? >

5b. Wrangling 2

5 a ii Duplicate Check

5 a iii Label Encoder,

12. Recommendation / Suggestion

<Recommendation / suggestion to solve the business problem.>

<Got constructive feedback into answering the question?>

9a. Results of ML technique 1

9 a i <Classification report>

9 a ii <Confusion matrix>

9 a iii i< F1 score>

6. Feature Selection

6 a Manual Attribute Selection

6 b Correlation Matrix

6 c Chi Square (for feature selection)

6 d Info Gain (for feature selection)

9b. Results of ML technique 2

9 b i <Classification report>

9 b ii <Confusion matrix>

9 b iii <Silhouette coefficient / F1 score>

4. Understanding

4 a <How many features?>

4 b < In what way can the data be visualized to get to the answer that is required?>

5a. Wrangling 1

5 a i <Missing data, imputation technique>

5 a ii Converting data types of filled NAN attributes

8. Modeling

8 a Segregating outcome & predictor variables

8 c I <Using Decision Tree Model>

8 c ii <Using Logistic Regression Model>

>

7b. Results of FE technique 2

7 c Handling Outliers (0.95 quantile cap)

7a. Results of FE technique 1

7 a <Feature ranking technique?> (KBest)

7 b <Normalization?> (Min-max-scaling)

10. Comparison

<Use a Classification Reports / ROC Curve to compare the 4 results>

10 a <FE1+ ML1 = result 1>

10 a <FE2+ ML1 = result 2>

10 b <FE1+ ML2 = result 3>

10 b <FE2+ ML2 = result 4>

11. Conclusion

Here model means the optimal FE + ML combination. <Which model is better>

<Can you put the model into practice?>