|  |  |  |
| --- | --- | --- |
| Criteria | Meets Expectations | Does Not Meet Expectations |
| Data Understanding, Preparation and EDA (30%) | All data quality checks are performed, and all data quality issues are addressed in the right way (missing value imputation, removing duplicate data and other kinds of data redundancies, etc.). Data quality issues are clearly explained in comments.    Dummy variables are created properly, wherever applicable.    New metrics are derived, if applicable, and are used for analysis and modelling.    The data is converted to a clean format suitable for analysis. | All quality checks are not done, and data quality issues are not addressed correctly to an appropriate level.      Dummy variables are not created properly.      New metrics are not derived or are not used for analysis.    The data is not converted to a clean format suitable for analysis or is not cleaned using commands. |
| Model Building and Evaluation (45%) | Model parameters are tuned using correct principles, and the approach is explained clearly. Both the technical and business aspects are considered while building the model.      Correct variable selection techniques are used. A reasonable number of different models are attempted, and the best one is chosen based on key performance metrics.      Model evaluation is done using the correct principles, and appropriate evaluation metrics are chosen.      The results are on par with the best possible model on the data set.    The model is interpreted and explained correctly. The commented code includes a brief explanation of the important variables and the model in simple terms. | Parameters are not tuned enough or are tuned incorrectly. Relevant business aspects are not considered during model building.    Variable selection techniques are used incorrectly / not conducted. A variety of models are not considered, or a suboptimal one is finalised.    The evaluation process deviates from the correct model selection principles, inappropriate metrics are evaluated or are incorrectly evaluated.      The results are not on par with the best possible model on the data set.    The model is not interpreted and explained correctly. |
| Coding Guidelines (5%) | Guidelines (5%)    Appropriate comments are written wherever applicable.        If new variables are created, the names are descriptive and unambiguous.      The code is written concisely, wherever possible.        Overall, the code is readable and has appropriate indentations. | Comments are not written, thereby rendering the code difficult to understand.      Variables are poorly or ambiguously named.      The code is more complex than what is required by the problem.    Code readability is poor because of poor indentation / other reasons. |

The rubrics for the second part are given below:

The subjective questions carry 20% marks.