1. Project topic: (5 pts)
    - Is there a clear explanation about what this project is about? Does it state clearly which type of problem (e.g. classification vs. regression)?
    - Does it state the motivation or the goal (or why it's important) clearly?
2. Data: (5 pts)
    - Is data source properly quoted and described? (including links, brief explanations)
    - Do they explain the data description properly? The data description can include the data size (number of samples/rows, number of features/columns, bytesize if a huge file), data type of each feature (or just a summary if too many features- e.g. 10 categorical, 20 numeric features), description of features (at least some key features if too many), whether the data is multi-table form or gathered from multiple data source.
3. Data cleaning and EDA (10 pts)
   Things to consider:
    - Does it include clear explanations on how and why a cleaning is performed? (e.g.) the author decided to drop a feature because it had too many NaN values and the data cannot be imputed (e.g.) the author decided to impute certain values in a feature because the number of missing values were small and he/she was able to find similar samples OR, he/she used an average value or interpolated value, etc. (e.g.) the author removed some features because there are too many of them and they are not relevant to the problem, or he/she knows only a few certain features are important based on their domain knowledge judgement. (e.g.) the author removed certain sample (row) or a value because it is an outlier.
    - Does it include clear explanations on how and why an certain analysis (EDA) is performed?
    - Does it have proper visualization?
    - Does it have proper analysis? E.g. histogram, correlation matrix, feature importance (if possible) etc.
    - Does it have conclusions or discussions? E.g. the EDA summary, findings, discussing foreseen difficulties and/or analysis strategy.

4. Model building / Model choice How many models have been used and compared?
   Things to consider:

   - Which models are suitable for my problem and why?
   - Compare with a baseline model(s); baseline means a simple, easy to check model.

5. Model training (15 pts) Things to consider:

   - How to find best hyperparameters of a model?
   - Is there overfitting? How to check, how to mitigate?
   - Are all features needed? Why or why not? If so, how can I choose features?
   - What are the important features?
   - Do I have imbalanced classes in my data? What to do?

6. Results and Analysis: 20 pts Things to consider:

   - Does it have a summary of results and analysis?
   - Does it have a proper visualization? (e.g. tables, graphs/plots, heat maps, statistics summary with interpretation etc)
   - Does it use different kinds of evaluation metric properly? (e.g. if your data is imbalance, there are other metrics F1, ROC or AUC better than mere accuracy). Also does it explain why they chose the metric?
   - Does it iterate the training and evaluation process and improve the performance? Does it address selecting features through the iteration process?
   - Did the author compare the results from the multiple models and did appropriate comparison?

7. Discussion and Conclusion: 10 pts
   Provide conclusions from your result. Discuss what went well or not well. Discuss any suggestions to improve.