# Final Project Rubrics

**Note:** The grading rubric for the final project includes 9 categories. Below has the breakdown of the scores. The raw max score is 100.

## 1. Project topic: 5 pts

- Is there a clear explanation of what this project is about? Does it state clearly which type of problem (e.g. type of learning and type of task)?
- Does it state the motivation or the goal (or why it's important, what goal the team wants to achieve, or want to learn) clearly?
- (Extra credit) Is the project topic creative? Requires collecting data (e.g. scraping)?

## 2. Data: 5 pts

- Is the 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
    - e.g. for tabulated data: number of samples/rows, number of features/columns, byte size 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 a multi-table form or gathered from multiple data source.
    - e.g. for images: you can include how many samples, number of channels (color or gray or more?) or modalities, image file format, whether images have the same dimension or not, etc.
    - e.g. sequential data: texts, sound file; please describe appropriate properties such as how many documents or words, how many sound files with typical length (are they the same or variable), etc.

## 3. Data cleaning: 10 pts

Example score breakdown for tabulated data format: no cleaning 0 pts (if the data was given perfectly cleaned, just give a default score of 5 pts), data types munging +1, drop NA +1, impute +1, identify imbalance +1, and identify data-specific potential problem +1.

- Does it include clear explanations on how and why 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 was 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 judgment.
  - ○ (e.g.) the author removed a certain sample (row) or a value because it is an outlier.
  - ○ (e.g.) if the project is on text data, is stopword filtering conducted? If no, why not?
- Does it have conclusions or discussions? E.g. the data cleaning summary, findings, discussing foreseen difficulties, and/or analysis strategy.
- Does it have a proper visualization?

## 4. Exploratory Data Analysis (EDA): 15 pts

Example score breakdown: 9 for simple plots (histogram, box plot), correlation mat (+2), extra EDA (e.g. statistical tests) +2, with some insights/conclusion +2

- Does it include clear explanations on how and why an analysis (EDA) is performed?
- Does it have a 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.

## 5. Models: 20 pts

Example score breakdown for typical supervised learning: 8 if a proper single model is used, +2 if addresses multilinear regression/collinearity for regression models, +2 feature engineering, +2 multiple ML models, +2 hyperparam tuning, +2 regularization, or other training techniques such as cross-validation, oversampling/undersampling or similar for managing data imbalance, +2 using models not covered from class.

- Is the choice of model(s) appropriate with the problem?
- Is the author aware of whether interaction/collinearity between features can be a problem for the choice of the model and properly treat if there is interaction or collinearity (e.g. linear regression)? Or confirms that there is no such effect with the choice of the model?

- Did the author use multiple (appropriate) models?
- Did the author investigate which ones are important features by looking at feature rankings or importance from the model? (Not by a judgment which we already covered in the EDA category) Did the author use techniques to reduce overfitting or data imbalance?
- Did the author use new techniques/models we didn't cover in the class?

## 6. Results and Analysis: 20 pts

Example score breakdown: showing basic result 12, with a good amount of visualization +2, try different evaluation metrics +2, iterate training/evaluating and improve performance +2, show/discuss model comparison +2

- 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 metrics properly? (e.g. if your data is imbalanced, 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 an appropriate comparison?

## 7. Discussion and Conclusion: 10 pts

Example score breakdown: basic reiteration of the result 6, discussion on what are the learnings and takeaways, discussion on why something didn't work +2, suggesting ways to improve +2.

## 8. Write-ups quality: 5 pts

- Is the write-up organized and clear?
- Are the codes commented and organized?
- Did the authors use git?

## 9. Midterm Presentation quality: 10 pts

- Is the presentation clear and organized?
- Did the presenter give explanations on how things work (e.g. concepts of a model or technique), rationales for analysis, walk-thru of a graph/plot, interpretation of results/metrics, and discussion/conclusion?