

Big Data Project Machine Learning

Outline

- All Big Data Project submissions should be in a font, style, and margins consistent with a professional looking document.
 - There are no strict rules with respect to line spacing or font size.
 - Recommended length of this milestone: 5 -10 pages. Please number the pages!
 - Also submit your Python code (Do not copy the code in to the report, submit the Jupyter notebook as a separate file)
 - » Use Jupyter Notebook

Machine Learning – Methodology

- Implement Supervised Learning
(Classification/Regression)
following the Machine learning template:
 1. Define Problem
(What are the features, what is the target?)
 2. Briefly Summarize Data
 3. Prepare Data (rescaling optional, see next page)
 4. Evaluate Algorithms
 5. Improve Results (optional, see next page)
 6. Present/Summarize Results

Requirements

- Minimum:
One Classification OR one Regression scenario
- Good:
One Classification AND one Regression scenario
- Even Better:
 - Additionally consider rescaling data
(see 3. Prepare Data)
 - Additionally include parameter tuning
(see 5. Improve Results)

Evaluation Criteria 1

- Data preparation (20%)
 - Preparing data to apply the Machine learning algorithms
 - Examples: Data encoding, normalization and binning
 - Feature selection, Test/Train splits
- Using Models (30%)
 - Using ML models (At least 1 model should be used.
But we encourage you to explore different models.)
 - Even Better: Improving performance of model
 - » E.g.: Model parameter tuning (Hyper Parameter Tuning)

Evaluation Criteria 2

- Model Evaluation (20%)
 - Use different evaluation metrics to explain the model performance
 - Examples: Confusion Matrix, Precision and Recall, RMSE etc.
- Insights and Summarizing results (30%)
 - Clearly explain your results (Are your predictions meaningful?) If not, how can you improve them?
 - Discuss how are your evaluation metrics helping you to better understand the model performance
 - How well did your model help you achieve the objective?

Submission

Submit:

- Jupyter Notebook (#comment code wherever necessary. Use markdowns if required.)
- Report (Should contain your insights and explanation of the model evaluation. Do not copy the code into the report. You can copy the tables, results or visualizations.)