

Data Science Career Track

Capstone: Pre-processing and Training Rubric

Pre-processing and Training Data Development

Learning Objectives

- Understand the importance of creating a model training development data set.
- Correctly identify when to create dummy features or one-hot encoded features.
- Understand the importance of magnitude standardization.
- Apply the train and test split to the development dataset effectively.

Criteria	Meets Expectations
Completion	<ul style="list-style-type: none"><input type="checkbox"/> Every step identified in the DSM has been addressed or completed, including:<ul style="list-style-type: none"><input type="checkbox"/> Creating dummy features<input type="checkbox"/> Scale standardization<input type="checkbox"/> Split data into training and testing subsets
Process and understanding	<ul style="list-style-type: none"><input type="checkbox"/> The submission shows that the student understands when to apply the steps identified in the completion block.
Presentation	<ul style="list-style-type: none"><input type="checkbox"/> Jupyter Notebook with all the applied code steps in working order and with notation or comments as needed.<input type="checkbox"/> The submission is complete and uploaded in full to the Github repo.