**CLO 2:** Train supervised learning models for classification tasks with categorical labels

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | **At Risk (1)** | **Initial Mastery (2)** | **Approaching Mastery (3)** | **Mastery (4)** |
| **Problem Identification**  Identify classification task  Selection of column as target label | * Misidentified problem as classification task * Selects inappropriate column as target label | * Correctly identifies problem as classification task * Selects appropriate column as target label | * Correctly identifies problem as classification task with binary or multiclass labels * Provides justification for feature columns selected to predict target label | |
| **Dataset Preparation**  Removal of target label column  Data types and encoding  Scaling or transformation of values | * Does not remove label column from training data * Uses incorrect data types or encoding * Does not scale values where appropriate | * Removes target label column from training data * Uses integer encoding where appropriate but may introduce side effects * Scales or normalizes values where appropriate | * Uses one-hot encoding to avoid side effects where appropriate * Scales, normalizes, or transforms values where appropriate | * Scales, normalizes, transforms, or bins values where appropriate * Strategically incorporates dimensionality reduction techniques (such as PCA) where appropriate |
| **Model Selection and Hyperparameter Tuning**  Selection of appropriate model type  Tuning of model hyperparameters | * Only tries one model type * Selects inappropriate model type * Does not tune model hyperparameters | * Tries at least two model types * Selects appropriate model type * Attempts to tune model hyperparameters with manual guesses | * Tries at least four distinct model types * Selects appropriate model type based on accuracy * Uses for-loops to tune model hyperparameters | * Justifies choice of model type based on accuracy and model type analysis * Uses best practices to tune hyperparameters (such as grid search or Bayesian optimization) |
| **Communicating Results**  Comments and explanations of process  Analysis and interpretation of results  Write-up and presentation | * No comments or explanation of process or decisions made * Very little or no analysis or interpretation of results * Write-up or presentation unclear or communicated ineffectively to audience | * Some comments and explanation of process and decisions made * Basic analysis and interpretation of results * Written or presented to communicate results to technical audience | * Frequent comments and explanation of process and decisions made * Thorough analysis and interpretation of results * Written or presented to communicate results to technical audience | * Thorough analysis and interpretation of results including discussion of trade-offs (e.g., inductive bias of model choice, limitations of dataset, etc) * Written or presented to communicate results to non-technical audience |