Milestone 2 – Task 5 Evaluation

Criterion	Decision Tree	PART	JRip
Rule Perspicacity	2	3	1
Mean Accuracy (Clean Data)	1	3	2
Stability	1	3	2
Robustness to Noise (10%)	1	3	2
Overall (Average Rating)	1.25	3.00	1.75

Overall, the Decision Tree model demonstrated the strongest balance of predictive performance and robustness. ANOVA indicated significant differences across models in both the clean ($p = 0.03$) and noisy ($p < 0.001$) datasets. Tukey post-hoc analysis confirmed that PART significantly underperformed the Decision Tree in both conditions, while JRip was not significantly different from the Decision Tree. Under 10% label noise, all models experienced performance degradation. The Decision Tree declined from approximately 72% accuracy to 68.6%, JRip declined from approximately 71% to 67%, and PART showed the largest decrease. This indicates that the Decision Tree maintains the highest robustness to corrupted supervision. Stability refers to how consistent each model's accuracy is across the repeated cross-validation resamples (i.e., how much the accuracy fluctuates from fold to fold). In this study, the Decision Tree and JRip showed relatively consistent resample performance, whereas PART was less stable and more sensitive to the particular training/validation split, which is consistent with its lower mean accuracy and larger degradation under label noise. In terms of rule perspicacity, JRip produced only 5 rules, making it the most concise and interpretable model. The Decision Tree generated 15 terminal rules, offering moderate complexity. PART produced 83 rules, resulting in the least interpretable rule base. Overall, the Decision Tree provides the strongest generalization performance, JRip offers the best interpretability, and PART demonstrated the weakest overall performance under both clean and noisy conditions.