

Assignment 5

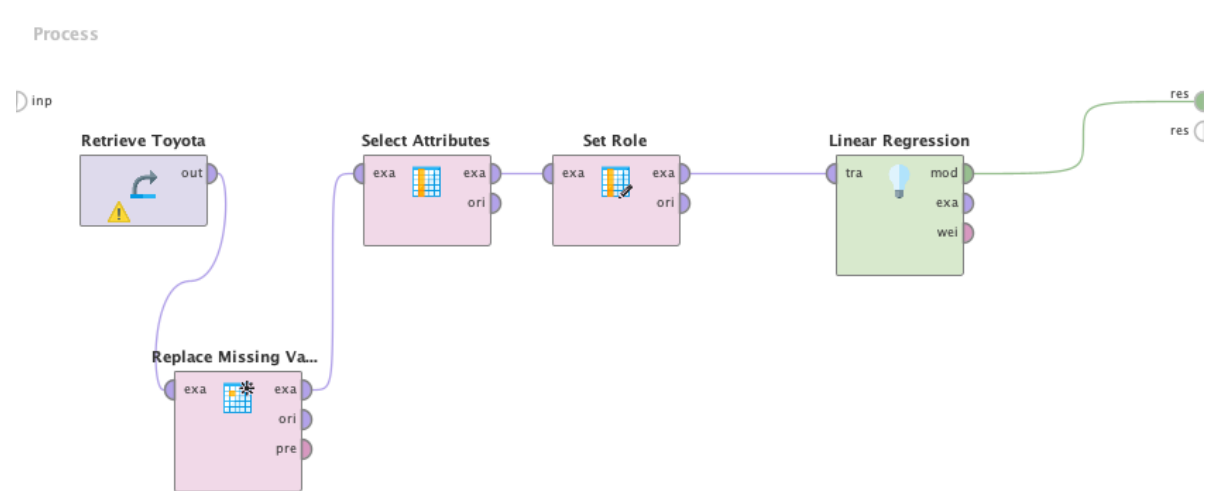
Name: Priyanshi Deliwala

Task 1: Car Sales Data

Target variable: price

Selected Attributes: age_08_04 cc doors gears hp km price quarterly_tax weight

Model 1: Linear Regression

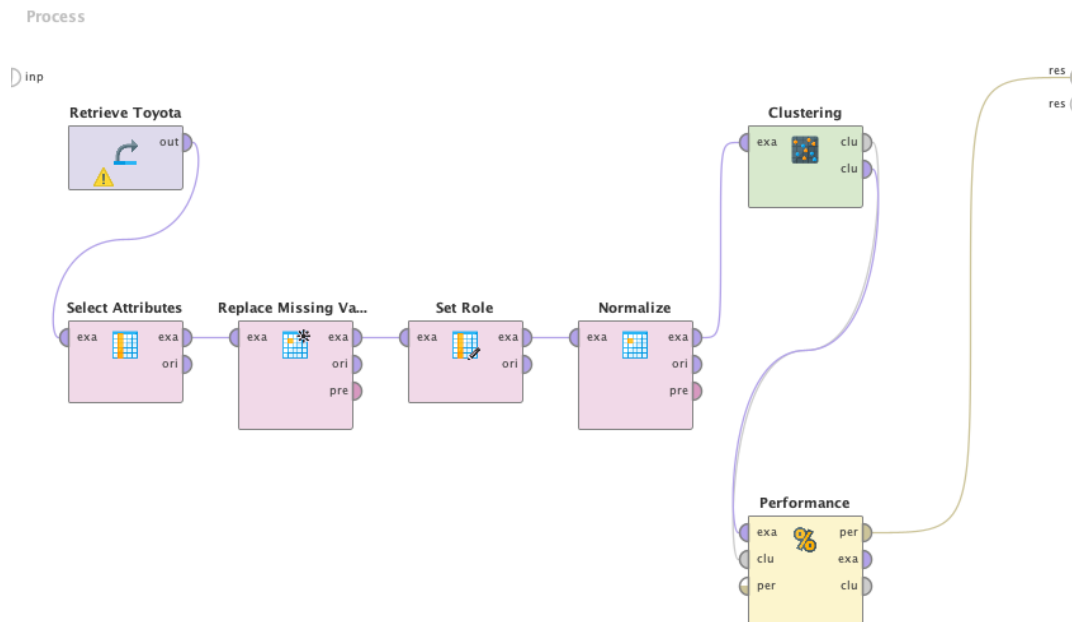


Result:

Views: Design Results Turbo Prep Auto Model Interactive Analysis							
Result History LinearRegression (Linear Regression) X							
Attribute	Coefficient	Std. Error	Std. Coefficient	Tolerance	t-Stat	p-Value	Code
age_08_04	-121.658	2.615	-0.624	0.547	-46.528	0	****
km	-0.021	0.001	-0.215	0.761	-16.636	0	****
hp	31.673	2.810	0.131	0.942	11.270	0	****
cc	-0.121	0.090	-0.014	0.978	-1.344	0.179	
gears	595.832	193.400	0.031	0.999	3.081	0.002	***
quarterly_tax	3.953	1.306	0.045	0.954	3.027	0.003	***
weight	16.948	1.033	0.246	0.729	16.401	0	****
(Intercept)	-5575.140	1409.999	?	?	-3.954	0.000	****

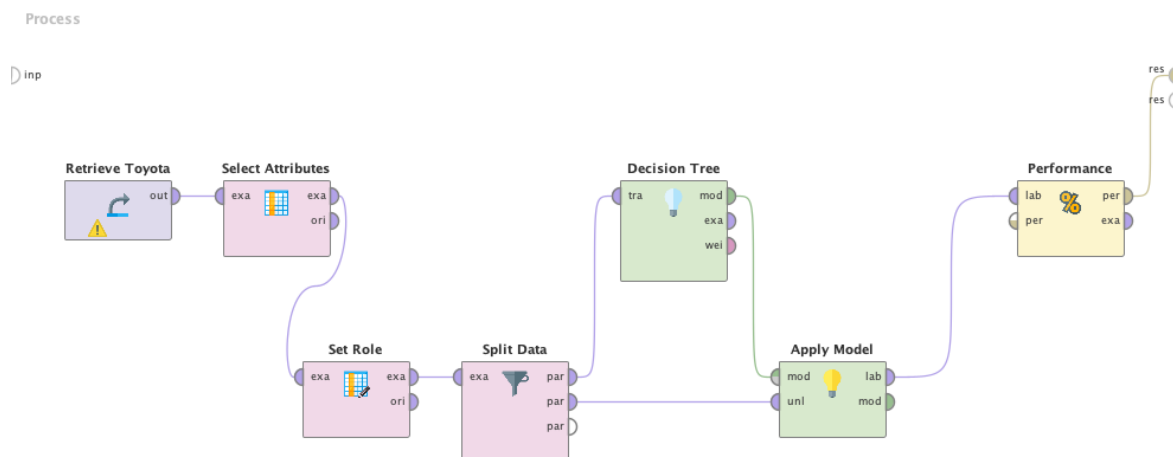
Interpretation: In summary, the results indicate that weight, kilometers driven, horsepower, the number of gears, and the age of the vehicle have statistically significant effects on the price of the vehicle. Engine displacement (cc) does not appear to have a statistically significant effect.

Model 2: K-means Clustering

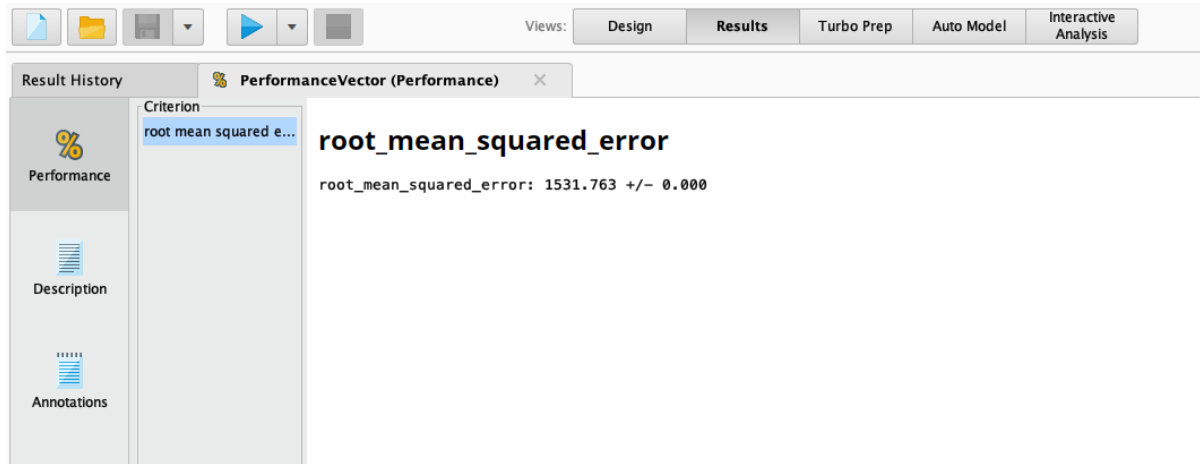


Interpretation: The "Avg. within centroid distance" of -6.549 suggests that the k-means clustering with K = 3 has resulted in relatively tight and well-separated clusters. The negative value, while unusual in interpretation, still indicates that the data points are closely associated with their respective cluster centroids.

Model 3: Decision Tree



Result:



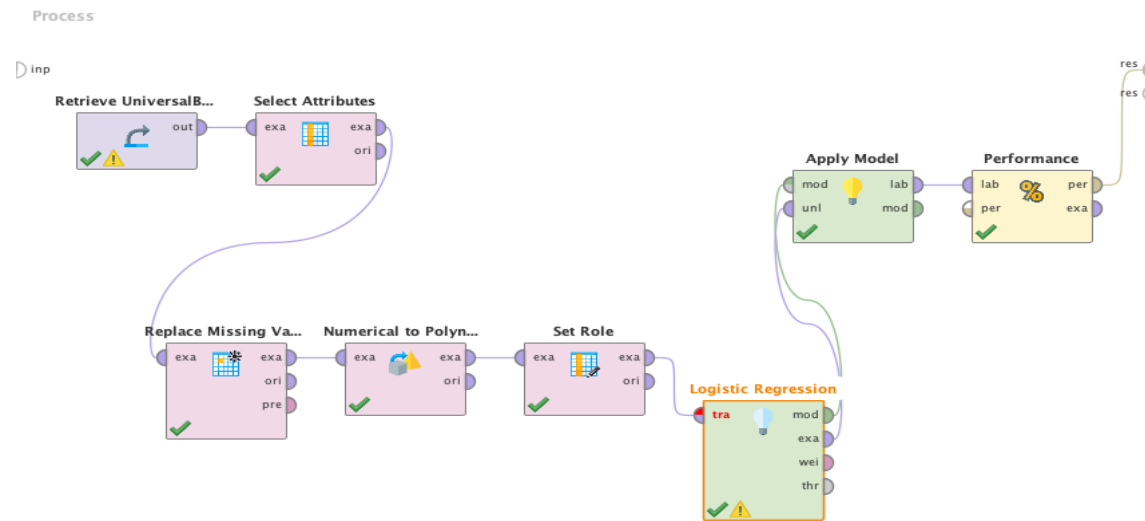
Interpretation: The RMSE (Root Mean Squared Error) value of approximately 1531.763 is a measure of the average error in the predictions made by your decision tree model. Specifically, it quantifies how much, on average, your model's predictions deviate from the actual values in your dataset. Lower RMSE values are desirable as they indicate that the model's predictions are, on average, closer to the actual values, which suggests better predictive accuracy.

Key Observation: Based on the results provided, the linear regression model appears to be the most suitable for predicting the price of a product or vehicle. It offers interpretability with statistically significant coefficients, allowing for a clear understanding of the impact of each attribute on the target variable. The k-means clustering result includes an unusual negative value in "Avg. within centroid distance," which may require further investigation, while the decision tree result presents an extremely low standard deviation in RMSE, suggesting very low variability in prediction errors. Therefore, linear regression is favored for its interpretability and relevance to price prediction.

Task 2: Bank Customer Data

Selected Attributed: Age CreditCard Experience Income Mortgage

Model 1: Logistic Regression



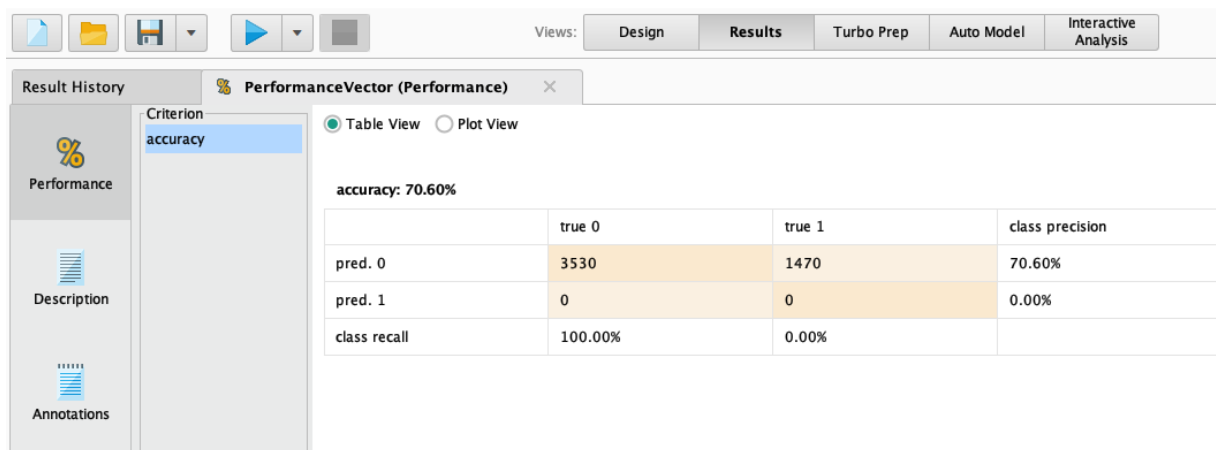
Result:

Views: Design Results Turbo Prep Auto Model Interactive Analysis

Result History

Logistic Regression Model (Logistic Regression)

Attribute	Coefficient	Std. Coefficient	Std. Error	z-Value	p-Value
Age	-0.021	-0.242	0.026	-0.828	0.408
Experience	0.023	0.260	0.026	0.890	0.374
Income	-0.000	-0.003	0.001	-0.104	0.917
Mortgage	-0.000	-0.016	0.000	-0.488	0.626
Intercept	-0.360	-0.876	0.657	-0.548	0.584

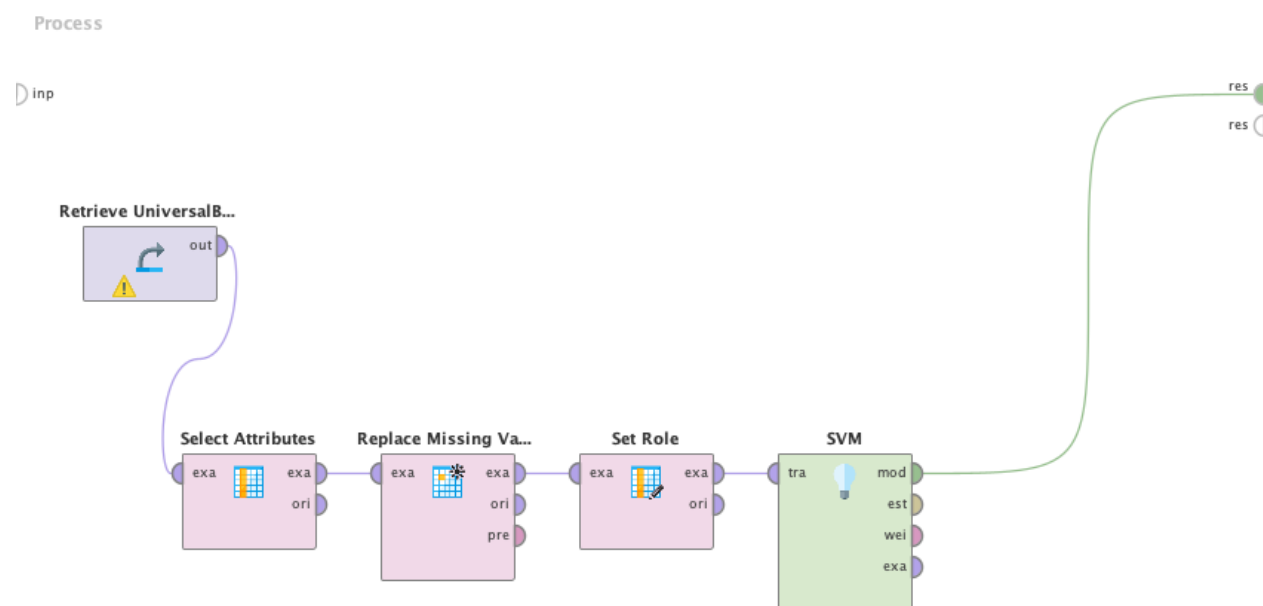


Interpretation:

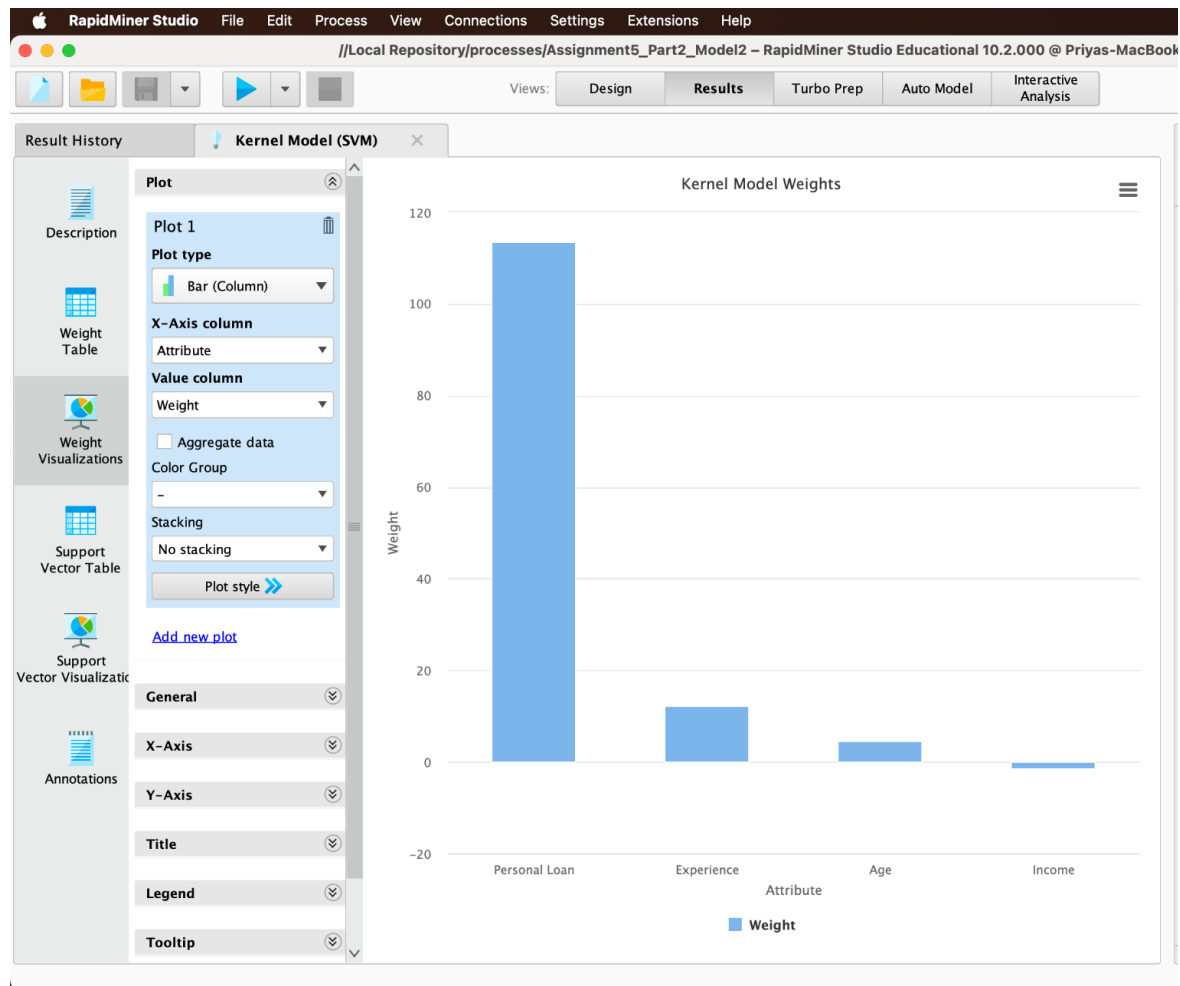
Confusion matrix and performance metrics suggest the following:

- The model correctly identified all cases of "Credit card approval" (recall of 100.00%), meaning that it has a high ability to identify individuals who should be approved for a credit card.
- The precision is 70.60%, indicating that out of all the instances predicted as "Credit card approval," approximately 70.60% are indeed approved. This suggests a relatively good balance between true positive predictions and false positive predictions.

Model 2: Support Vector Machine(SVM)



Result:

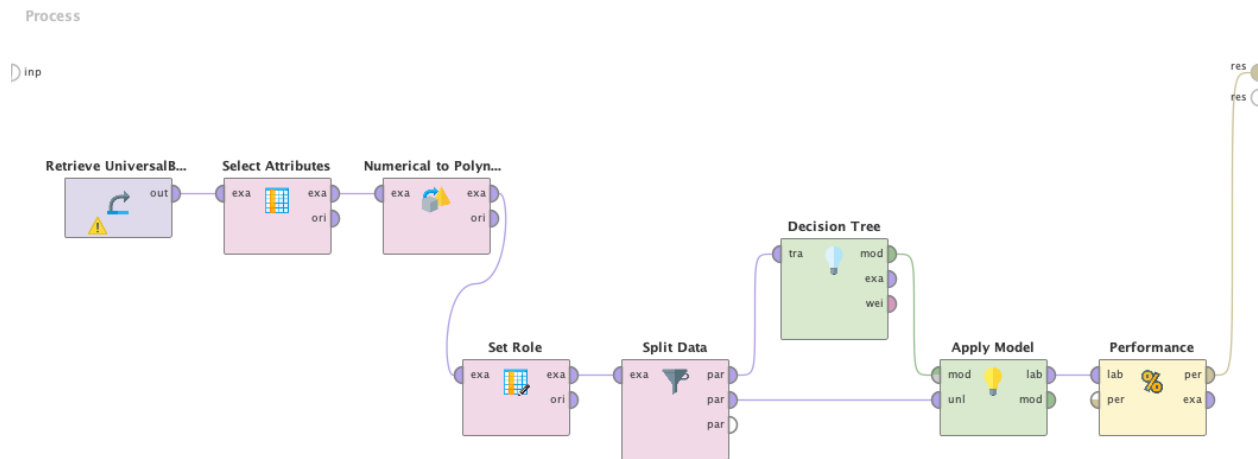


Interpretation:

- The positive weights for "Age" and "Experience" indicate that as individuals get older and gain more experience, they are more likely to be approved for a credit card.
- The negative weight for "Income" suggests that higher income may be associated with a reduced likelihood of credit card approval.
- The substantial positive weight for "Personal Loan" indicates that the presence of a personal loan application is a strong predictor of credit card approval, with a positive impact.
-

Overall, the SVM model appears to consider "Age," "Experience," "Income," and the presence of a "Personal Loan" application as significant factors in determining credit card approval. These insights can be valuable in understanding the driving factors behind credit card approval decisions and can guide business strategies related to lending and risk assessment.

Model 3: Decision Tree



Result:

Views: Design Results Turbo Prep Auto Model Interactive Analysis

Result History

PerformanceVector (Performance)

Table View Plot View

accuracy: 70.53%

	true 0	true 1	class precision
pred. 0	1058	441	70.58%
pred. 1	1	0	0.00%
class recall	99.91%	0.00%	

Interpretation:

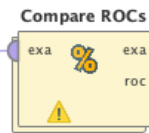
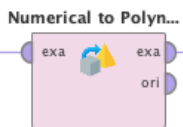
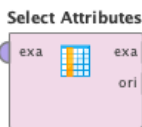
The decision tree model's performance can be summarized as follows:

- **High Recall:** The model has a very high recall (99.91%), which indicates its ability to correctly identify nearly all actual negative cases ("true 0"). This suggests that the model is effective at minimizing false negatives, meaning it correctly identifies the vast majority of non-approval cases.
- **Moderate Precision:** The precision is 70.58%, indicating that 70.58% of the instances predicted as "true 0" are indeed true negatives. While the precision is not as high as the recall, it still suggests that a majority of the negative predictions are correct.

ROC CURVE

Process

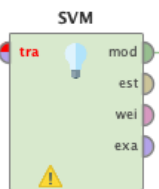
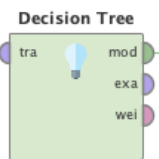
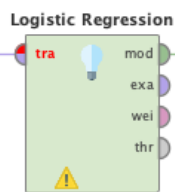
inp



res
res
res

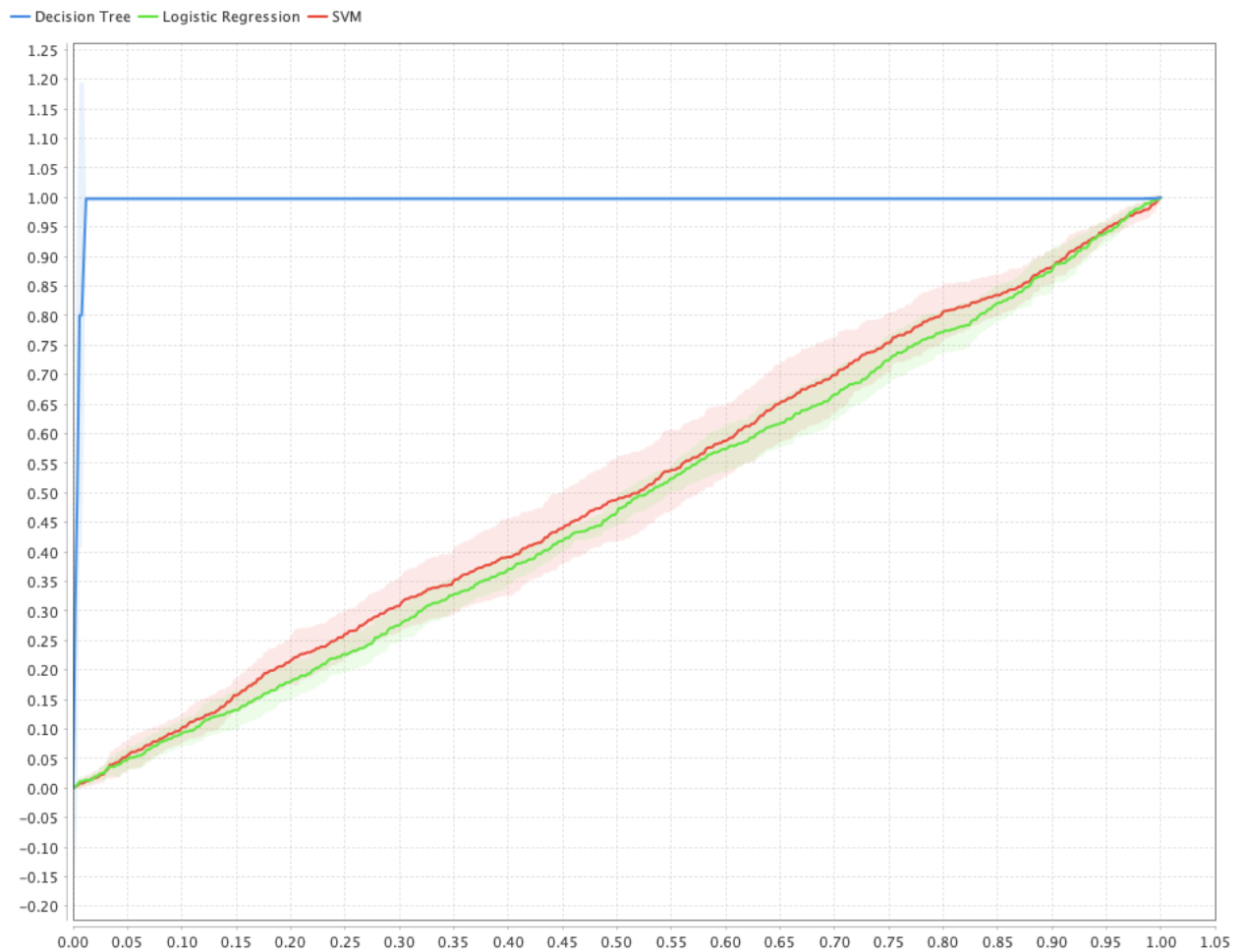
Compare ROCs

tra
tra
tra
tra



mod
mod
mod
mod

Result:



Interpretation: The order of the ROC curves suggests that, in terms of classifying positive and negative cases, the decision tree performs the best, followed by the SVM, and then logistic regression. As the AUC is highest for Decision Tree.