

MACHINE LEARNING

using RapidMiner and Weka
on Indian Liver Patient Dataset

Artificial Intelligence
CPCS-331 Individual Project
Saturday, 10 June 2023

Student name	ID	Section
Noor [REDACTED]	[REDACTED]	B2
Nsreen Hujjatullah Asadullah	[REDACTED]	B2
Hano [REDACTED]	[REDACTED]	B2
Sara [REDACTED]	[REDACTED]	B2



Table of Contents

Task Assignment	3
1. Introduction	4
1.1. Introduce the topic of the report.....	4
1.2. Explain the disease selected for the project.....	4
1.3. Purpose of the project.....	4
1.4. Outline approach	4
2. Technical Description.....	5
2.1. Description of the Dataset.....	5
2.2. Description of the Chosen Machine Learning Algorithms.....	6
2.2.1. Naïve Bayes.....	6
2.2.2. Random Forest	6
2.2.3. Support Vector Machine	6
2.2.4. Decision Stump	7
3. Results	8
3.1. The Experimental Results of Naïve Bayes.....	8
3.1.1. Naïve Bayes Results Analysis.....	8
3.2. The Experimental Results of Random Forest.....	8
3.2.1. Random Forest Results Analysis.....	8
3.3. The Experimental Results of Support Vector Machine	9
3.3.1. Support Vector Machine Results Analysis	9
3.4. The Experimental Results of Decision Stump.....	9
3.4.1. Decision Stump Results Analysis	9
4. Conclusion	10
5. References.....	11
6. Appendix A: Naïve Bayes Application.....	12
7. Appendix B: Random Forest Application.....	17
8. Appendix C: Supportive Vector Machine	23
9. Appendix D: Decision Stump.....	29



Task Assignment

Team Member	Task
All team members.	Introduction, Technical Description, and Conclusion
Nsreen Hujjatullah Asadullah	Naïve Bayes Alg., its experiment, and result
Noor [REDACTED]	Random Forest Alg., its experiment, and result
Hanoof [REDACTED]	Support Vector Machine Alg., its experiment, and result
Sara Turki Alyami	Decision Stump Alg., its experiment, and result



1. Introduction

1.1. Introduce the topic of the report.

Artificial intelligence affects the future of every industrial sector and every human being on this planet. It is the main engine of all emerging technologies such as mass data collection, robotics and the Internet of things, and artificial intelligence is expected to play a greater role in the coming years.

Machine learning is a subset of artificial intelligence (AI) that focuses on creating systems that learn - or improve performance - based on the data they consume. Artificial intelligence is a broad term that refers to systems or devices that simulate human intelligence. Machine learning and artificial intelligence are often discussed together, and the terms are sometimes used interchangeably, but they don't mean the same thing. It is important here to mention that although all machine learning techniques are AI, not all AI is machine learning. In this report, we will show results for different algorithms on the same dataset with different software Weka, RapidMiner (2014)

1.2. Explain the disease selected for the project.

The disease selected for the project is liver disease. Liver disease refers to a range of conditions that impair the normal functioning of the liver. Common symptoms include fatigue, jaundice (yellowing of the skin and eyes), and abdominal pain. The project aims to utilize machine learning algorithms, such as forest tree, naive Bayes, decision stump, and support vector machine, to analyze the Indian liver patient dataset. By comparing the accuracy of these algorithms, the project seeks to identify the most suitable algorithm for diagnosing liver disease based on the dataset's performance evaluation. (Liver Diseases, 2019)

1.3. Purpose of the project

This project aims to teach students the basics of machine learning while attempting to execute it in the simplest way possible. The primary focus was on separating the dataset into testing and training data. Aside from that, we hope to discover the relationship and impact of certain a group of different medical analyzes or variables on the outcome. As a result, we chose the Selector attribute as our label.

1.4. Outline approach

The approach followed to find the appropriate algorithms is as follows, first, we have chosen a dataset with clear behaviors and inputs such as age, gender, Total Bilirubin, Direct Bilirubin, ...etc., to reach the goal to be deduced which is whether the person has liver disease or not based on the data analyzed beforehand. Then we studied some ML algorithms whose uses match the specifications of the dataset and chose them for testing



2. Technical Description

2.1. Description of the Dataset

This data set contains 416 liver patient records and 167 non-liver patient records. The data set was collected from the northeast of Andhra Pradesh, India.

The selector is a class label used to divide into groups (liver patient or not). This data set contains 441 male patient records and 142 female patient records.

1. Selector (Label)
2. Age of the patient (Integer)
3. Gender of the patient (Female, Male)
4. TB Total Bilirubin (Real)
5. DB Direct Bilirubin (Real)
6. Alkphos Alkaline Phosphatase (Integer)
7. Sgpt Alamine Aminotransferase (Integer)
8. Sgot Aspartate Aminotransferase (Integer)
9. TP Total Proteins (Real)
10. ALB Albumin (Real)
11. A/G Ratio Albumin and Globulin Ratio (Real)

ExampleSet (/Local Repository/data/Indian Liver Patient Dataset (ILPD))

Open in [Turbo Prep](#) [Auto Model](#) Filter (583 / 583 examples): all

Row No.	Selector field	Age of the p...	Gender of th...	TB Total Bili...	DB Direct B...	Alkphos Alk...	Sgpt Alamin...	Sgot Aspa...
1	yes	65	Female	0.700	0.100	187	16	18
2	yes	62	Male	10.900	5.500	699	64	100
3	yes	62	Male	7.300	4.100	490	60	68
4	yes	58	Male	1	0.400	182	14	20
5	yes	72	Male	3.900	2	195	27	59
6	yes	46	Male	1.800	0.700	208	19	14
7	yes	26	Female	0.900	0.200	154	16	12
8	yes	29	Female	0.900	0.300	202	14	11
9	no	17	Male	0.900	0.300	202	22	19
10	yes	55	Male	0.700	0.200	290	53	58
11	yes	57	Male	0.600	0.100	210	51	59
12	yes	72	Male	2.700	1.300	260	31	56
13	no	64	Male	0.900	0.300	310	61	58

ExampleSet (583 examples, 1 special attribute, 10 regular attributes)



2.2. Description of the Chosen Machine Learning Algorithms

2.2.1. Naïve Bayes

Naïve Bayes is a classification algorithm that takes input data and categorizes it into predefined classes. The algorithm calculates the probability of each class given the input data using Bayes' theorem, which states that the probability of a hypothesis (in this case, a class) is equal to the prior probability of that hypothesis multiplied by the likelihood of the data given that hypothesis, normalized by the probability of the data. The "naïve" part in the algorithm's name comes from its assumption that the features or attributes in the input data are independent, simplifying the calculation of the class probabilities. Despite its simplicity, Naïve Bayes is effective in many applications, such as text classification and spam filtering. (Ray, 2019).

2.2.2. Random Forest

Random Forest is one of the most popular and commonly used algorithms by Data Scientists. Random forest is a Supervised Machine Learning Algorithm that is used widely in Classification and Regression problems. It builds decision trees on different samples and takes their majority vote for classification and average in case of regression (E R, 2021).

One of the most important features of the Random Forest Algorithm is that it can handle the data set containing continuous variables, as in the case of regression, and categorical variables, as in the case of classification. It performs better for classification and regression tasks (E R, 2021).

2.2.3. Support Vector Machine

Support Vector Machines are powerful and flexible supervised machine-learning algorithms used for classification and regression. Recently, it has become very popular due to its ability to handle many continuous and categorical variables. The work of SVM is basically a representation of different classes of a hyperplane in a multidimensional space. The super level will be generated in an iterative manner by the SVM so that the error can be minimized. The goal of the support vector machine is to divide data sets into classes to find the maximum marginal hyperlevel (MMH).



There are main concepts in SVM:

- Support Vector: The data points closest to the hyperlevel are called support vectors.
- Hyperplane: A decision plane or space divided between a group of objects of different classes.
- Margin: It can be defined as the gap between two lines on Treasury data points from different categories. (accessed Jun. 10, 2023)

2.2.4. Decision Stump

Decision stump is one of the most straightforward classification algorithms. It is a binary classification algorithm. Binary classification is for classifying two categories, such as 0 and 1. And only focus on one feature each time and find a point that can separate data the most. Decision stumps are light-weighted, incremental, and easy to implement and understand, but more importantly, an elegant approach to identifying which feature maximizes a purity criterion and selecting a feature accordingly. (Hsu, 2022)



3. Results

3.1. The Experimental Results of Naïve Bayes

Percentage / Folds		Weka Accuracy	Rapid Miner Accuracy
Split Validation	65% Data, 35% Test	54.41%	60.78%
	70% Data, 30% Test	55.42%	61.71%
	85% Data, 15% Test	55.17%	54.02%
Cross Validation	4 folds	55.23%	55.74%
	8 folds	55.40%	56.11%
	16 folds	55.23%	55.93%

3.1.1. Naïve Bayes Results Analysis

As we can see in the table above, the results from both Weka & RapidMiner range between 55% - 61%, but we also note that RapidMiner results in both split & cross-validation are better than Weka. Therefore, we can say that RapidMiner can be a better choice when using the Naïve Bayes algorithm.

In addition, when looking at the best 4 accuracies for each tool and taking the average, we can conclude that the accuracy of the Naïve Bayes algorithm is 57%.

3.2. The Experimental Results of Random Forest

Percentage / Folds		Weka Accuracy	Rapid Miner Accuracy
Split Validation	65% Data, 35% Test	67.6471 %	71.57%
	70% Data, 30% Test	68 %	72.57%
	85% Data, 15% Test	66.6667 %	73.56%
Cross Validation	4 folds	70.8405 %	72.21%
	8 folds	69.9828 %	71.18%
	16 folds	68.6106 %	71.36%

3.2.1. Random Forest Results Analysis

As we can see in the table above, the results from both Weka & RapidMiner range between 66% - 73%, but we also note that RapidMiner results in both split & cross-validation are better than Weka. Therefore, we can say that RapidMiner can be a better choice when using the Random Forest algorithm.

In addition, when looking at the best 4 accuracies for each tool and taking the average, we can conclude that the accuracy of the Random Forest algorithm is 71% which is good.



3.3. The Experimental Results of Support Vector Machine

Percentage / Folds		Weka Accuracy	Rapid Miner Accuracy
Split Validation	65% Data, 35% Test	71.56%	71.57%
	70% Data, 30% Test	72.57%	71.43%
	85% Data, 15% Test	70.11%	71.26%
Cross Validation	4 folds	71.35%	71.35%
	8 folds	71.35%	71.35%
	16 folds	71.35%	70.83%

3.3.1. Support Vector Machine Results Analysis

Using the SVM algorithm and based on our results, we observed that the results were slightly similar, so using Weka were more accurate than those obtained with RapidMiner. We also observed that both validations provided results close to some, but that split validation was better than cross-validation for accuracy.

3.4. The Experimental Results of Decision Stump

Percentage / Folds		Weka Accuracy	Rapid Miner Accuracy
Split Validation	65% Data, 35% Test	68.62%	72.55%
	70% Data, 30% Test	68%	72.57%
	85% Data, 15% Test	73.56%	72.41%
Cross Validation	4 folds	71.35%	72.38%
	8 folds	71.35%	72.38%
	16 folds	71.35%	72.38%

3.4.1. Decision Stump Results Analysis

As we can see in the table above, the results from both Weka & RapidMiner range between 68% - 73%, but we also note that RapidMiner results in both split & cross-validation are better than Weka. Therefore, we can say that RapidMiner can be a better choice when using the Decision Stump algorithm.

In addition, when looking at the best 4 accuracies for each tool and taking the average, we can conclude that the accuracy of the Decision Stump algorithm is 72% which is good.



4. Conclusion

Algorithm	Average accuracy
Decision Stump	72%
Random Forest	71%
Support Vector Machine	71%
Naïve Bayes	57%

When comparing the results between the four classifiers, we see that the highest results were for the Random Forest algorithm, the Decision Stump algorithm, and the Support Vector Machine algorithm, ranging from 71% - 72%, while the naive Bayes algorithm was the lowest, with a range of 57%. But the highest average accuracy was obtained by the decision stem algorithm, so it is the most accurate and suitable algorithm for building the prediction model we aim for.



5. References

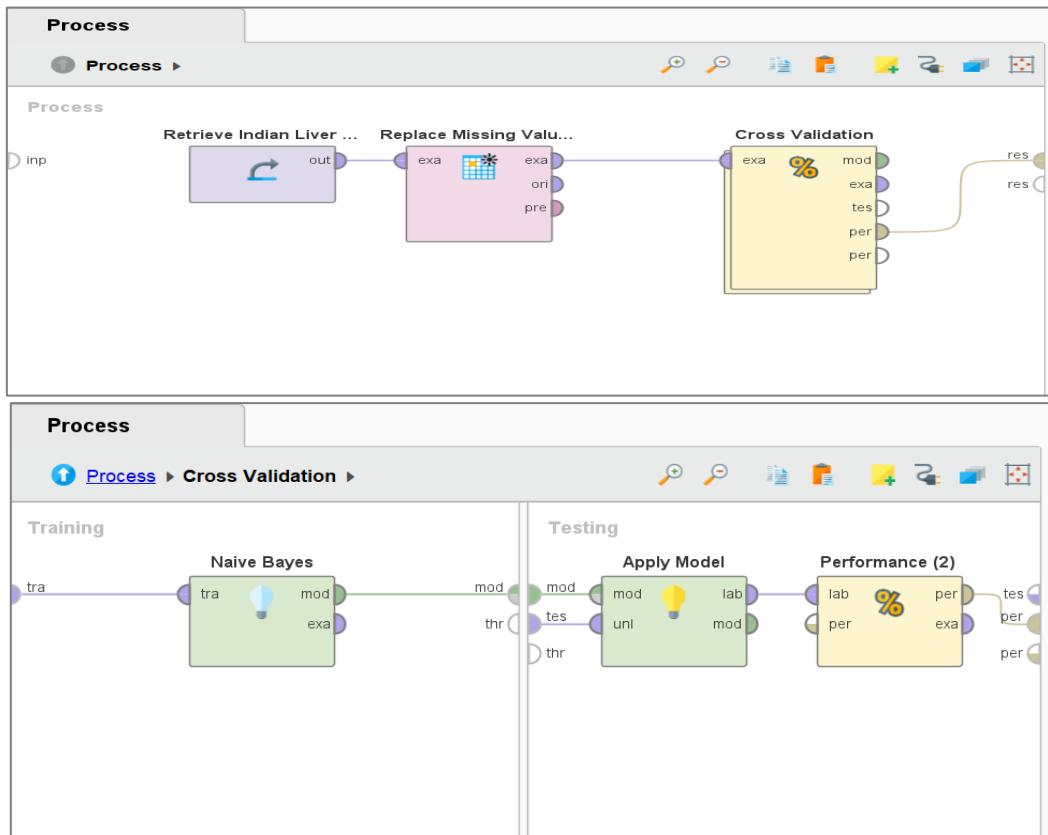
- Ray, S. (2019, September 3). *6 Easy Steps to Learn Naive Bayes Algorithm (with code in Python)*. Analytics Vidhya.
<https://www.analyticsvidhya.com/blog/2017/09/naive-bayes-explained/> (Ray, 2019)
- E R, S. (2021, June 17). *Random Forest / Introduction to Random Forest Algorithm*. Analytics Vidhya.
<https://www.analyticsvidhya.com/blog/2021/06/understanding-random-forest/>
- ما هو التعلم الآلي؟ (2014). Oracle.com.
<https://www.oracle.com/middleeast-ar/artificial-intelligence/machine-learning/what-is-machine-learning/> (2014)
- R. Hsu, “Decision Stump,” *Geek Culture*, May 02, 2022.
<https://medium.com/geekculture/decision-stump-b8e93c1f54d7> (Hsu, 2022)
- موقع كرسي للتعليم، حسن خنيري، ”آلة المتوجهات الداعمة أو SVM،“ Apr. 15, 2021.
<https://coursee.org/blog/artificial-intelligence/support-vector-machine/> (accessed Jun. 10, 2023).
- Liver Diseases. (2019). Medlineplus.gov; National Library of Medicine. <https://medlineplus.gov/liverdiseases.html> (Liver Diseases, 2019)



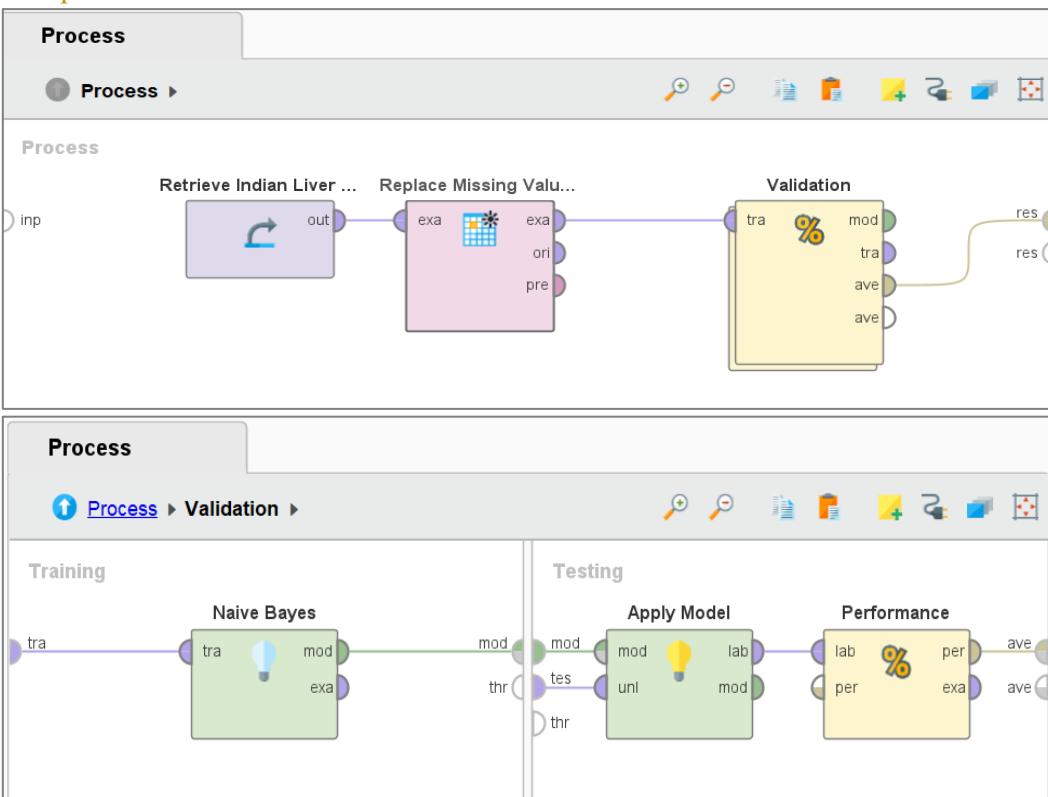
6. Appendix A: Naïve Bayes Application

6.1. RapidMiner

6.1.1. Cross-Validation



6.1.2 Split Validation





RapidMiner 65% Split Validation using Naïve Bayes

The screenshot shows the RapidMiner Studio interface with the 'Results' tab selected. A 'PerformanceVector (Performance)' result is displayed, showing accuracy of 60.78%. The confusion matrix table is as follows:

	true yes	true no	class precision
pred. yes	70	4	94.59%
pred. no	76	54	41.54%
class recall	47.95%	93.10%	

RapidMiner 70% Split Validation using Naïve Bayes

The screenshot shows the RapidMiner Studio interface with the 'Results' tab selected. A 'PerformanceVector (Performance)' result is displayed, showing accuracy of 61.71%. The confusion matrix table is as follows:

	true yes	true no	class precision
pred. yes	62	4	93.94%
pred. no	63	46	42.20%
class recall	49.60%	92.00%	

RapidMiner 85% Split Validation using Naïve Bayes

The screenshot shows the RapidMiner Studio interface with the 'Results' tab selected. A 'PerformanceVector (Performance)' result is displayed, showing accuracy of 54.02%. The confusion matrix table is as follows:

	true yes	true no	class precision
pred. yes	22	0	100.00%
pred. no	40	25	38.46%
class recall	35.48%	100.00%	



RapidMiner 4 folds Cross Validation using Naïve Bayes

accuracy: 55.74% +/- 5.61% (micro average: 55.75%)

	true yes	true no	class precision
pred. yes	167	9	94.89%
pred. no	249	158	38.82%
class recall	40.14%	94.61%	

RapidMiner 8 folds Cross Validation using Naïve Bayes

accuracy: 56.11% +/- 7.18% (micro average: 56.09%)

	true yes	true no	class precision
pred. yes	167	7	95.98%
pred. no	249	160	39.12%
class recall	40.14%	95.81%	

RapidMiner 16 folds Cross Validation using Naïve Bayes

accuracy: 55.93% +/- 8.32% (micro average: 55.92%)

	true yes	true no	class precision
pred. yes	166	7	95.95%
pred. no	250	160	39.02%
class recall	39.90%	95.81%	



6.2. Weka

Weka 65% Split Validation using Naïve Bayes

The screenshot shows the Weka interface with the Naïve Bayes classifier selected. The 'Test options' section has 'Percentage split % 65' selected. The 'Classifier output' pane displays the evaluation results for a test split:

```
==== Evaluation on test split ====
Time taken to test model on test split: 0 seconds
==== Summary ====
Correctly Classified Instances      111      54.4118 %
Incorrectly Classified Instances   93       45.5882 %
Kappa statistic                   0.2394
Mean absolute error               0.4529
Root mean squared error           0.6639
Relative absolute error            10.7721 %
Root relative squared error       147.1742 %
Total Number of Instances         204

==== Detailed Accuracy By Class ====

```

TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
0.370	0.017	0.982	0.370	0.537	0.358	0.756	0.904	yes
0.983	0.630	0.383	0.983	0.551	0.358	0.756	0.480	no
Weighted Avg.	0.544	0.191	0.811	0.544	0.358	0.756	0.783	

```
==== Confusion Matrix ====
a b  <-- classified as
54 92 | a = yes
1 57 | b = no
<
```

Weka 70% Split Validation using Naïve Bayes

The screenshot shows the Weka interface with the Naïve Bayes classifier selected. The 'Test options' section has 'Percentage split % 70' selected. The 'Classifier output' pane displays the evaluation results for a test split:

```
==== Evaluation on test split ====
Time taken to test model on test split: 0 seconds
==== Summary ====
Correctly Classified Instances      97      55.4286 %
Incorrectly Classified Instances   78       44.5714 %
Kappa statistic                   0.2563
Mean absolute error               0.4465
Root mean squared error           0.66
Relative absolute error            109.8597 %
Root relative squared error       147.8039 %
Total Number of Instances         175

==== Detailed Accuracy By Class ====

```

TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
0.396	0.000	1.000	0.386	0.557	0.383	0.768	0.913	yes
1.000	0.614	0.381	1.000	0.552	0.383	0.768	0.488	no
Weighted Avg.	0.554	0.168	0.830	0.554	0.555	0.383	0.768	0.796

```
==== Confusion Matrix ====
a b  <-- classified as
49 78 | a = yes
0 48 | b = no
<
```

Weka 85% Split Validation using Naïve Bayes

The screenshot shows the Weka interface with the Naïve Bayes classifier selected. The 'Test options' section has 'Percentage split % 85' selected. The 'Classifier output' pane displays the evaluation results for a test split:

```
==== Evaluation on test split ====
Time taken to test model on test split: 0.01 seconds
==== Summary ====
Correctly Classified Instances      48      55.1724 %
Incorrectly Classified Instances   39       44.8276 %
Kappa statistic                   0.2521
Mean absolute error               0.4463
Root mean squared error           0.6599
Relative absolute error            107.9181 %
Root relative squared error       144.0876 %
Total Number of Instances         87

==== Detailed Accuracy By Class ====

```

TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
0.361	0.000	1.000	0.361	0.530	0.380	0.786	0.916	yes
1.000	0.639	0.400	1.000	0.571	0.380	0.786	0.538	no
Weighted Avg.	0.552	0.191	0.821	0.552	0.542	0.380	0.786	0.803

```
==== Confusion Matrix ====
a b  <-- classified as
22 39 | a = yes
0 26 | b = no
<
```



Weka 4 folds Cross Validation using Naïve Bayes

The screenshot shows the Weka Explorer interface with the NaiveBayes classifier selected. The 'Test options' panel indicates 'Cross-validation' with 'Folds' set to 4. The 'Classifier output' pane displays the results of the stratified cross-validation:

Time taken to build model: 0 seconds

==== Stratified cross-validation ====
==== Summary ====
Correctly Classified Instances 322 55.2316 %
Incorrectly Classified Instances 261 44.7684 %
Kappa statistic 0.2371
Mean absolute error 0.4439
Root mean squared error 0.6556
Relative absolute error 108.405 %
Root relative squared error 145.0112 %
Total Number of Instances 583

==== Detailed Accuracy By Class ====
TP Rate FP Rate Precision Recall F-Measure MCC ROC Area FRC Area Class
0.394 0.054 0.948 0.394 0.557 0.337 0.720 0.879 yes
0.946 0.606 0.385 0.946 0.548 0.337 0.721 0.428 no
Weighted Avg. 0.552 0.212 0.787 0.552 0.554 0.337 0.720 0.750

==== Confusion Matrix ====
a b <-- classified as
164 252 | a = yes
9 158 | b = no

Weka 8 folds Cross Validation using Naïve Bayes

The screenshot shows the Weka Explorer interface with the NaiveBayes classifier selected. The 'Test options' panel indicates 'Cross-validation' with 'Folds' set to 8. The 'Classifier output' pane displays the results of the stratified cross-validation:

Time taken to build model: 0 seconds

==== Stratified cross-validation ====
==== Summary ====
Correctly Classified Instances 323 55.4031 %
Incorrectly Classified Instances 260 44.5969 %
Kappa statistic 0.2391
Mean absolute error 0.4449
Root mean squared error 0.6572
Relative absolute error 108.7264 %
Root relative squared error 145.3553 %
Total Number of Instances 583

==== Detailed Accuracy By Class ====
TP Rate FP Rate Precision Recall F-Measure MCC ROC Area FRC Area Class
0.397 0.054 0.948 0.397 0.559 0.339 0.721 0.879 yes
0.946 0.603 0.386 0.946 0.549 0.339 0.722 0.435 no
Weighted Avg. 0.554 0.211 0.787 0.554 0.556 0.339 0.721 0.752

==== Confusion Matrix ====
a b <-- classified as
165 251 | a = yes
9 158 | b = no

Weka 16 folds Cross Validation using Naïve Bayes

The screenshot shows the Weka Explorer interface with the NaiveBayes classifier selected. The 'Test options' panel indicates 'Cross-validation' with 'Folds' set to 16. The 'Classifier output' pane displays the results of the stratified cross-validation:

Time taken to build model: 0 seconds

==== Stratified cross-validation ====
==== Summary ====
Correctly Classified Instances 322 55.2316 %
Incorrectly Classified Instances 261 44.7684 %
Kappa statistic 0.2371
Mean absolute error 0.4465
Root mean squared error 0.6585
Relative absolute error 109.1289 %
Root relative squared error 145.639 %
Total Number of Instances 583

==== Detailed Accuracy By Class ====
TP Rate FP Rate Precision Recall F-Measure MCC ROC Area FRC Area Class
0.394 0.054 0.948 0.394 0.557 0.337 0.726 0.881 yes
0.946 0.606 0.385 0.946 0.548 0.337 0.727 0.439 no
Weighted Avg. 0.552 0.212 0.787 0.552 0.554 0.337 0.726 0.754

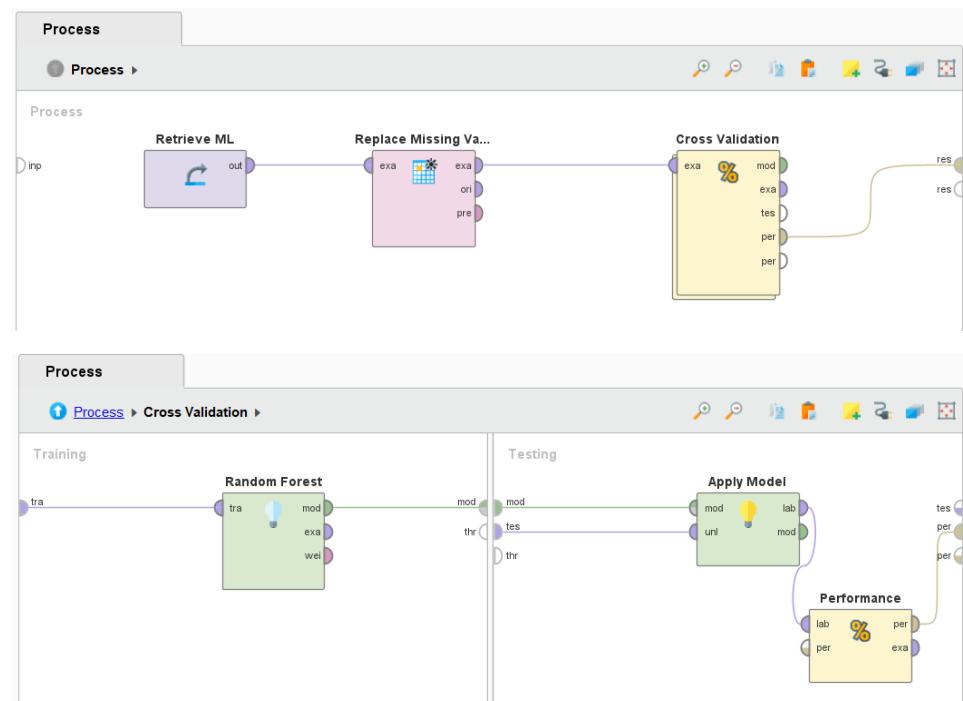
==== Confusion Matrix ====
a b <-- classified as
164 252 | a = yes
9 158 | b = no



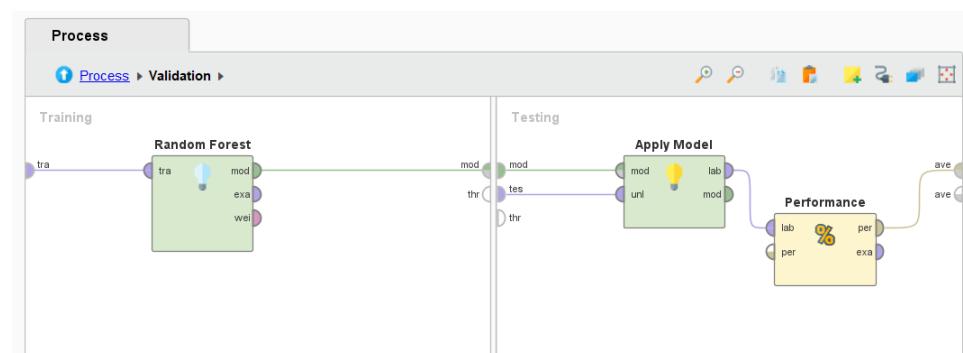
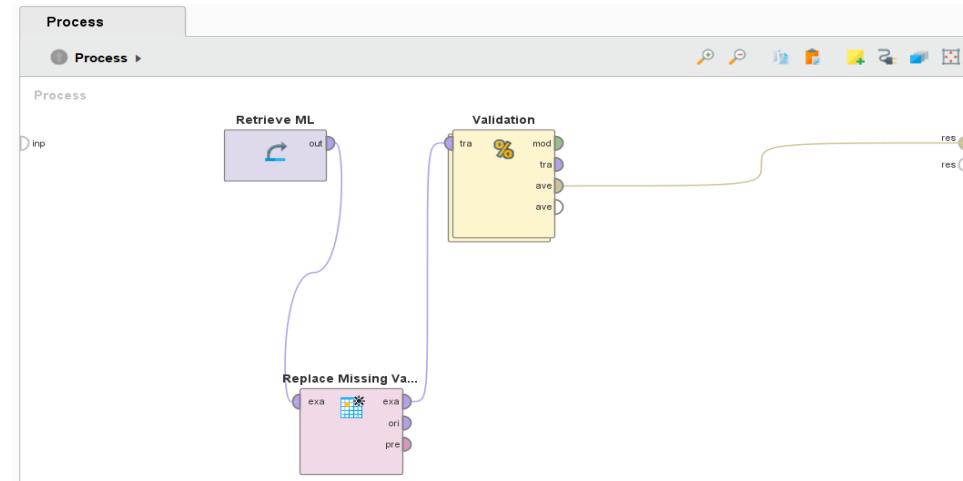
7. Appendix B: Random Forest Application

7.1. RapidMiner

7.1.1. Cross-Validation



7.1.2 Split Validation





RapidMiner 65% Split Validation using Random Forest

accuracy: 71.57%

	true yes	true no	class precision
pred yes	142	64	72.45%
pred no	4	4	50.00%
class recall	97.20%	6.90%	

RapidMiner 70% Split Validation using Random Forest

accuracy: 72.57%

	true yes	true no	class precision
pred yes	125	48	72.26%
pred no	0	2	100.00%
class recall	100.00%	4.00%	

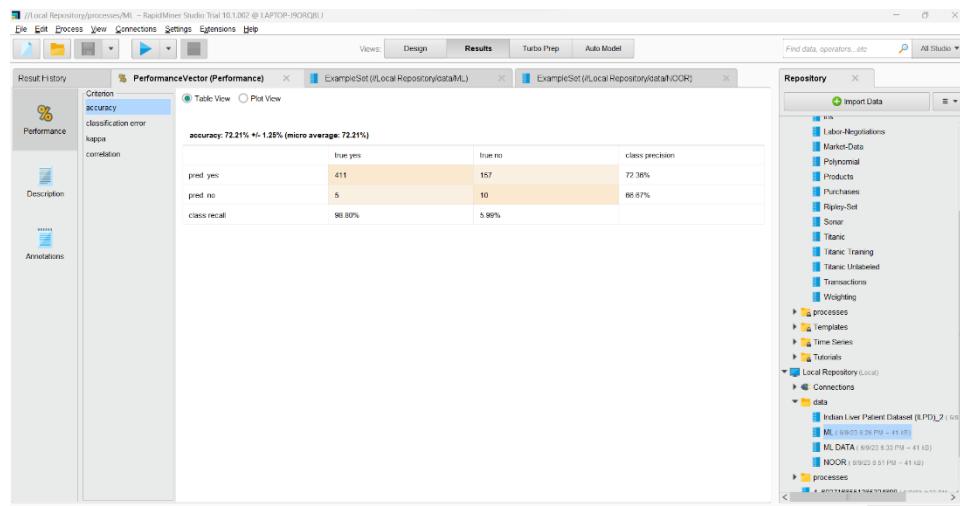
RapidMiner 85% Split Validation using Random Forest

accuracy: 73.56%

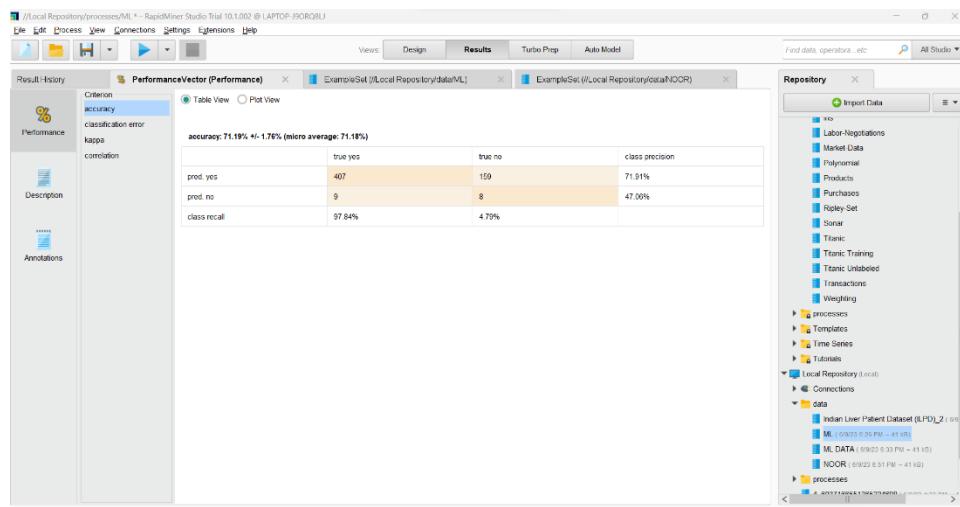
	true yes	true no	class precision
pred yes	62	23	72.94%
pred no	0	2	100.00%
class recall	100.00%	8.00%	



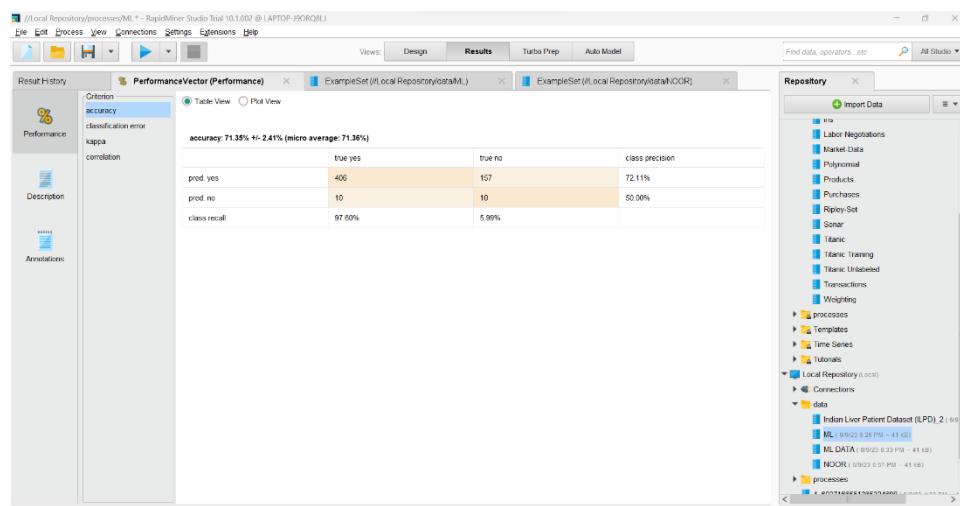
RapidMiner 4 folds Cross Validation using Random Forest



RapidMiner 8 folds Cross Validation using Random Forest



RapidMiner 16 folds Cross Validation using Random Forest





7.2. Weka

Weka 65% Split Validation using Random Forest

The screenshot shows the Weka Explorer interface. Under the 'Classifier' tab, 'RandomForest' is selected with parameters: -P 100 -I 100 -num-slots 1 -K 0 -M 1.0 -V 0.001 -S 1. In the 'Test options' section, 'Percentage split' is chosen with 65%. The 'Classifier output' pane displays the results of a 65% validation split. Summary statistics include:

	Value
Correctly Classified Instances	130
Incorrectly Classified Instances	66
Kappa statistic	0.1614
Mean absolute error	0.3496
Root mean squared error	0.4243
Relative absolute error	85.5086 %
Root relative squared error	94.0669 %
Total Number of Instances	204

Detailed accuracy by class:

Class	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area
yes	0.808	0.655	0.756	0.808	0.781	0.163	0.743	0.895
no	0.345	0.192	0.417	0.345	0.377	0.163	0.743	0.473

Confusion Matrix:

	a	b
a	118	28
b	38	20

Weka 70% Split Validation using Random Forest

The screenshot shows the Weka Explorer interface. Under the 'Classifier' tab, 'RandomForest' is selected with parameters: -P 100 -I 100 -num-slots 1 -K 0 -M 1.0 -V 0.001 -S 1. In the 'Test options' section, 'Percentage split' is chosen with 70%. The 'Classifier output' pane displays the results of a 70% validation split. Summary statistics include:

	Value
Correctly Classified Instances	119
Incorrectly Classified Instances	56
Kappa statistic	0.1637
Mean absolute error	0.3461
Root mean squared error	0.4186
Relative absolute error	85.1636 %
Root relative squared error	93.7472 %
Total Number of Instances	175

Detailed accuracy by class:

Class	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area
yes	0.803	0.646	0.767	0.803	0.785	0.164	0.745	0.902
no	0.354	0.197	0.405	0.354	0.378	0.164	0.745	0.441

Confusion Matrix:

	a	b
a	102	25
b	31	17



Weka 85% Split Validation using Random Forest

Weka Explorer

Preprocess Classify Cluster Associate Select attributes Visualize

Classifier Choose **RandomForest -P 100 -I 100 -num-slots 1 -K 0 -M 1.0 -V 0.001 -S 1**

Test options Use training set Supplied test set Set... Cross-validation Folds 16 Percentage split % 85 More options...

(Nom) Selector field used to split the d...

Start Stop

Result list (right-click for options)

20:57:49 - trees.RandomForest
06:12:32 - trees.RandomForest
06:13:30 - trees.RandomForest
06:14:29 - trees.RandomForest
06:15:34 - trees.RandomForest
06:16:23 - trees.RandomForest

Classifier output

RandomForest

Bagging with 100 iterations and base learner

weka.classifiers.trees.RandomTree -K 0 -M 1.0 -V 0.001 -S 1 -do-not-check-capabilities

Time taken to build model: 0.1 seconds

*** Evaluation on test split ***

Time taken to test model on test split: 0 seconds

*** Summary ***

	Correctly Classified Instances	58	66.6667 %
Incorrectly Classified Instances	29	33.3333 %	
Kappa statistic	0.1582		
Mean absolute error	0.3475		
Root mean squared error	0.4217		
Relative absolute error	84.0191 %		
Root relative squared error	92.0806 %		
Total Number of Instances	87		

*** Detailed Accuracy By Class ***

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
0.603	0.654	0.742	0.603	0.772	0.6160	0.757	0.503	0.491	yes
0.346	0.197	0.429	0.346	0.383	0.160	0.757	0.491	0.491	no
Weighted Avg.	0.667	0.517	0.649	0.667	0.655	0.160	0.757	0.780	

*** Confusion Matrix ***

a	b	<-- Classified as
49	12	a = yes
17	9	b = no

Status OK

Weka 4 folds Cross Validation using Random Forest

Weka Explorer

Preprocess Classify Cluster Associate Select attributes Visualize

Classifier Choose **RandomForest -P 100 -I 100 -num-slots 1 -K 0 -M 1.0 -V 0.001 -S 1**

Test options Use training set Supplied test set Set... Cross-validation Folds 4 Percentage split % 66 More options...

(Nom) Selector field used to split the d...

Start Stop

Result list (right-click for options)

20:57:49 - trees.RandomForest

Classifier output

RandomForest

Bagging with 100 iterations and base learner

weka.classifiers.trees.RandomTree -K 0 -M 1.0 -V 0.001 -S 1 -do-not-check-capabilities

Time taken to build model: 0.34 seconds

*** Stratified cross-validation ***

*** Summary ***

	Correctly Classified Instances	413	70.8405 %
Incorrectly Classified Instances	170	29.1595 %	
Kappa statistic	0.2103		
Mean absolute error	0.3508		
Root mean squared error	0.4912		
Relative absolute error	85.7152 %		
Root relative squared error	95.3701 %		
Total Number of Instances	583		

*** Detailed Accuracy By Class ***

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
0.861	0.671	0.762	0.861	0.808	0.217	0.718	0.675	0.444	yes
0.329	0.139	0.487	0.329	0.393	0.217	0.718	0.444	0.444	no
Weighted Avg.	0.708	0.518	0.683	0.708	0.689	0.217	0.718	0.751	

*** Confusion Matrix ***

a	b	<-- classified as
358	58	a = yes
112	55	b = no



Weka 8 folds Cross Validation using Random Forest

Weka Explorer

Preprocess Classify Cluster Associate Select attributes Visualize

Classifier Choose **RandomForest** -P 100 -I 100 -num-slots 1 -K 0 -M 1.0 -V 0.001 -S 1

Test options Use training set Supplied test set Set... Cross-validation Folds 8 Percentage split % 66 More options...

(Nom) Selector field used to split the data...

Start Stop

Result list (right-click for options)
205749 - trees.RandomForest
061232 - trees.RandomForest

Classifier output

==== Classifier model (full training set) ====
RandomForest

Bagging with 100 iterations and base learner

weka.classifiers.trees.RandomTree -K 0 -M 1.0 -V 0.001 -S 1 -do-not-check-capabilities

Time taken to build model: 0.16 seconds

==== Stratified cross-validation ====
==== Summary ====

Correctly Classified Instances 400 69.9828 %
Incorrectly Classified Instances 175 30.0172 %
Kappa statistic 0.1769
Mean absolute error 0.3423
Root mean squared error 0.4241
Relative absolute error 63.665 %
Root relative squared error 93.8119 %
Total Number of Instances 583

==== Detailed Accuracy By Class ====

TP Rate FP Rate Precision Recall F-Measure MCC ROC Area FRC Area Class
0.861 0.701 0.754 0.861 0.808 0.186 0.733 0.880 yes
0.299 0.139 0.463 0.299 0.364 0.186 0.733 0.471 no
Weighted Avg. 0.700 0.540 0.670 0.700 0.678 0.186 0.733 0.763

==== Confusion Matrix ====

a b <-- classified as
350 58 | a = yes
117 50 | b = no

Status OK

Weka 16 folds Cross Validation using Random Forest

Weka Explorer

Preprocess Classify Cluster Associate Select attributes Visualize

Classifier Choose **RandomForest** -P 100 -I 100 -num-slots 1 -K 0 -M 1.0 -V 0.001 -S 1

Test options Use training set Supplied test set Set... Cross-validation Folds 16 Percentage split % 66 More options...

(Nom) Selector field used to split the data...

Start Stop

Result list (right-click for options)
205749 - trees.RandomForest
061232 - trees.RandomForest
061330 - trees.RandomForest

Classifier output

==== Classifier model (full training set) ====
RandomForest

Bagging with 100 iterations and base learner

weka.classifiers.trees.RandomTree -K 0 -M 1.0 -V 0.001 -S 1 -do-not-check-capabilities

Time taken to build model: 0.09 seconds

==== Stratified cross-validation ====
==== Summary ====

Correctly Classified Instances 400 68.6106 %
Incorrectly Classified Instances 183 31.3894 %
Kappa statistic 0.1549
Mean absolute error 0.3496
Root mean squared error 0.4308
Relative absolute error 85.4377 %
Root relative squared error 95.289 %
Total Number of Instances 583

==== Detailed Accuracy By Class ====

TP Rate FP Rate Precision Recall F-Measure MCC ROC Area FRC Area Class
0.841 0.701 0.749 0.841 0.793 0.159 0.715 0.876 yes
0.259 0.159 0.431 0.259 0.353 0.159 0.715 0.430 no
Weighted Avg. 0.686 0.545 0.658 0.686 0.667 0.159 0.715 0.748

==== Confusion Matrix ====

a b <-- classified as
350 66 | a = yes
117 50 | b = no

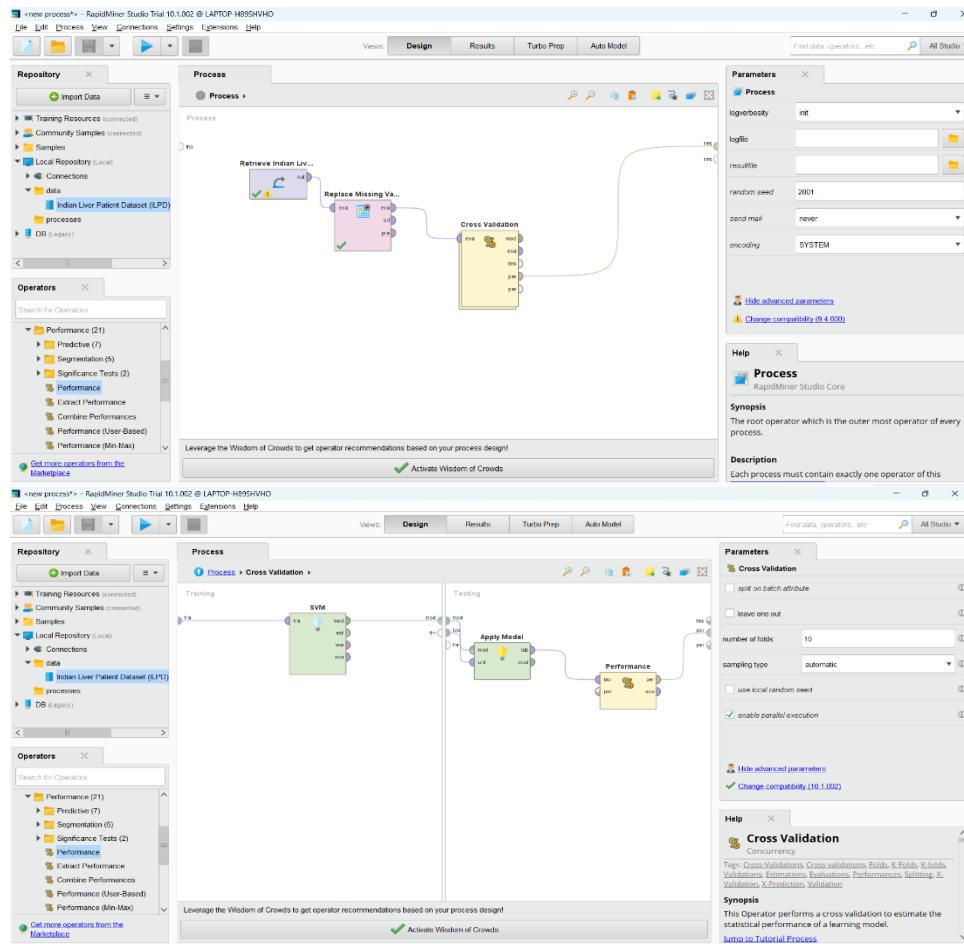
Status OK



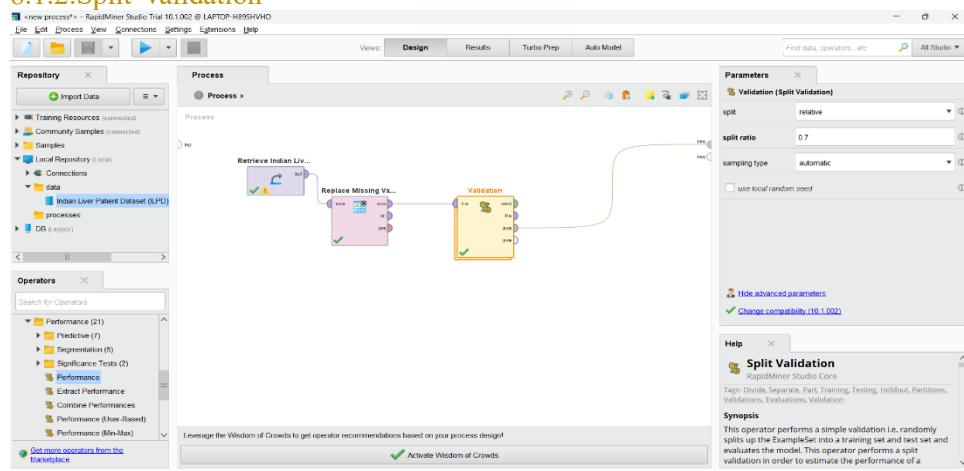
8. Appendix C: Supportive Vector Machine

8.1. RapidMiner

8.1.1. Cross-Validation



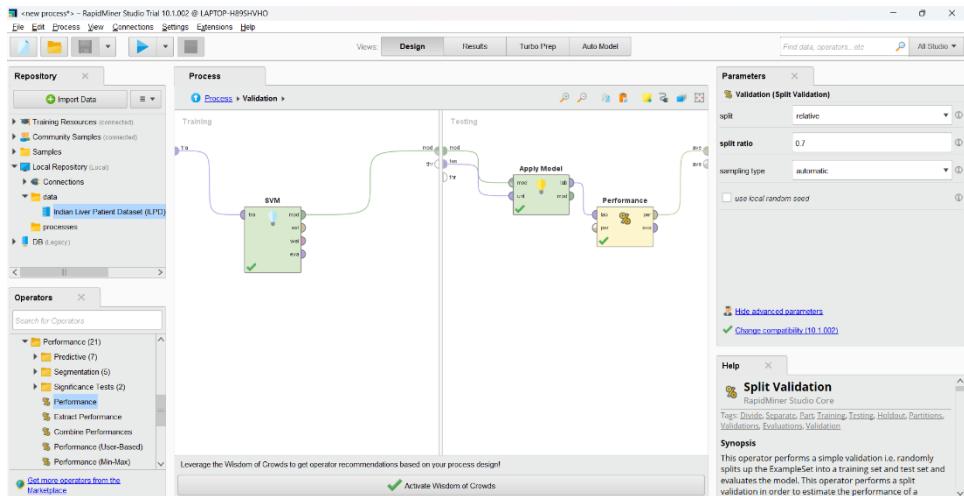
8.1.2. Split-Validation



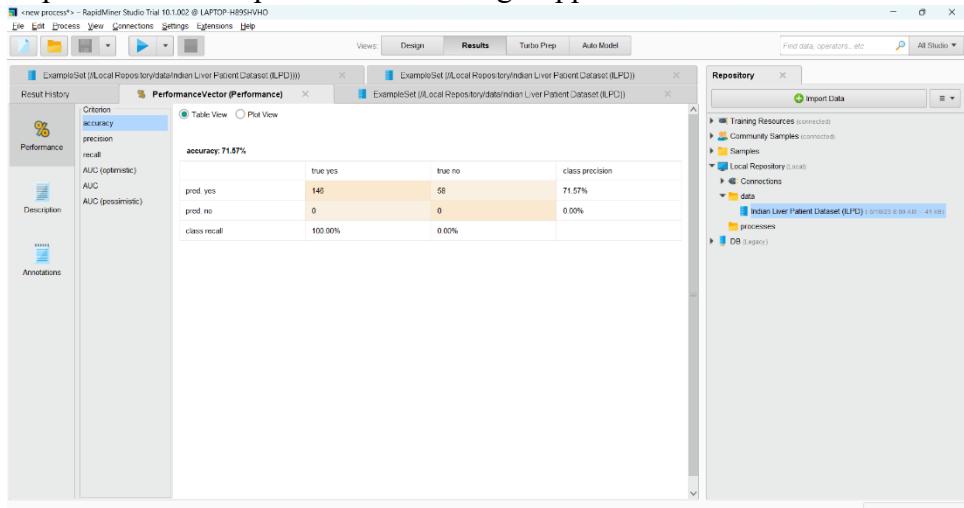


MACHINE LEARNING

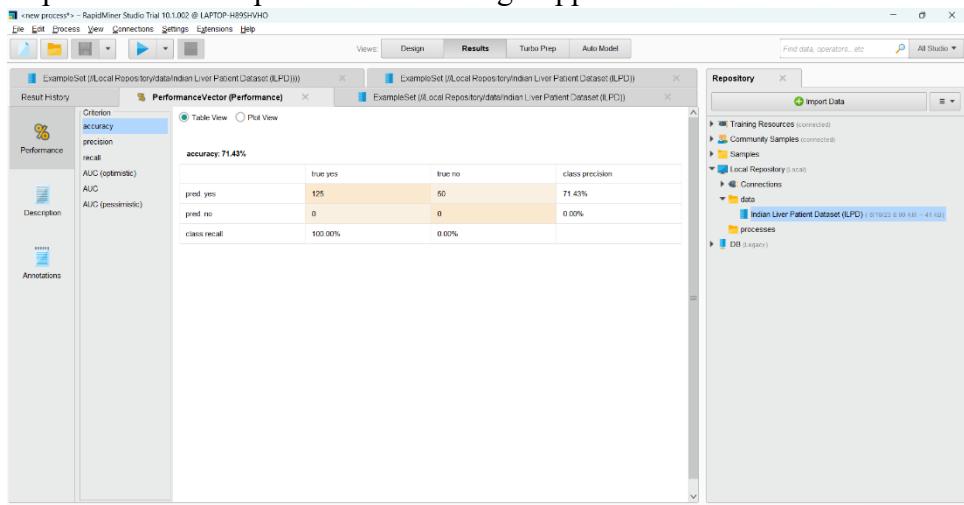
using RapidMiner and Weka
on Indian Liver Patient Dataset



RapidMiner 65% Split Validation using Support Vector Machine



RapidMiner 70% Split Validation using Support Vector Machine





RapidMiner 85% Split Validation using Support Vector Machine

The screenshot shows the RapidMiner interface with the following details:

- File menu:** File, Edit, Process, View, Connections, Settings, Extensions, Help.
- Toolbars:** Standard (File, Open, Save, etc.) and a Repository toolbar.
- Result History:** Shows three entries: ExampleSet (Local Repository/Indian Liver Patient Dataset (ILPD)), PerformanceVector (Performance), and ExampleSet (Local Repository/Indian Liver Patient Dataset (ILPD)).
- Performance View:** Selected under the Results tab. It displays accuracy: 71.28%. A confusion matrix table is shown:

	true yes	true no	class precision
pred. yes	62	25	71.26%
pred. no	0	0	0.00%
class recall	100.00%	0.00%	

- Repository:** Shows the Local Repository (Local) with the Indian Liver Patient Dataset (ILPD) selected.

RapidMiner 4 folds Cross- Validation using Support Vector Machine

The screenshot shows the RapidMiner interface with the following details:

- File menu:** File, Edit, Process, View, Connections, Settings, Extensions, Help.
- Toolbars:** Standard (File, Open, Save, etc.) and a Repository toolbar.
- Result History:** Shows three entries: ExampleSet (Local Repository/Indian Liver Patient Dataset (ILPD)), PerformanceVector (Performance), and ExampleSet (Local Repository/Indian Liver Patient Dataset (ILPD)).
- Performance View:** Selected under the Results tab. It displays accuracy: 71.35% +/- 0.39% (micro average: 71.36%). A confusion matrix table is shown:

	true yes	true no	class precision
pred. yes	416	167	71.36%
pred. no	0	0	0.00%
class recall	100.00%	0.00%	

- Repository:** Shows the Local Repository (Local) with the Indian Liver Patient Dataset (ILPD) selected.

RapidMiner 8 folds Cross- Validation using Support Vector Machine

The screenshot shows the RapidMiner interface with the following details:

- File menu:** File, Edit, Process, View, Connections, Settings, Extensions, Help.
- Toolbars:** Standard (File, Open, Save, etc.) and a Repository toolbar.
- Result History:** Shows three entries: ExampleSet (Local Repository/Indian Liver Patient Dataset (ILPD)), PerformanceVector (Performance), and ExampleSet (Local Repository/Indian Liver Patient Dataset (ILPD)).
- Performance View:** Selected under the Results tab. It displays accuracy: 71.38% +/- 0.52% (micro average: 71.36%). A confusion matrix table is shown:

	true yes	true no	class precision
pred. yes	416	167	71.36%
pred. no	0	0	0.00%
class recall	100.00%	0.00%	

- Repository:** Shows the Local Repository (Local) with the Indian Liver Patient Dataset (ILPD) selected.



RapidMiner 16 folds Cross- Validation using Support Vector Machine

The screenshot shows the RapidMiner interface with the following details:

- File menu:** <new process>, RapidMiner Studio Trial 10.1.02 @ LATOP-HB95/H0
- Toolbars:** File, Edit, Process, View, Connections, Settings, Extensions, Help
- Views:** Views, Design, Results, Turbo Prep, Auto Model
- Performance Vector (Performance) window:**
 - Criterion: accuracy
 - Table View (selected), Plot View
 - accuracy: 70.83% +/- 1.86% (micro average: 70.84%)
 - Confusion Matrix:

	true yes	true no	class precision
pred yes	413	167	71.21%
pred no	3	0	0.00%
class recall	99.28%	0.00%	
- Repository window:**
 - Training Resources (connected)
 - Community Samples (connected)
 - Samples
 - Local Repository (local)
 - data
 - Indian Liver Patient Dataset (ILPD) (61922.8.99 kB - 41 MB)
 - processes
 - DB (empty)

8.2.Weka

Weka 65% Split Validation using Support Vector Machine

The screenshot shows the Weka Explorer interface with the following details:

- Classifier settings:**
 - SMO-C 1.0 L 0.001 P 1.0E-12 N 0 V -1 W K "weka.classifiers.functions.supportVector.PolyKernel" E 1.0 C 2500007+ calibrator "weka.classifiers.functions.Logit" R 1.0E-8 M -1 num decimal places 4"
 - Test options: Use training set, Supplied test set, Cross-validation (Folds: 4), Percentage split % 65 (selected)
- Result list:** Shows multiple SMO instances, with the last one selected: 094101 - functions.SMO.
- Classifier output:**
 - Number of kernel evaluations: 10122 (77.095% cached)
 - Time taken to build model: 0.01 seconds
 - Evaluation on test split:
 - Time taken to test model on test split: 0 seconds
 - Summary:

	Correctly Classified Instances	71.5606 %
Incorrectly Classified Instances	59	28.4314 %
Kappa statistic	0	
Mean absolute error	0.2543	
Root mean squared error	0.5392	
Relative absolute error	69.5987 %	
Root relative squared error	118.2001 %	
Coverage of cases (0.95 level)	71.5606 %	
Mean rel. region size (0.95 level)	50	
Total Number of Instances	204	
 - Detailed Accuracy By Class:

Class	TP Rate	FP Rate	Recall	ROC Area	PNU Area
yes	1.000	1.000	1.000	0.500	0.716
no	0.000	0.000	0.000	0.500	0.284
Weighted Avg.	0.716	0.716	0.716	0.500	0.593
 - Confusion Matrix:

	a	b	== classified as
a	148	5	a = yes
b	0	59	b = no

Weka 70% Split Validation using Support Vector Machine

The screenshot shows the Weka Explorer interface with the following details:

- Classifier settings:**
 - SMO-C 1.0 L 0.001 P 1.0E-12 N 0 V -1 W K "weka.classifiers.functions.supportVector.PolyKernel" E 1.0 C 2500007+ calibrator "weka.classifiers.functions.Logit" R 1.0E-8 M -1 num decimal places 4"
 - Test options: Use training set, Supplied test set, Cross-validation (Folds: 4), Percentage split % 70 (selected)
- Result list:** Shows multiple SMO instances, with the last one selected: 094101 - functions.SMO.
- Classifier output:**
 - Number of kernel evaluations: 18122 (77.895% cached)
 - Time taken to build model: 0.01 seconds
 - Evaluation on test split:
 - Time taken to test model on test split: 0.01 seconds
 - Summary:

	Correctly Classified Instances	72.5714 %
Incorrectly Classified Instances	48	27.4286 %
Kappa statistic	0	
Mean absolute error	0.2743	
Root mean squared error	0.5537	
Relative absolute error	67.4857 %	
Root relative squared error	111.2046 %	
Coverage of cases (0.95 level)	72.5714 %	
Mean rel. region size (0.95 level)	50	
Total Number of Instances	175	
 - Detailed Accuracy By Class:

Class	TP Rate	FP Rate	Recall	ROC Area	PNU Area
yes	1.000	1.000	1.000	0.500	0.726
no	0.000	0.000	0.000	0.500	0.274
Weighted Avg.	0.726	0.726	0.726	0.500	0.602
 - Confusion Matrix:

	a	b	== classified as
a	127	5	a = yes
b	0	48	b = no



MACHINE LEARNING

using RapidMiner and Weka
on Indian Liver Patient Dataset

Weka 85% Split Validation using Support Vector Machine

```
Weka Explorer
Preprocess Classify Cluster Associate Select attributes Visualize
Classifier
Choose: SMO - C:1.0 - L:0.001 P:1.0E-12 N:0 V:-1 W:-1 "weka.classifiers.functions.supportVector.PolyKernel_E:1.0-C:2.00007"-calibrator "weka.classifiers.functions.Logistic_R:1.0E-8 M:-1-num decimal places:4"
Test options
 Use training set
 Supplied test set Set...
 Cross-validation Folds: 4
 Percentage split %: 85
 More options...
(Nom) Selector
Start Stop
Result list (right-click for options)
091555 - functions.SMO
091349 - functions.SMO
091415 - functions.SMO
091504 - functions.SMO
092324 - functions.SMO
092352 - functions.SMO
092101 - functions.SMO
092444 - functions.SMO
092559 - functions.SMO
092557 - functions.SMO
094317 - functions.SMO
094128 - functions.SMO
094126 - functions.SMO
Classifier output
Number of kernel evaluations: 18122 (77.895% cached)

Time taken to build model: 0.01 seconds
*** Evaluation on test split ***
Time taken to test model on test split: 0 seconds
*** Summary ***
Correctly Classified Instances 61 70.1149 %
Incorrectly Classified Instances 26 29.8851 %
Kappa statistic 0
Mean absolute error 0.2099
Root mean squared error 0.2447
Relative absolute error 72.2605 %
Root relative squared error 119.9723 %
Coverage of cases (0.95 level) 70.1149 %
Mean rel. region size (0.95 level) 50 %
Total Number of Instances 87

*** Detailed Accuracy By Class ***
      TP Rate  FP Rate  Recall  ROC Area  PRC Area  Class
  1.000   1.000   1.000  0.500  0.701  yes
  0.000   0.000   0.000  0.500  0.299  no
Weighted Avg.  0.701  0.701  0.701  0.500  0.591

*** Confusion Matrix ***
  a | b <- classified as
  61 |  a = yes
  26 |  b = no


```

Weka 4 folds Cross Validation using Support Vector Machine

```
Weka Explorer
Preprocess Classify Cluster Associate Select attributes Visualize
Classifier
Choose: SMO - C:1.0 - L:0.001 P:1.0E-12 N:0 V:-1 W:-1 "weka.classifiers.functions.supportVector.PolyKernel_E:1.0-C:2.00007"-calibrator "weka.classifiers.functions.Logistic_R:1.0E-8 M:-1-num decimal places:4"
Test options
 Use training set
 Supplied test set Set...
 Cross-validation Folds: 4
 Percentage split %: 85
 More options...
(Nom) Selector
Start Stop
Result list (right-click for options)
091014 - functions.SMO
091555 - functions.SMO
091349 - functions.SMO
091415 - functions.SMO
091504 - functions.SMO
092324 - functions.SMO
092352 - functions.SMO
092101 - functions.SMO
092444 - functions.SMO
092559 - functions.SMO
092557 - functions.SMO
094317 - functions.SMO
094128 - functions.SMO
094126 - functions.SMO
Classifier output
Number of kernel evaluations: 18122 (77.895% cached)

Time taken to build model: 0.01 seconds
*** Stratified cross-validation ***
*** Summary ***
Correctly Classified Instances 416 71.3551 %
Incorrectly Classified Instances 147 28.6449 %
Kappa statistic 0
Mean absolute error 0.2064
Root mean squared error 0.5335
Relative absolute error 70.0007 %
Root relative squared error 118.2813 %
Coverage of cases (0.95 level) 71.3551 %
Mean rel. region size (0.95 level) 50 %
Total Number of Instances 563

*** Detailed Accuracy By Class ***
      TP Rate  FP Rate  Recall  ROC Area  PRC Area  Class
  1.000   1.000   1.000  0.500  0.714  yes
  0.000   0.000   0.000  0.500  0.286  no
Weighted Avg.  0.714  0.714  0.714  0.500  0.591

*** Confusion Matrix ***
  a | b <- classified as
  416 |  a = yes
  147 |  b = no


```

Weka 8 folds Cross Validation using Support Vector Machine

```
Weka Explorer
Preprocess Classify Cluster Associate Select attributes Visualize
Classifier
Choose: SMO - C:1.0 - L:0.001 P:1.0E-12 N:0 V:-1 W:-1 "weka.classifiers.functions.supportVector.PolyKernel_E:1.0-C:2.00007"-calibrator "weka.classifiers.functions.Logistic_R:1.0E-8 M:-1-num decimal places:4"
Test options
 Use training set
 Supplied test set Set...
 Cross-validation Folds: 8
 Percentage split %: 85
 More options...
(Nom) Selector
Start Stop
Result list (right-click for options)
091514 - functions.SMO
091555 - functions.SMO
091349 - functions.SMO
091415 - functions.SMO
091504 - functions.SMO
092324 - functions.SMO
092352 - functions.SMO
092101 - functions.SMO
092444 - functions.SMO
092559 - functions.SMO
092557 - functions.SMO
094317 - functions.SMO
094128 - functions.SMO
094126 - functions.SMO
094154 - functions.SMO
094402 - functions.SMO
Classifier output
Number of kernel evaluations: 18122 (77.895% cached)

Time taken to build model: 0 seconds
*** Stratified cross-validation ***
*** Summary ***
Correctly Classified Instances 416 71.3551 %
Incorrectly Classified Instances 147 28.6449 %
Kappa statistic 0
Mean absolute error 0.2064
Root mean squared error 0.5335
Relative absolute error 70.0108 %
Root relative squared error 118.2814 %
Coverage of cases (0.95 level) 71.3551 %
Mean rel. region size (0.95 level) 50 %
Total Number of Instances 563

*** Detailed Accuracy By Class ***
      TP Rate  FP Rate  Recall  ROC Area  PRC Area  Class
  1.000   1.000   1.000  0.500  0.714  yes
  0.000   0.000   0.000  0.500  0.286  no
Weighted Avg.  0.714  0.714  0.714  0.500  0.591

*** Confusion Matrix ***
  a | b <- classified as
  416 |  a = yes
  147 |  b = no


```



Weka 16 folds Cross Validation using Support Vector Machine

Weka Explorer

Reprocess Classify Cluster Associate Select attributes Visualize

Classifier

Choose: SMO C=1.0 L=0.001 P=1.0E-12 N=0 V=1 W=1 K="weka.classifiers.functions.supportVector.PolyKernel E=1.0 C=250000" calibrator="weka.classifiers.functions.Logistic R=1.0E-8 M=-1 num decimal places=4"

Test options

Use training set

Supplied test set Set...

Cross-validation Folds: 16

Percentage split % 85

More options...

(Nom) Selector

Start Stop

Result list (right click for options)

091455 - functions.SMO

091349 - functions.SMO

091415 - functions.SMO

091350 - functions.SMO

091451 - functions.SMO

091351 - functions.SMO

092352 - functions.SMO

092301 - functions.SMO

092444 - functions.SMO

092459 - functions.SMO

094427 - functions.SMO

094428 - functions.SMO

094429 - functions.SMO

094430 - functions.SMO

094431 - functions.SMO

Classifier output

+ -0.002 * (normalized) A/G Ratio Albumin and Globulin Ratio

- 0.9994

Number of kernel evaluations: 10122 (77.095% cached)

Time taken to build model: 0.01 seconds

*** Stratified cross-validation ***

*** Summary ***

	Correctly Classified Instances	71.3551 %
Incorrectly Classified Instances	187	28.6449 %
Kappa statistic	0	
Mean absolute error	0.0064	
Root mean squared error	0.0352	
Relative absolute error	70.0137 %	
Root relative squared error	118.3706 %	
Overage of cases (0.95 level)	71.3551 %	
Mean rel. region size (0.95 level)	50	%
Total Number of Instances	503	

*** Detailed Accuracy By Class ***

	TP Rate	FP Rate	Fallout	ROC Area	PRC Area	Class
1.000	1.000	1.000	0.500	0.714	yes	
0.000	0.000	0.000	0.500	0.286	no	
Weighted Avg.	0.714	0.714	0.714	0.500	0.591	

*** Confusion Matrix ***

	a	b
a	416	0
b	0	187

--- classified as

a = yes

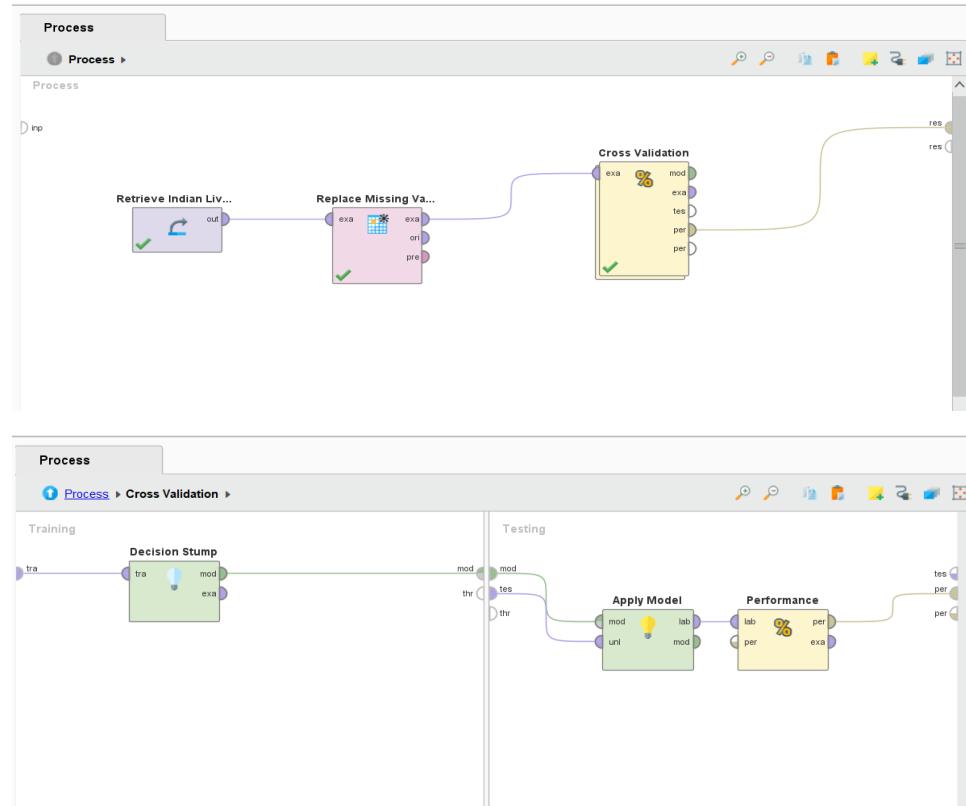
b = no



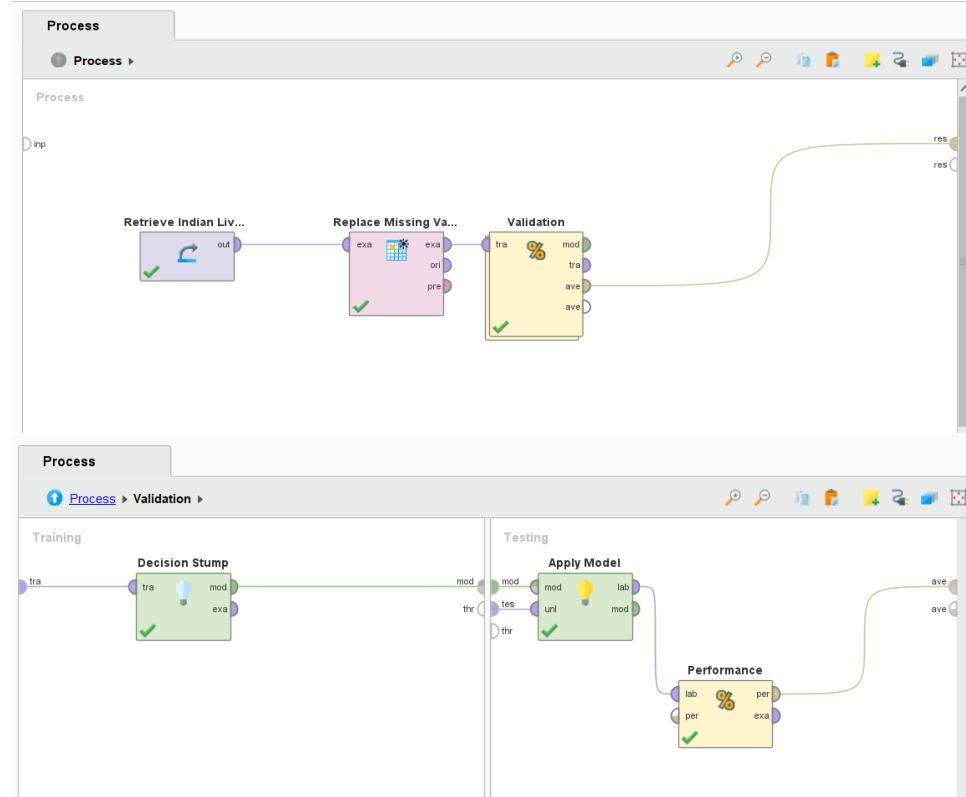
9. Appendix D: Decision Stump

9.1. RapidMiner

9.1.1. Cross-Validation



9.1.2 Split Validation





RapidMiner 65% Split Validation using Decision Stump

The screenshot shows the RapidMiner Studio interface with the 'Results' tab selected. The central panel displays a 'PerformanceVector (Performance)' table view with the following data:

	true yes	true no	class precision
pred yes	146	56	72.28%
pred no	0	2	100.00%
class recall	100.00%	3.42%	

The accuracy is listed as 72.55%. The repository sidebar shows the project structure, including the 'Indian Liver Patient Dataset (ILPD)' and the 'RapidMiner-Liver' process.

RapidMiner 70% Split Validation using Decision Stump

The screenshot shows the RapidMiner Studio interface with the 'Results' tab selected. The central panel displays a 'PerformanceVector (Performance)' table view with the following data:

	true yes	true no	class precision
pred yes	125	48	72.25%
pred no	0	2	100.00%
class recall	100.00%	4.00%	

The accuracy is listed as 72.47%. The repository sidebar shows the project structure, including the 'Indian Liver Patient Dataset (ILPD)' and the 'RapidMiner-Liver' process.

RapidMiner 85% Split Validation using Decision Stump

The screenshot shows the RapidMiner Studio interface with the 'Results' tab selected. The central panel displays a 'PerformanceVector (Performance)' table view with the following data:

	true yes	true no	class precision
pred yes	62	24	72.09%
pred no	0	1	100.00%
class recall	100.00%	4.00%	

The accuracy is listed as 72.41%. The repository sidebar shows the project structure, including the 'Indian Liver Patient Dataset (ILPD)' and the 'RapidMiner-Liver' process.



RapidMiner 4 folds Cross Validation using Decision Stump

The screenshot shows the RapidMiner Studio interface. The top menu bar includes File, Edit, Process, View, Connections, Settings, Extensions, Help, Views, Design, Results, Turbo Prep, Auto Model, Find data, operators, etc., and All Studio. The main window has tabs for Result History, PerformanceVector (Performance), and ExampleSet (/Local Repository/catalndan Liver Patient Dataset (ILPD)). The PerformanceVector tab displays a table view of performance metrics. The table shows accuracy, class precision, and class recall for two classes: true yes and true no. The overall accuracy is 72.38% +/- 0.95% (micro average: 72.38%). The right side of the interface shows the Repository pane with sections for Training Resources, Samples, Community Samples, Local Repository (local), Connections, data, processes, and DB (legacy).

RapidMiner 8 folds Cross Validation using Decision Stump

This screenshot is identical to the one above, showing the RapidMiner Studio interface with the same results for an 8-fold cross-validation using a Decision Stump model. The accuracy remains at 72.38% +/- 1.14% (micro average: 72.38%). The repository structure is also the same.

RapidMiner 16 folds Cross Validation using Decision Stump

This screenshot is identical to the previous ones, showing the RapidMiner Studio interface with the same results for a 16-fold cross-validation using a Decision Stump model. The accuracy is 72.38% +/- 1.66% (micro average: 72.38%). The repository structure is consistent with the other versions.



9.2. Weka

Weka 65% Split Validation using Decision Stump

The screenshot shows the Weka Explorer interface with the 'Classify' tab selected. Under 'Classifier', 'DecisionStump' is chosen. In the 'Test options' section, 'Percentage split' is selected with 65% chosen. The 'Classifier output' pane displays the following results:

```
Time taken to build model: 0 seconds
===
Evaluation on test split ===
Time taken to test model on test split: 0 seconds
===
Summary ===
Correctly Classified Instances      140          68.6275 %
Incorrectly Classified Instances   64           31.3725 %
Kappa statistic                   0
Mean absolute error               0.3809
Root mean squared error           0.4449
Relative absolute error            91.6966 %
Root relative squared error       96.3944 %
Coverage of cases (0.95 level)    97.0588 %
Mean rel. region size (0.95 level) 86.5196 %
Total Number of Instances         204

===
Detailed Accuracy By Class ===
      TP Rate   FP Rate   Recall   ROC Area   PRC Area   Class
1.000     1.000     1.000    0.628     0.758     yes
0.000     0.000     0.000    0.628     0.382     no
Weighted Avg.      0.686     0.686     0.686    0.628     0.640

===
Confusion Matrix ===
a   b   <-- classified as
140  0 |  a = yes
64   0 |  b = no
```

Status: OK

Weka 70% Split Validation using Decision Stump

The screenshot shows the Weka Explorer interface with the 'Classify' tab selected. Under 'Classifier', 'DecisionStump' is chosen. In the 'Test options' section, 'Percentage split' is selected with 70% chosen. The 'Classifier output' pane displays the following results:

```
Time taken to build model: 0.01 seconds
===
Evaluation on test split ===
Time taken to test model on test split: 0 seconds
===
Summary ===
Correctly Classified Instances      119          68. %
Incorrectly Classified Instances   56           32. %
Kappa statistic                   0
Mean absolute error               0.3838
Root mean squared error           0.4539
Relative absolute error            91.7345 %
Root relative squared error       96.807 %
Coverage of cases (0.95 level)    96.5714 %
Mean rel. region size (0.95 level) 86.2857 %
Total Number of Instances         175

===
Detailed Accuracy By Class ===
      TP Rate   FP Rate   Recall   ROC Area   PRC Area   Class
1.000     1.000     1.000    0.623     0.749     yes
0.000     0.000     0.000    0.623     0.386     no
Weighted Avg.      0.680     0.680     0.680    0.623     0.633

===
Confusion Matrix ===
a   b   <-- classified as
119  0 |  a = yes
56   0 |  b = no
```

Status: OK



Weka 85% Split Validation using Decision Stump

Weka Explorer

Preprocess Classify Cluster Associate Select attributes Visualize

Classifier

Choose DecisionStump

Test options

Use training set

Supplied test set Set...

Cross-validation Folds 16

Percentage split % 85

More options...

(Nom) Selector

Start Stop

Result list (right-click for options)

15:2803 - trees.DecisionStump
15:3056 - trees.DecisionStump
15:3211 - trees.DecisionStump
15:3303 - trees.DecisionStump
15:3314 - trees.DecisionStump
15:3327 - trees.DecisionStump
15:3333 - trees.DecisionStump
15:3340 - trees.DecisionStump
15:3815 - trees.DecisionStump
15:4540 - trees.DecisionStump
15:4630 - trees.DecisionStump

Classifier output

```
Time taken to build model: 0.01 seconds
===
Evaluation on test split ===
Time taken to test model on test split: 0 seconds
===
Summary ===
Correctly Classified Instances      64          73.5632 %
Incorrectly Classified Instances   23          26.4368 %
Kappa statistic                   0
Mean absolute error               0.3882
Root mean squared error           0.45
Relative absolute error            96.6559 %
Root relative squared error       101.9479 %
Coverage of cases (0.95 level)   95.4023 %
Mean rel. region size (0.95 level) 86.7816 %
Total Number of Instances        87
===
Detailed Accuracy By Class ===
      TP Rate   FP Rate   Recall   ROC Area   PRC Area   Class
        1.000     1.000     1.000    0.561     0.762     yes
        0.000     0.000     0.000    0.561     0.291     no
Weighted Avg.                    0.736     0.736     0.736    0.561     0.638
===
Confusion Matrix ===
      a   b   <-- classified as
  64  0 |  a = yes
  23  0 |  b = no
```

Status OK

Log x 0

Weka 4 folds Cross Validation using Decision Stump

Weka Explorer

Preprocess Classify Cluster Associate Select attributes Visualize

Classifier

Choose DecisionStump

Test options

Use training set

Supplied test set Set...

Cross-validation Folds 4

Percentage split % 85

More options...

(Nom) Selector

Start Stop

Result list (right-click for options)

15:2803 - trees.DecisionStump
15:3056 - trees.DecisionStump
15:3211 - trees.DecisionStump
15:3303 - trees.DecisionStump
15:3314 - trees.DecisionStump
15:3327 - trees.DecisionStump
15:3333 - trees.DecisionStump
15:3340 - trees.DecisionStump
15:3815 - trees.DecisionStump
15:4540 - trees.DecisionStump
15:4630 - trees.DecisionStump
15:4725 - trees.DecisionStump

Classifier output

```
0.7135506003430532    0.2864493996569468
Time taken to build model: 0 seconds
===
Stratified cross-validation ===
===
Summary ===
Correctly Classified Instances      416          71.3551 %
Incorrectly Classified Instances   167          28.6449 %
Kappa statistic                   0
Mean absolute error               0.3698
Root mean squared error           0.4342
Relative absolute error            90.3653 %
Root relative squared error       96.0312 %
Coverage of cases (0.95 level)   98.1132 %
Mean rel. region size (0.95 level) 92.7101 %
Total Number of Instances        583
===
Detailed Accuracy By Class ===
      TP Rate   FP Rate   Recall   ROC Area   PRC Area   Class
        1.000     1.000     1.000    0.647     0.803     yes
        0.000     0.000     0.000    0.647     0.375     no
Weighted Avg.                    0.714     0.714     0.714    0.647     0.681
===
Confusion Matrix ===
      a   b   <-- classified as
  416  0 |  a = yes
  167  0 |  b = no
```

Status OK

Log x 0



Weka 8 folds Cross Validation using Decision Stump

The screenshot shows the Weka Explorer interface. The 'Classify' tab is selected. Under 'Classifier', 'DecisionStump' is chosen. In the 'Test options' section, 'Cross-validation' is selected with 'Folds' set to 8. The 'Result list' pane shows a list of 16 instances, with the last one, '15:48:15 - trees.DecisionStump', highlighted. The 'Classifier output' pane displays the following results:

```
0.7135506003430532 0.2864493996569468
Time taken to build model: 0 seconds
== Stratified cross-validation ==
== Summary ==
Correctly Classified Instances      416          71.3551 %
Incorrectly Classified Instances   167          28.6449 %
Kappa statistic                      0
Mean absolute error                 0.3749
Root mean squared error             0.4353
Relative absolute error             91.6333 %
Root relative squared error        96.2818 %
Coverage of cases (0.95 level)    98.6278 %
Mean rel. region size (0.95 level) 94.8542 %
Total Number of Instances           583

== Detailed Accuracy By Class ==
      TP Rate  FP Rate  Recall  ROC Area  PRC Area  Class
1.000   1.000   1.000   0.646   0.808   yes
0.000   0.000   0.000   0.646   0.376   no
Weighted Avg.  0.714   0.714   0.714   0.646   0.684

== Confusion Matrix ==
a   b   <-- classified as
416  0 |  a = yes
167  0 |  b = no
```

Weka 16 folds Cross Validation using Decision Stump

The screenshot shows the Weka Explorer interface. The 'Classify' tab is selected. Under 'Classifier', 'DecisionStump' is chosen. In the 'Test options' section, 'Cross-validation' is selected with 'Folds' set to 16. The 'Result list' pane shows a list of 16 instances, with the last one, '15:48:15 - trees.DecisionStump', highlighted. The 'Classifier output' pane displays the following results:

```
0.7135506003430532 0.2864493996569468
Time taken to build model: 0 seconds
== Stratified cross-validation ==
== Summary ==
Correctly Classified Instances      416          71.3551 %
Incorrectly Classified Instances   167          28.6449 %
Kappa statistic                      0
Mean absolute error                 0.3749
Root mean squared error             0.4353
Relative absolute error             91.6333 %
Root relative squared error        96.2818 %
Coverage of cases (0.95 level)    98.6278 %
Mean rel. region size (0.95 level) 94.8542 %
Total Number of Instances           583

== Detailed Accuracy By Class ==
      TP Rate  FP Rate  Recall  ROC Area  PRC Area  Class
1.000   1.000   1.000   0.646   0.808   yes
0.000   0.000   0.000   0.646   0.376   no
Weighted Avg.  0.714   0.714   0.714   0.646   0.684

== Confusion Matrix ==
a   b   <-- classified as
416  0 |  a = yes
167  0 |  b = no
```