

- 1) UFlorida only training + moose testing
- 2) UFlorida + 2/3 moose training + 1/3 moose testing
- 3) 2/3 moose training + 1/3 moose testing

## Full feature set

(1) ML method and its parameters (2) any data filtering and our labeling parameters and (3) confusion matrix (for 66-34split and not 10cv)

-----

1. **Scenario 3**, 2 class labelling, only moose data as train and test (66-34 split) solver + pc combined , with all features : 83.8 % good accuracy , Remove Useless filter applied, b=30 labelling used

BayesNet

	TP Rate	FP Rate	Precision	Recall	F-Measure	ROC Area	Class
	0.838	0.269	0.465	0.838	0.599	0.881	good
	0.731	0.162	0.942	0.731	0.823	0.881	bad
Weighted Avg.	0.754	0.185	0.838	0.754	0.774	0.881	

=== Confusion Matrix ===

```

a  b  <-- classified as
337 65 |  a = good
387 1052 |  b = bad

```

-----

RF :

Correctly Classified Instances	1830	99.4025 %
Incorrectly Classified Instances	11	0.5975 %
Kappa statistic	0.9826	
Mean absolute error	0.0307	
Root mean squared error	0.0805	
Relative absolute error	9.1177 %	
Root relative squared error	19.4803 %	
Total Number of Instances	1841	

=== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	ROC Area	Class
	0.995	0.006	0.978	0.995	0.986	1	good
	0.994	0.005	0.999	0.994	0.996	1	bad
Weighted Avg.	0.994	0.005	0.994	0.994	0.994	1	

=== Confusion Matrix ===

```
a  b <-- classified as
400  2 |  a = good
9 1430 |  b = bad
```

-----

ADT

=== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	ROC Area	Class
	0.975	0.003	0.987	0.975	0.981	0.999	good
	0.997	0.025	0.993	0.997	0.995	0.999	bad
Weighted Avg.	0.992	0.02	0.992	0.992	0.992	0.999	

=== Confusion Matrix ===

```
a  b <-- classified as
392  10 |  a = good
5 1434 |  b = bad
```

-----

KNN

== Detailed Accuracy By Class ==

	TP Rate	FP Rate	Precision	Recall	F-Measure	ROC Area	Class
	0.863	0.024	0.911	0.863	0.886	0.988	good
	0.976	0.137	0.962	0.976	0.969	0.988	bad
Weighted Avg.	0.952	0.112	0.951	0.952	0.951	0.988	

=== Confusion Matrix ===

```
a  b <-- classified as
347  55 |  a = good
34 1405 |  b = bad
```

-----

J48

=== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	ROC Area	Class
	0.985	0.004	0.985	0.985	0.985	0.994	good
	0.996	0.015	0.996	0.996	0.996	0.994	bad
Weighted Avg.	0.993	0.013	0.993	0.993	0.993	0.994	

=== Confusion Matrix ===

a b <-- classified as  
 396 6 | a = good  
 6 1433 | b = bad

Scenario 3: All features

All features, 5414 moose (train + test)

RemoveUseless Filter 2class labelling (solver + pc combined)

b=30

66-34%train test split

1152 good

4262 bad

ML mtd	Noise filtering			
B= 30	10-cv		66-34%split	
5414 datapts	Overall accuracy	Good solver accuracy	Overall accuracy	Good solver accuracy
LibSVM	44.7	67.8	21.8	100(do not consider, as all labelled as good , 0% bad accuracy)
RF(100)	99.4	99.4	99.4	99.5
BN	75.7	81.3	75.4	83.8
knn(K=10)	96.6	91.2	95.1	86.3
ADT	98.7	95.1	99.1	97.5
J48	99.4	99.4	99.3	98.5

## Scenario 1: All features 5437 UFlorida (train) + 5414 moose (test)

RemoveUseless Filter + <10,000 nnz excluded from training set

2class labelling (solver + pc combined)

b=30

2109 good, 8742 bad

50.11 % train, 49.89 % test

ML mtd	Noise filtering			
B= 30	10-cv(No more Uflorida as train, moose as test; its a mix)		50.11 % train, 49.89 % test	
10851 datapts	Overall accuracy	Good solver accuracy	Overall accuracy	Good solver accuracy
LibSVM	Too slow	Too slow	19.56	100(do not consider, as all labelled as good , 0% bad accuracy)
RF(100)	95.6	83.2	94.5	80.5
BN	78.7	57.7	76.5	63.6
knn(K=10)	90.4	65.3	89.3	64.7
ADT	89.8	54.3	88.5	53.1
J48	96.0	89.4	95.0	87.2

RF

Correctly Classified Instances	5120	94.5696 %
Incorrectly Classified Instances	294	5.4304 %
Kappa statistic	0.8198	
Mean absolute error	0.1007	
Root mean squared error	0.2019	
Relative absolute error	32.154 %	
Root relative squared error	50.9052 %	
Total Number of Instances	5414	

=== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	ROC Area	Class
	0.805	0.02	0.906	0.805	0.853	0.981	good
	0.98	0.195	0.954	0.98	0.967	0.981	bad
Weighted Avg.	0.946	0.16	0.945	0.946	0.944	0.981	

=== Confusion Matrix ===

```
a  b  <-- classified as
853 206 | a = good
88 4267 | b = bad
```

-----  
BayesNet

Correctly Classified Instances	4142	76.5054 %
Incorrectly Classified Instances	1272	23.4946 %

Kappa statistic	0.367
Mean absolute error	0.2395
Root mean squared error	0.4639
Relative absolute error	76.4703 %
Root relative squared error	116.9465 %
Total Number of Instances	5414

=== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	ROC Area	Class
	0.636	0.204	0.432	0.636	0.515	0.807	good
	0.796	0.364	0.9	0.796	0.845	0.807	bad
Weighted Avg.	0.765	0.332	0.808	0.765	0.78	0.807	

=== Confusion Matrix ===

```

a  b  <-- classified as
674 385 |  a = good
887 3468 |  b = bad

```

---

LibSVM

=== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	ROC Area	Class
	1	1	0.196	1	0.327	0.5	good
	0	0	0	0	0.5		bad
Weighted Avg.	0.196	0.196	0.038	0.196	0.064	0.5	

=== Confusion Matrix ===

```

a  b  <-- classified as
1059 0 |  a = good
4355 0 |  b = bad

```

---

Knn

=== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	ROC Area	Class
	0.647	0.047	0.771	0.647	0.704	0.919	good
	0.953	0.353	0.917	0.953	0.935	0.919	bad
Weighted Avg.	0.893	0.293	0.889	0.893	0.89	0.919	

=== Confusion Matrix ===

```

a  b  <-- classified as
685 374 |  a = good
203 4152 |  b = bad

```

---

ADT

Correctly Classified Instances	4792	88.5113 %
Incorrectly Classified Instances	622	11.4887 %

Kappa statistic	0.5789
Mean absolute error	0.2028
Root mean squared error	0.2967
Relative absolute error	64.735 %
Root relative squared error	74.7886 %
Total Number of Instances	5414

=== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	ROC Area	Class
	0.531	0.029	0.818	0.531	0.644	0.927	good
	0.971	0.469	0.895	0.971	0.932	0.927	bad
Weighted Avg.	0.885	0.383	0.88	0.885	0.875	0.927	

=== Confusion Matrix ===

```

a  b  <-- classified as
562 497 |  a = good
125 4230 |  b = bad

```

-----

J48

=== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	ROC Area	Class
	0.872	0.03	0.875	0.872	0.873	0.934	good
	0.97	0.128	0.969	0.97	0.969	0.934	bad
Weighted Avg.	0.95	0.109	0.95	0.95	0.95	0.934	

=== Confusion Matrix ===

```

a  b  <-- classified as
923 136 |  a = good
132 4223 |  b = bad

```

-----

Confusion Matrix For 10cv for scenario 1

J48

Correctly Classified Instances	10426	96.0833 %
Incorrectly Classified Instances	425	3.9167 %
Kappa statistic	0.8744	
Mean absolute error	0.0474	
Root mean squared error	0.189	
Relative absolute error	15.1381 %	
Root relative squared error	47.7559 %	
Total Number of Instances	10851	

=== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	ROC Area	Class
	0.894	0.023	0.904	0.894	0.899	0.949	good
	0.977	0.106	0.974	0.977	0.976	0.949	bad
Weighted Avg.	0.961	0.09	0.961	0.961	0.961	0.949	

=== Confusion Matrix ===

```
a  b  <-- classified as
1885 224 | a = good
201 8541 | b = bad
```

-----

ADT

Correctly Classified Instances	9749	89.8443 %
Incorrectly Classified Instances	1102	10.1557 %
Kappa statistic	0.6191	
Mean absolute error	0.2028	
Root mean squared error	0.2857	
Relative absolute error	64.7604 %	
Root relative squared error	72.2104 %	
Total Number of Instances	10851	

=== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	ROC Area	Class
	0.543	0.016	0.892	0.543	0.675	0.933	good
	0.984	0.457	0.899	0.984	0.94	0.933	bad
Weighted Avg.	0.898	0.371	0.898	0.898	0.888	0.933	

=== Confusion Matrix ===

```
a  b  <-- classified as
1145 964 | a = good
138 8604 | b = bad
```

-----

KNN

Correctly Classified Instances	9812	90.4248 %
Incorrectly Classified Instances	1039	9.5752 %
Kappa statistic	0.6691	
Mean absolute error	0.1254	
Root mean squared error	0.2604	
Relative absolute error	40.0296 %	
Root relative squared error	65.7977 %	
Total Number of Instances	10851	

=== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	ROC Area	Class
	0.653	0.035	0.817	0.653	0.726	0.94	good
	0.965	0.347	0.92	0.965	0.942	0.94	bad
Weighted Avg.	0.904	0.286	0.9	0.904	0.9	0.94	

=== Confusion Matrix ===

```
a  b  <-- classified as
```

1378 731 | a = good  
308 8434 | b = bad

-----  
BN

Correctly Classified Instances	8550	78.7946 %
Incorrectly Classified Instances	2301	21.2054 %
Kappa statistic	0.3805	
Mean absolute error	0.2152	
Root mean squared error	0.4449	
Relative absolute error	68.7035 %	
Root relative squared error	112.4431 %	
Total Number of Instances	10851	

=== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	ROC Area	Class
	0.577	0.161	0.463	0.577	0.514	0.814	good
	0.839	0.423	0.892	0.839	0.864	0.814	bad
Weighted Avg.	0.788	0.372	0.808	0.788	0.796	0.814	

=== Confusion Matrix ===

a b <-- classified as  
1217 892 | a = good  
1409 7333 | b = bad

-----  
RF

Correctly Classified Instances	10382	95.6778 %
Incorrectly Classified Instances	469	4.3222 %
Kappa statistic	0.8558	
Mean absolute error	0.0843	
Root mean squared error	0.1803	
Relative absolute error	26.9219 %	
Root relative squared error	45.5681 %	
Total Number of Instances	10851	

=== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	ROC Area	Class
	0.832	0.013	0.939	0.832	0.882	0.99	good
	0.987	0.168	0.961	0.987	0.974	0.99	bad
Weighted Avg.	0.957	0.138	0.956	0.957	0.956	0.99	

=== Confusion Matrix ===

a b <-- classified as  
1755 354 | a = good  
115 8627 | b = bad



## Scenario 2: All features [5437 UFlorida + 3609 moose] (train) +[1805 moose (test)]

(UFlorida + 2/3rd moose as train) + (1/3rd moose as test)

RemoveUseless Filter + <10,000 nnz excluded from training set

2class labelling (solver + pc combined)

b=30

2109 good, 8742 bad

Same data file as scenario 1, just diff. Train-test split

ML mtd	Noise filtering			
B= 30	10-cv		83.37 train-16.63 test split	
10851 datapts	Overall accuracy	Good solver accuracy	Overall accuracy	Good solver accuracy
LibSVM	Same as scenario 1, 10cv	Same as scenario 1, 10cv	Too slow	Too slow
RF(100)	Same as scenario 1, 10cv	Same as scenario 1, 10cv	95.6	83.7
BN	Same as scenario 1, 10cv	Same as scenario 1, 10cv	77.8	60.8
knn(K=10)	Same as scenario 1, 10cv	Same as scenario 1, 10cv	90.4	66.3
ADT	Same as scenario 1, 10cv	Same as scenario 1, 10cv	90.3	58.0
J48	Same as scenario 1, 10cv	Same as scenario 1, 10cv	95.9	90.1

RF

Correctly Classified Instances 1726 95.6233 %

Incorrectly Classified Instances 79 4.3767 %

Kappa statistic 0.8578

Mean absolute error 0.087

Root mean squared error 0.1858

Relative absolute error 27.5243 %

Root relative squared error 46.4002 %

Total Number of Instances 1805

=== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	ROC Area	Class
	0.837	0.014	0.938	0.837	0.885	0.986	good
	0.986	0.163	0.96	0.986	0.973	0.986	bad
Weighted Avg.	0.956	0.133	0.956	0.956	0.955	0.986	

=== Confusion Matrix ===

```
a  b  <-- classified as
303 59 | a = good
20 1423 | b = bad
```

-----  
BN

Correctly Classified Instances	1405	77.8393 %
Incorrectly Classified Instances	400	22.1607 %
Kappa statistic	0.383	
Mean absolute error	0.2237	
Root mean squared error	0.4548	
Relative absolute error	70.7395 %	
Root relative squared error	113.5543 %	
Total Number of Instances	1805	

=== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	ROC Area	Class
	0.608	0.179	0.46	0.608	0.524	0.815	good
	0.821	0.392	0.893	0.821	0.856	0.815	bad
Weighted Avg.	0.778	0.349	0.806	0.778	0.789	0.815	

=== Confusion Matrix ===

```

a   b  <-- classified as
220 142 |  a = good
258 1185 |  b = bad

```

-----

Knn

Correctly Classified Instances	1632	90.4155 %
Incorrectly Classified Instances	173	9.5845 %
Kappa statistic	0.6774	
Mean absolute error	0.1261	
Root mean squared error	0.2636	
Relative absolute error	39.8868 %	
Root relative squared error	65.8184 %	
Total Number of Instances	1805	

=== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	ROC Area	Class
	0.663	0.035	0.825	0.663	0.735	0.935	good
	0.965	0.337	0.919	0.965	0.941	0.935	bad
Weighted Avg.	0.904	0.277	0.9	0.904	0.9	0.935	

=== Confusion Matrix ===

```

a   b  <-- classified as
240 122 |  a = good
51 1392 |  b = bad

```

-----

ADT

Correctly Classified Instances	1630	90.3047 %
Incorrectly Classified Instances	175	9.6953 %
Kappa statistic	0.6511	
Mean absolute error	0.2027	

Root mean squared error	0.2918
Relative absolute error	64.0997 %
Root relative squared error	72.8743 %
Total Number of Instances	1805

=== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	ROC Area	Class
	0.58	0.016	0.901	0.58	0.706	0.932	good
	0.984	0.42	0.903	0.984	0.942	0.932	bad
Weighted Avg.	0.903	0.339	0.903	0.903	0.895	0.932	

=== Confusion Matrix ===

```
a  b  <-- classified as
210 152 | a = good
23 1420 | b = bad
```

-----  
J48

Correctly Classified Instances	1732	95.9557 %
Incorrectly Classified Instances	73	4.0443 %
Kappa statistic	0.874	
Mean absolute error	0.0467	
Root mean squared error	0.1908	
Relative absolute error	14.7791 %	
Root relative squared error	47.648 %	
Total Number of Instances	1805	

=== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	ROC Area	Class
	0.901	0.026	0.898	0.901	0.899	0.945	good
	0.974	0.099	0.975	0.974	0.975	0.945	bad
Weighted Avg.	0.96	0.085	0.96	0.96	0.96	0.945	

=== Confusion Matrix ===

```
a  b  <-- classified as
326 36 | a = good
37 1406 | b = bad
```

## Analysis for Reduced sets

More data than for all feature analysis

RS1 (8 features)

**Scenario 3:** [ moose] (train) +[ moose (test)]

2class labelling (solver + pc combined)

b=30

479 good, 9270 bad

ML mtd	Noise filtering			
B= 30	10-cv		66-34 train-test split	
9749 datapts	Overall accuracy	Good solver accuracy	Overall accuracy	Good solver accuracy
LibSVM	TOO SLOW	TOO SLOW	TOO SLOW	TOO SLOW
RF(100)	99.9	99	99.9	98.8
BN	95.1	96.5	94.7	95.7
knn(K=10)	99.3	94.6	99.1	92.7
ADT	99.7	95.2	99.9	98.8
J48	99.9	99.6	99.9	98.8

CONFUSION MATRICES FOR 66-34%SPLIT

RF

Correctly Classified Instances      3313      99.9397 %

Incorrectly Classified Instances      2      0.0603 %

Kappa statistic      0.9935

Mean absolute error      0.0021

Root mean squared error      0.0214

Relative absolute error      2.2436 %

Root relative squared error      9.8898 %

Total Number of Instances      3315

=== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	ROC Area	Class
	0.988	0	1	0.988	0.994	1	good
	1	0.012	0.999	1	1	1	bad
Weighted Avg.	0.999	0.012	0.999	0.999	0.999	0.999	1

=== Confusion Matrix ===

```
a  b  <-- classified as
162  2 |  a = good
0 3151 |  b = bad
```

-----  
BN

Correctly Classified Instances	3141	94.7511 %
Incorrectly Classified Instances	174	5.2489 %
Kappa statistic	0.6184	
Mean absolute error	0.0379	
Root mean squared error	0.1519	
Relative absolute error	40.4802 %	
Root relative squared error	70.0257 %	
Total Number of Instances	3315	

=== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	ROC Area	Class
	0.957	0.053	0.485	0.957	0.643	0.995	good
	0.947	0.043	0.998	0.947	0.972	0.995	bad
Weighted Avg.	0.948	0.043	0.972	0.948	0.955	0.995	

=== Confusion Matrix ===

```
a  b  <-- classified as
157  7 |  a = good
167 2984 |  b = bad
```

-----  
ADT

Correctly Classified Instances	3313	99.9397 %
Incorrectly Classified Instances	2	0.0603 %
Kappa statistic	0.9935	
Mean absolute error	0.0123	
Root mean squared error	0.0469	
Relative absolute error	13.0951 %	
Root relative squared error	21.609 %	
Total Number of Instances	3315	

=== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	ROC Area	Class
	0.988	0	1	0.988	0.994	0.993	good
	1	0.012	0.999	1	1	0.993	bad
Weighted Avg.	0.999	0.012	0.999	0.999	0.999	0.993	

=== Confusion Matrix ===

```
a  b  <-- classified as
162  2 |  a = good
0 3151 |  b = bad
```

-----  
J48

Correctly Classified Instances	3313	99.9397 %
Incorrectly Classified Instances	2	0.0603 %
Kappa statistic	0.9935	

Mean absolute error	0.0006
Root mean squared error	0.0246
Relative absolute error	0.6438 %
Root relative squared error	11.3269 %
Total Number of Instances	3315

=== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	ROC Area	Class
	0.988	0	1	0.988	0.994	0.994	good
	1	0.012	0.999	1	1	0.994	bad
Weighted Avg.	0.999	0.012	0.999	0.999	0.999	0.994	

=== Confusion Matrix ===

```

a  b  <-- classified as
162  2 |  a = good
0 3151 |  b = bad

```

---

KNN

Correctly Classified Instances	3286	99.1252 %
Incorrectly Classified Instances	29	0.8748 %
Kappa statistic	0.9083	
Mean absolute error	0.0128	
Root mean squared error	0.0793	
Relative absolute error	13.6323 %	
Root relative squared error	36.5632 %	
Total Number of Instances	3315	

=== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	ROC Area	Class
	0.927	0.005	0.899	0.927	0.913	0.999	good
	0.995	0.073	0.996	0.995	0.995	0.999	bad
Weighted Avg.	0.991	0.07	0.991	0.991	0.991	0.999	

=== Confusion Matrix ===

```

a  b  <-- classified as
152  12 |  a = good
17 3134 |  b = bad

```

## RS2 (6 features)

### **Scenario 3:** [ moose] (train) +[ moose (test)]

2class labelling (solver + pc combined)

b=30

479 good, 9270 bad

ML mtd	Noise filtering			
B= 30	10-cv		66-34% train-test split	
9749 datapts	Overall accuracy	Good solver accuracy	Overall accuracy	Good solver accuracy
LibSVM	TOO SLOW	TOO SLOW	TOO SLOW	TOO SLOW
RF(100)	99.9	99.9	99.9	98.8
BN	95.1	96.5	94.7	95.7
knn(K=10)	99.3	95.4	99.1	93.3
ADT	99.7	95.0	99.9	98.8
J48	99.9	99.6	99.9	98.8

Confusion Matrix for 66-34%train-test split

J48

Correctly Classified Instances	3313	99.9397 %
Incorrectly Classified Instances	2	0.0603 %
Kappa statistic	0.9935	
Mean absolute error	0.0006	
Root mean squared error	0.0246	
Relative absolute error	0.6438 %	
Root relative squared error	11.3269 %	
Total Number of Instances	3315	

=== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	ROC Area	Class
	0.988	0	1	0.988	0.994	0.994	good
	1	0.012	0.999	1	1	0.994	bad
Weighted Avg.	0.999	0.012	0.999	0.999	0.999	0.994	

=== Confusion Matrix ===

a	b	<-- classified as
162	2	a = good
0	3151	b = bad

RF

Correctly Classified Instances	3313	99.9397 %
Incorrectly Classified Instances	2	0.0603 %
Kappa statistic	0.9935	
Mean absolute error	0.0015	
Root mean squared error	0.0224	
Relative absolute error	1.5966 %	
Root relative squared error	10.3329 %	
Total Number of Instances	3315	

=== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	ROC Area	Class
	0.988	0	1	0.988	0.994	1	good
	1	0.012	0.999	1	1	1	bad
Weighted Avg.	0.999	0.012	0.999	0.999	0.999	0.999	1

=== Confusion Matrix ===

```
a  b  <-- classified as
162  2 |  a = good
0 3151 |  b = bad
```

-----

BN

Correctly Classified Instances	3141	94.7511 %
Incorrectly Classified Instances	174	5.2489 %
Kappa statistic	0.6184	
Mean absolute error	0.0379	
Root mean squared error	0.1519	
Relative absolute error	40.4802 %	
Root relative squared error	70.0257 %	
Total Number of Instances	3315	

=== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	ROC Area	Class
	0.957	0.053	0.485	0.957	0.643	0.995	good
	0.947	0.043	0.998	0.947	0.972	0.995	bad
Weighted Avg.	0.948	0.043	0.972	0.948	0.955	0.995	

=== Confusion Matrix ===

```
a  b  <-- classified as
157  7 |  a = good
167 2984 |  b = bad
```

-----

KNN

Correctly Classified Instances	3288	99.1855 %
Incorrectly Classified Instances	27	0.8145 %
Kappa statistic	0.9146	
Mean absolute error	0.0125	



```
Root mean squared error      0.0782
Relative absolute error      13.3748 %
Root relative squared error   36.0508 %
Total Number of Instances    3315
```

=== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	ROC Area	Class
	0.933	0.005	0.905	0.933	0.919	0.999	good
	0.995	0.067	0.997	0.995	0.996	0.999	bad
Weighted Avg.	0.992	0.064	0.992	0.992	0.992	0.999	

=== Confusion Matrix ===

```
  a   b  <-- classified as
153  11 |  a = good
16 3135 |  b = bad
```

-----

ADT

Correctly Classified Instances	3313	99.9397 %
Incorrectly Classified Instances	2	0.0603 %
Kappa statistic	0.9935	
Mean absolute error	0.0123	
Root mean squared error	0.0469	
Relative absolute error	13.0951 %	
Root relative squared error	21.609 %	
Total Number of Instances	3315	

=== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	ROC Area	Class
	0.988	0	1	0.988	0.994	0.993	good
	1	0.012	0.999	1	1	0.993	bad
Weighted Avg.	0.999	0.012	0.999	0.999	0.999	0.993	

=== Confusion Matrix ===

```
  a   b  <-- classified as
162   2 |  a = good
0 3151 |  b = bad
```

-----

### Scenario 3: All features [ moose] (train) +[ moose (test)]

2class labelling (solver + pc combined)

b=30

15757 datapoints

1007 good,14750 bad

ML mtd	Noise filtering			
B= 30	10-cv	66-34% train-test split		
15757 datapts	Overall accuracy	Good solver accuracy	Overall accuracy	Good solver accuracy
LibSVM	TOO SLOW	TOO SLOW	TOO SLOW	TOO SLOW
RF(100)	99.7	98.1	99.7	96.8
BN	74.5	76.5	76.7	78.0
knn(K=10)	96.8	69.3	96.2	63.8
ADT	99.6	97.7	99.6	97.4
J48	99.8	99.7	99.7	98.0

Confusion matrix for 66-34% train-test split

ADT

Correctly Classified Instances 5339 99.664 %

Incorrectly Classified Instances 18 0.336 %

Kappa statistic 0.9721

Mean absolute error 0.0116

Root mean squared error 0.0515

Relative absolute error 9.658 %

Root relative squared error 20.987 %

Total Number of Instances 5357

=== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	ROC Area	Class
	0.974	0.002	0.974	0.974	0.974	0.997	good
	0.998	0.026	0.998	0.998	0.998	0.997	bad
Weighted Avg.	0.997	0.025	0.997	0.997	0.997	0.997	

=== Confusion Matrix ===

```
a  b  <-- classified as
336  9 |  a = good
 9 5003 |  b = bad
```

-----  
RF

Correctly Classified Instances	5342	99.72 %
Incorrectly Classified Instances	15	0.28 %
Kappa statistic	0.9765	
Mean absolute error	0.0126	
Root mean squared error	0.0535	
Relative absolute error	10.4722 %	
Root relative squared error	21.8044 %	
Total Number of Instances	5357	

=== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	ROC Area	Class
	0.968	0.001	0.988	0.968	0.978	1	good
	0.999	0.032	0.998	0.999	0.999	1	bad
Weighted Avg.	0.997	0.03	0.997	0.997	0.997	1	

=== Confusion Matrix ===

```

a  b  <-- classified as
334  11 |  a = good
 4 5008 |  b = bad

```

-----

J48

Correctly Classified Instances	5345	99.776 %
Incorrectly Classified Instances	12	0.224 %
Kappa statistic	0.9814	
Mean absolute error	0.0032	
Root mean squared error	0.0464	
Relative absolute error	2.6936 %	
Root relative squared error	18.896 %	
Total Number of Instances	5357	

=== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	ROC Area	Class
	0.98	0.001	0.985	0.98	0.983	0.99	good
	0.999	0.02	0.999	0.999	0.999	0.99	bad
Weighted Avg.	0.998	0.019	0.998	0.998	0.998	0.99	

=== Confusion Matrix ===

```

a  b  <-- classified as
338   7 |  a = good
 5 5007 |  b = bad

```

-----

BN

Correctly Classified Instances	4112	76.7594 %
Incorrectly Classified Instances	1245	23.2406 %
Kappa statistic	0.2208	

Mean absolute error	0.2346
Root mean squared error	0.4789
Relative absolute error	195.5926 %
Root relative squared error	195.1143 %
Total Number of Instances	5357

=== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	ROC Area	Class
	0.78	0.233	0.187	0.78	0.302	0.888	good
	0.767	0.22	0.981	0.767	0.861	0.888	bad
Weighted Avg.	0.768	0.221	0.93	0.768	0.825	0.888	

=== Confusion Matrix ===

```

a  b  <-- classified as
269 76 |  a = good
1169 3843 |  b = bad

```

---

KNN

Correctly Classified Instances	5157	96.2666 %
Incorrectly Classified Instances	200	3.7334 %
Kappa statistic	0.6678	
Mean absolute error	0.0521	
Root mean squared error	0.1631	
Relative absolute error	43.4662 %	
Root relative squared error	66.427 %	
Total Number of Instances	5357	

=== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	ROC Area	Class
	0.638	0.015	0.746	0.638	0.688	0.978	good
	0.985	0.362	0.975	0.985	0.98	0.978	bad
Weighted Avg.	0.963	0.34	0.961	0.963	0.961	0.978	

=== Confusion Matrix ===

```

a  b  <-- classified as
220 125 |  a = good
75 4937 |  b = bad

```

---

## Scenario 3 : Reduced Sets [ moose] (train) +[ moose (test)]

2class labelling (solver + pc combined)

### RS1: 8 FEATURES

b=30

15757 datapoints

1007 good,14750 bad

ML mtd	Noise filtering			
B= 30	10-cv		66-34% train-test split	
15757 datapts	Overall accuracy	Good solver accuracy	Overall accuracy	Good solver accuracy
LibSVM	TOO SLOW	TOO SLOW	TOO SLOW	TOO SLOW
RF(100)	99.8	99.2	99.8	98.8
BN	97.8	85.3	98.0	83.8
knn(K=10)	98.9	94.7	98.7	90.1
ADT	99.8	99.0	99.8	98.6
J48	99.8	99.5	99.7	97.4

### C FOR 66-34% TRAIN TEST SPLIT

#### KNN

```

Correctly Classified Instances      5291      98.768 %
Incorrectly Classified Instances    66         1.232 %
Kappa statistic                    0.8975
Mean absolute error                 0.0179
Root mean squared error             0.0945
Relative absolute error             14.9417 %
Root relative squared error         38.4775 %
Total Number of Instances          5357

```

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

```

      TP Rate  FP Rate  Precision  Recall  F-Measure  ROC Area  Class
      0.901   0.006   0.907   0.901   0.904   0.997   good
      0.994   0.099   0.993   0.994   0.993   0.997   bad
Weighted Avg.  0.988   0.093   0.988   0.988   0.988   0.997

```

#### === Confusion Matrix ===

```

a  b  <-- classified as
311 34 | a = good

```

32 4980 | b = bad

-----  
ADT

Correctly Classified Instances	5347	99.8133 %
Incorrectly Classified Instances	10	0.1867 %
Kappa statistic	0.9845	
Mean absolute error	0.0134	
Root mean squared error	0.0508	
Relative absolute error	11.1516 %	
Root relative squared error	20.6983 %	
Total Number of Instances	5357	

=== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	ROC Area	Class
	0.986	0.001	0.986	0.986	0.986	0.995	good
	0.999	0.014	0.999	0.999	0.999	0.995	bad
Weighted Avg.	0.998	0.014	0.998	0.998	0.998	0.995	

=== Confusion Matrix ===

a	b	<-- classified as
340	5	a = good
5	5007	b = bad

-----

J48

Correctly Classified Instances	5343	99.7387 %
Incorrectly Classified Instances	14	0.2613 %
Kappa statistic	0.9782	
Mean absolute error	0.0039	
Root mean squared error	0.0505	
Relative absolute error	3.2474 %	
Root relative squared error	20.5598 %	
Total Number of Instances	5357	

=== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	ROC Area	Class
	0.974	0.001	0.985	0.974	0.98	0.989	good
	0.999	0.026	0.998	0.999	0.999	0.989	bad
Weighted Avg.	0.997	0.024	0.997	0.997	0.997	0.989	

=== Confusion Matrix ===

a	b	<-- classified as
336	9	a = good
5	5007	b = bad

-----

BN

Correctly Classified Instances	5252	98.0399 %
Incorrectly Classified Instances	105	1.9601 %
Kappa statistic	0.8358	
Mean absolute error	0.0279	
Root mean squared error	0.1314	
Relative absolute error	23.2528 %	
Root relative squared error	53.5244 %	
Total Number of Instances	5357	

=== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	ROC Area	Class
	0.838	0.01	0.855	0.838	0.846	0.989	good
	0.99	0.162	0.989	0.99	0.99	0.989	bad
Weighted Avg.	0.98	0.152	0.98	0.98	0.98	0.989	

=== Confusion Matrix ===

```

a   b  <-- classified as
289 56 |  a = good
49 4963 |  b = bad

```

---

## Scenario 3 : Reduced Sets [ moose] (train) +[ moose (test)]

2class labelling (solver + pc combined)

### RS2: 6 FEATURES

b=30

15757 datapoints

1007 good,14750 bad

ML mtd	Noise filtering			
B= 30	10-cv		66-34% train-test split	
15757 datapts	Overall accuracy	Good solver accuracy	Overall accuracy	Good solver accuracy
LibSVM	TOO SLOW	TOO SLOW	TOO SLOW	TOO SLOW
RF(100)	99.8	99.1	99.9	99.1
BN	98.4	89.7	98.3	83.8
knn(K=10)	99.0	95.2	98.8	90.7
ADT	99.8	99.1	99.8	98.8
J48	99.8	99.5	99.7	97.7

CM for 66-34% split

Knn

Correctly Classified Instances      5293      98.8053 %

Incorrectly Classified Instances      64      1.1947 %

Kappa statistic      0.9009

Mean absolute error      0.0177

Root mean squared error      0.0933

Relative absolute error      14.7238 %

Root relative squared error      37.9995 %

Total Number of Instances      5357

=== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	ROC Area	Class
	0.907	0.006	0.907	0.907	0.907	0.997	good
	0.994	0.093	0.994	0.994	0.994	0.997	bad
Weighted Avg.	0.988	0.087	0.988	0.988	0.988	0.997	

=== Confusion Matrix ===

```
a  b  <-- classified as
313 32 |  a = good
32 4980 |  b = bad
```



-----  
RF

Correctly Classified Instances	5352	99.9067 %
Incorrectly Classified Instances	5	0.0933 %
Kappa statistic	0.9922	
Mean absolute error	0.0016	
Root mean squared error	0.0264	
Relative absolute error	1.2981 %	
Root relative squared error	10.7647 %	
Total Number of Instances	5357	

=== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	ROC Area	Class
	0.991	0	0.994	0.991	0.993	1	good
	1	0.009	0.999	1	1	1	bad
Weighted Avg.	0.999	0.008	0.999	0.999	0.999	1	

=== Confusion Matrix ===

```
a  b  <-- classified as
342  3 |  a = good
 2 5010 |  b = bad
```

-----  
ADT

Correctly Classified Instances	5348	99.832 %
Incorrectly Classified Instances	9	0.168 %
Kappa statistic	0.9861	
Mean absolute error	0.0133	
Root mean squared error	0.0494	
Relative absolute error	11.0936 %	
Root relative squared error	20.1399 %	
Total Number of Instances	5357	

=== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	ROC Area	Class
	0.988	0.001	0.986	0.988	0.987	0.996	good
	0.999	0.012	0.999	0.999	0.999	0.996	bad
Weighted Avg.	0.998	0.011	0.998	0.998	0.998	0.996	

=== Confusion Matrix ===

```
a  b  <-- classified as
341  4 |  a = good
 5 5007 |  b = bad
```

-----  
BN

Correctly Classified Instances	5266	98.3013 %
Incorrectly Classified Instances	91	1.6987 %

Kappa statistic	0.8549
Mean absolute error	0.0216
Root mean squared error	0.1124
Relative absolute error	17.9943 %
Root relative squared error	45.772 %
Total Number of Instances	5357

=== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	ROC Area	Class
	0.838	0.007	0.892	0.838	0.864	0.993	good
	0.993	0.162	0.989	0.993	0.991	0.993	bad
Weighted Avg.	0.983	0.152	0.983	0.983	0.983	0.993	

=== Confusion Matrix ===

```

a   b  <-- classified as
289 56 |  a = good
35 4977 |  b = bad

```

---

J48

Correctly Classified Instances	5344	99.7573 %
Incorrectly Classified Instances	13	0.2427 %
Kappa statistic	0.9798	
Mean absolute error	0.0037	
Root mean squared error	0.0486	
Relative absolute error	3.0917 %	
Root relative squared error	19.792 %	
Total Number of Instances	5357	

=== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	ROC Area	Class
	0.977	0.001	0.985	0.977	0.981	0.99	good
	0.999	0.023	0.998	0.999	0.999	0.99	bad
Weighted Avg.	0.998	0.022	0.998	0.998	0.998	0.99	

=== Confusion Matrix ===

```

a   b  <-- classified as
337  8 |  a = good
5 5007 |  b = bad

```

---



---