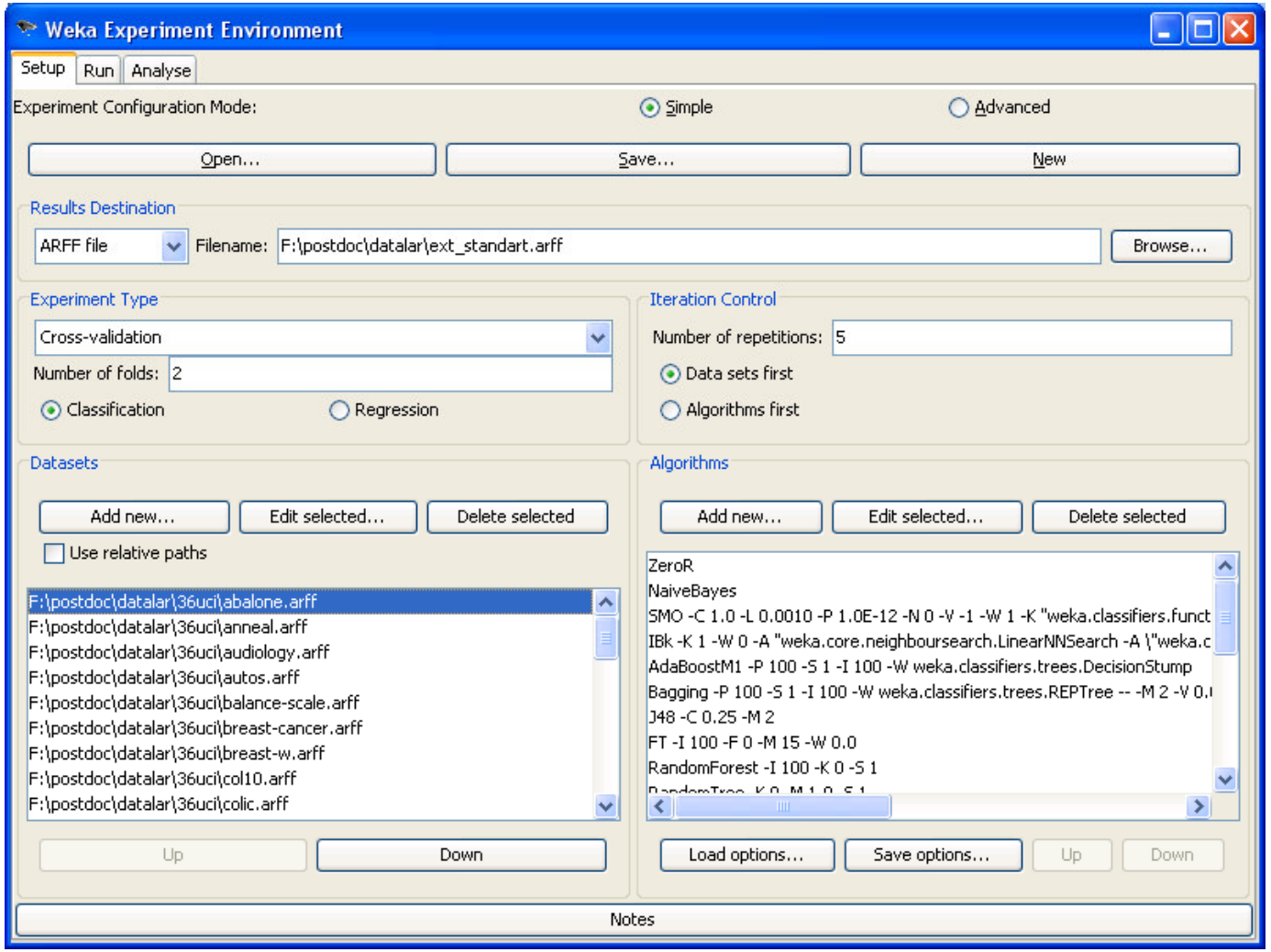
Kolektif Öğrenme Modelleri Karşılaştırma Yöntemleri

- Weka'nın tanıtımı
- Bir Deneyim

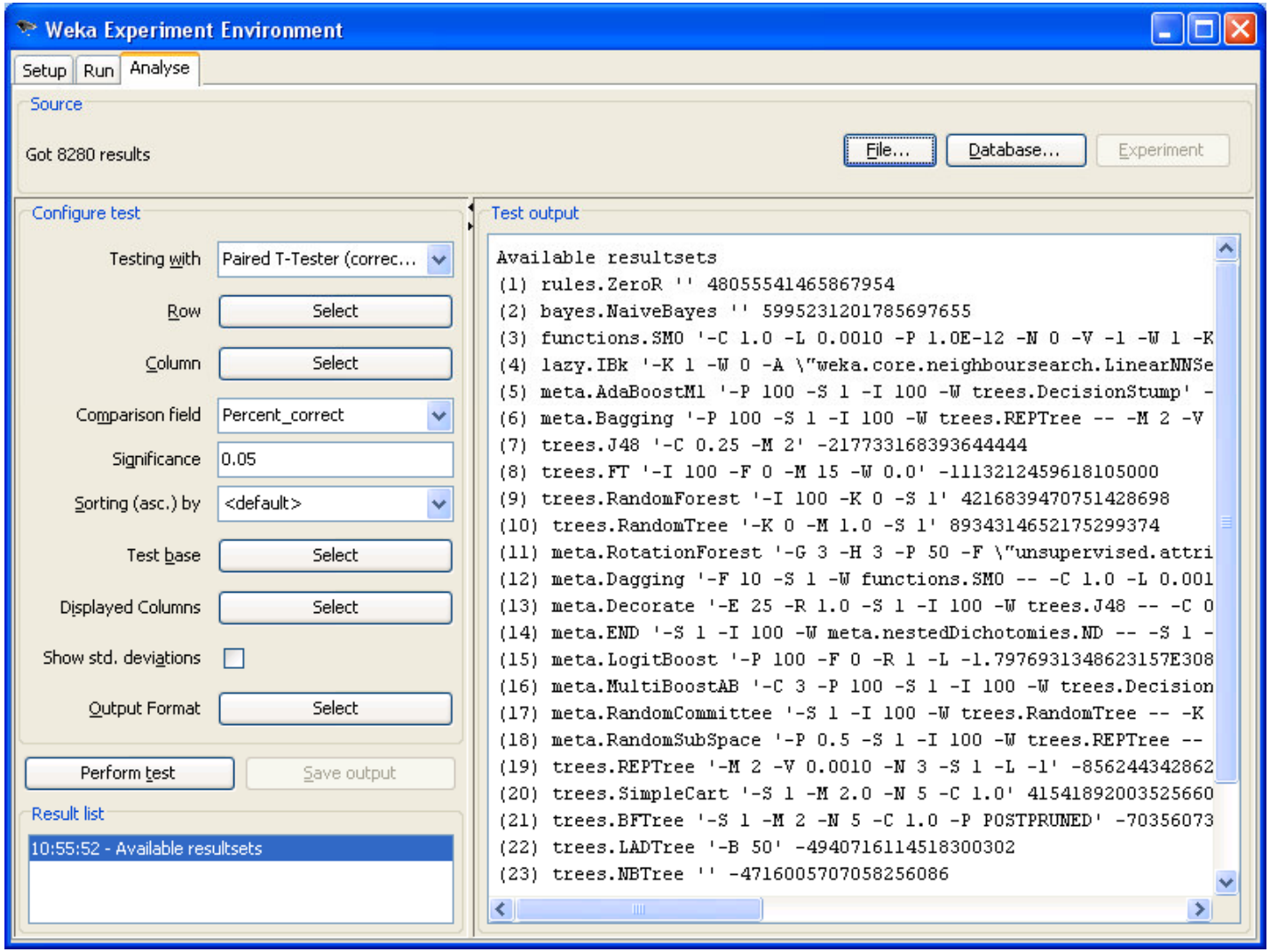
Weka Experimenter

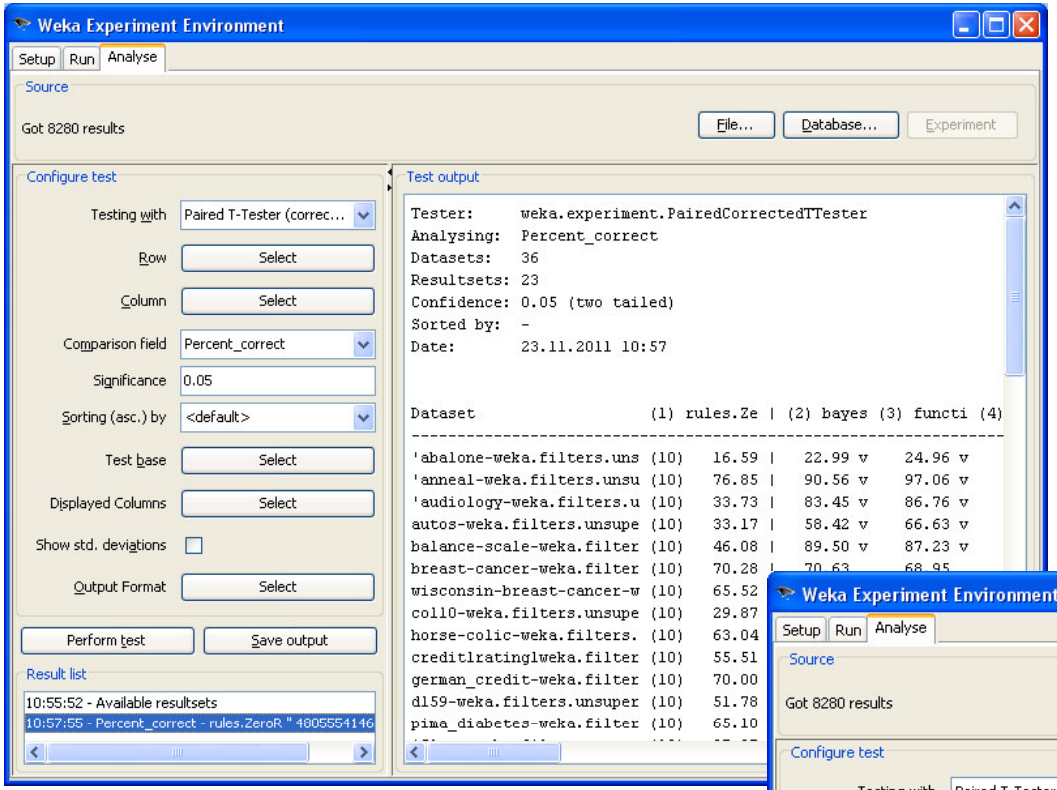
- İhtiyacımız olanlar:
- Weka 3.6.0 sürümü
- Dersin web sitesindeki
  - 36 veri kümesi
  - ext_standart.exp dosyası (konfigürasyon)
  - ext_standart.arff dosyası (sonuçlar)

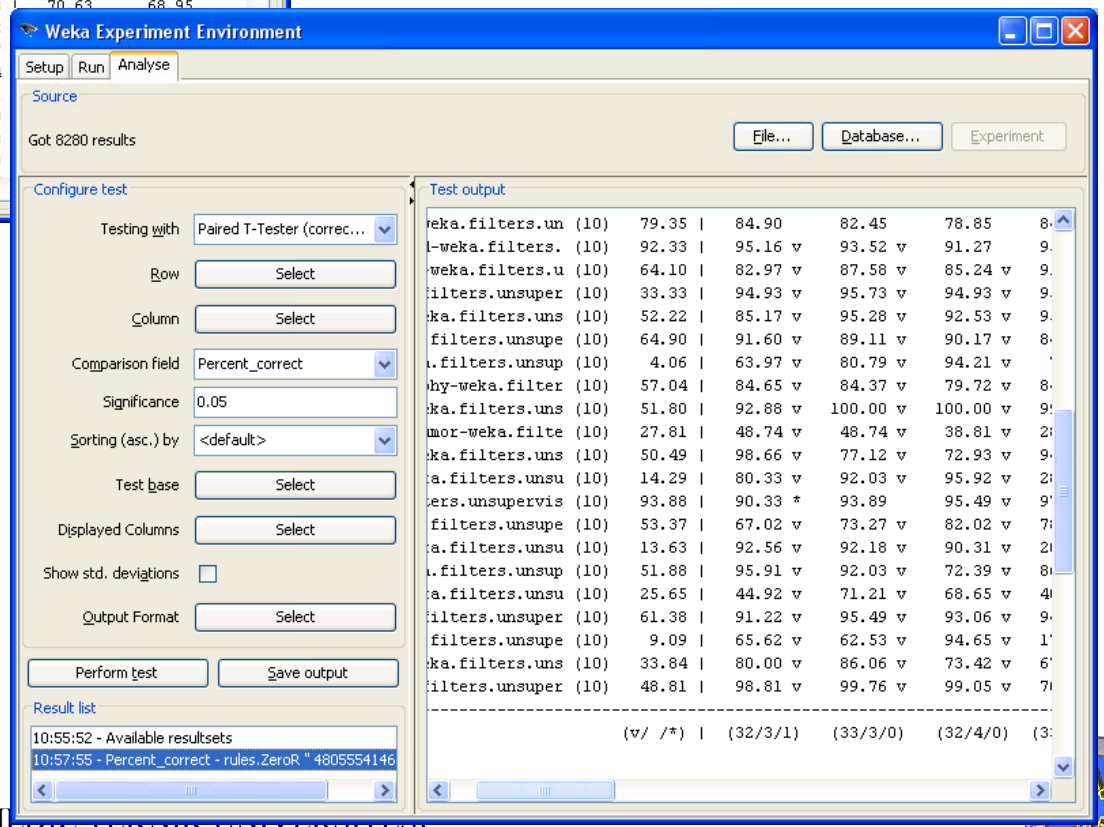
ext_standart.exp dosyasını yüklediğinizde



ext_standart.arff dosyasını yüklediğinizde

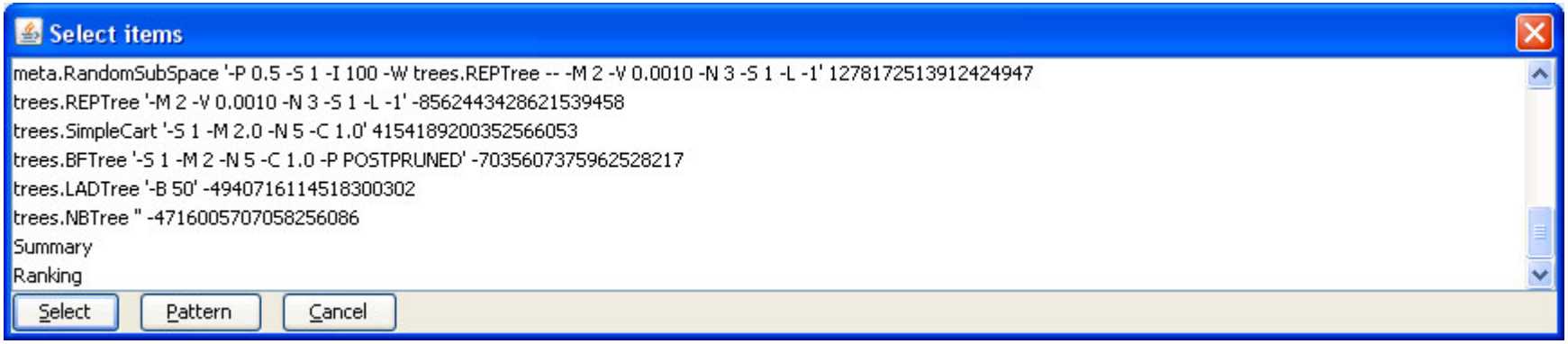






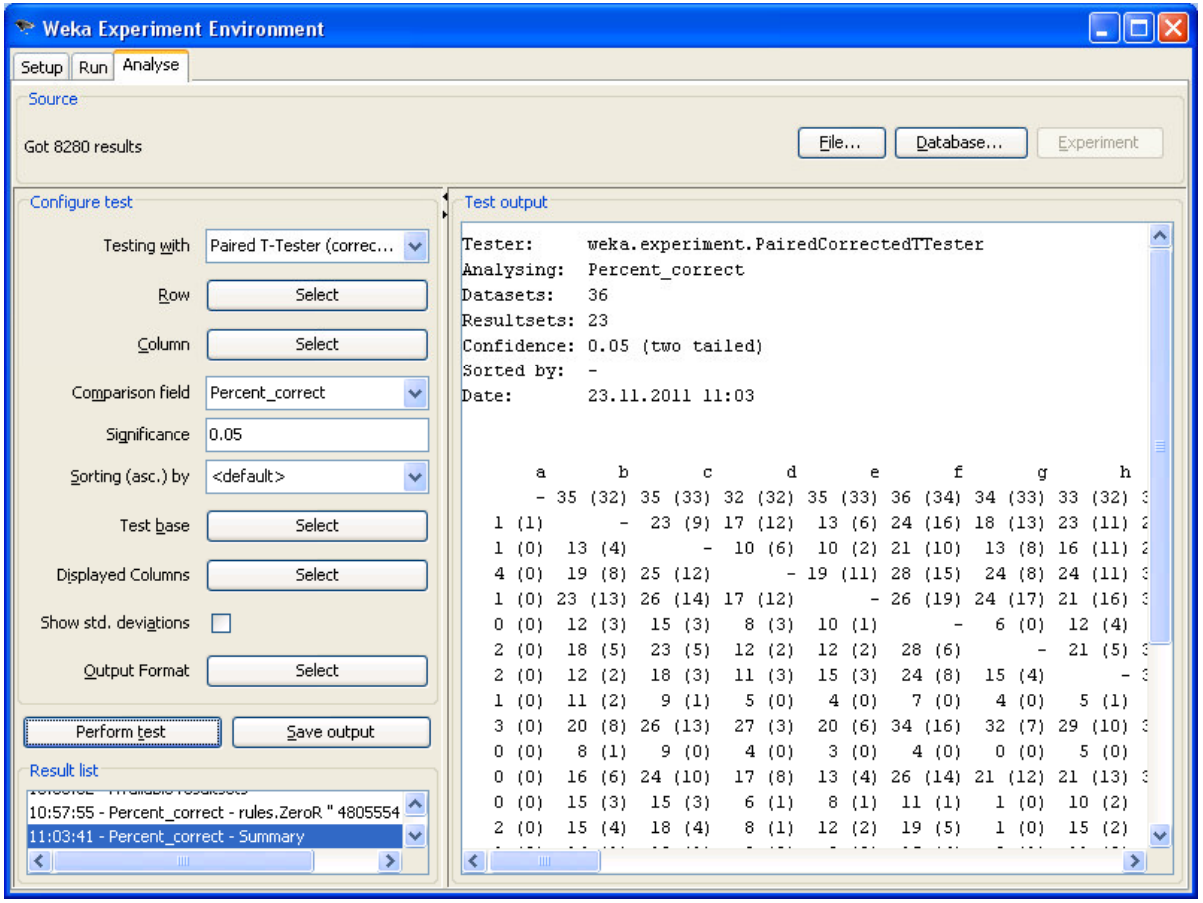
Perform test dediğimizde: ilk sıradaki alg. a göre diğer alg.lerin başarısını bulur

- İlk alg. yerine başka bir alg. a göre başarıları görmek istersek test base i açıp istediğimiz algoritmayı seçip perform test deriz.

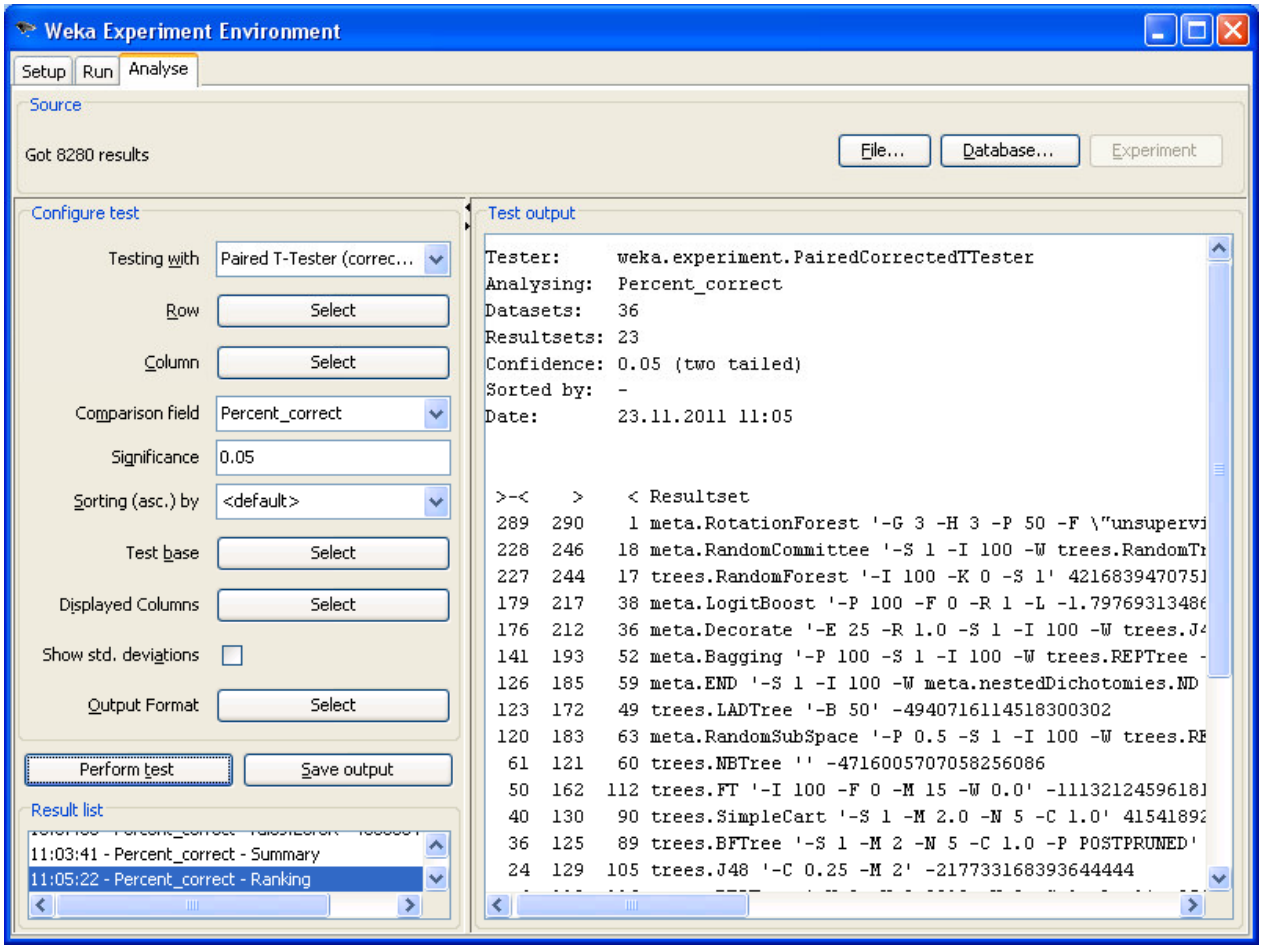


- Toplu karşılaştırma yapmak içinse, test base den summary ya da ranking i seçip perform test deriz.

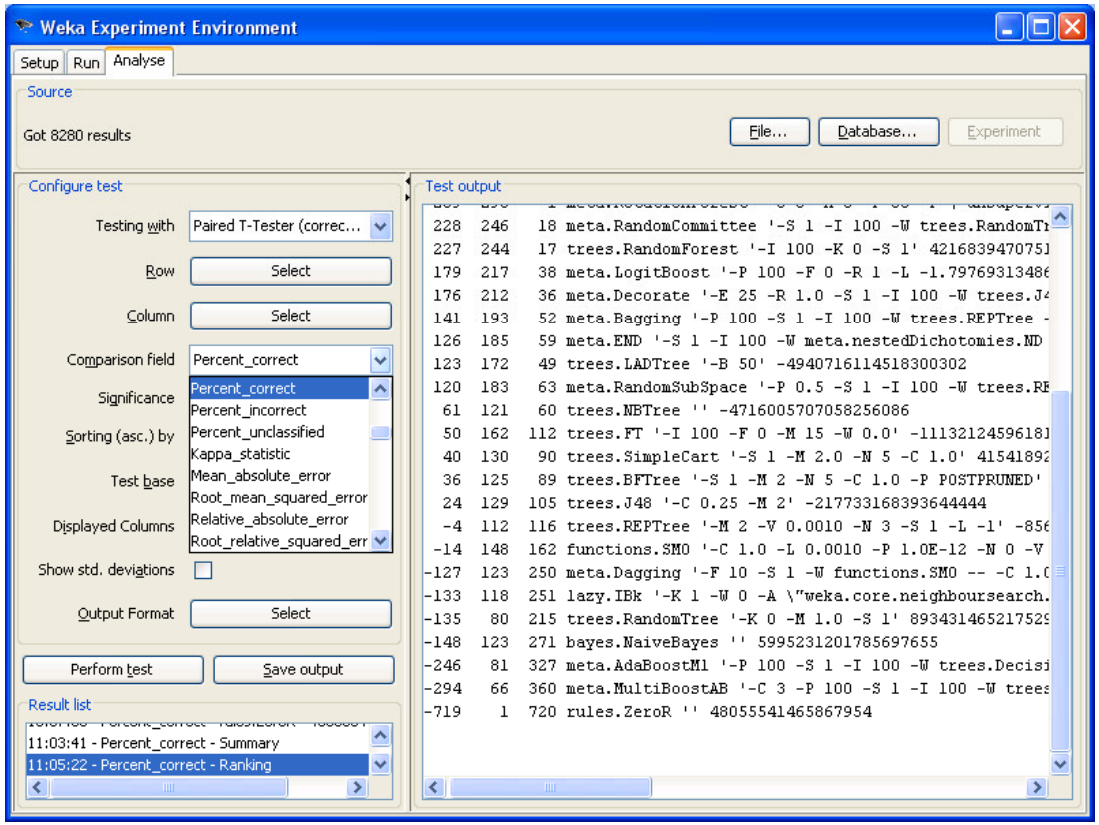
Summary i seçersek



Ranking i seçersek



- Algoritmaları şu ana dek percent correct leri üzerinden karşılaştırdık. Bunu değiştirmek için comparison field i kullanırız. Buna göre algoritmaları eğitim, test zamanı vb. ölçütlere göre benzer şekilde karşılaştırabiliriz.



Our Motivation

- In literature,
  - Several works show higher performances of ensemble algorithms over single algorithms.
  - Lack of run time comparisons
- In this work, we compared single and ensemble algorithms in terms of accuracy and run time.
- Our results can provide clues to the algorithm selection process when the run time is another important factor.
- In this work, we also investigate the similarities of the algorithms and the datasets according to the performances.

Experimental Settings

- 36 UCI dataset
- 11 single, 12 ensemble algorithms
- 5*2 cross validation
- Default parameters
- The number of base learners in ensemble algorithms: 100

36 Datasets

Dataset name	The number of features	The number of classes	The number of Samples
abalone	11	19	4153
Anneal	63	4	890
audiology	70	5	169
Autos	72	5	202
balance-scale	5	3	625
breast-cancer	39	2	286
breast-w	10	2	699
col10	8	10	2019
Colic	61	2	368
credit-a	43	2	690
credit-g	60	2	1000
d159	33	2	7182
diabetes	9	2	768
Glass	10	5	205
heart-statlog	14	2	270
hepatitis	20	2	155
hypothyroid	32	3	3770
ionosphere	34	2	351

Dataset name	The number of features	The number of classes	The number of Samples
iris	5	3	150
kr-vs-kp	40	2	3196
labor	27	2	57
letter	17	26	20000
lymph	38	2	142
mushroom	113	2	8124
primary-tumor	24	11	302
ringnorm	21	2	7400
segment	19	7	2310
sick	32	2	3772
sonar	61	2	208
soybean	84	18	675
splice	288	3	3190
vehicle	19	4	846
vote	17	2	435
vowel	12	11	990
waveform	41	3	5000
Zoo	17	4	84

Single Algorithms

Algorithm Name	Abbreviation
Zero Rule	ZR
Naïve Bayes	NB
Support Vector Machines	SMO
One Nearest Neighbor	KNN
C4.5 Decision Tree	J48
Functional Trees	FT
Random Tree	RT
Fast decision tree learner	REPT
Classification and Regression Trees	CART
Best First Tree	BFT
Multiclass Alternating Decision Tree	LADT
Naïve Bayes Tree	NBT

Ensemble Algorithms

Algorithm Name	Abbre-viation	Base Learner
AdaBoost	ADB	Decision Stump
Bagging	BG	REPT
Random Forest	RNDF	Random Tree
Rotation Forest	ROTF	J48
Dagging	DG	SMO
Decorate	DEC	J48
Ensemble of nested dichotomies	END	nested dichotomies
LogitBoostAB	LB	Decision Stump
MultiBoostAB	MB	Decision Stump
Random Committee	RC	Random Tree
Random Subspace	RS	REPT

Performance Measures

- Average Rank: For each dataset, the performances of ensembles are ranked from 1 (the best) to 23 (the worst). Then all ranks are averaged over all datasets for each algorithm.
- Ranking test count: The ranking test ranks the algorithms according to the total significant wins and significant losses against the other algorithms. Each algorithm is compared with all the other algorithms. The ranking test count is the difference between the number of significant wins and the number of significant losses. The significances is determined by using a “t-test”.

Algorithms' Average Ranks and Ranking Test Counts

Algorithm Abbreviation	Average Rank	Ranking Test Count
ZR	21.9130	-719
NB	12.7391	-148
SMO	**10.2174**	**-14**
KNN	16.4348	-133
J48	14.2174	24
FT	12.4348	50
RT	18.5217	-135
REPT	15.3478	-4
CART	13.1304	40
BFT	13.6957	36
LADT	**10.8696**	**123**
NBT	14.0435	61

Algorithm Abbreviation	Average Rank	Ranking Test Count
ADB	13.5652	-246
BG	**8.5217**	**141**
RNDF	**5.3913**	**227**
ROTF	**4.1304**	**289**
DG	12.5652	-127
DEC	**7.8261**	**176**
END	11.5652	126
LB	**8.3043**	**179**
MB	13.3913	-294
RC	**6.9130**	**228**
RS	10.2609	120

The Best Algorithms

- According to average rank: Rotation Forest (ROTF), Random Forest (RNDF), Random Committees (RC), Decorate (DEC), Bagging (BG), Logit Boost (LB)
- Meaningful difference according to Nemenyi test: **5.781**= 3.616*sqrt((23*24)/(6*36))
- According to average ranking test count: Rotation Forest (ROTF), Random Committees (RC), Random Forest (RNDF), Logit Boost (LB), Decorate (DEC), Bagging (BG)

The Best Algorithms

- All of the 6 best performed algorithms are ensemble algorithms. (Parallel with the current literature)
- The best performed single algorithm is Support Vector Machines (SMO) according to average rank
- The best performed single algorithm is Multiclass Alternating Decision Tree (LADT) according to ranking test count

t-pair test results of 6 best performing algorithms

The first number represents the number of wins for the column with regard to the row. The second number is the number of significant wins.

	BG	**RNDF**	**ROTF**	**DEC**	**LB**	**RC**
BG	-	29(7)	32(8)	25(3)	20(5)	28(6)
RNDF	7(0)	-	25(6)	11(0)	8(2)	15(1)
ROTF	4(0)	10(0)	-	3(0)	9(0)	10(0)
DEC	11(1)	25(3)	32(6)	-	17(4)	25(4)
LB	15(4)	27(4)	27(5)	19(3)	-	25(4)
RC	8(0)	20(0)	26(5)	11(1)	10(2)	-

Speed is another important criterion

Training / Testing Times of the Algorithms on 3 datasets (the values are in seconds)

Algorithm	letter	mushroom	splice
ZR	0.01/0.09	0.00/0.03	0.00/0.01
NB	0.09/3.46	0.20/1.02	0.16/1.25
SMO	40.52/1.17	3.64/0.06	7.20/0.06
KNN	0.00/21.85	0.00/19.41	0.00/31.67
J48	1.84/0.17	0.65/0.03	0.96/0.01
FT	682.7/228.71	14.86/0.09	35.79/6.86
RT	0.29/0.06	0.07/0.04	0.11/0.02
REPT	0.52/0.05	0.40/0.03	0.55/0.01
CART	12.80/0.04	4.56/0.01	5.76/0.01
BFT	27.63/0.05	3.45/0.01	5.23/0.01
LADT	1064.09/0.20	139.43/0.02	144.21/0.01

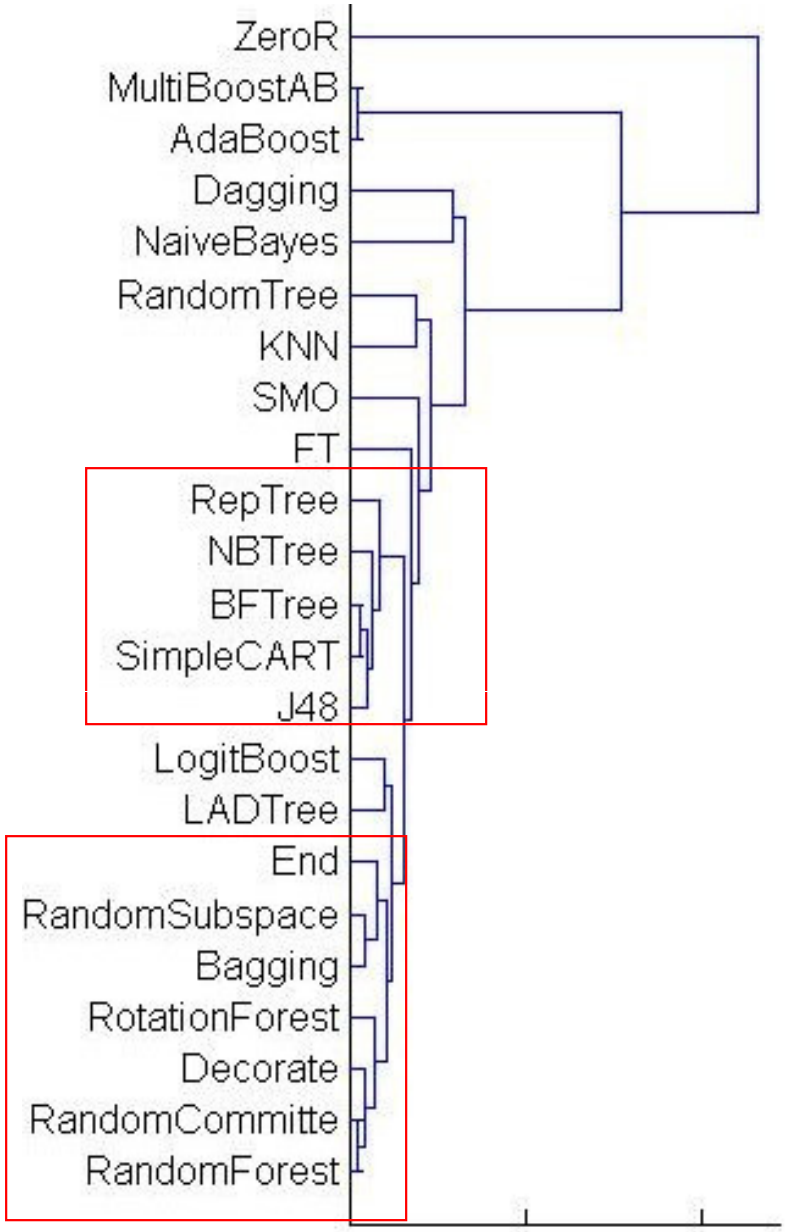
Algorithm	letter	mushroom	splice
NBT	247.59/12.03	169.94/0.25	620.42/0.39
ADB	0.17/0.07	20.49/0.05	2.06/0.02
BG	36.98/1.13	30.86/0.06	55.49/0.04
RNDF	25.94/2.62	6.66/0.20	7.90/0.27
ROTF	**289.57/69.86**	**250.6/348.7**	**324.2/442.7**
DG	330.54/12.14	3.17/0.53	4.57/0.55
DEC	**2160.04/3.51**	**200.82/0.08**	**587.15/0.05**
END	525.17/222.5	0.79/0.66	4.01/0.64
LB	266.90/1.71	48.20/0.08	66.84/0.05
MB	0.18/0.05	20.04/0.04	2.05/0.01
RC	27.12/2.87	7.01/0.22	9.15/0.30
RS	27.74/3.25	20.76/1.24	37.60/0.89

Run Times

- Among the 6 best performing algorithms (DEC, ROTF, LB, BG, RC, and RNDF), DEC and ROTF are the slowest algorithms over 3 largest datasets in term of training time. ROTF also needs very long testing time.
- RNDF and RC algorithms are the fastest algorithms over datasets among the 6 best performing algorithms.
- As a result, RNDF and RC algorithms can be considered as the best algorithms when accuracy and execution time are considered together.

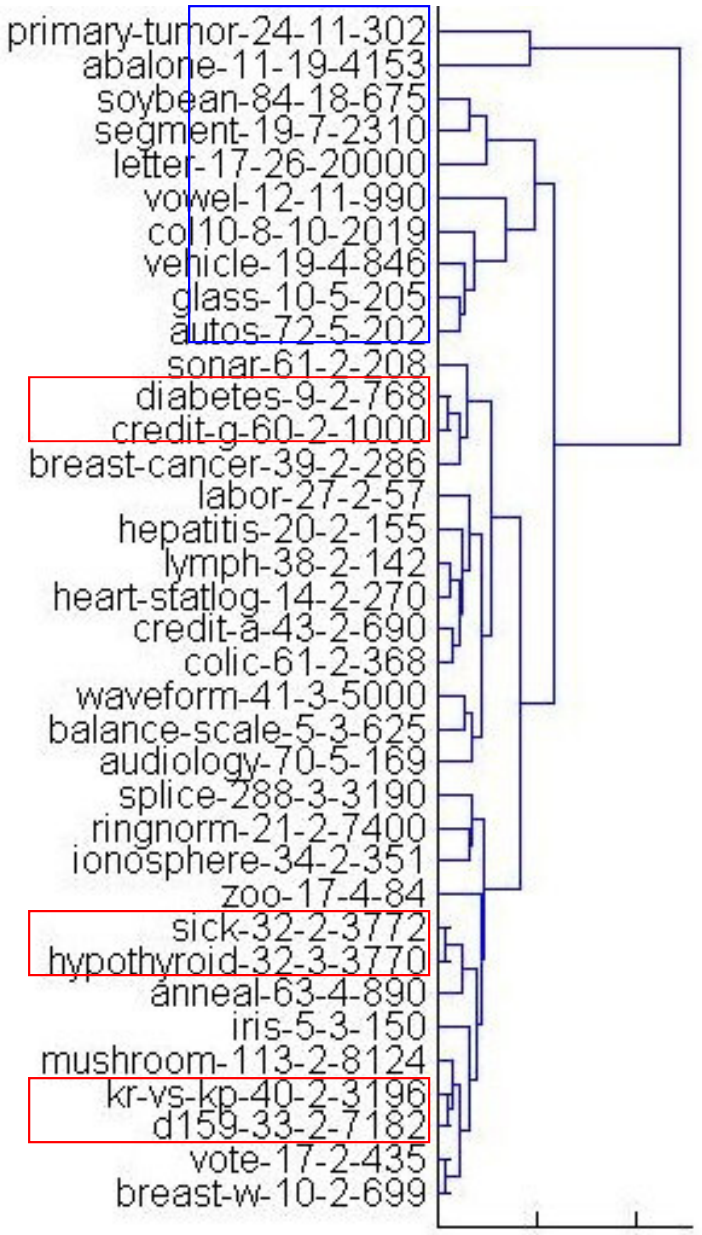
Similarities of Algorithms

- Approach: considering each algorithm as a 36 dimensional point. 36 is the number of datasets.
- Each dimension corresponds to a performance of algorithm over a datasets.
- The similarities of the points were calculated using Euclidian distance metric. After the similarity values were obtained, the hierarchical clustering process was applied to have similar hierarchical groups.



Similarities of Algorithms

- Ensemble algorithms were grouped together.
- Decision tree based algorithms were grouped together.
- The similar ideas generate similar performances.



Similarities of Datasets

- Same approach with the Similarities of Algorithms. Each dimension of a dataset corresponds to the performance obtained by an algorithm with the dataset.
- Format: dataset name, # of features, # of classes, # of samples
- The most similar dataset pairs have similar sample, feature and class numbers. (red rectangles)
- Generally, datasets are grouped together according to their class numbers. (blue rectangle)

Conclusions

- 12 different ensemble algorithms, and 11 single classifiers are compared according to the accuracy and run times using a big dataset collection.
- Rotation Forest has the highest accuracy but it requires long run times.
- However, when accuracy and run time are considered together, Random Forest and Random Committees can be the best choices.