

Nama : Alfi Nurrahmah

NIM : 221810140

No. Absen : 05

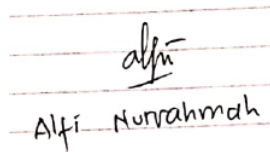
Kelas : 3SD1

Dosen : Ibnu Santoso SST, M.T

Mata Kuliah : Data Mining and Knowledge Management

Hari, Tanggal : Selasa, 27 Oktober 2020

“Saya mengatakan bahwa ujian ini saya kerjakan dengan jujur sesuai dengan kemampuan saya sendiri dan tidak mengutip sebagian atau seluruh pekerjaan orang lain. Jika suatu saat ditemukan saya melanggar ketentuan ujian, saya siap menerima konsekuensi yang berlaku”



Handwritten signature of Alfi Nurrahmah on lined paper.

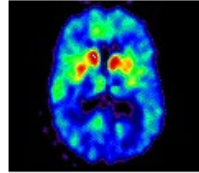
## Dataset

<https://archive.ics.uci.edu/ml/datasets/Parkinsons>

### Parkinsons Data Set

Download: [Data Folder](#), [Data Set Description](#)

Abstract: Oxford Parkinson's Disease Detection Dataset



Data Set Characteristics:	Multivariate	Number of Instances:	197	Area:	Life
Attribute Characteristics:	Real	Number of Attributes:	23	Date Donated	2008-06-26
Associated Tasks:	Classification	Missing Values?	N/A	Number of Web Hits:	275469

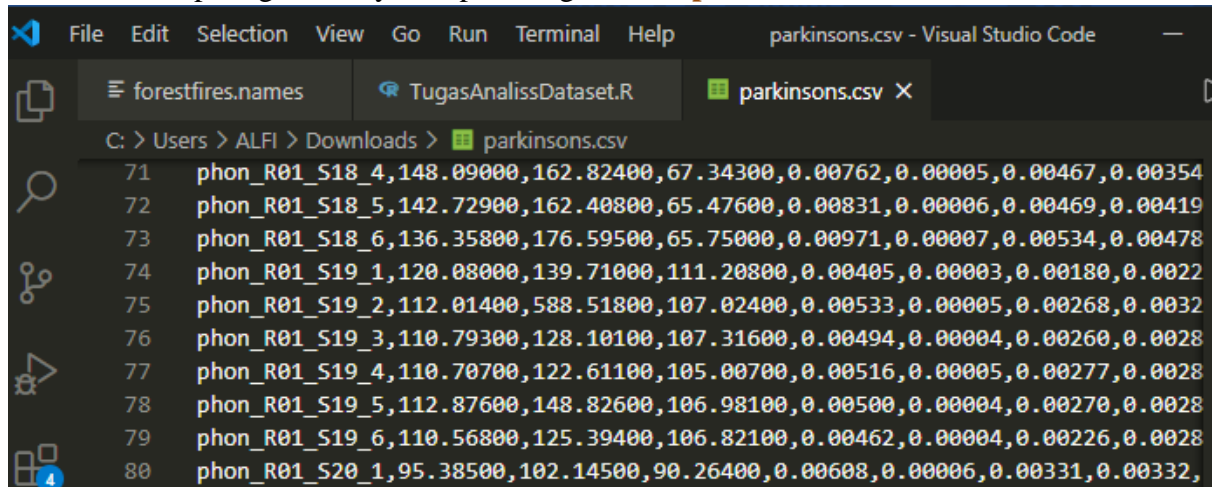
Pada Dataset ini mempunyai tujuan (aim) untuk mengidentifikasi apakah dari 197 orang terindikasi penyakit Parkinson (PD) atau sehat berdasarkan **kolom “status” dengan 0 untuk “healthy” dan 1 untuk “PD”**.

#### Attribute Information:

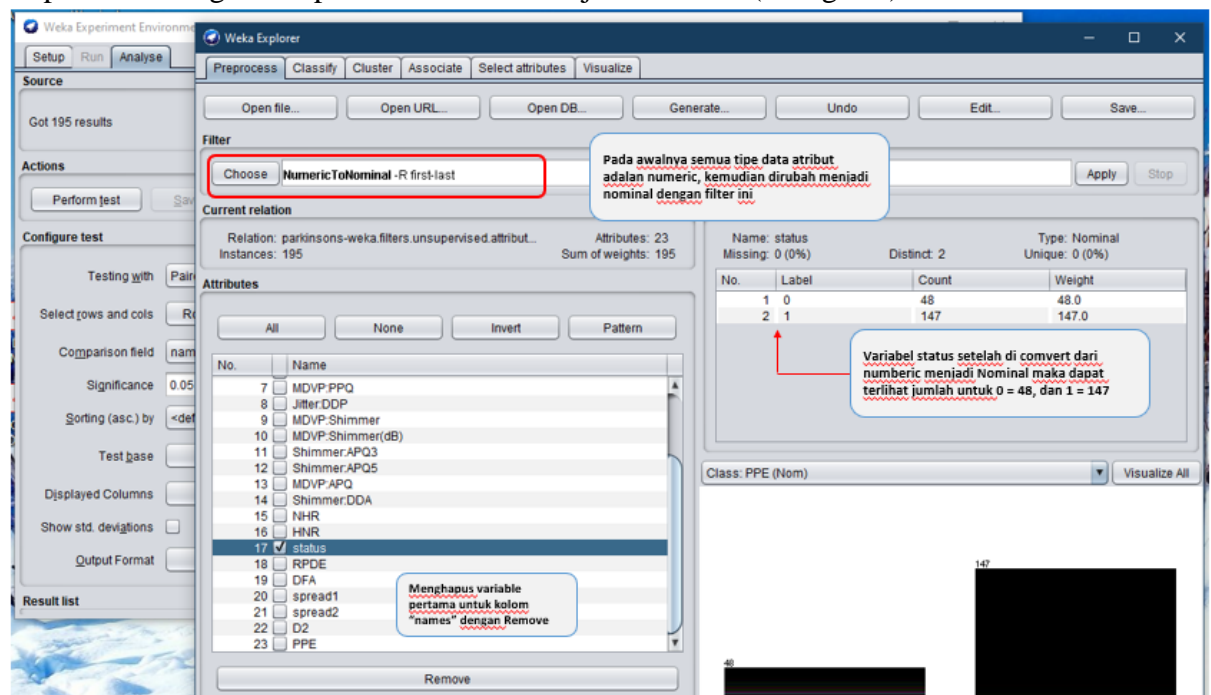
- Matrix column entries (attributes):
- name - ASCII subject name and recording number
- MDVP:F0(Hz) - Average vocal fundamental frequency
- MDVP:F1(Hz) - Maximum vocal fundamental frequency
- MDVP:F2(Hz) - Minimum vocal fundamental frequency
- MDVP:Jitter(%),MDVP:Jitter(Abs),MDVP:RAP,MDVP:PPQ,Jitter:DDP - Several measures of variation in fundamental frequency
- MDVP:Shimmer,MDVP:Shimmer(dB),Shimmer:APQ3,Shimmer:APQ5,MDVP:APQ,Shimmer:DDA - Several measures of variation in amplitude
- NHR,HNR - Two measures of ratio of noise to tonal components in the voice
- status - Health status of the subject (one) - Parkinson's, (zero) - healthy
- RPDE,D2 - Two nonlinear dynamical complexity measures
- DFA - Signal fractal scaling exponent
- spread1,spread2,PPE - Three nonlinear measures of fundamental frequency variation

## TOOLS : WEKA

1. Download dataset parkinsons → parkinsons.data
2. Ubah format parkinsons.data menjadi format file CSV kemudian menambahkan nama atribut dibaris paling atas, saya simpan dengan nama **parkinsons.csv**

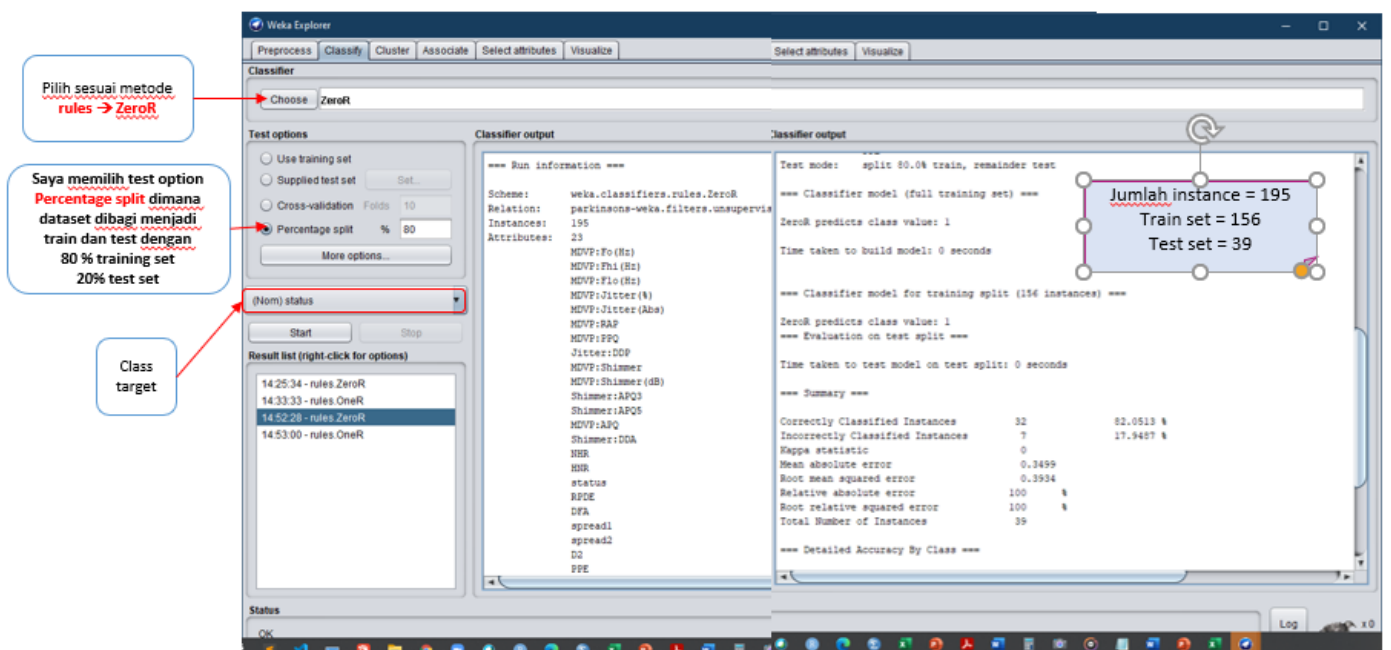
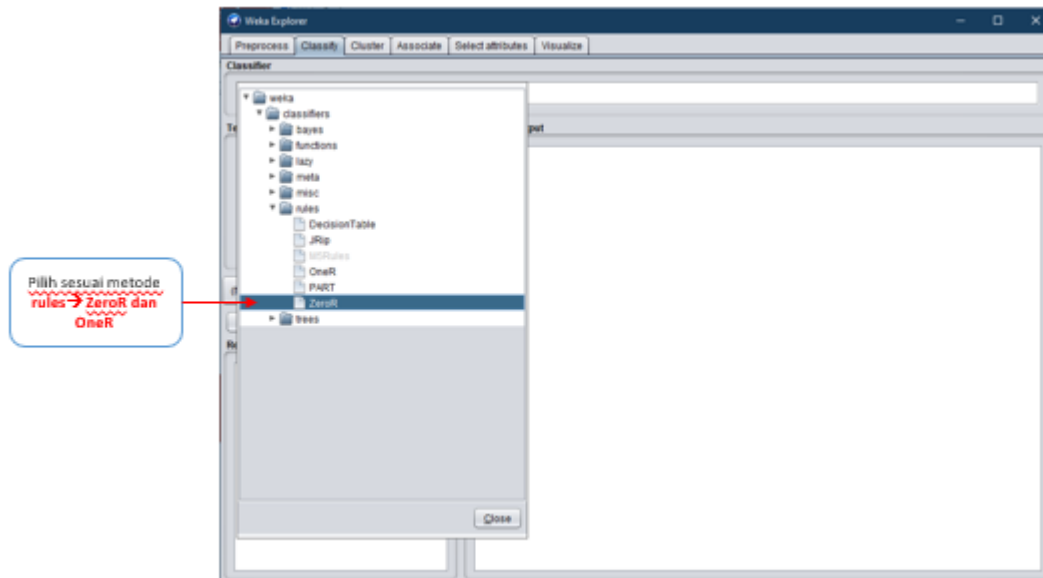


3. Mengimport data CSV  
Buka aplikasi Weka lalu Experimenter → Analyse → open file datazoo.csv → Open Explorer → Mengubah tipe data numerik menjadi Nominal (Kategorik)



4. Lalu untuk memulai klasifikasi klik Explorer pada halaman awal Weka

- Masukkan dataset yang sudah berformat arff. Klik tab Classify → pilih metode yang diinginkan.



```

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

ZeroR predicts class value: 1

Time taken to build model: 0 seconds

=== Classifier model for training split (156 instances) ===

ZeroR predicts class value: 1
=== Evaluation on test split ===

Time taken to test model on test split: 0 seconds

```

Dapat dilihat bahwa ZeroR predict 1 sebab nilai mayoritas adalah 1 dengan instance 147, dan untuk 0 = 48

```

=== Summary ===

Correctly Classified Instances      32      82.0513 %
Incorrectly Classified Instances    7      17.9487 %
Kappa statistic                     0
Mean absolute error                 0.3499
Root mean squared error             0.3934
Relative absolute error             100 %
Root relative squared error         100 %
Total Number of Instances          39

=== Detailed Accuracy By Class ===

```

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
	0.000	0.000	?	0.000	?	?	0.500	0.179	0
	1.000	1.000	0.821	1.000	0.901	?	0.500	0.821	1
Weighted Avg.	0.821	0.821	?	0.821	?	?	0.500	0.705	

```

=== Confusion Matrix ===

 a  b  <-- classified as
0  7  |  a = 0
0 32  |  b = 1

```

Didapatkan output seperti gambar diatas

Yaitu didapatkan nilai Confusion matrix

	a	b	dengan a =0=healthy , b= 1=PD
a	0	7	a
b	0	32	b

akurasi □ didapatkan pada output correctly classified instance sebesar 82.0513%

interpretasi : maka model ZeroR dapat mengklasifikasikan dengan akurat sebesar 82.0513%

sensitivity == recall

interpretasi : banyaknya prediksi positif yang benar per total positif sebenarnya (aktual).

Specificity adalah nilai incorrect pada gambar diatas

Precision → banyaknya prediksi positif yang benar per total prediksi positif

interpretasi :

F-1 score ada ditabel

F1 adalah rata-rata harmonic dari precision dan recall

rata-rata harmonik =  $n/(1/x_1 + 1/x_2 + 1/x_n)$

$f1 = 2/(1/precision + 1/recall)$

makin tinggi f1 maka precision dan recallnya juga tinggi.

Interpretasi : jika nilai f1 lebih dari 1 maka model berhasil memprediksi setiap kelasnya dengan baik, tetapi kita lihat dari hasil diatas bahwa kelas a dibawah 1 maka model tidak baik untuk klasifikasi dataset ini

☐ Use training set
 ☐ Supplied test set 
☐ Cross-validation Folds 10
 ☒ Percentage split % 80

(Nom) status

Result list (right-click for options)

14:25:34 - rules.ZeroR

14:33:33 - rules.OneR

14:52:28 - rules.ZeroR

14:53:00 - rules.OneR

Classifier output

```

Jitter:DDP
MDVP:Shimmer
MDVP:Shimmer (dB)
Shimmer:APQ3
Shimmer:APQ5
MDVP:APQ
Shimmer:DDA
NHR
HNR
status
RPDE
DFA
spread1
spread2
D2
PPE

Test mode: split 80.0% train, remainder test

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

MDVP:F0(Hz):
88.333 -> 1
91.904 -> 1
95.056 -> 1
95.385 -> 1
95.605 -> 1
-- -- :

```

Ditunjukkan bahwa predictor yang paling baik adalah MDVP:F0(Hz)

=== Evaluation on test split ===

Time taken to test model on test split: 0 seconds

=== Summary ===

Correctly Classified Instances	7	17.9487 %
Incorrectly Classified Instances	32	82.0513 %
Kappa statistic	0	
Mean absolute error	0.8205	
Root mean squared error	0.9058	
Relative absolute error	234.5083 %	
Root relative squared error	230.283 %	
Total Number of Instances	39	

=== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area
	1.000	1.000	0.179	1.000	0.304	?	0.500	0.179
	0.000	0.000	?	0.000	?	?	0.500	0.821
Weighted Avg.	0.179	0.179	?	0.179	?	?	0.500	0.705

=== Confusion Matrix ===

```

a b  <-- classified as
7  0 | a = 0
32 0 | b = 1

```

Yaitu didapatkan nilai Confusion matrix

	a	b	dengan a =0=healthy , b= 1=PD
a	7	0	a
b	32	0	b

akurasi □ didapatkan pada output correctly classified instance sebesar 17.98%

interpretasi : maka model ZeroR dapat mengklasifikasikan dengan akurat sebesar 17.98%