



YAYASAN NURUL JADID PAITON  
**LEMBAGA PENERBITAN, PENELITIAN, &  
PENGABDIAN KEPADA MASYARAKAT**  
**UNIVERSITAS NURUL JADID**  
PROBOLINGGO JAWA TIMUR

PP. Nurul Jadid  
Karanganyar Paiton  
Probolinggo 67291  
☎ 0888-3077-077  
e: [lp3m@unuja.ac.id](mailto:lp3m@unuja.ac.id)  
w: <https://lp3m.unuja.ac.id>

**SURAT KETERANGAN**

Nomor : NJ-To6/0324/A.03/LP3M/02.2021

Lembaga Penerbitan, Penelitian, dan Pengabdian Masyarakat (LP3M) Universitas Nurul Jadid Probolinggo menerangkan bahwa artikel/karya tulis dengan identitas berikut ini:

Judul : *Perbandingan Kinerja Tool Data Mining Weka dan Rapidminer  
Dalam Algoritma Klasifikasi*  
Penulis : MOCHAMMAD FAID, S.Kom., M.Kom.  
Identitas : TEKNIKA, Vol. VIII, No.1, Juli 2019, E-ISSN : 2549-8045, Institut  
Informatika Indonesia  
No. Pemeriksaan : 012041

Telah selesai dilakukan *similarity check* dengan menggunakan perangkat lunak **Turnitin** pada 01 Januari 2021 dengan hasil sebagai berikut: Tingkat kesamaan diseluruh artikel (*Similarity Index*) adalah 23 % dengan publikasi yang telah diterbitkan oleh penulis pada TEKNIKA

(<http://repository.ikado.ac.id/72/1/FINAL%20-%20Paper%202%20-%20ID%2095%20-%20Rev%203.pdf>)

Demikian Surat Keterangan ini dibuat untuk digunakan sebagaimana mestinya.

Probolinggo, 02 Februari 2021

Kepala LP3M,



**ACHMAD FAWAID, M.A., M.A.**

NIDN. 2123098702

# Perbandingan Kinerja Tool Data Mining Weka dan Rapidminer

*by* mohammad faid

---

**Submission date:** 01-Jan-2021 04:45AM (UTC+0100)

**Submission ID:** 1482329541

**File name:** Perbandingan\_Kinerja\_Tool\_Data\_Mining\_Weka\_dan\_Rapidminer.pdf (1.27M)

**Word count:** 681

**Character count:** 4419

# Perbandingan Kinerja Tool Data Mining Weka dan Rapidminer Dalam Algoritma Klasifikasi

<sup>2</sup> Mochammad Faid  
Program Studi Teknik Informatika  
Universitas Nurul Jadid  
mfaid@unuja.ac.id

Moh. Jasri  
Program Studi Teknik Informatika  
Universitas Nurul Jadid  
jasri@unuja.ac.id

Titasari Rahmawati  
Program Studi Sistem Informasi  
Institut Informatika Indonesia  
tita@ikado.ac.id

**Abstrak-** Ilmu Data mining adalah ilmu mempelajari dan menjelaskan cara mengekstrak sebuah pengetahuan yang tidak terlihat dari sebuah data yang banyak atau besar. Klasifikasi adalah bagian rumpun ilmu data mining. Algoritma klasifikasi pada data mining mempunyai model - model. Oleh Karena itu pada tiap model memiliki algoritma yang tidak sama, hal ini akan membuat tingkat akurasi pada tiap algoritma tidak akan sama. Untuk mengetahui dan melihat hasil kinerja algoritma dari data mining, salah satu indikatornya yaitu mengetahui tingkat presisi atau akurasi dari sebuah algoritma. Algoritma mempunyai perhitungan sulit, rumit serta membutuhkan banyak waktu, sehingga dibutuhkan tools atau software data mining yang bisa memproses, mengelolah dan mempresentasikan sebuah data menjadi lebih efisien. Tools yang digunakan adalah tools weka dan tools rapidminer, tujuan adanya penelitian perbandingan tools ini ialah membandingkan kinerja tools weka dan tools rapidminer dengan menggunakan algoritma dan data yang sama.

**Kata Kunci:** Software Rapidminer, Software Weka, Klasifikasi

## I. PENDAHULUAN

Dalam sebuah disiplin ilmu, Data merupakan kumpulan dari beberapa fakta, yang selanjutnya data akan diproses sehingga bisa dideskripsikan secara utuh, jelas dan valid sehingga mudah untuk dimengerti pihak luar yang sama sekali tidak mengalaminya. sedangkan klasifikasi itu sendiri adalah pemilihan data yang sesuai persamaan dan perbedaan yang ada pada data tersebut. Proses merubah data menjadi informasi sangat bermanfaat sebagai acuan dalam mengambil sebuah keputusan. Namun dengan berkembangannya ilmu pengetahuan terkait teknologi, data yang sudah terkumpul dari tahun ke tahun berikutnya, data yang sudah dianggap tidak berguna ternyata masih dapat diekstraksi sehingga akan menghasilkan ilmu dan pengetahuan yang akan berguna pihak pemilik data. Ilmu yang menjelaskan terkait dengan ekstraksi atau penambahan adalah data mining atau sering disebut dengan Knowledge discovery (mining) in databases. Data mining mengembangkan basis data agar bisa mencari beberapa pola atau rule yang masih terhidden, atau mungkin mencari

informasi untuk memprediksi kemungkinan yang terlupakan. Dengan adanya data mining, lahirah algoritma-algoritma data mining yang mempunyai kekurangan serta kelebihan

Pada penelitian komparasi tools ini mencoba membandingkan algoritma klasifikasi dengan memakai tool yang berkaitan dengan data mining, dalam penelitian ini tools yang digunakan yaitu weka serta rapidminer.

Dari hasil penelitian sebelumnya yang dilakukan Masriyah yaitu penelitian tentang "Komparasi Algoritma Klasifikasi Data Mining untuk Evaluasi Pemberian Kredit" penelitian memanfaatkan rapidminer sebagai tools data miningnya, penelitian ini berusaha untuk membandingkan dua model data mining klasifikasi diantaranya model c4.5 dan model naive bayes. Sesudah melakukan perbandingan bisa diambil kesimpulan ternyata model c4.5 mempunyai prediksi lebih baik dari model Naive Bayes[1]. Sedangkan penelitian berikutnya dilakukan vinita chandani dengan judul "Komparasi Algoritma Klasifikasi Machine Learning dan Feature Selection Pada Analisis Sentimen Review Film". ada beberapa algoritma yang dikomparasikan diantaranya : ANN, NB, Dan SVM. Dari hasil dapat diambil kesimpulan ternyata model SVM mempunyai kinerja yang lebih baik ketimbang model lainnya yang memiliki nilai akurasi sebesar 81,10% [2]. Pada penelitian selanjutnya dilakukan rizal A.S. yang memiliki judul penelitian "Komparasi Algoritma Klasifikasi Data Mining Untuk Memprediksi Penyakit Tuberculosis (TB) Studi Kasus Puskesmas Karawang Sukabumi". Penelitian ini memkomparasikan 4 algoritma diantaranya C4.5, Neural Network, Logistic Regression dan Naive Bayes. Hasil dari penelitian model Naive bayes mempunyai kinerja yang lebih baik dibandingkan dengan model lainnya sebesar 91.61%[3]. Pada beberapa penelitian yang sudah dilakukan untuk mengukur akurasi, dan untuk mengukurnya semua peneliti menggunakan tool data mining. Perbedaan penelitian ini berusaha untuk memkomparasikan tools data mining, dalam penelitian ini tool yang akan dikomparasikan yaitu weka dan rapidminer.

Tujuan adanya penelitian komporasi tools ini adalah untuk melihat, mengukur kinerja serta memkomparasikan tools data mining antara tools weka dan tools rapidminer, sehingga dapat dimanfaatkan demi kepentingan penelitian terkait data mining.

## II. STUDI PUSTAKA

### A. Data Mining

Data mining merupakan suatu istilah yang dipakai untuk menjelaskan penemuan pengetahuan pada sebuah database

Data mining adalah proses yang memakai teknik statistik, kecerdasan buatan, machine learning dan matematika untuk mengekstrak dan mengidentifikasi sebuah informasi dan pengetahuan yang berhubungan dengan database besar yang tentunya bermanfaat[4]. Sedangkan menurut peneliti yang lain yaitu Gatner Grup, data mining merupakan proses pencarian hubungan yang berarti atau memiliki pola dan kecenderungan dengan cara memeriksa kumpulan data besar dengan memakai teknik pengenalan sebuah pola seperti teknik Matematika dan Statistik[5].

### B. Klasifikasi

Klasifikasi merupakan teknik terapan dalam memprediksi sebuah class atau properti pada tiap instance data, adapun klasifikasi digunakan pada data setelah proses preprocessing. Proses preprocessing dilakukan sebuah pengecekan data duplikat dengan levenhteain string metric, penanganan missing value, dan pemilihan fitur [6]. Model-model algoritma klasifikasi diantaranya adalah sebagai berikut:

#### a. Algoritma C4.5

C4.5 diciptakan oleh Quinlan merupakan versi baru dari Algoritma ID3. Pada pembentukan pohon keputusan hanya bisa diterapkan pada data yang mempunyai tipe kategori artinya bersifat nominal ataupun ordinal. Untuk data yang mempunyai tipe numerik belum bisa diterapkan. perbaikan dari ID3 sudah bisa menangani fitur data yang memiliki tipe numerik, dan juga dapat melakukan pemangkasan pohon keputusan. Model c4.5 memakai kriteria pada mencari data yang menjadikan sebagai pemecah pada pohon keputusan yang telah diinduksi [7].

$$Gain(S, A) = Entropy(S) - \sum_{i=1}^n \frac{|S_i|}{|S|} * Entropy(S_i) \quad (1)$$

Keterangan :

- S : Himpunan Semesta
- A : Atribut Data
- N : Jumlah jenis atribut A
- |S<sub>i</sub>| : Jumlah semesta pada jenis ke i
- |S| : Jumlah jenis dalam Semesta

Sedangkan nilai *Entropy* bisa didapatkan pada persamaan 2-3:

$$entropy(A) = - \sum_{i=1}^n \frac{|S_i|}{|S|} \log_2 \frac{|S_i|}{|S|} \quad (2)$$

Keterangan :

- S : Himpunan Semesta
- A : Atribut Data
- N : Jumlah jenis atribut A
- |S<sub>i</sub>| : Jumlah semesta pada jenis ke i
- |S| : Jumlah jenis dalam semesta

#### b. Algoritma Naive Bayes

Algoritma Naive Bayes beracuan dalam asumsi sederhana dari nilai atribut dengan cara kondisional dan saling bebas jika diberi nilai output. Model ini adalah sebuah model yang berasal dari teorema Bayes. Adapun teorema Bayes dapat dilihat pada persamaan (3):

$$P(B|A) = \frac{P(A|B)P(B)}{P(B)} \quad (3)$$

Keterangan:

- P(B|A) : Probabilitas-*posterior*, probabilitas-muncul-B jika diketahui nilai A
- P(A|B) : Probabilitas-*posterior*, probabilitas-muncul-A jika diketahui nilai B
- P(A) : Probabilitas-*prior*, probabilitas-kejadian-A
- P(B) : Probabilitas-*prior*, probabilitas-kejadian-B

#### c. Algoritma K-NN

K-Nearest Neighbor adalah model algoritma yang dapat mengklasifikasikan objek yang beracuan pada data training dan mengidentifikasi dengan melihat jarak terdekat pada objek. Training diproyeksikan pada ruang yang memiliki banyak dimensi, dimana pada beberapa dimensi merepresentasikan pada fitur yang ada. Ruang yang terbentuk dipecah menjadi beberapa jenis yang beracuan pada klasifikasi trainingnya. Nilai k terbaik dari model algoritma bergantung dengan data trainingnya. pada umumnya, nilai k tinggi dapat mengurangi kekaburan pada sebuah model klasifikasi, dalam batasan pada tiap klasifikasi akan menjadi sulit diidentifikasi. ada begitu banyak model dalam menghitung jarak kemiripan sebuah data, diantaranya sebagai berikut euclidean, distance, dan manhattan-distance.

$$R = \sqrt{(a_1 - b_1)^2 + (a_2 - b_2)^2 + \dots + (a_n - b_n)^2} \quad (4)$$

Keterangan:

- R : Jarak-data
- a<sub>1</sub> : Atribut-pertama-data-*testing*
- b<sub>1</sub> : Atribut-pertama-data-*training*
- a<sub>2</sub> : Atribut-kedua-data-*testing*
- b<sub>2</sub> : Atribut-kedua-data-*training*

#### d. Algoritma

##### e. Random Forest

Random Forest merupakan kumpulan classifier yang membentuk banyak pohon keputusan dan melakukan model klasifikasi yang beracuan pada keluaran dari sebuah hasil model klasifikasi pada tree keputusan anggota, metode random forest merupakan pengembangan dari sebuah algoritma yang bernama algoritma CART, yaitu dengan menggunakan model bootstrap aggregating dan random featur selection[9]. Pada random forest memiliki banyak pohon keputusan yang dihasilkan sehingga tercipta hutan (forest).

Data mining adalah proses yang memakai teknik statistik, kecerdasan buatan, machine learning dan matematika untuk mengekstrak dan mengidentifikasi sebuah informasi dan pengetahuan yang berhubungan dengan database besar yang tentunya bermanfaat[4]. Sedangkan menurut peneliti yang lain yaitu Gatner Grup, data mining merupakan proses pencarian hubungan yang berarti atau memiliki pola dan kecenderungan dengan cara memeriksa kumpulan data besar dengan memakai teknik pengenalan sebuah pola seperti teknik Matematika dan Statistik[5].

**B. Klasifikasi**

Klasifikasi merupakan teknik terapan dalam memprediksi sebuah class atau properti pada tiap instance data, adapun klasifikasi digunakan pada data setelah proses preprocessing. Proses preprocessing dilakukan sebuah pengecekan data duplikat dengan levenhteain string metric, penanganan missing value, dan pemilihan fitur [6]. Model-model algoritma klasifikasi diantaranya adalah sebagai berikut:

**a. Algoritma C4.5**

C4.5 diciptakan oleh Quinlan merupakan versi baru dari Algoritma ID3. Pada pembentukan pohon keputusan hanya bisa diterapkan pada data yang mempunyai tipe kategori artinya bersifat nominal ataupun ordinal. Untuk data yang mempunyai tipe numerik belum bisa diterapkan. perbaikan dari ID3 sudah bisa menangani fitur data yang memiliki tipe numerik, dan juga dapat melakukan pemangkasan pohon keputusan. Model c4.5 memakai kriteria pada mencari data yang menjadikan sebagai pemecah pada pohon keputusan yang telah diinduksi [7].

$$Gain(S, A) = Entropy(S) - \sum_{i=1}^n \frac{|S_i|}{|S|} * Entropy(S_i) \quad (1)$$

Keterangan :

- S : Himpunan Semesta
- A : Atribut Data
- N : Jumlah jenis atribut A
- |Si| : Jumlah semesta pada jenis ke i
- |S| : Jumlah jenis dalam Semesta

Sedangkan nilai *Entropy* bisa didapatkan pada persamaan 2-3:

$$entropy(A) = - \sum_{i=1}^n \frac{|S_i|}{|S|} \log_2 \frac{|S_i|}{|S|} \quad (2)$$

Keterangan :

- S : Himpunan Semesta
- A : Atribut Data
- N : Jumlah jenis atribut A
- |Si| : Jumlah semesta pada jenis ke i
- |S| : Jumlah jenis dalam semesta

**b. Algoritma Naive Bayes**

Algoritma Naive Bayes beracuan dalam asumsi sederhana dari nilai atribut dengan cara kondisional dan saling bebas jika diberi nilai output. Model ini adalah sebuah model yang berasal dari teorema Bayes. Adapun teorema Bayes dapat dilihat pada persamaan (3):

$$P(B|A) = \frac{P(A|B)P(B)}{P(B)} \quad (3)$$

Keterangan:

- P(B|A) : Probabilitas-posterior,probabilitas-muncul-B jika diketahui nilai A
- P(A|B) : Probabilitas-posterior,probabilitas-muncul-A jika diketahui nilai B
- P(A) : Probabilitas-prior,probabilitas-kejadian-A
- P(B) : Probabilitas-prior,probabilitas-kejadian-B

**c. Algoritma K-NN**

K-Nearest Neighbor adalah model algoritma yang dapat mengklasifikasikan objek yang beracuan pada data training dan mengidentifikasi dengan melihat jarak terdekat pada objek. Training diproyeksikan pada ruang yang memiliki banyak dimensi, dimana pada beberapa dimensi merepresentasikan pada fitur yang ada. Ruang yang terbentuk dipecah menjadi beberapa jenis yang beracuan pada klasifikasi trainingnya. Nilai k terbaik dari model algoritma bergantung dengan data trainingnya. pada umumnya, nilai k tinggi dapat mengurangi kekaburan pada sebuah model klasifikasi, dalam batasan pada tiap klasifikasi akan menjadi sulit diidentifikasi. ada begitu banyak model dalam menghitung jarak kemiripan sebuah data, diantaranya sebagai berikut euclidean, distance, dan manhattan-distance.

$$R = \sqrt{(a_1 - b_1)^2 + (a_2 - b_2)^2 + \dots + (a_n - b_n)^2} \quad (4)$$

Keterangan:

- R : Jarak-data
- a<sub>1</sub> : Atribut-pertama-data-testing
- b<sub>1</sub> : Atribut-pertama-data-training
- a<sub>2</sub> : Atribut-kedua-data-testing
- b<sub>2</sub> : Atribut-kedua-data-training

**d. Algoritma Random Forest**

Random Forest merupakan kumpulan classifier yang membentuk banyak pohon keputusan dan melakukan model klasifikasi yang beracuan pada keluaran dari sebuah hasil model klasifikasi pada tree keputusan anggota. metode random forest merupakan pengembangan dari sebuah algoritma yang bernama algoritma CART, yaitu dengan menggunakan model bootstrap aggregating dan random featur selection[9]. Pada random forest memiliki banyak pohon keputusan yang dihasilkan sehingga tercipta hutan (forest).

C. Tool Data Mining

Tool merupakan software yang dipakai untuk mempermudah proses seorang akademis dan peneliti digunakan untuk mengolah, menambang data, selain itu bisa juga dijadikan library sehingga bisa ditanam dalam sebuah program, sehingga program yang dibuat bisa meniru apa yang bisa dikerjakan oleh tool.

1. Weka

Weka adalah rangkaian software pembelajaran yang mesin yang dibuat dalam pemrograman java, dikembangkan pada Universitas waikato, dan selandia baru. Software ini mempunyai banyak model algoritma dari machine learning untuk kepentingan data. weka mempunyai banyak tool yang digunakan untuk mengolah data, dari proses preprocessing, association rules, classification, and visualization.

2. Rapidminer

Rapidminer merupakan software untuk mengolah data. Pekerjaan yang bisa dikerjakan oleh rapidminer, contoh text mining berkaitan dengan teks analisis, mengekstrak beberapa pola dari dataset besar serta mengkombinasikan dengan model statistika, database, dan juga kecerdasan buatan.

3. Confusion Matrix

Confusion marupaka model yang bisa digunakan untuk membantu perhitungan tingkat akurasi pada data mining[11]. dengan melakukan split ratio 0.1dan data yang akan digunakan yaitu data playing tenis.

```

accuracy 69.2308 %
=== Confusion Matrix ===
 a b  <-- classified as
 0 4 | a = No
 0 9 | b = Yes
    
```

Gambar 1. Confusion Matrix Weka.

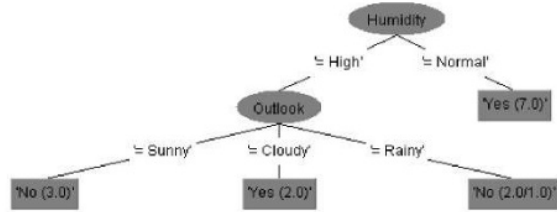
accuracy: 69.23%

	true No	true Yes
pred. No	0	0
pred. Yes	4	9
class recall	0.00%	100.00%

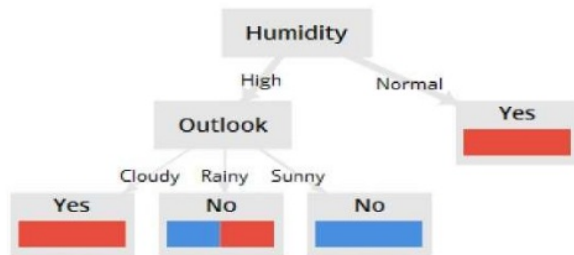
Gambar 2. Confusion Matrix Rapidminer.

4. Decision Tree

Decision tree adalah pohon keputusan dihasilkan pada hasil ekstraksi data dengan algoritma klasifikasi.



Gambar 3. Decision Tree Weka.



Gambar 4. Decision Tree Rapidminer.

III. METODOLOGI PENELITIAN

Beberapa Instrumen yang dipakai dalam penelitian, Instrumen yang dipakai dibagi jadi dua bagian diantaranya:

1. Instrumen Pengumpulan terkait data dan
2. Instrumen penelitian.

1. Instrumen Pengumpulan Data

Dalam proses ini melakukan pengumpulan data untuk digunakan sebagai bahan menganalisis kinerja model algoritma klasifikasi, ada tool yang akan digunakan dalam data mining ini ada dua yaitu weka dan rapidminerl. berikut ini adalah dataset dalam penelitian ini:

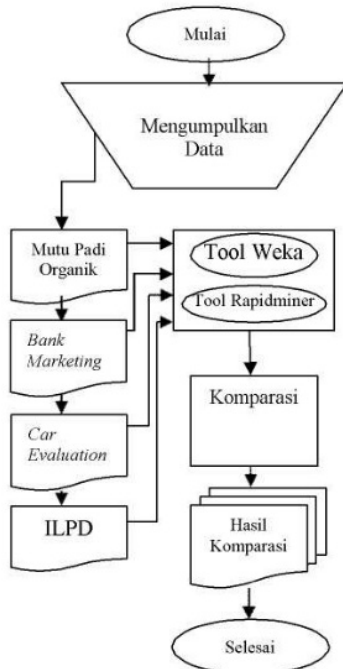
- Bank Marketing
- Mutu Padi Organik
- Indian Liver Patient Dataset (ILPD)
- Car Evaluation

2. Instrumen Penelitian

Dalam mendesain sebuah model sistem juga membutuhkan instrumen yang mendukung dalam menganalisis serta mengkomparasi model algoritma klasifikasi:

- Tool Data Mining (Weka, Rapidminer)
- Excel
- Timer

Adapun tahapan penelitian bisa dilihat seperti *Flowchart* dibawah ini.a

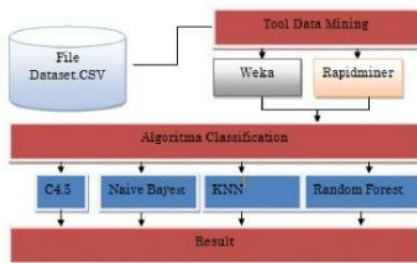


Gambar 5. *Flowchart* Tahapan Penelitian.

#### IV. HASIL DAN PEMBAHASAN

##### A. Prosedur Penelitian

Pada proses ini melakukan proses pengumpulan 2 tools terkait data mining yaitu weka dan rapidminer serta untuk menentukan kumpulan data yang diterapkan, dan memilih satu model algoritma klasifikasi dalam hal menguji kinerja tool, dan menulis hasil tingkat akurasinya pada tabel dan grafik



Gambar 6. Skema Penelitian

Hasil percobaan untuk melakukan penelitian ini menggunakan sample data Indian Liver Patient Dataset (ILPD), Bank Marketing, Car Evaluation, dan padi organik. Dalam menentukan atribut yang dipakai dalam dataset bisa dilihat pada tabel1. sedangkan hasil tingkat akurasi dan kecepatan prosesnya bisa dilihat pada tabel2. sedangkan untuk melihat kinerja terbaik dari tool data mining maka diperlukan rata-rata selisih dari masing-masing model algoritma. untuk melihat detail performa bisa dilihat pada tabel3.

Data yang digunakan untuk melakukan testing adalah data mutu padi organik, car evaluation, bank marketing, dan Indian Liver Patient Dataset (ILPD). Alasan beberapa database itu dipilih karena datanya sesuai model algoritma klasifikasi yang dipakai antara lain C4.5/j48, K-NN, Naive Bayes, Dan random Forest, adapun Gamba7 sampai gambar10 adalah representasi dari tabel2.

Tabel 1. Atribut Yang Ada Pada *Dataset*.

<i>Dataset</i>	<i>Atribut</i>	<i>Sumber</i>
<i>Mutu Padi Organik</i>	<i>Varietas</i> <i>Panjang</i> <i>Bentuk</i> <i>Warna</i> <i>Rasa</i> <i>Teknik</i> <i>Musim</i> <i>Hama</i> <i>Ph</i>	<i>Dinas Pertanian Bondowoso</i>
<i>Bank Marketing</i>	<i>Age</i> <i>Job</i> <i>Marital</i> <i>Education</i> <i>Default</i> <i>Balance</i> <i>Housing</i> <i>Loan</i> <i>Contact</i> <i>Day</i> <i>Month</i> <i>Duration</i> <i>Campaign</i> <i>Pdays</i> <i>Previous</i> <i>Poutcome</i>	<i>archive.ics.uci.edu</i>
<i>Car Evaluation</i>	<i>Buying</i> <i>Maint</i> <i>Doors</i> <i>Persons</i> <i>Ug_boot</i> <i>Safety</i> <i>Varian</i>	<i>archive.ics.uci.edu</i>

Indian Liver Patient Dataset (ILPD)	Age Gender Tb_total_bilirubin Db_direct_bilirubin Alkphos_alkaline_p_hosphotase Sgpt_alamine_aminotransferase Sgot_aspartate_aminotransferase Tp_total_protiens Alb_albumin A/g_ratio_albumin_and_globulin_ratio	archive.ics.uci.edu
-------------------------------------	---	---------------------

Tabel 2. Detail Performa Tool Data Mining.

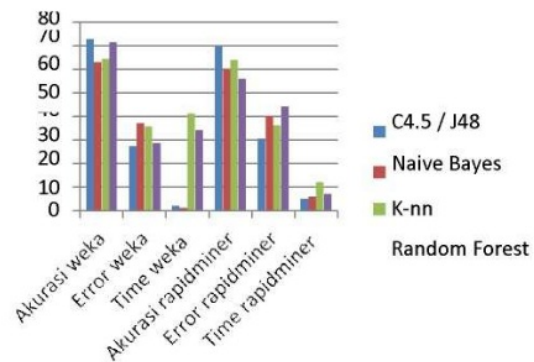
Dataset	Algoritma Classification	Weka			Rapidminer		
		Akurasi	Error	Time	Akurasi	Error	Time
Mutu Padi Organik	C4.5 / J48	72,79	27,21	2	69,47	30,53	5
	Naive Bayes	62,90	37,10	1	60,06	39,94	6
	K-nn	64,29	35,71	41	63,94	36,06	12
	Random Forest	71,27	28,73	34	55,95	44,05	7
Bank Marketing	C4.5 / J48	87,81	12,19	3	88,28	11,72	3
	Naive Bayes	88,82	11,18	1	87,44	12,56	4
	K-nn	87,05	12,95	60	85,82	14,18	8
	Random Forest	89,01	10,99	19	88,47	11,53	15
Car Evaluation	C4.5 / J48	80,51	19,49	1	79,36	20,64	5
	Naive Bayes	79,04	20,96	1	80,26	19,74	3
	K-nn	77,75	22,25	4	76,78	23,22	3
Patient Dataset (ILPD) Liver	Random Forest	83,15	16,85	8	70,68	29,32	4
	Naive Bayes	59,24	40,76	1	56,68	43,32	4
	K-nn / J48	84,98	32,28	3	67,94	32,88	5
	Random Forest	70,86	29,14	3	69,66	30,34	3
Rata Rata Akurasi dan Waktu		75,41	11,44		72,80688		5,8125

Tabel3. Rata Akurasi dan Waktu penggunaan Tool Data Mining.

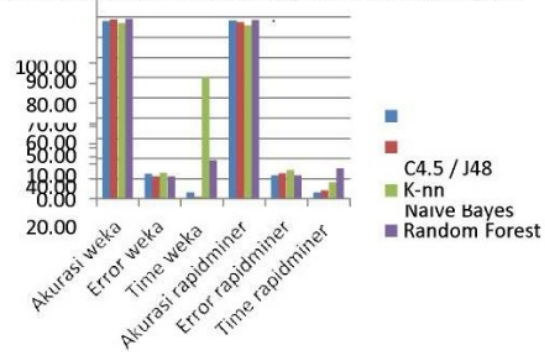
Weka		Rapidminer	
Akurasi	Time	Akurasi	Time
75,41	11,44	72,80688	5,8125

Perbandingan antar tool data mining weka dan rapidminer dari tabel2 dapat dilihat hasilnya dalam bentuk rata-rata. Rata-rata kecepatan yang dimiliki oleh rapidminer 5,8125 detik dan weka 11,44 detik. selisih rata-rata antara Rapidminer dan weka adalah 5,625 sedangkan dari tingkat akurasi weka lebih unggul dari rapidminer dengan jarak rata-rata sebanyak 2,60%.

Tabel 2 kemudian direpresentasikan dalam bentuk grafik untuk memudahkan membaca kinerja tool data mining serta algoritmanya. Hal ini nampak pada Gambar 7-10 sebagai berikut

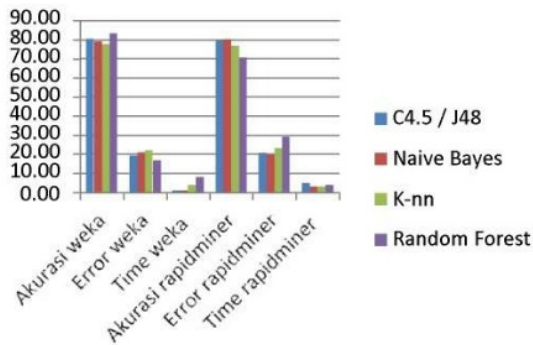


Gambar 7. Grafik Performa Dengan Dataset Padi Organik

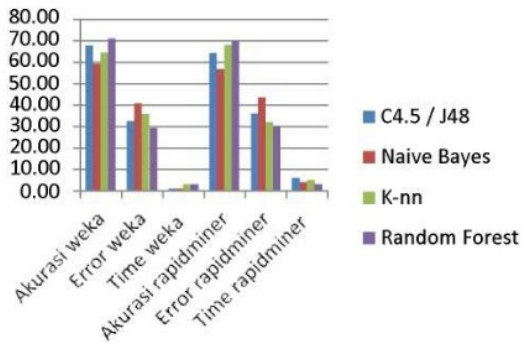


Gambar 8. Grafik Performa Dengan Dataset Bank Marketing.





Gambar 9. Grafik Performa Dengan Dataset Car Evaluation.



### V. KESIMPULAN

Mengenai kesimpulan pada hasil penelitian untuk kecepatan pemrosesan data pada tool data mining rapidminer memiliki kecepatan yang lebih unggul dari Weka, Sedangkan disisi tingkat akurasi weka lebih unggul dari rapidminer, sedangkan data yang memiliki waktu pemrosesan tercepat dari yang lain ialah data Indian Liver Patient Dataset (ILPD).

### REFERENSI

[1] Masripah, S. (2016). Komparasi Algoritma Klasifikasi Data Mining untuk Evaluasi Pemberian Kredit. *Bina Insani ICT Journal*, Vol. 3, No. 1, pp. 187-193.  
 [2] Chandani, V., Wahono, R.S. & Purwanto. (2015). Komparasi Algoritma Klasifikasi Machine Learning dan Feature Selection pada Analisis Sentimen Review Film. *Journal of Intelligent Systems*, Vol. 1, No. 1, pp. 56-60.

[3] Amegia, R. (2014). Komparasi Algoritma Klasifikasi Data Mining untuk memprediksi Penyakit Tuberculosis (TB) Studi Kasus Puskesmas Karawang Sukabumi. *Proceedings SNIT 2014*.  
 [4] Turban, E., dkk., (2005). *Decision Support Systems and Intelligent Systems*. Yogyakarta: Andi Offset.  
 [5] Furnkranz, J. (1994). A Comparison of Pruning Methods for Relational Concept Learning. Austria: AAAI.  
 [6] Han, J. & Kamber, M. (2006). *Data Mining Concept and Tehniques*. San Fransisco: Morgan Kauffman.  
 [7] Nithya, A. & Sundaram, V. (2011). Classification Rules for Indian Race Diseases. *IJCSI*.  
 [8] Bramer, M. (2007). *Principles of Data Mining*. London: Springer.  
 [9] Breiman, L., (2001). *Random Forests*. California: University of California Berkeley.  
 [10] Larose, D.T. (2005). *Discovering Knowledge in Data*. Canada. Wiley-Interscience.  
 [11] Berry & Linoff (2004). *Data Mining Techniques for Marketing, Sales and CRM*. Wiley.

# Perbandingan Kinerja Tool Data Mining Weka dan Rapidminer

---

## ORIGINALITY REPORT

---

**23%**

SIMILARITY INDEX

**23%**

INTERNET SOURCES

**19%**

PUBLICATIONS

**5%**

STUDENT PAPERS

---

## PRIMARY SOURCES

---

**1**

**repository.ikado.ac.id**

Internet Source

**19%**

---

**2**

**programstudi11.blogspot.com**

Internet Source

**2%**

---

**3**

**journal.ipb.ac.id**

Internet Source

**1%**

---

**4**

**www.scribd.com**

Internet Source

**1%**

---

Exclude quotes    On

Exclude matches    < 1%

Exclude bibliography    On