Seminar Nasional Hasil Penelitian dan Pengabdian Masyarakat 2021
Institut Informatika dan Bisnis Darmajaya, 19 Agustus 2021
ISSN: 2598-0256, E-ISSN: 2598-0238

Perbandingan Data Set IRIS Dengan Aplikasi Rapid Miner dan Orange menggunakan Algoritma Klasifikasi

Yulia Verawati[1*], Muhammad Said Hasibuan[2]

[1,2]Magister Teknik Informatika,Institut Informatika dan Bisnis Darmajaya Lampung
Email: [1]*yulia.verawati.2021211035@mail.darmajaya.ac.id*, [2]*msaid@darmajaya.ac.id*

Abstract

Data mining is a data visualization that provides data information resulting from the processed data. The data mining method is determined by the data set to be used. Iris data set has numeric attribute and nominal class so that the right algorithm is classification with a general algorithm that can be used, namely decision tree. In this study, we want to compare the performance of the decision tree algorithm using rapid miner and orange tools. Where the results obtained are Accuracy: 86.67%, classification_error: 13.33%, kappa: 0.800, weighted_mean_recall: 86.67%, weights: 1, 1, 1 and weighted_mean_precision: 90.48%, weights: 1, 1, 1 for rapidminer applications while for the application orange petal length iris sentosa 33.3%, petal length iris version color 90.7%, petal width iris versicolor 90.7% and petal width iris virginica 66.7%.

Keywords*: Classification, Decision tree ,Rapid Miner, Orange.*

Abstrak

Data mining merupakan visualisasi data yang memberikan informasi data hasil dari data yang diolah. Metode data mining ditentukan oleh data set yang akan digunakan. Data set iris memiliki attribute numeric dan class nominal sehingga algoritma yang tepat yaitu klasifikasi dengan algoritma yang umum bisa digunakan yaitu decision tree. Dalam penelitian ini ingin dibandingkan performa dari algoritma decision tree dengan menggunakan tools rapid miner dan orange. Dimana hasil yang didapatkan adalah Menghasilkan nilai accuracy: 86.67%, classification_error: 13.33%, kappa: 0.800, weighted_mean_recall: 86.67%, weights: 1, 1, 1 dan weighted_mean_precision: 90.48%, weights: 1, 1, 1 untuk aplikasi rapidminer sedangkan untuk pada aplikasi orange petal length iris sentosa 33,3%, petal length iris versi color 90,7%, petal width iris versicolor 90,7% dan petal width iris virginica 66,7%.

Keywords: Klasifikasi, Decision tree , Rapid Miner, Orange

1. PENDAHULUAN

Klasifikasi merupakan salah satu peran utama dari data mining. Klasifikasi termasuk kedalam supervised learning karena dalam proses klasifikasi terdapat proses pembelajaran dengan data lampau. Data iris merupakan salah satu dataset yang banyak digunakan dalam proses klasifikasi. Pada data iris ada 4 atribut yang dapat mempengaruhi klasifikasi yaitu: sepal length, sepal width, petal length, serta petal width. Atribut tujuan atau label dari data iris memiliki 3 kelas yaitu: iris setosa, iris versicolour, serta iris virginica.

Munculnya beragam aplikasi data mining juga menentukan hasil performa dari metode yang digunakan saat menggunakan data set yang telah ditentukan sebagai objek eksperimen. Bentuk visualisasi data set yang ditampilkan pada aplikasi akan membantu untuk menyajikan performa data mining. Tujuan dari penelitian ini adalah untuk melihat performa dari data iris dengan menggunakan aplikasi rapid miner dan orange.

2. KERANGKA TEORI

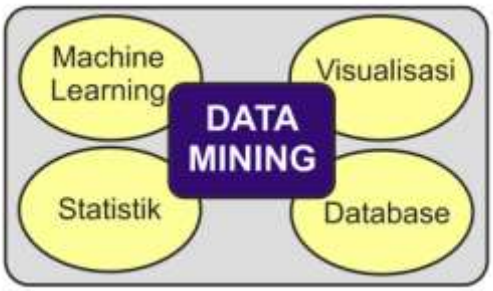
2.1. Data Mining

Data Mining atau sering juga disebut *Knowledge Discovery* in Database (KDD) adalah sebuah bidang ilmu yang banyak membahas tentang pola sebuah data. Serangkaian proses guna mendapatkan pengetahuan atau pola dari kumpulan data disebut dengan data mining (Witten, 2011). Sebuah data yang besar bisa saja

ISSN: 2598-0256, E-ISSN: 2598-0238

tidak berguna dan hanya akan menjadi sampah bila kita tidak dapat memanfaatkannya. Data mining menjawab masalah ini dengan menganalisa data yang besar tersebut kemudian membuat sebuah aturan, pola, ataupun model tertentu untuk mengenali data baru yang tidak berada dalam baris data yang tersimpan (Prasetyo, 2012).

Data *mining* merupakan kegiatan yang meliputi pengumpulan dan pemakaian data historis untuk menemukan keteraturan, pola atau hubungan dalam set data (Santosa, 2007). Output dari data *mining* dapat dipakai untuk memperbaiki pengambilan sebuah keputusan di masa depan. Data mining memiliki kaitan dengan berbagai bidang ilmu yang lain seperti *Machine Learning*, Statistik, Visualisasi serta *database*. Gambar 1. merupakan posisi data *mining* dengan berbagai disiplin ilmu lain.



Gambar 1. Posisi Data Mining dengan Berbagai Disiplin Ilmu (Santosa , 2007)

Beberapa peran utama data mining adalah: *Estimation*, *Prediction*, *Classification*, *Clustering* dan *Association*. Dari semua peranan data mining tersebut terbagi menjadi 2 berdasarkan metode pembelajarannya (Santosa, 2007) yaitu *Supervised Learning*, *Unsupervised* Learning. Perbedaan dari kedua metode pembelajaran pada algoritma data mining tersebut adalah jika dalam *supervised* learning harus memiliki data sampel atau sering disebut juga dengan data *training*. Sedangkan dalam *unsupervised learning* tidak membutuhkan data *training*. Salah satu contoh peran data *mining* dengan metode *supervised learning* adalah klasifikasi.

2.2 Klasifikasi

Klasifikasi merupakan salah satu peran utama dari data *mining*. Klasifikasi termasuk kedalam *supervised learning* karena dalam proses klasifikasi terdapat proses pembelajaran dengan data lampau. Proses ini digunakan algoritma untuk mengenali pola dari data yang nantinya dapat diterapkan kepada data baru yang belum diketahui kelompoknya. Teknik klasifikasi banyak diterapkan dalam dunia nyata seperti halnya dalam dunia medis (Christobel dan Sivaprakasam, 2011), Pendidikan (Ragab, 2014) (hastuti, 2012), teknik bangunan (Ashari, 2011), jaringan komputer (Widiastuti , 2007), serta banyak digunakan dalam bidang lain.

Dalam melakukan suatu klasifikasi dibutuhkan data lampau yang nantinya akan diolah menjadi sebuah aturan ataupun sebuah pengetahuan baru. Masalah klasifikasi pada dasarnya adalah sebagai berikut (Susanto dan Suryadi, 2010) :
1. Masalah Klasifikasi berangkat dari data training yang tersedia.
2. Data training akan diolah dengan menggunakan algoritma klasifikasi.
3. Masalah klasifikasi berakhir dengan dihasilkannya sebuah pengetahuan yang direpresentasikan dalam bentuk diagram, aturan atau pengetahuan.

2.3 Decision Tree

Decision tree sendiri merupakan metode klasifikasi dan prediksi yang sangat kuat dan terkenal (Kusrini dan Taufiq, 2009). Dalam decision tree ini data yang berupa fakta dirubah menjadi sebuah pohon keputusan yang berisi aturan dan tentunya dapat lebih mudah dipahami dengan bahasa alami. Model pohon keputusan banyak digunakan pada kasus data dengan output yang bernilai diskrit (Santosa, 2007). Walaupun tidak menutup kemungkinan dapat juga digunakan untuk kasus data dengan atribut numeric.

ISSN: 2598-0256, E-ISSN: 2598-0238

3. METODOLOGI

Penelitian ini dilakukan dengan menggunakan metode eksperimen dan dengan alat bantu yaitu rapid miner dan orange. Komparasi akan dilakukan untuk mengetahui alat bantu mana yang memiliki visual terbaik untuk menyajikan data mining dari penerapan algoritma *decision tree* dengan menggunakan dataset iris.

3.1 Dataset

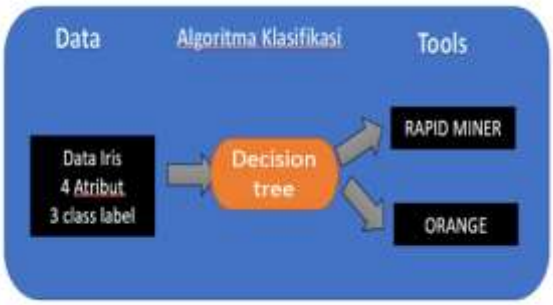
Dataset yang digunakan dalam penelitian ini adalah data iris. Data iris merupakan salah satu dataset yang banyak digunakan dalam proses klasifikasi. Data iris merupakan data publik yang dikeluarkan oleh uci repository. dalam data iris ada 4 atribut yang dapat mempengaruhi klasifikasi yaitu: sepal length, sepal width, petal length, serta petal width. Atribut tujuan atau label dari data iris memiliki 3 kelas yaitu: iris setosa, iris versicolour, serta iris virginica. Gambar 3.1 menunjukkan potongan dataset data iris.

	Name	Type	Role	Values
1	sepal length	N numeric	feature	
2	sepal width	N numeric	feature	
3	petal length	N numeric	feature	
4	petal width	N numeric	feature	
5	iris	C categorical	target	Iris-setosa, Iris-versicolor, Iris-virginica

Gambar 2. Data iris

3.2 Kerangka pemikiran

Sebagai acuan dalam melakukan penelitian maka kegiatan penelitian ini dirinci dalam sebuah kerangka pemikiran seperti pada Gambar 3.2. dimana pada gamabr tersebut menjelaskan bahwa algoritma klasifikasi yang digunakan yaitu decision tree dan dataset yang digunakan adalah data iris yang nantinya akan dilihat performa klasifikasinya menggunakan tools rapid miner dan orange.



Gambar 3. Kerangka pemikiran

4. HASIL DAN PEMBAHASAN

Hasil dari penelitian ini menunjukkan bahwa algoritma *decision tree* merupakan algoritma yang tepat untuk mengklasfikasikan sesuai dengan atribut dan label yang terdapat pada data iris. Sedangkan hasil performa dari klasifikasi menggunakan aplikasi rapid miner adalah nilai accuracy: 86.67%, classification_error: 13.33%, kappa: 0.800, weighted_mean_recall: 86.67%, weights: 1, 1, 1 dan weighted_mean_precision: 90.48%, weights: 1, 1, 1. Seperti tampilan pada gambar 4.1 sedangkan pada aplikasi orange performa klasfikasi yang didapat yaitu :

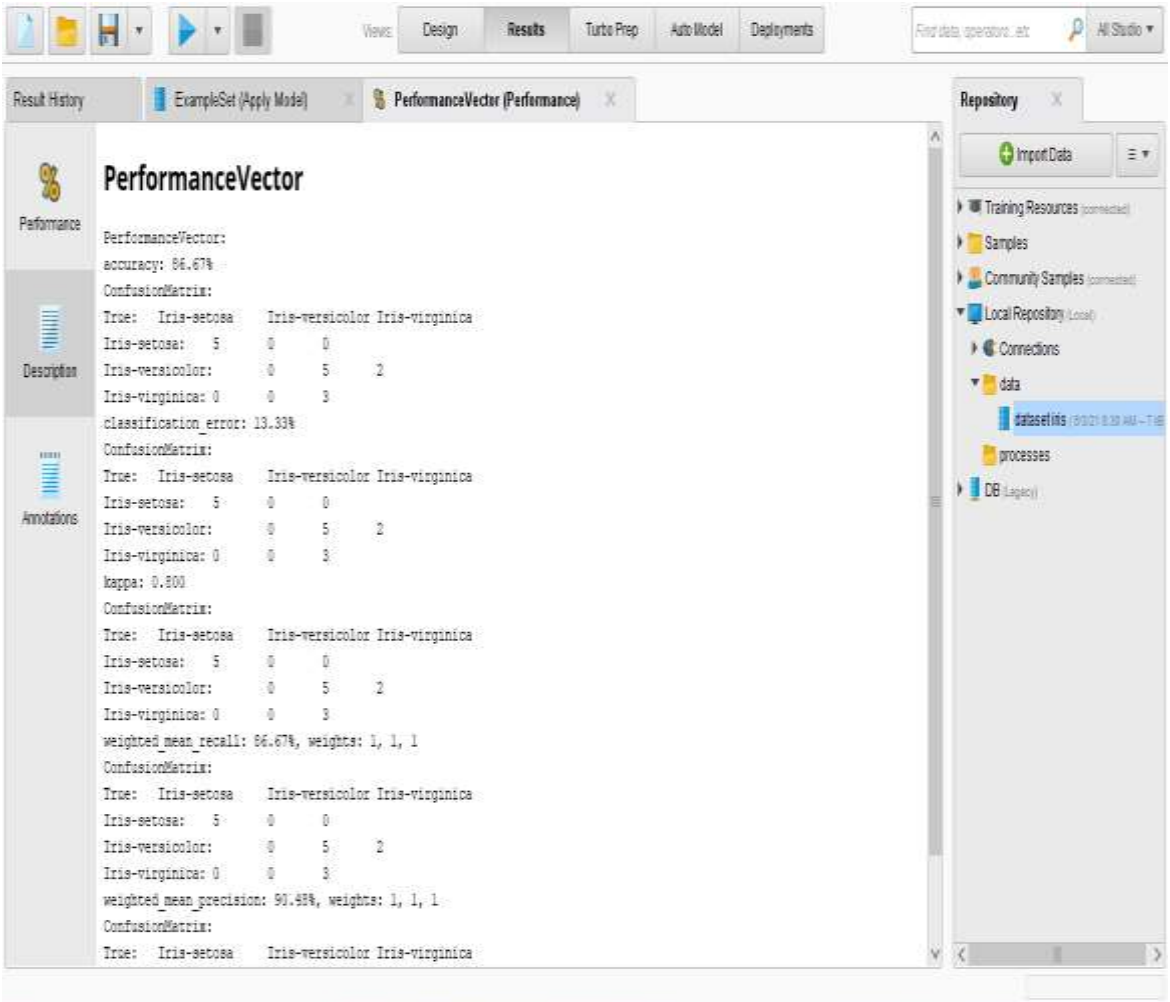
ISSN: 2598-0256, E-ISSN: 2598-0238

a. Berdasarkan petal length Jumlah Bunga iris setosa = 50 atau 33,3%.
dan dari data di atas diklasifikasikan bahwa untuk petal length ≤ 1,9 cm adalah 100% bunga iris setosa, dan untuk petal length ≥ 1,9 Iris versicolor.

b. Dari Iris versicolor dengan petal length ≥ 1,9 cm, jumlah bunga iris versicolor = 50/100 atau 50%. Akan tetapi dari data di atas diklasifikasikan lagi bahwa untuk petal length ≤ 1,7 cm sebanyak 90,7% = 49 kuntum bunga iris Versicolor, dan petal length > 1,7 cm sebanyak 97,8% = 45 kuntum bunga iris virginica.

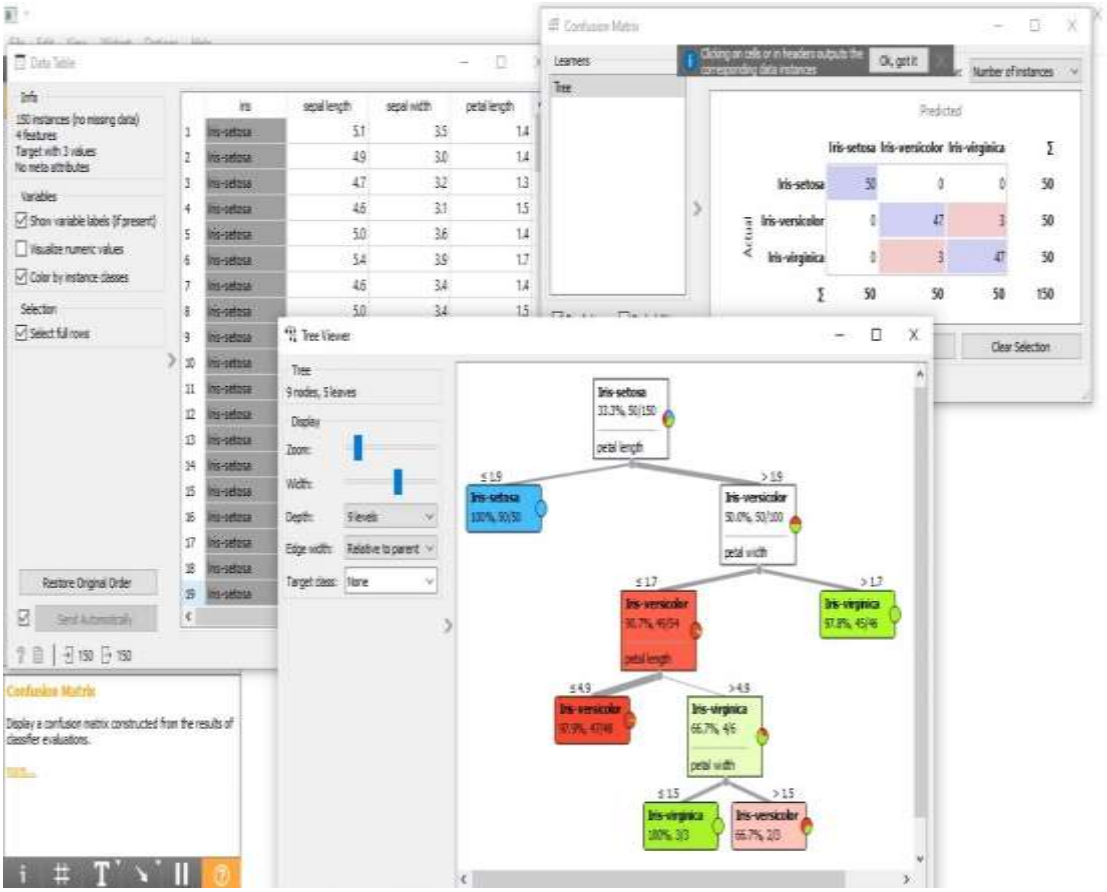
c. Akan tetapi dari data bunga Iris versicolor dengan petal length ≤ 1,7 cm dengan jumlah 49 kuntum, terklasifikasikan lagi bahwa untuk petal length ≤ 4,9 cm sebanyak 97,9% = 47 kuntum bunga iris Versicolor, dan petal length > 4,9 cm sebanyak 66,7% = 4 kuntum bunga iris virginica.

d. Selanjutnya dari data bunga Iris virginica dengan petal length > 4,9 cm sejumlah 4 kuntum, terklasifikasikan lagi bahwa untuk petal length ≤ 1,5 cm sebanyak 100% = 3 kuntum bunga iris virginica, dan petal length > 1,5 cm sebanyak 66,7% = 2 kuntum bunga iris virginica. Seperti yang tampilan pada gambar 5.



Gambar 4. Performa algoritma *decision tree* untuk dataset iris dengan tools rapid miner

ISSN: 2598-0256, E-ISSN: 2598-0238



Gambar 5. Performa algoritma *decision* tree untuk dataset iris dengan *tools* orange

5. KESIMPULAN

Dari data set iris yang berjumlah 150 data dengan penerapan algoritma *decision* tree maka data tersebut dibagi menjadi dua 90% data Testing dan 10% data training yang menghasilkan nilai accuracy: 86.67%, classification_error: 13.33%, kappa: 0.800, weighted_mean_recall: 86.67%, weights: 1, 1, 1 dan weighted_mean_precision: 90.48%, weights: 1, 1, 1. Pada aplikasi rapid miner dan petal length iris sentosa 33,3%, petal length iris versi color 90,7%, petal width iris versicolor 90,7% dan petal width iris virginica 66,7%.

UCAPAN TERIMA KASIH

Ucapan terima kasih penulis sampaikan kepada suami dan dua anak kembar penulis yang waktunya kebersamaannya sudah direlakan berkurang karena penyelesaian paper ini.

DAFTAR PUSTAKA

Ashari, A., Paryudi, I., & Min, A. (2013). Performance Comparison between Naïve Bayes, Decision Tree and k-Nearest Neighbor in Searching Alternative Design in an Energy Simulation Tool. International Journal of Advanced Computer Science and Applications, 4(11). https://doi.org/10.14569/ijacsa.2013.041105

A. Christobel and D. . Sivaprakasam. (2011). An Empirical Comparison of Data Mining Classification Methods. (vol. 3, no. 2, pp. 24–28).

D. Widiastuti. (2007). Analisa Perbandingan Algoritma SVM, Naïve Bayes, dan Decission Tree dalam Mengklasifikasikan Serangan (Attack) pada Sistem Pendeteksi Intrusi. Jur. Sist. Inf. Univ. Gunadarma, pp. 1–8, 2007.

ISSN: 2598-0256, E-ISSN: 2598-0238

I. H., Frank, E., & Mark A. Hall. (2011). Data Mining: Practical Machine learning 3rd ed. Morgan Kaufmann Publishers (p. 630).

K. Hastuti. (2012). Analisis komparasi algoritma klasifikasi data mining untuk prediksi mahasiswa non aktif. (vol. 2012, no. Semantik, pp. 241–249).

Ragab, A. H. M., Noaman, A. Y., AL-Ghamdi, A. S., & Madbouly, A. I. (2014). A comparative analysis of classification algorithms for students college enrollment approval using data mining. In ACM International Conference Proceeding Series (Vol. 09-June-2014, pp. 106–113). Association for Computing Machinery. https://doi.org/10.1145/2643604.2643631

E. Prasetyo. (2012). Data Mining Konsep Dan Aplikasi Menggunakan Matlab. Yogyakarta: Andi Offset p. 353.

B. Santosa. (2007). Data Mining Teknik Pemanfaatan Data untuk Keperluan Bisnis, Edisi Pert. Yogyakarta: Graha Ilmu.

Kusrini and L. E. Taufiq. (2009). Algoritma Data Mining. Yogyakarta: Andi Offset.

S. Susanto and D. Suryadi. (2010). Pengantar Data Mining: Menggali Pengetahuan dari Bongkahan Data. Yogyakarta: Andi Offset. p. 116.