

## VISUALISASI DATA PADA DATA MINING MENGGUNAKAN METODE KLASIFIKASI NAÏVE BAYES

Windi Irmayani

Program Studi Sistem Informasi Akuntansi Kampus Kota Pontianak  
Universitas Bina Sarana Informatika  
www.bsi.ac.id  
windi.wnr@bsi.ac.id

### INFO ARTIKEL

**Diajukan :**  
8 Desember 2020

**Diterima :**  
17 Mei 2021

**Diterbitkan :**  
14 Juni 2021

### Kata Kunci :

Visualisasi Data, Data Mining,  
Klasifikasi Naïve Bayes

### INTISARI

Visualisasi data diperlukan untuk membuat penerima data memahami dengan baik hasil dari pengolahan data. Terutama ketika data yang akan diolah berjumlah sangat banyak. Pengolahan data dengan jumlah yang sangat banyak dikenal dengan data mining. Dengan dataset mengenai kebun binatang metode klasifikasi Naïve Bayes digunakan untuk memprediksi klasifikasi hewan yang ada dengan 7 label, antara lain amphibian, bird, fish, insect, invertebrate, mammal dan reptile. Perhitungan klasifikasi Naïve Bayes menggunakan aplikasi data mining Orange. Didapatlah Proportion of prediction menunjukkan ada tiga prediksi klasifikasi dengan nilai 100%, satu dengan nilai 92,9%, satu dengan nilai 80%, satu dengan nilai 66,7% dan satu dengan nilai 33,3%. Dan Orange mampu memvisualisasikan hasil klasifikasi Naïve Bayes dengan grafik garis pada Receiver Operating Characteristics Analysis.

## I. PENDAHULUAN

Selama dua dekade terjadi peningkatan yang luar biasa terhadap jumlah data yang disimpan secara elektronik. Data akan bertambah terus menerus dan berpotensi menjadi sampah. Data akan berguna jika dilakukan pengolahan sehingga terjadi penemuan pengetahuan. "Untuk mengolah data yang jumlahnya besar tersebut menjadi sebuah informasi atau pengetahuan, diperlukan suatu teknik/metode yang dinamakan dengan data mining"(Suntoro, 2019).

Salah satu metode data mining adalah klasifikasi naïve bayes. "Naïve Bayes Classifier merupakan sebuah metoda klasifikasi yang berakar pada teorema Bayes" (Widianto, 2019). Naïve Bayes menghasilkan data yang terklasifikasi dengan metode probabilitas dan statistik untuk memprediksi masa datang berdasarkan masa lalu.

Hasil dari perhitungan data mining menggunakan metode klasifikasi naïve bayes ini akan makin berguna ketika penyajiannya menarik dan dapat dipahami dengan baik oleh penerima data. Terdapat beberapa cara menyajikan data diantaranya adalah dengan visualisasi data. Visualisasi data merupakan penyajian data yang menarik dan mudah untuk dipahami menggunakan grafik. Sehingga visualisasi data dapat membantu mempercepat pengambilan keputusan yang tepat. "Visualisasi data dapat menggambarkan relasi dan pola antara variable yang ada dalam data" (<http://lbi.si.fti.unand.ac.id/>).

Jenis-jenis visualisasi data antara lain:

1. Grafik

Grafik merupakan visualisasi data yang menggambarkan hubungan antar antara dua variable. Grafik diantaranya garis, batang, lingkaran luasan, solid dan pictorial.

2. Diagram  
Visualisasi data yang menggambarkan gambaran umum suatu proses.
3. Chart  
visualisasi yang menggambarkan hubungan yang ringkas. Chart diantaranya tree, flow, outline dan tabular.

Dataset yang digunakan adalah bawaan dari aplikasi data mining Orange. Dataset berisikan data hewan yang ada pada kebun binatang. Dengan metode klasifikasi naïve bayes data hewan yang ada akan diklasifikasikan menjadi beberapa jenis hewan seperti burung, serangga dan sebagainya. Setelah diklasifikasi hasil perhitungan disajikan dengan visualisasi berupa grafik. Grafik ini disajikan dengan harapan penerima data memahami dengan baik situasi yang ada sehingga membantu dalam pengambilan keputusan.

## II. BAHAN DAN METODE

Perhitungan klasifikasi naïve bayes maupun visualisasi data menggunakan aplikasi Orange.

1. Persiapan Dataset

Dataset yang digunakan merupakan dataset bawaan dari aplikasi Orange dengan nama file zoo.tab

2. Select Columns

Menentukan features, target variable dan meta attribute.

3. Naïve Bayes  
Perhitungan klasifikasi naïve bayes menggunakan persamaan dari teorema Bayes:

$$P(H|X) = \frac{P(X|H) \cdot P(H)}{P(X)}$$

Keterangan:

$X$  : Data dengan class yang belum diketahui

$H$  : Hipotesis data  $X$  merupakan suatu class

spesifik

$P(H|X)$  : Probabilitas hipotesis  $H$  berdasar kondisi

$X$

$P(H)$  : Probabilitas hipotesis  $H$

$P(X|H)$  : Probabilitas  $X$  berdasarkan kondisi pada

hipotesis  $H$

$P(X)$  : Probabilitas  $X$

Proses klasifikasi memerlukan sejumlah petunjuk untuk menentukan label/kelas apa yang cocok dengan sampel. Maka rumus di atas disesuaikan menjadi:

$$P(C|F_1 \dots F_n) = \frac{P(C)P(F_1 \dots F_n|C)}{P(F_1 \dots F_n)}$$

4. Test and Score

Evaluasi hasil model dengan mendapatkan data numerik.

- a. Area Under the Curve (AUC)  
Prediksi yang terdapat dalam penggunaan model AUC memudahkan dalam membandingkan model satu dengan yang lainnya

- b. Classification Accuracy (CA)  
Menghitung tingkat akurasi klasifikasi.

- c. F1  
Rata-rata terimbang dari presisi dan daya ingat dengan rumus:  
 $F1 = 2 \times (\text{presisi} \times \text{recall}) / (\text{presisi} + \text{recall})$

- d. Precision  
Kemampuan *classifier* untuk tidak memberikan label positif kepada sampel negative begitupun sebaliknya. Precision memiliki rumus sebagai berikut:

$$\text{Presisi} = \text{tp} / (\text{tp} + \text{fp})$$

Keterangan:

tp = jumlah positif sebenarnya

fp = jumlah positif palsu

- e. Recall

Kemampuan *classifier* untuk menemukan dan menggolongkan semua sampel yang bernilai positif.

- f. *Confusion Matrix*

Pengukuran performa masalah klasifikasi data mining dengan keluaran berupa dua kelas atau lebih.

- g. ROC Analysis

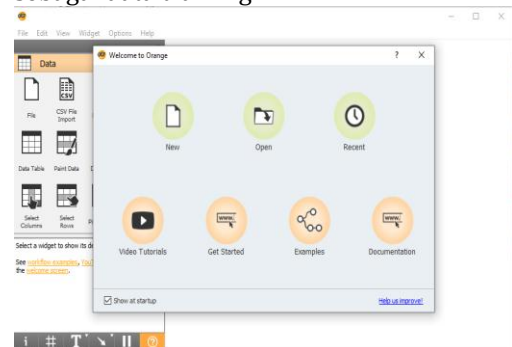
Analisa hasil dari alat ukur performance untuk mengklasifikasikan masalah dalam menentukan threshold dari suatu model.

### III. HASIL DAN PEMBAHASAN

Aplikasi Orange merupakan aplikasi data mining yang menghitung secara otomatis sesuai widget yang dipilih.

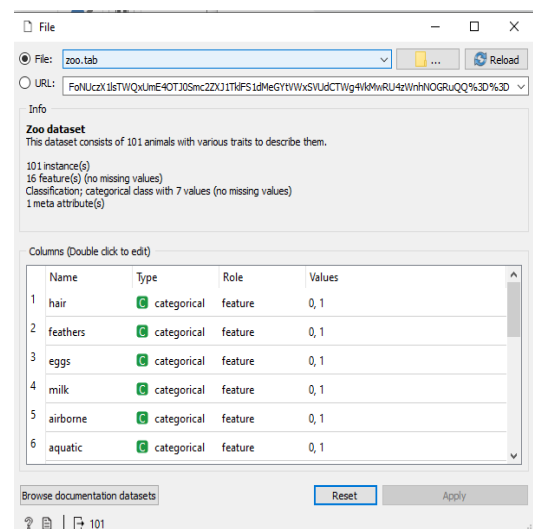
1. Persiapan Dataset

Dataset yang digunakan bernama zoo.tab. dataset telah dipersiapkan aplikasi orange sebagai data training.



Sumber: Hasil penelitian 2020

**Gambar 1 Menu New untuk memulai perhitungan data mining baru**

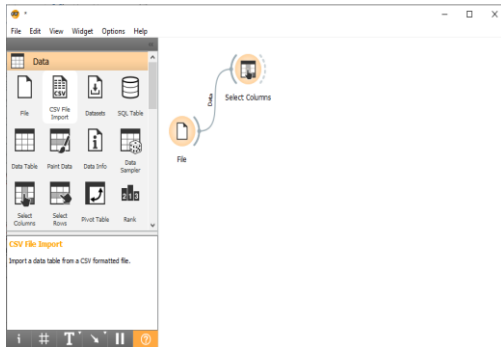


Sumber: Hasil penelitian 2020

**Gambar 2. Memilih dataset dengan nama file zoo.tab**

2. Select Columns

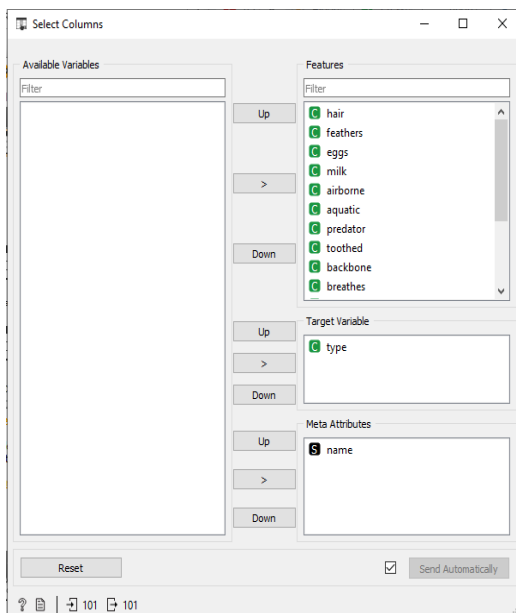
Select Columns mendapat masukan dari data dan features. Dan menghasilkan keluaran berupa data dan features. Data didapat setelah menghubungkan widget file dengan select columns seperti di bawah ini.



Sumber: Hasil Penelitian 2020

**Gambar 3. Select Columns**

Klik dua kali maka *form select columns* menampilkan *features*, target variable dan meta attribute.

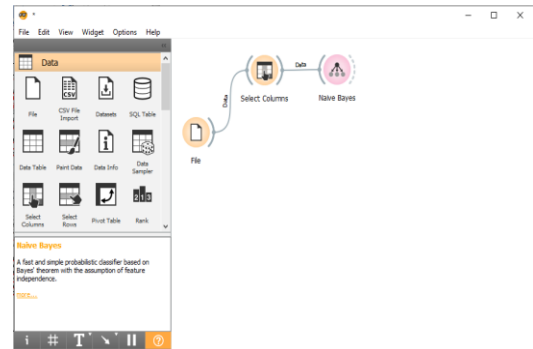


Sumber: Hasil Penelitian 2020

**Gambar 4. Select Columns**

3. Naïve Bayes

Masukan berupa Data dan Preprocessor dan menghasilkan model dan learner.

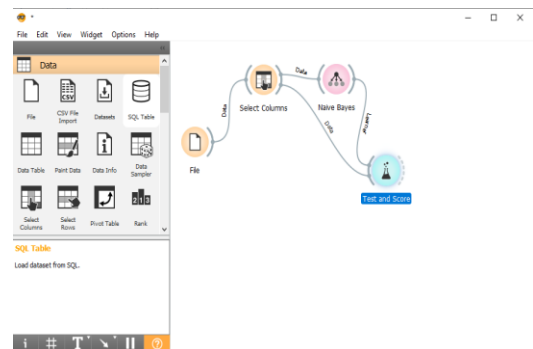


Sumber: Hasil penelitian 2020

**Gambar 5. Naïve Bayes**

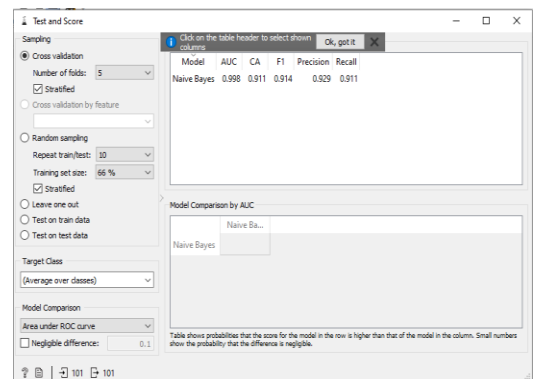
4. Test and Score

Masukan berupa Data, Test Data, Preprocessor dan Learner. Menghasilkan Prediksi dan Evaluasi Hasil.



Sumber: Hasil Penelitian 2020

**Gambar 6. Test and Score**



Sumber: Hasil Penelitian 2020

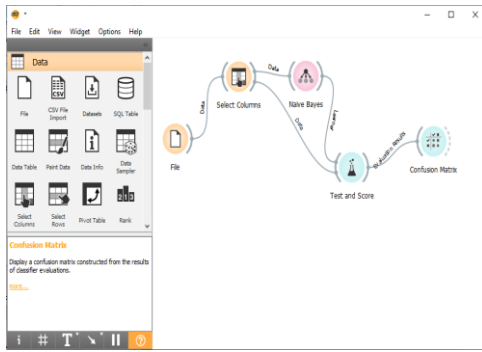
**Gambar 7. Output Test and Score**

Berdasarkan hasil perhitungan Bayes didapatkan:

- a. AUC = 0,998
- b. CA = 0,911
- c. F1 = 0,914
- d. Precision = 0,929
- e. Recall = 0,911

5. Confusion Matrix

Berdasarkan Hasil Evaluasi dari Test and Score *Confusion Matrix* menghasilkan Selected Data dan Data



Sumber: Hasil Penelitian 2020  
**Gambar 8. Confusion Matrix**

		Predicted						Σ	
		amphibian	bird	fish	insect	invertebrate	mammal		reptile
Actual	amphibian	4	0	0	0	0	0	0	4
	bird	0	20	0	0	0	0	0	20
	fish	0	0	13	0	0	0	0	13
	insect	0	0	0	8	0	0	0	8
	invertebrate	0	0	0	2	7	0	1	10
	mammal	0	0	0	0	0	38	3	41
reptile	2	0	1	0	0	0	2	5	
Σ		6	20	14	10	7	38	6	101

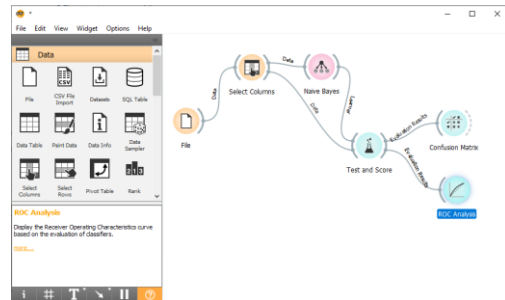
Sumber: Hasil Penelitian 2020  
**Gambar 9. Output Confusion Matrix**

Dari *Confusion Matrix* terdapat hasil data antara lain:

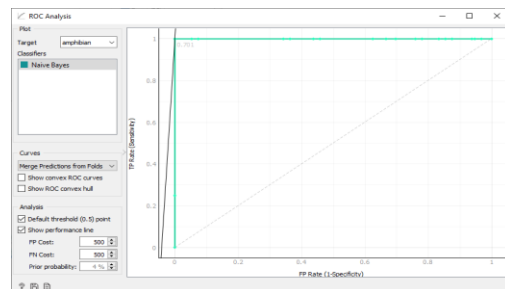
- a. Dari 6 jumlah data hewan dengan klasifikasi amphibian terdapat 4 data hewan dengan prediksi yang sesuai dengan data actual.
- b. Dari 20 jumlah data hewan dengan klasifikasi bird terdapat 20 data hewan dengan prediksi yang sesuai dengan data actual.
- c. Dari 14 jumlah data hewan dengan klasifikasi fish terdapat 13 data hewan dengan prediksi yang sesuai dengan data actual.
- d. Dari 10 jumlah data hewan dengan klasifikasi insect terdapat 8 data hewan dengan prediksi yang sesuai dengan data actual.
- e. Dari 7 jumlah data hewan dengan klasifikasi invrtebrate terdapat 7 data hewan dengan prediksi yang sesuai dengan data actual.
- f. Dari 38 jumlah data hewan dengan klasifikasi mammal terdapat 38 data hewan dengan prediksi yang sesuai dengan data actual
- g. Dari 6 jumlah data hewan dengan klasifikasi reptile terdapat 2 data hewan dengan prediksi yang sesuai dengan data actual

6. ROC Analysis  
ROC Analysis didapat dari Hasil Evaluasi. Menggambarkan hubungan Sensifitas

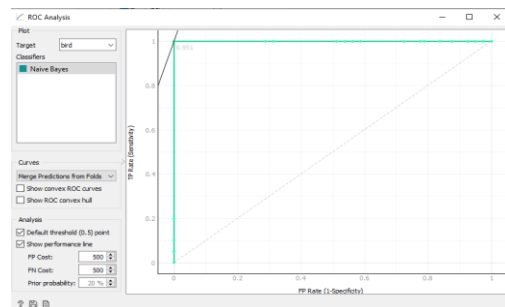
dengan spesifikasi terhadap model naïve bayes berupa grafik garis.



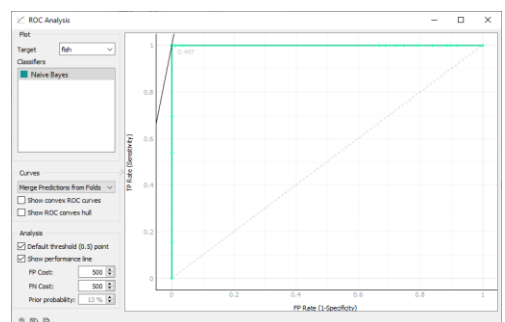
Sumber: Hasil Penelitian 2020  
**Gambar 10. ROC Analysis**



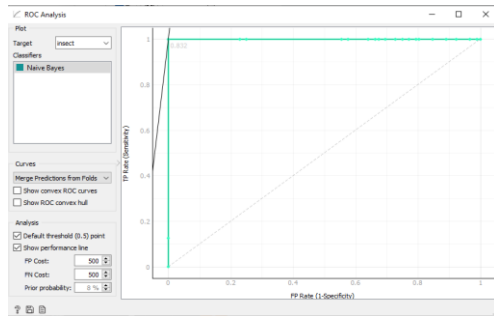
Sumber: Hasil Penelitian 2020  
**Gambar 11. Grafik ROC Klasifikasi Amphibian**



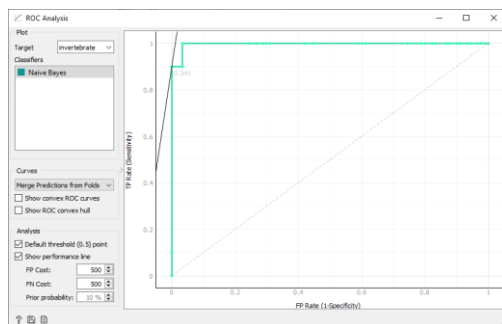
Sumber: Hasil Penelitian 2020  
**Gambar 12. Grafik ROC Klasifikasi Bird**



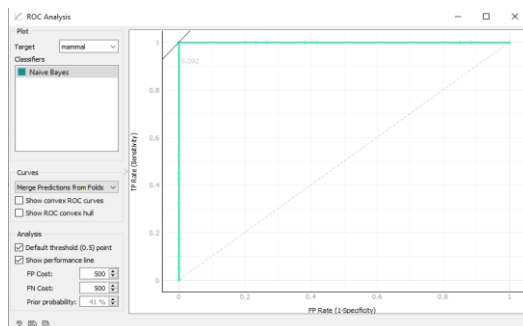
Sumber: Hasil Penelitian 2020  
**Gambar 13. Grafik ROC Klasiifikasi Fish**



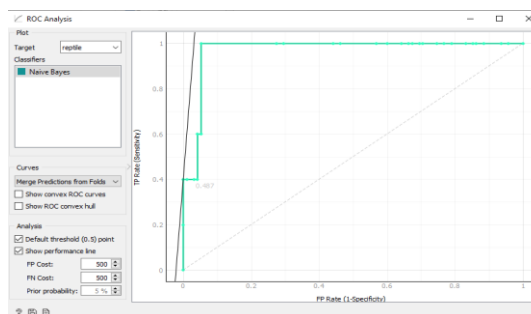
Sumber: Hasil Penelitian 2020  
**Gambar 14. Grafik ROC Klasifikasi Insect**



Sumber: Hasil Penelitian 2020  
**Gambar 15. Grafik ROC Klasifikasi Invertebrata**



Sumber: Hasil Penelitian 2020  
**Gambar 16. Grafik ROC Klasifikasi Mammal**



Sumber: Hasil Penelitian 2020  
**Gambar 17. Grafik ROC Klasifikasi Reptile**

yang terdapat pada dataset zoo.tab. Yaitu dengan Classification Accuracy sebesar 0,911. Klasifikasi dibagi menjadi 7 label antara lain amphibian, bird, fish, insect, invertebrate, mammal dan reptile. Proportion of prediction menunjukkan ada tiga prediksi klasifikasi dengan nilai 100%, satu dengan nilai 92,9%, satu dengan nilai 80%, satu dengan nilai 66,7% dan satu dengan nilai 33,3%.

Aplikasi Orange dapat memvisualisasikan data pada data mining dengan metode Naive Bayes menggunakan grafik garis menggunakan ROC Analysis.

## REFERENSI

Chairunnisa, Syifa. 2019. *Pentingnya visualisasi data untuk informasi yang menarik*. (Online), (<https://http://lbi.si.fti.unand.ac.id/pentingnya-visualisasi-data-untuk-informasi-yang-menarik/>, diakses 08 Desember 2020).

Suntoro, Joko (2019). *Data Mining Algoritma dan Implementasi dengan Pemrograman PHP*. Jakarta: Elex Media Komputindo.

Widiyanto, Haldi Mochammad. 2019. *Algoritma Naive Bayes*. (Online), (<https://binus.ac.id/bandung/2019/12/algoritma-naive-bayes/>, diakses 08 Desember 2020).

## IV. KESIMPULAN

Data Mining dengan metode Naive Bayes mampu dengan baik mengklasifikasi 101 hewan