

# Propensity To Buy Modelling Menggunakan Metode Klasifikasi Naïve Bayes

Ridho Suharis

Program Studi Magister Teknik Informatika Universitas Pamulang

*e-mail*: ridhosuharis95@gmail.com

**Abstrak**—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 *propensity to buy* metode klasifikasi Naïve Bayes digunakan untuk memprediksi klasifikasi kecenderungan user untuk membeli atau tidak membeli suatu produk. Perhitungan klasifikasi Naïve Bayes menggunakan aplikasi data mining Orange. Didapatlah Proportion of prediction menunjukkan Classification Accuracy sebesar 0,982. Klasifikasi dibagi menjadi 2 label antara lain membeli dan tidak membeli. Proportion of prediction menunjukkan prediksi klasifikasi dengan nilai 99,97% untuk tidak membeli dan dengan nilai 69,61% untuk membeli Orange mampu memvisualisasikan hasil klasifikasi Naïve Bayes dengan grafik garis pada Receiver Operating Characteristics Analysis.

**Kata Kunci**— Propensity To Buy; Naive Bayes; Orange.

## I. PENDAHULUAN

Propensity to buy adalah istilah yang digunakan dalam ilmu pemasaran dan ekonomi untuk menggambarkan sejauh mana seseorang atau kelompok orang cenderung untuk membeli produk atau layanan tertentu. Ini mencerminkan kecenderungan atau keinginan untuk melakukan pembelian. Faktor-faktor yang dapat memengaruhi propensity to buy termasuk kebutuhan, preferensi, anggaran, pengalaman pembelian sebelumnya, dan faktor-faktor psikologis. Dalam analisis pemasaran, perusahaan sering menggunakan data konsumen dan penelitian pasar untuk mengidentifikasi dan memahami propensity to buy potensial pelanggan, yang dapat membantu dalam merancang strategi pemasaran yang efektif.

Dalam konteks analisis data, propensity to buy dapat menjadi bagian dari model prediktif untuk memprediksi kemungkinan seseorang membeli produk atau layanan berdasarkan sejumlah variabel yang relevan. Ini dapat membantu perusahaan dalam mengarahkan upaya pemasaran mereka dengan lebih efektif dan meningkatkan peluang sukses dalam mengonversi prospek menjadi pelanggan. Untuk membangun model propensity to buy salah satu metode machine learning yang bisa digunakan adalah metode Naïve-Bayes.

Alasan menggunakan metode Naïve-Bayes pada makalah ini adalah Naïve Bayes adalah sederhana dan mudah diimplementasikan. Ini membuatnya cepat dalam pelatihan dan pengujian, sehingga cocok untuk masalah dengan jumlah fitur yang besar. Selain itu, Kinerja yang Baik dalam Kasus Klasifikasi Bernoulli dan Multinomial: Naïve Bayes cocok untuk masalah klasifikasi yang melibatkan data biner (klasifikasi Bernoulli) atau data berjumlah (klasifikasi Multinomial).

## II. METODE PENELITIAN

Perhitungan klasifikasi naïve bayes maupun visualisasi data menggunakan aplikasi Orange.

### A. *Persiapan Dataset*

Dataset yang digunakan merupakan data pengunjung sebuah data pengunjung website marketplace.

### B. *Select Colums*

*Select colums* digunakan untuk menentukan variabel features dan variabel target.

### C. *Naïve Bayes*

Perhitungan klasifikasi naïve bayes menggunakan persamaan dari teorema Bayes:

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

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:

$$P(H|X_1 \dots X_n) = \frac{P(X_1 \dots X_n|H) * P(H)}{P(X_1 \dots X_n)}$$

**D. Test and Score**

Evaluasi hasil model dengan mendapatkan data numerik

- 1) Area Under the Curve (AUC): Prediksi yang terdapat dalam penggunaan model AUC memudahkan dalam membandingkan model satu dengan yang lainnya.
- 2) Classification Accuracy (CA): Menghitung tingkat akurasi klasifikasi
- 3) F1: Rata-rata terimbang dari presisi dan daya ingat dengan rumus:

$$F1 = \frac{2 * (precision * recall)}{(precision + recall)}$$

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

$$Precision = \frac{tp}{tp + fp}$$

Keterangan:

*tp* : Jumlah positif sebenarnya

*fp* : Jumlah positif palsu

- 5) Recall: Kemampuan classifier untuk menemukan dan menggolongkan semua sampel yang bernilai positif

**E. Confusion Matrix**

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

**F. ROC Analysis**

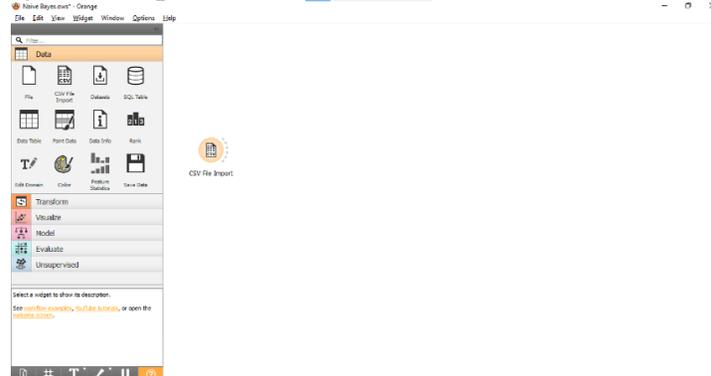
Analisa hasil dari alat ukur performance untuk mengklasifikasikan masalah dalam menentukan threshold dari suatu model.

**III. HASIL DAN PEMBAHASAN**

**A. Orange Work Flow**

- 1) Import Dataset

Import dataset kedalam aplikasi Orange.



**Gambar 1.**

Import dataset ke dalam aplikasi Orange

- 2) Select Columns

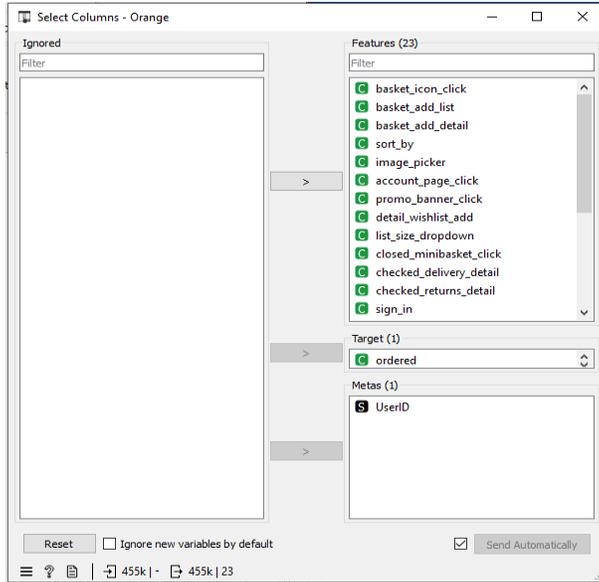
Select Columns dapat membantu kita dalam memilah data antara data target dan features. Data didapat setelah menghubungkan widget file dengan select columns seperti pada gambar di bawah ini.



**Gambar 2.**

Select Columns

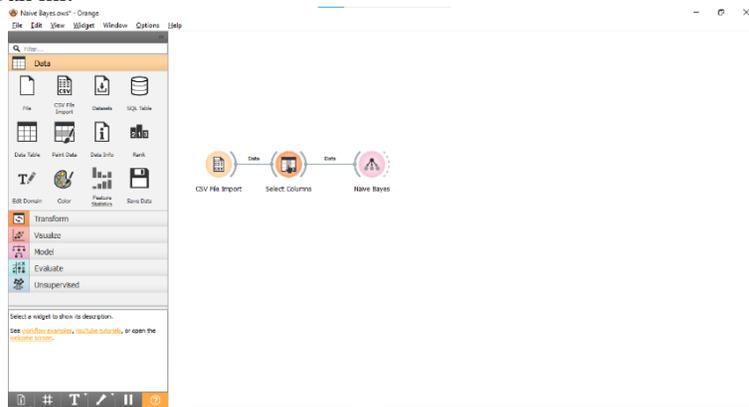
Klik dua kali maka form select columns akan muncul dan kita dapat memilih variabel yang menjadi target dan features seperti pada gambar di bawah ini.



**Gambar 3.**  
Form Select Colume

3) Naïve Bayes

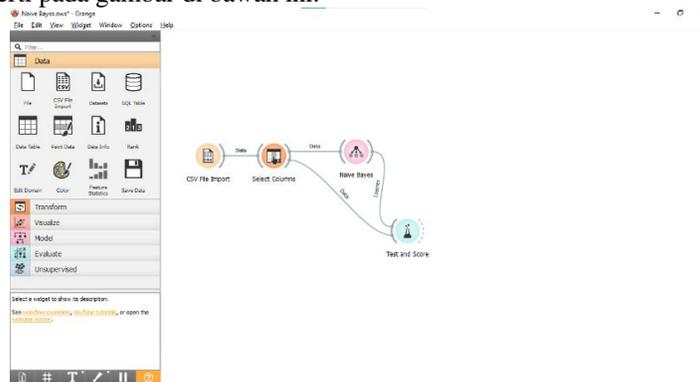
Untuk menggunakan metode Naïve Bayes, dapat dengan menghubungkan widget select columns ke widget naïve bayes seperti gambar di bawah ini:



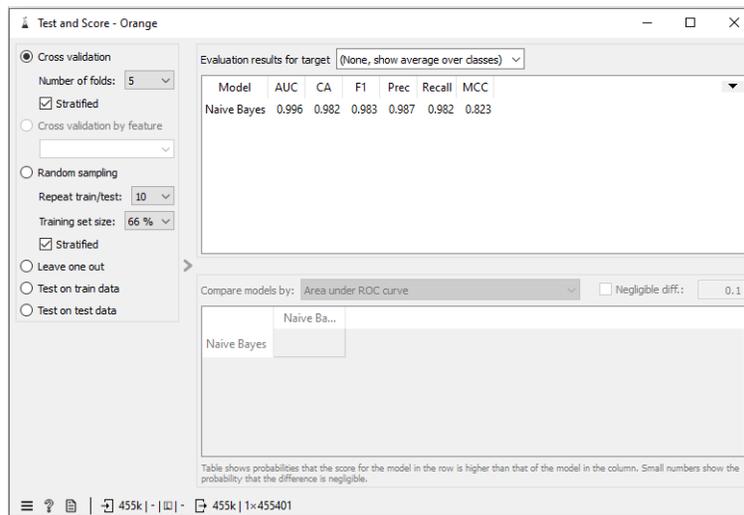
**Gambar 4.**  
Naïve Bayes

4) Test and Score

Test and score untuk menghasilkan prediksi dan evaluasi hasil. Data yang didapat merupakan data dari select columns dan widget naïve bayes seperti pada gambar di bawah ini.



**Gambar 5.**  
Test and Score



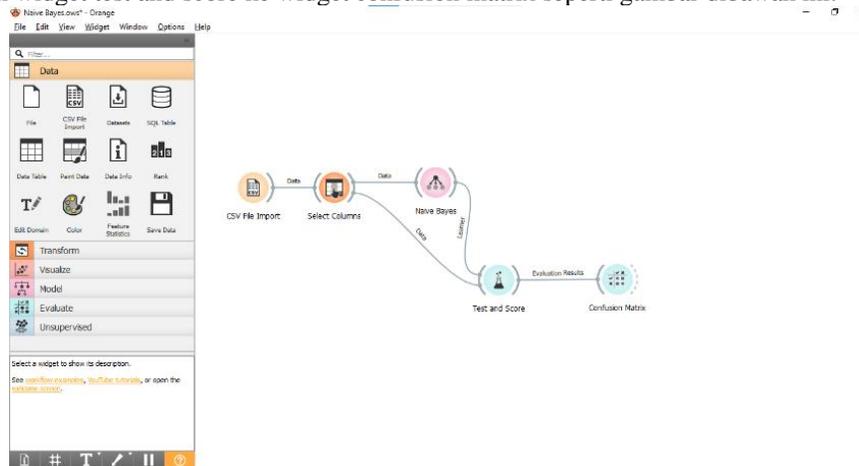
**Gambar 6.**  
 Hasil Test and Score

Berdasarkan test and score didapatkan hasil perhitungan sebagai berikut:

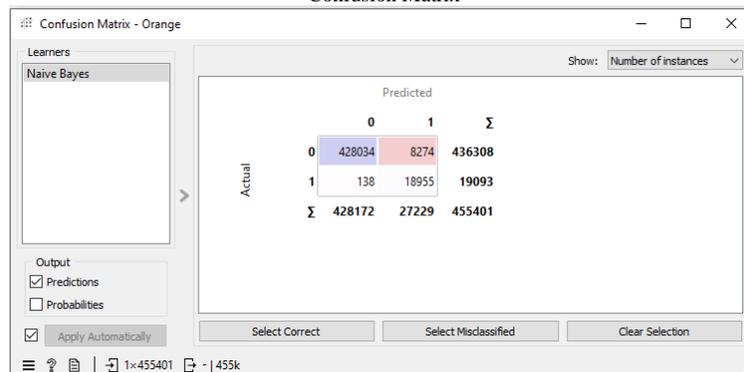
- AUC = 0.996
- CA = 0.982
- F1 = 0.983
- Precision = 0.987
- Recall = 0.982

5) Confusion Matrix

Confusion matrix dapat dibuat setelah hasil dari test and score didapatkan. Cara membuat confusion matrix adalah dengan menghubungkan widget test and score ke widget confusion matrix seperti gambar dibawah ini.



**Gambar 7.**  
 Confusion Matrix



**Gambar 8.**  
 Output Confusion Matrix

Dari confusion matrix terdapat hasil data sebagai berikut:

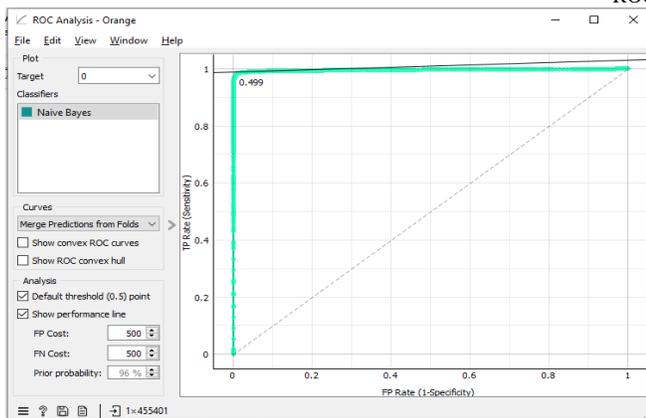
- Dari 436308 data pengunjung web yang tidak membeli, 428034-nya diprediksi tidak membeli.
- Dari 19093 data pengunjung yang membeli, 18955-nya diprediksi membeli.

6) ROC analysis

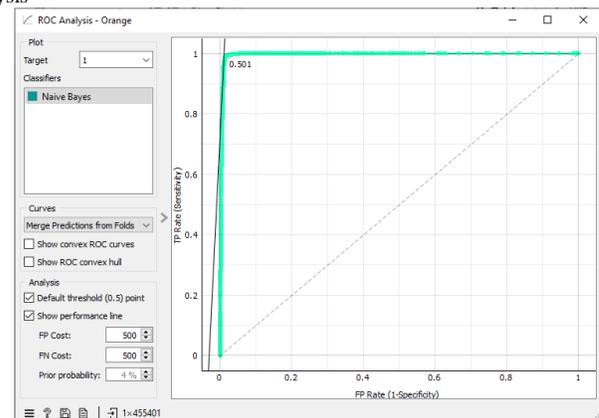
ROC analysis didapat dari hasil evaluasi. Menggambar hubungan sensitifitas dengan spesififikasi terhadap model naïve bayes berupa grafik garis.



**Gambar 9.**  
 ROC Analysis



**Gambar 10.**  
 Grafik ROC klasifikasi tidak membeli



**Gambar 11.**  
 Grafik ROC klasifikasi membeli

**IV. KESIMPULAN**

Data Mining dengan metode Naïve Bayes mampu dengan baik mengklasifikasi pengunjung yang membeli dan tidak membeli saat mengunjungi sebuah website market place. Yaitu dengan Classification Accuracy sebesar 0,982. Klasifikasi dibagi menjadi 2 label antara lain membeli dan tidak membeli. Proportion of prediction menunjukkan prediksi klasifikasi dengan nilai 99,97% untuk tidak membeli dan dengan nilai 69,61% untuk membeli. Aplikasi Orange dapat memvisualisasikan data pada data mining dengan metode Naïve Bayes menggunakan grafik garis menggunakan ROC Analysis.

**DAFTAR PUSTAKA**

- [1] Chairunnisa, Syifa. 2019. Pentingnya visualisasi data untuk informasi yang menarik. (Online), ([https:// lbi.si.fti.unand.ac.id/important-visualisasi-data-untuk-informasi-yang menarik/](https://lbi.si.fti.unand.ac.id/important-visualisasi-data-untuk-informasi-yang-menarik/), diakses 25 Oktober 2023).
- [2] Suntoro, Joko (2019). Data Mining Algoritma dan Implementasi dengan Pemrograman PHP. Jakarta: Elex Media Komputindo.
- [3] Widiyanto, Haldi Mochammad. 2019. Algoritma Naïve Bayes. (Online), (<https://binus.ac.id/bandung/2019/12/alg-oritma-naive-bayes/>, diakses 25 Oktober 2023).
- [4] Irmayani, Windi. 2021. Visualisasi Data Pada Data Mining Menggunakan Metode Klasifikasi Naive Bayes. Jakarta: Jurnal Khatulistiwa Informatika.