

## ANALISIS SENTIMEN TERHADAP APLIKASI LINKAJA MENGGUNAKAN METODE NAÏVE BAYES CLASSIFIER

Marheni Sitepu, Margaretha Yohanna✉, Samuel Van Basten H. Manurung

Program Studi Teknik Informatika, Universitas Methodist Indonesia, Medan, Indonesia

Email: [yohanna.na2@gmail.com](mailto:yohanna.na2@gmail.com)

DOI: <https://doi.org/10.46880/jmika.Vol8No1.pp44-50>

### ABSTRACT

*Electronic money is increasing, causing more and more service innovations to emerge. People carry out online transactions using electronic technology, or Fintech. The fintechs widely used today are e-wallets or digital wallets such as Dana, Ovo, LinkAja, etc. In this research, LinkAja as a fintech application will be analyzed using 100 review data samples and summarized into two classes: positive and negative. This research was carried out using the Naïve Bayes Classifier classification method. Sentiment analysis of reviews on the LinkAja application obtained results of 75% accuracy, 83% Precision, 75% Recall, and 73% F1\_score.*

**Keyword:** *Sentiment Analysis, LinkAja, Naïve Bayes Classifier.*

### ABSTRAK

*Uang elektronik yang semakin berkembang pesat membuat inovasi layanan yang muncul semakin banyak. Masyarakat melakukan transaksi online dengan teknologi elektronik atau disebut dengan Fintech. Fintech yang banyak digunakan saat ini adalah e-wallet atau dompet digital seperti Dana, Ovo, LinkAja dan lain sebagainya. Pada penelitian ini, LinkAja sebagai aplikasi fintech yang akan dianalisis dengan 100 sampel data ulasan dan dikelompokkan kedalam dua kelas yaitu positif dan negatif. Penelitian ini dilakukan dengan metode klasifikasi Naïve Bayes Classifier. Analisis sentimen ulasan pada aplikasi Linkaja didapatkan hasil akurasi sebesar 75%, Precicion 83%, Recall 75% dan F1\_score 73%.*

**Kata Kunci:** *Analisis Sentimen, LinkAja, Naïve Bayes Classifier.*

### PENDAHULUAN

Fintech merupakan singkatan dari financial technology. Fintech merupakan inovasi layanan dari sektor *payment* ialah *e-wallet*. Masyarakat saat ini lebih menyukai pembayaran non-tunai dibandingkan dengan uang tunai. Masyarakat akan mempertimbangkan kenyamanan, keamanan, ketepatan transaksi, dan banyaknya promosi (Ivan et al., 2018).

LinkAja merupakan *e-wallet* Indonesia yang dibuat oleh PT. Fintek Karya Nusantara (Finaraya) yang merupakan perusahaan BUMN yang telah diresmikan sejak 21 Februari 2019. LinkAja adalah layanan keuangan elektronik yang telah menghadirkan beragam fitur. LinkAja saat ini menempati posisi ke-6 berdasarkan Google Play Store (Turjaman & Budi, 2022).

Analisis sentimen merupakan proses memahami, pada suatu kalimat opini perlu melakukan mengolah data serta mengekstrak data secara otomatis untuk mendapatkan informasi sentimen. Analisis sentimen sering digunakan untuk melihat kecenderungan opini dan pendapat seseorang pada sebuah masalah, apakah orang tersebut cenderung

berpandangan positif atau negatif. Penelitian akan menganalisis permasalahan apa saja yang berhubungan dengan ulasan tentang aplikasi LinkAja dalam Google Play Store dan menentukan hasil ataupun nilai akurasi analisis sentimen yang dihasilkan algoritma *Naïve Bayes Classifier* (Surohman et al., 2020).

### TINJAUAN PUSTAKA

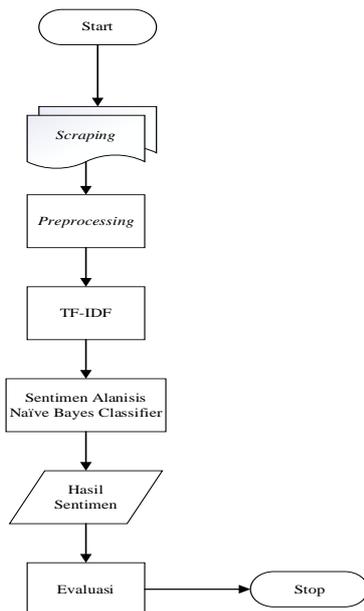
Ilmu komputer yang daya gunanya berperilaku seperti manusia disebut kecerdasan buatan. Ini merupakan ilmu komputer yang dapat membuat komputer melakukan pekerjaan seperti manusia (Restari et al., 2020). *Machine Learning* adalah kumpulan algoritma pemrograman yang dapat digunakan dalam sistem yang berdasarkan data sampel yang sudah ada dan mengoptimalkan cara kerja komputer. Keunggulannya adalah mampu memodifikasi dan mengadaptasi dalam merespon pada perubahan data (Putra & Santika, 2020). Pada penelitian ini menggunakan *Supervised Learning* yang merupakan teknik dalam melatih mesin menggunakan data yang dilabel atau proses pembelajaran untuk mencapai tingkat akurasi yang mengawasi berperan

sebagai ‘supervisor’ atau ‘guru’ (Kristiawan & Widjaja, 2021). Pada *Supervised Learning* ini menggunakan algoritma *Naïve Bayes Classifier* yang merupakan klasifikasi dan perhitungan, yang menemukan algoritma ini ialah *Thomas Bayes* adalah seorang ilmuwan yang berasal dari inggris dan menghasilkan prediksi peluang kedepannya dengan pengalaman yang telah terjadi (Rachman & Handayani, 2021). Pada tahun 2021 dilakukan penelitian yang untuk menganalisis sentimen review Twitter BMKG Nasional dimana peneliti membahas media sosial *twitter* yang jumlah peningkatan penggunaanya tercatat di seluruh dunia. Penelitian mendapatkan hasil akurasi pada pengujian yang telah dilakukan sebesar 68,97% (Darwis et al., n.d.). Penelitian lainnya membahas kebijakan kampus Merdeka yang sangat ramai disorot publik terutama pada platform youtube berhubungan dengan video yang di upload Mendikbud, opini Masyarakat yang terdapat pada kolom komentar youtube akan diklasifikasi peneliti ini kedalam sentimen positif dan negatif. Hasil yang didapat untuk akurasinya sebesar 97% dengan data yang digunakan 900 data latih dan 100 data uji (Zhafira et al., 2021)

**METODE PENELITIAN**

**Alur Proses Penelitian**

Pada Penelitian ini dilakukan tahapan yang akan dilakukan, pengolahan diagram alir dalam penelitian dapat diperhatikan pada Gambar 1.



**Gambar 1.** Proses Penelitian

**Scraping**

*Scraping* merupakan tahap pengambilan data. Penelitian ini memperoleh data ulasan google play

store menggunakan tools *Python*. File tersebut dapat dibuka dengan Excel dan mengolah data sesuai yang penelitian.

**Preprocessing**

*Preprocessing* merupakan tahapan penghilangan atribut yang tidak penting atau diperlukan dalam proses pengelompokan sentimen. Pada penelitian ini akan melakukan empat tahapan preprocessing yaitu *case folding, tokenizing, stopword removal* dan *stemming*.

**Pembobotan Kata**

Pembobotan kata pada penelitian ini menggunakan algoritma *TF-IDF*. Algoritma *TF-IDF* dibutuhkan dalam mendapatkan nilai bobot pada suatu dokumen, agar dapat dilakukan poses selanjutnya bentuk kalimat dirubah kedalam bentuk numerik ataupun bentuk angka (Robison Manalu et al., 2022). Rumus untuk menentukan algoritma *TF-IDF* dapat dilihat pada persamaan (1).

$$W_{dt} = tf_t * IDF_t \quad \dots(1)$$

Keterangan:

$W_{dt}$  : Bobot *term* ke *t* dokumen *d*.

$TF_t$  : Jumlah kemunculan *term* *t* pada dokumen *t*.

$IDF_t$  : Banyak jumlah kata dalam dokumen

$$IDF = \log \left( \frac{D}{df_t} \right)$$

**Klasifikasi Naïve Bayes Clasifier**

*Naïve Bayes Classifier* merupakan klasifikasi kemungkinan yang sederhana untuk dapat menghitung semua kemungkinan dengan cara digabungkan semua jumlah kombinasi dan frekuensi pada nilai dari basis data yang diperoleh. *Naïve Bayes* merupakan klasifikasi dan perhitungan, yang menemukan ialah *Thomas Bayes* seorang ilmuwan yang berasal dari inggris dan menghasilkan prediksi peluang kedepannya dengan pengalaman yang telah terjadi (Rachman & Handayani, 2021). Pengklasifikasi bayes adalah klasifikasi statistik, dimana klasifikasi dapat mempredikasi probabilitas anggota kelas suatu data dan ditentukan kelas yang akan dimasuki.

Algoritma *Naïve Bayes* merupakan algoritma yang klasifikasinya dapat digunakan dalam data mining dan text mining. Langkah - langkah ini berdasarkan teorema *Bayes* bahwa pemilihan kelas tertentu itu semua kegiatan dalam memberikan suatu kontribusi yang saling bebas dan sama penting. *Naïve Bayes Classifier* merupakan metode klasifikasi dalam menentukan bagaimana bayangan persepsi masyarakat

pada text mining(Darwis et al., n.d.). Adapun langkah-langkah dalam penerapan Algoritma *Naïve Bayes Classifier*, yaitu:

- 1) Menyimpan dataset
- 2) Menghitung jumlah kelas.
- 3) Menghitung jumlah kasus. Jika terdapat data atribut nominal maka dapat digunakan perhitungan *Teorema Bayes* pada persamaan (2).

$$P(A\setminus B) = \frac{P(B\setminus A)P(A)}{P(B)} \quad \dots(2)$$

Keterangan:

- B : Data kelas belum diketahui
- A : Hipotesis data
- P(A\B) : Probabilitas A berdasarkan kondisi B
- P(A) : Probabilitas hipotesis A (prior probabilitas)
- P(B\A) : Probabilitas B berdasarkan kondisi A
- P(B) : Probabilitas B

Sedangkan jika data atribut bertipe numerik maka dapat digunakan perhitungan distribusi *Gaussian* pada persamaan (3).

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \quad \dots(3)$$

Keterangan:

- f(x) : gaussian
- x : data
- μ : rata-rata (mean)
- σ : standar deviasi
- π : phi
- e : 2,7183

Mengkalikan seluruh hasil dengan data testing yang akan ditentukan kelasnya dengan persamaan (4).

$$P(A\setminus B) = P(A) \prod_{i=1}^n P(B_i\setminus A) \quad \dots(4)$$

Keterangan :

- B : Data kelas belum diketahui
- A : Hipotesis data
- P(A\B) : Probabilitas A berdasarkan kondisi B (
- P(A) : Probabilitas hipotesis A (prior probabilitas)
- P(B\A) : Probabilitas B berdasarkan kondisi A
- P(B) : Probabilitas B
- N : Jumlah data

Selanjutnya mengkalikan dengan hasil pada Langkah ke 2 yang merupakan jumlah kelas dalam data training. Membandingkan hasil per kelas, kelas baru dilihat dari nilai yang tertinggi.

### Evaluasi

Penelitian ini akan dilakukan evaluasi terhadap 100 dataset. Ukuran umum akurasi untuk klasifikasi teks adalah *precision*, *recall*, *accuracy* dan *F-measure (F1-Score)*. Peneliti akan menghitung akurasinya menggunakan *Confusion Matrix*. *Confusion matrix* merupakan table yang digunakan dalam menganalisis seberapa baik keakuratan suatu cara klasifikasi dalam memprediksi data. Penelitian ini, data dilakukan klasifikasi menjadi positif dan negatif(Kurniawan & Adinugroho, 2019) . Tabel *confusion matrix* dapat diperhatikan pada Tabel 1.

**Tabel 1.** Confusion Matrix

| Kelas   | Positif             | Negatif             |
|---------|---------------------|---------------------|
| Positif | True Positive (TP)  | False Negative (FN) |
| Negatif | False Positive (FP) | True Negative (TN)  |

Untuk menghitung akurasinya adalah dengan persamaan (5).

$$accuracy = \frac{tp+tn}{tp+tn+fp+fn} \quad \dots(5)$$

### HASIL DAN PEMBAHASAN

#### Scraping Data

Pada scraping ini data yang digunakan adalah ulasan komentar aplikasi LinkAja dari google play store. Teknik tersebut dilakukan dalam pengumpulan ataupun pengambilan data. Data ulasan yang ditarik ada 100 data ulasan. Sampel data ulasan dapat diperhatikan pada Tabel 2.

**Tabel 2.** Sampel Data Ulasan

| No | Ulasan  |
|----|---|
| 1  | Tolong dibantu akun saya di blokir  |
| 2  | Aplikasi sudah di update tapi mau ganti password aja ngga bisa. Aplikasi apaan ini? Bikin aplikasi yg bener |
| 3  | Gak bisa aktif  |
| 4  | mau reset pin kok gak bisa . terus solusinya gimana ya .  |
| 5  | mau reset pin kok gak bisa . terus solusinya gimana ya .  |

#### Preprocessing Data

*Preprocessing* dataset diperlukan untuk mengolah kumpulan data, dimana data yang tidak terstruktur akan diproses menjadi data yang telah

terstruktur, *preprocessing* terdiri dari empat tahapan yaitu:

- Case Folding*: merupakan proses perubahan seluruh jenis huruf *uppercase* menjadi *lowercase*,
- Tokenizing*: adalah tahapan pemecahan yang berupa kalimat menjadi suatu kata atau memutus string menjadi potongan kata.
- Stopword Removal*: adalah proses penghapusan kata yang umum, imbuhan yang tidak penting dan menghilangkan kata gaul dan lainnya.
- Stemming*: merupakan proses yang mengembalikan kata-kata menjadi bentuk akarnya (*root*).

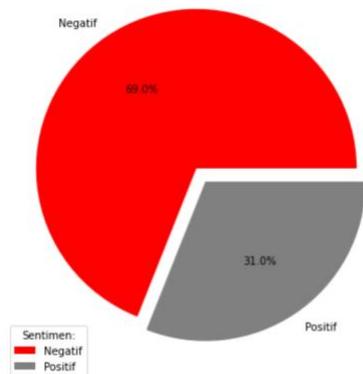
Adapun sampel data dari hasil *preprocessing* dapat dilihat pada Tabel 3.

**Tabel 3.** Sampel Data Ulasan Hasil Preprocessing

| No | Ulasan                                   |
|----|--|
| 1  | akun blokir                              |
| 2  | puas                                     |
| 3  | aktif                                    |
| 4  | pin solusi                               |
| 5  | susah verifikasi kirim uang terima kasih |

**Pelabelan Data Ulasan**

Tahapan selanjutnya dilakukan proses pelabelan sentimen data ulasan secara manual, dalam pelabelan sentimen data peneliti hanya membagi kedalam nilai 1 dan -1. Sentimen kelas positif akan diberi nilai 1 dan sentimen kelas negatif diberi nilai -1. Hasil dari pelabelan sentimen sebanyak 100 sampel data secara manual oleh peneliti adalah sebesar 31 data bersentimen positif dan 69 data bersentimen negatif. Hasil pelabelan data ulasan dapat dilihat diagram lingkaran pada Gambar 2.



**Gambar 2.** Diagram Pelabelan Manual Data Ulasan

**Pembobotan TF-IDF**

Pembobotan data menggunakan algoritma *TF-IDF* tujuannya untuk memperoleh nilai/bobot terhadap dokumen. Untuk proses perhitungan *TF-IDF* manual menggunakan persamaan (1). Sampel data ulasan yang digunakan adalah 5 data ulasan hasil *preprocessing* sebelumnya, hasil pembobotan *TF-IDF* dapat diperhatikan pada Tabel 4.

**Tabel 4.** Hasil Pembobotan TF-IDF

| Term | W(TF*IDF) |       |       |       |       |
|------|-----------|-------|-------|-------|-------|
|      | D1        | D2    | D3    | D4    | D5    |
| T1   | 0,698     | 0     | 0     | 0     | 0     |
| T2   | 0,698     | 0     | 0     | 0     | 0     |
| T3   | 0         | 0,698 | 0     | 0     | 0     |
| T4   | 0         | 0     | 0,698 | 0     | 0     |
| T5   | 0         | 0     | 0     | 0,698 | 0     |
| T6   | 0         | 0     | 0     | 0,698 | 0     |
| T7   | 0         | 0     | 0     | 0     | 0,698 |
| T8   | 0         | 0     | 0     | 0     | 0,698 |
| T9   | 0         | 0     | 0     | 0     | 0,698 |
| T10  | 0         | 0     | 0     | 0     | 0,698 |
| T11  | 0         | 0     | 0     | 0     | 0,698 |
| T12  | 0         | 0     | 0     | 0     | 0,698 |

**Klasifikasi Metode Naïve Bayes Classifier**

Sebelum melakukan tahapan klasifikasi pada metode *naïve bayes Classifier* terlebih dahulu melakukan dataset *Splitting* yang merupakan membagi data menjadi dua, yaitu data *training* dan data *testing*. Skenario yang dipakai oleh penelitian adalah data *training* sebesar 80% (80 data) dan data *testing* 20% (20 data). Pengklasifikasian dengan *Naïve Bayes Classifier* merupakan memanfaatkan hasil dari pembobotan kata *TF-IDF* yang terdapat 5 sampel ulasan sebelumnya terdapat kumpulan kata dari kelima dokumen dengan *term* sebanyak 12 dan variabel *y* adalah label sentimen dari kelima dokumen, D1 dan D2 merupakan sentimen positif dan D3, D4, D5 merupakan sentimen negatif. Kelima dokumen akan dijadikan sebagai data *training* dan dokumen untuk data *testing* akan diambil secara acak untuk membentuk suatu model dan dijadikan pembelajaran bagi sistem yang akan menentukan kelas atau label pada data uji. Berikut adalah langkah-langkah algoritma *naïve bayes classifier*.

- Menyimpan dataset, dataset dapat diperhatikan pada Tabel 5.

Tabel 5 . Dataset

| Term       | D1 | D2 | D3 | D4 | D5 |
|------------|----|----|----|----|----|
| akun       | 1  | 0  | 0  | 0  | 0  |
| blokir     | 1  | 0  | 0  | 0  | 0  |
| puas       | 0  | 1  | 0  | 0  | 0  |
| aktif      | 0  | 0  | 1  | 0  | 0  |
| pin        | 0  | 0  | 0  | 1  | 0  |
| solusi     | 0  | 0  | 0  | 1  | 0  |
| susah      | 0  | 0  | 0  | 0  | 1  |
| verifikasi | 0  | 0  | 0  | 0  | 1  |
| kirим      | 0  | 0  | 0  | 0  | 1  |
| uang       | 0  | 0  | 0  | 0  | 1  |
| terima     | 0  | 0  | 0  | 0  | 1  |
| kasih      | 0  | 0  | 0  | 0  | 1  |

- Hitung jumlah kelas pada data training, dari Tabel 5 dapat dilihat sentimen dari kelima dokumen.
  - Jumlah kelas positif = 2
  - Jumlah kelas negatif = 3
- Menghitung jumlah kasus yang sama, pada perhitungan ini menggunakan persamaan (2).

Dalam tahapan perhitungan *term* kemungkinan tidak terdapat dalam suatu kelas, jika *term* tersebut tidak terdapat suatu kelas maka *term* tersebut akan bernilai 0. Hal ini bisa membuat hasil perhitungan *Naïve Bayes Classifier* menjadi tidak tepat. Terdapat sebuah metode bernama *Laplace Smoothing* yang bisa dipergunakan untuk mengantisipasi hal seperti ini. *Laplacian Smoothing* merupakan metode *smoothing* yang dapat diterapkan pada *Naïve Bayes Classifier*, memiliki konsep yang sederhana dengan cara menambahkan nilai positif yang bernilai 1 pada setiap probabilitas untuk menghindari nilai nol pada model probabilitas dan mendapatkan nilai agar hasil yang didapatkan maksimal dan lebih akurat.

Jadi berdasarkan pernyataan diatas, kurangnya data pelatihan pada teori *Naïve Bayes Classifier* dapat menyebabkan nilai probabilitas 0, untuk mengatasinya dengan menggunakan metode *Laplace Smoothing* dengan menambahkan setiap perhitungan dengan angka 1. Penambahan tersebut tidak akan membuat perbedaan yang berarti dan akan menghindari nilai probabilitas 0. Maka rumus *Naïve Bayes Classifier* akan diubah menjadi:

$$P(A|B) = \frac{P(B|A) P(A) + 1}{P(B)}$$

Berikut merupakan perhitungan untuk *term* dataset pada kelas positif dan negatif.

$$P(\text{akun}|\text{positif}) = \frac{1+1}{3+12} = \frac{2}{15} = 0,133$$

$$P(\text{akun}|\text{negatif}) = \frac{0+1}{9+12} = \frac{1}{21} = 0,047$$

$$P(\text{blokir}|\text{positif}) = \frac{1+1}{3+12} = \frac{2}{15} = 0,133$$

$$P(\text{blokir}|\text{negatif}) = \frac{0+1}{9+12} = \frac{1}{21} = 0,047$$

$$P(\text{puas}|\text{positif}) = \frac{1+1}{3+12} = \frac{2}{15} = 0,133$$

$$P(\text{puas}|\text{negatif}) = \frac{0+1}{9+12} = \frac{1}{21} = 0,047$$

$$P(\text{aktif}|\text{positif}) = \frac{0+1}{3+12} = \frac{1}{15} = 0,067$$

$$P(\text{aktif}|\text{negatif}) = \frac{1+1}{9+12} = \frac{2}{21} = 0,095$$

$$P(\text{pin}|\text{positif}) = \frac{0+1}{3+12} = \frac{1}{15} = 0,067$$

$$P(\text{pin}|\text{negatif}) = \frac{1+1}{9+12} = \frac{2}{21} = 0,095$$

$$P(\text{solusi}|\text{positif}) = \frac{0+1}{3+12} = \frac{1}{15} = 0,067$$

$$P(\text{solusi}|\text{negatif}) = \frac{1+1}{9+12} = \frac{2}{21} = 0,095$$

$$P(\text{susah}|\text{positif}) = \frac{0+1}{3+12} = \frac{1}{15} = 0,067$$

$$P(\text{susah}|\text{negatif}) = \frac{1+1}{9+12} = \frac{2}{21} = 0,095$$

$$P(\text{verifikasi}|\text{positif}) = \frac{0+1}{3+12} = \frac{1}{15} = 0,067$$

$$P(\text{verifikasi}|\text{negatif}) = \frac{1+1}{9+12} = \frac{2}{21} = 0,095$$

$$P(\text{kirим}|\text{positif}) = \frac{0+1}{3+12} = \frac{1}{15} = 0,067$$

$$P(\text{kirим}|\text{negatif}) = \frac{1+1}{9+12} = \frac{2}{21} = 0,095$$

$$P(\text{uang}|\text{positif}) = \frac{0+1}{3+12} = \frac{1}{15} = 0,067$$

$$P(\text{uang}|\text{negatif}) = \frac{1+1}{9+12} = \frac{2}{21} = 0,095$$

$$P(\text{terima}|\text{positif}) = \frac{0+1}{3+12} = \frac{1}{15} = 0,067$$

$$P(\text{terima}|\text{negatif}) = \frac{1+1}{9+12} = \frac{2}{21} = 0,095$$

$$P(\text{kasih}|\text{positif}) = \frac{0+1}{3+12} = \frac{1}{15} = 0,067$$

$$P(\text{kasih}|\text{negatif}) = \frac{1+1}{9+12} = \frac{2}{21} = 0,095$$

Hasil semua perhitungan *conditional probabilitas* dapat diperhatikan pada Tabel 6.

Tabel 6. Hasil Perhitungan *Conditional Probabilitas*

| Term   |        |         |
|--------|--------|---------|
|        | P(w 1) | P(w -1) |
| akun   | 0,133  | 0,047   |
| blokir | 0,133  | 0,047   |
| puas   | 0,133  | 0,047   |
| aktif  | 0,067  | 0,095   |
| pin    | 0,067  | 0,095   |
| solusi | 0,067  | 0,095   |

|            |       |       |
|------------|-------|-------|
| susah      | 0,067 | 0,095 |
| verifikasi | 0,067 | 0,095 |
| kirim      | 0,067 | 0,095 |
| uang       | 0,067 | 0,095 |
| terima     | 0,067 | 0,095 |
| kasih      | 0,067 | 0,095 |

Nilai probabilitas kata pada masing-masing kelas akan digunakan untuk menguji data yang ada, sedangkan probabilitas setiap kelas ditentukan dengan menggunakan rumus dibawah ini.

$$P(c) = \frac{Nc}{N}$$

Keterangan :

Nc = Jumlah data training dalam kelas c

N = Jumlah seluruh data training

Nilai prior probabilitas dapat diperhatikan pada Tabel 7.

**Tabel 7.** Nilai Probabilitas Kelas

| Kelas   | Probabilitas Kelas         |
|---------|----------------------------|
| Positif | $P(n) = \frac{2}{5} = 0,4$ |
| Negatif | $P(n) = \frac{3}{5} = 0,6$ |

4. Kalikan seluruh hasil dengan data *testing* yang akan digali kelasnya menggunakan rumus tersebut.

$$P(A \setminus B) = P(A) \prod_{i=1}^n P(B_i \setminus A)$$

Data testing (D6) = ‘cara verifikasinya sangat susah’  
 D6 = ‘verifikasi’ ‘sangat’ ‘susah’

$$P(\text{Positif} | D6) = P(\text{Positif}) * P(\text{verifikasi} | \text{Positif}) * P(\text{susah} | \text{Positif})$$

$$P(\text{Positif} | D6) = 0,4 * 0,067 * 0,067 = 0,0017956$$

$$P(\text{Negatif} | D6) = P(\text{Negatif}) * P(\text{verifikasi} | \text{Negatif}) * P(\text{susah} | \text{Negatif})$$

$$P(\text{Negatif} | D6) = 0,6 * 0,095 * 0,095 = 0,005415$$

Karena posterior sentimen negatif terbesar, maka D6 masuk ke sentimen Negatif.

**Pengujian Model Naïve Bayes Classifier**

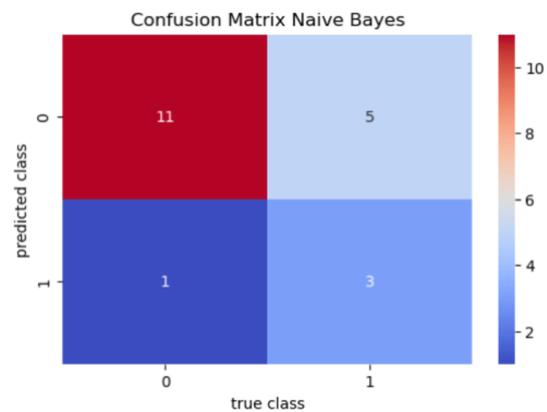
Pengujian model dilakukan untuk mengetahui baik performa model *Naïve Bayes Classifier* yang sudah dilakukan sebelumnya dalam data *testing* sebanyak 80% data *training* (80 data) dan data *training*

sebanyak 20% data *testing* (20 data). Hasil dari pengujian data uji pada model *Naïve Bayes Classifier* menggunakan *confusion matrix* didapatkan akurasi sebesar 70% seperti Gambar 3.

```
MultinomialNB Accuracy: 0.7
MultinomialNB Precision: 0.7125
MultinomialNB Recall: 0.7
MultinomialNB f1_score: 0.6714285714285715
confusion matrix:
[[11  1]
 [ 5  3]]
```

**Gambar 3.** Hasil Pengujian Naïve Bayes Classifier

Tahapan terakhir merupakan melakukan evaluasi pada Naïve Bayes Classifier untuk mengukur atau meninjau hasil klasifikasi dari model yang sudah ditentukan dengan klasifikasi yang sebenarnya. *Confusion matrix* yang berordo 2 x 2 hasil dari evaluasi dapat diperhatikan pada Gambar 4.



**Gambar 4.** Hasil Visualisasi

Pengujian pada akurasi model *Naïve Bayes Classifier* sebelumnya didapat hasil akurasi sebesar 70%, untuk mengetahui kebenaran dari nilai akurasi model Naïve Bayes Classifier maka akan dilakukan perhitungan akurasi manual menggunakan persamaan (5) sebagai berikut:

$$accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

$$= \frac{11 + 3}{11 + 3 + 5 + 1}$$

$$= \frac{14}{20} \times 100\%$$

$$= 70\%$$

Perhitungan dengan cara manual akan dilakukan oleh penelitian mendapatkan hasil yang sama dengan perhitungan melalui sistem, untuk mengetahui seberapa besar kemampuan model dan sistem pada klasifikasi

data dengan benar adalah mengetahui nilai akurasi suatu model. Model *Naïve Bayes Classifier* yang dibuat dalam penelitian ini membuktikan bahwa dapat mengklasifikasi sebuah data ulasan dengan benar dan baik dan tingkat keakurasiannya didapat sebesar 70%.

#### **KESIMPULAN**

Berdasarkan hasil dari penelitian Analisis Sentimen Terhadap Ulasan Aplikasi LinkAja Menggunakan Metode *Naïve Bayes Classifier* yang telah dilakukan oleh peneliti dapat disimpulkan bahwa Metode *Naïve Bayes Classifier* dapat memprediksi analisis sentimen dengan baik berdasarkan hasil dari sistem analisis yang telah dibuat. Hasil tingkat keakurasian pada 100 data sampel sebesar 70%. Dari hasil pelabelan secara manual untuk 100 data sampel terdapat 69 sentimen negatif dan 31 sentimen positif, dapat dilihat bahwa kebanyakan ulasan bernilai negatif yang berisi keluhan masyarakat untuk transaksi.

#### **DAFTAR PUSTAKA**

- Darwis, D., Siskawati, N., & Abidin, Z. (n.d.). *Penerapan Algoritma Naive Bayes untuk Analisis Sentimen Review Data Twitter BMKG Nasional*. 15(1), 131–145.
- Ivan, Arkaan, N. R., Venir, K. R., Ananda, A. S., & A., B. K. (2018). Fintech Indonesia 101. In *Digits Insight 3* (pp. 63–67).
- Kristiawan, K., & Widjaja, A. (2021). Perbandingan Algoritma Machine Learning dalam Menilai Sebuah Lokasi Toko Ritel. *Jurnal Teknik Informatika Dan Sistem Informasi*, 7(1), 35–46. <https://doi.org/10.28932/jutisi.v7i1.3182>
- Kurniawan, A., & Adinugroho, S. (2019). Analisis Sentimen Opini Film Menggunakan Metode *Naïve Bayes* dan *Lexicon Based Features*. *Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputer*, 3(9), 8335–8342.
- Putra, A. I., & Santika, R. R. (2020). Implementasi Machine Learning dalam Penentuan Rekomendasi Musik dengan Metode *Content-Based Filtering*. *Edumatic : Jurnal Pendidikan Informatika*, 4(1), 121–130. <https://doi.org/10.29408/edumatic.v4i1.2162>
- Rachman, R., & Handayani, R. N. (2021). Klasifikasi Algoritma *Naive Bayes* Dalam Memprediksi Tingkat Kelancaran Pembayaran Sewa Teras UMKM. *Jurnal Informatika*, 8(2), 111–122. <https://doi.org/10.31294/ji.v8i2.10494>
- Restari, R. H., Sinurat, S., & Suginam, S. (2020). Rancangan Aplikasi Sistem Pakar Diagnosa Penyakit Mononukleosis Dengan Metode *Naive Bayes*. *JURIKOM (Jurnal Riset Komputer)*, 7(3), 403. <https://doi.org/10.30865/jurikom.v7i3.2179>
- Robison Manalu, D., Christofell, M., Tobing, L., & Yohanna, M. (2022). *METHOMIKA: Jurnal*

- Manajemen Informatika & Komputerisasi Akuntansi Analisis Sentimen Twitter Terhadap Wacana Penundaan Pemilu Dengan Metode Support Vector Machine*. 6(2), 149–156.
- Surohman, S., Aji, S., Rousyati, R., & Wati, F. F. (2020). Analisa Sentimen Terhadap Review Fintech Dengan Metode *Naive Bayes Classifier* Dan *K- Nearest Neighbor*. *EVOLUSI : Jurnal Sains Dan Manajemen*, 8(1), 93–105. <https://doi.org/10.31294/evolusi.v8i1.7535>
- Turjaman, R. M., & Budi, I. (2022). Analisis Sentimen Berbasis Aspek Marketing Mix Terhadap Ulasan Aplikasi Dompot Digital (Studi Kasus: Aplikasi Linkaja Pada Twitter). *Jurnal Darma Agung*, 30(2), 266. <https://doi.org/10.46930/ojsuda.v30i2.1672>
- Zhafira, D. F., Rahayudi, B., & Indriati, I. (2021). Analisis Sentimen Kebijakan Kampus Merdeka Menggunakan *Naive Bayes* dan Pembobotan *TF-IDF* Berdasarkan Komentar pada Youtube. *Jurnal Sistem Informasi, Teknologi Informasi, Dan Edukasi Sistem Informasi*, 2(1), 55–63. <https://doi.org/10.25126/justsi.v2i1.24>