KLIK: Kajian Ilmiah Informatika dan Komputer
ISSN 2723-3898 (Media Online)
Vol 4, No 4, Februari 2024, Hal 1962-1971
DOI 10.30865/klik.v4i4.1600
https://djournals.com/klik

Analisis Sentimen Ulasan Aplikasi Digital Korlantas POLRI Menggunakan Naïve Bayes pada Google Play Store

Syarif Hidayatulloh Wazir Putra, Dimas Febriawan*

Fakultas Teknologi Industri dan Informatika, Teknik Informatika, Universitas Muhammadiyah Prof. Dr. Hamka, Jakarta, Indonesia
Email: [1]hidayatullohwajir13@gmail.com, [2,*]dimas.febriawan@uhamka.ac.id
Email Penulis Korespondensi: dimas.febriawan@uhamka.ac.id

Abstrak−Digital Korlantas POLRI merupakan aplikasi yang dirilis pada tanggal, 13 April 2021 oleh Korlantas POLRI. Aplikasi ini bertujuan untuk memberikan pelayanan perpanjangan SIM kepada masyarakat secara digital. Aplikasi ini telah mencapai lebih 5 juta unduhan, dengan ulasan sebanyak 102 ribu dan rating sebesar 3,7. Ulasan-ulasan tersebut mencakup beragam komentar positif dan negatif. Banyak pengguna yang mengeluhkan terkait performa aplikasi tersebut. Maka dari itu dibutuhkannya, analisis sentimen dengan tujuan untuk mengklasifikasikan sentimen seseorang secara otomatis kedalam kelompok positif dan negatif. Penelitian ini menggunakan algoritma Naïve Bayes sebagai metode klasifikasi dengan dataset sebanyak 1500, terdiri dari 536 data sentimen positif dan 964 data sentimen negatif. Dalam proses pengolahan data menggunakan pembagian data dengan rasio 80:20, menghasilkan 1200 data latih dan 300 data uji. Hasil akhir dari penelitian ini adalah mendapat nilai accuracy, precision, dan recall yang bertujuan untuk mengevaluasi seberapa baik kinerja algoritma *Naïve Bayes* dalam proses klasifikasi. Nilai yang didapatkan dari evaluasi dengan menggunakan data sebanyak 300 data latih, menunjukkan nilai *accuracy* sebesar 70.33%, nilai *precision* sebesar 56.34%, dan nilai *recall* sebesar 74.77%. Hasil pengujian evaluasi menunjukkan bahwa algoritma *Naïve bayes* berhasil dalam proses klasifikasi.

Kata Kunci: Naïve Bayes; Klasifikasi; Analisis Sentimen; Digital Korlantas POLRI; Google Play Store

Abstract−Digital Korlantas POLRI is an application released on April 13, 2021 by Korlantas POLRI. This application aims to provide SIM extension services to the public digitally. This application has reached more than 5 million downloads, with 102 thousand reviews and a rating of 3.7. The reviews include a variety of positive and negative coments. Many users complained about the app's performance. Therefore, sentiment analysis is needed with the aim of utomatically classifying a person's sentiments into positive and negative groups. This research uses the Naïve Bayes algorithm as a classification method with a dataset of 1500, consisting of 536 positive sentiment data and 964 negative sentiment data. In a data processing process using data division with a ratio of 80:20, resulting in 1200 training data and 300 test data. The final result of this research is to get accuracy, precision and recall values which aim to evaluate how well the Naïve Bayes algorithm performs in the classification process. The values obtained from the evaluation by using data as much as 300 training data, shows an accuracy value of 70.33%, a precision value of 56.34%, and a recall value of 74.77%. The evalution test results show that the Naïve Bayes algorithm is successful in the classification process.

Keywords: Naïve Bayes; Classification; Sentiment Analysis; Digital Korlantas POLRI; Google Play Store

1. PENDAHULUAN

Pemerintah tengah berupaya mempercepat transformasi digital dan mengembangkan infrastruktur digital sebagai bagian dari program nasional yang mencakup hampir semua sektor industri [1]. Salah satunya adalah sektor digitalisasi layanan publik merupakan penyediaan layanan masyarakat yang memanfaatkan teknologi digital untuk memberikan pelayanan publik yang lebih unggul, efesien, dan efektif sesuai dengan kebutuhan masyarakat [2].

Kepolisian Republik Indonesia adalah bagian pemerintahan Indonesia. Menurut UU tahun 2002 no 2, kepolisian bertanggung jawab dalam menjaga keamanan serta ketertiban masyarakat, melakukan penegakan hukum, dan memberikan perlindungan, pengayoman, serta pelayanan kepada masyarakat [3]. Kepolisian Republik Indonesia juga bertanggung jawab dalam administrasi lalu lintas, termasuk penerbitan Buku Pemilik Kendaraan Bermotor (BPKB), layanan perpanjangan Surat Tanda Nomor Kendaraan (STNK) selama lima tahun, serta pembuatan dan perpanjangan Surat Izin Mengemudi (SIM).

Dalam rangka memberi layanan administratif kepada masyarakat, Korlantas POLRI meluncurkan aplikasi digital bernama Digital Korlantas POLRI pada tanggal, 13 April 2021. Aplikasi ini mencakup beberapa fungsi, termasuk SINAR (SIM Nasional Presisi). Aplikasi Digital Korlantas POLRI dapat di *download* pada *Google Play Store* [4].

Digital Korlantas POLRI untuk saat ini memiliki fokus utama dalam menyediakan layanan untuk perpanjangan SIM [5]. Di *Google Play Store,* aplikasi ini telah diunduh mencapai lima juta kali, mendapatkan ulasan sebanyak 102 ribu dan rating sebesar 3,7. Ulasan-ulasan tersebut mencakup beragam komentar positif dan negatif. Banyak pengguna yang mengeluhkan terkait performa aplikasi tersebut seperti aplikasinya suka error, kesulitan dalam proses verifikasi dan gagal terus dalam mengunngah berkas. Hasil ini, menunjukkan bahwa ada aspek-aspek yang perlu ditingkatkan dalam memenuhi ekspetasi pengguna terhadap layanan aplikasi tersebut. Oleh karena itu, analisis sentimen menjadi sebuah alat yang digunakan untuk secara otomatis mengklasifikasikan sentimen seseorang kedalam kelompok positif dan negatif.

Analisis sentimen merupakan cabang ilmu, bertujuan untuk mengkaji pandangan, sikap, perasaan, penilaian, dan evaluasi individu terhadap berbagai aspek seperti produk, individu, organisasi, isu, peristiwa, atau topik tertentu [6]. Analisis sentimen juga merupakan salah satu contoh yang paling poplar dibidang *Natural Language Processing* (NLP) yang merupakan cabang dari *Artifical Intellegence* [7]. Secara keseluruhan, terdapat lima langkah untuk melakukan analisis sentimen, yaitu mengumpulkan sejumlah data melalui dengan cara *scraping*, *preprocessing*, *feature selection*, *classification*, dan *evaluation* [8]. Ada beberapa metode klasifikasi data yang dapat digunakan seperti *K-Nearest Neighbors*, *Support Vector Machine*, dan *Naïve Bayes*.

Naïve Bayes merupakan salah satu teknik *supervised learning*, melibatkan proses pembelajaran dari data trainig. Metode ini umumnya digunakan dalam klasifikasi objek karena tingkat akurasi yang tinggi, kemudahan penggunaan, efesien, dan sederhana dalam pemahaman [9]. Metode ini termasuk dalam keluarga metode Bayes, dimana klasifikasi Bayesian bergantung pada penerapan Teorema Bayes. Teorema ini memainkan peran penting dalam menghitung atau memprediksi probabilitas suatu hipotesisyang dianalis untuk mencapai keputusan terbaik dalam statistik [10].

Dalam penelitian ini, metode *Naïve Bayes* digunakan untuk melakukan analisis sentimen terhadap ulasan aplikasi Digital Korlantas POLRI. Data didapatkan pada *Google Play Store* melalui teknik *web scraping* menggunakan *Google Colab*, dengan total 1500 data yang relevan. Kemudian, dataset tersebut disimpan dalam format csv dan dikelompokkan ke dalam dua kategori, yaitu positif dan negatif. Selanjutnya dataset akan diolah menggunakan aplikasi RapidMiner Studio.

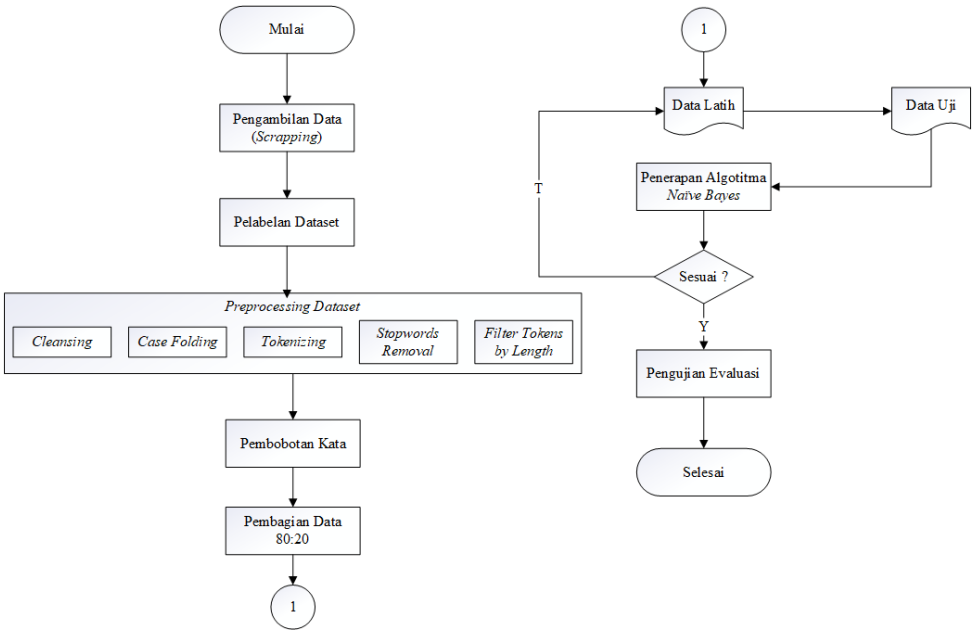
Melihat penelitian Siti Nurwahyuni yang membahas tentang analisis sentimen terhadap aplikasi Transportasi Online KRL Access, mendapatkan akurasi 84% dari 300 data ulasan berhasil diprediksi secara akurat menggunakan pendeketan *Naïve Bayes.* Selain itu juga, dapat efektif dalam mengklasifikasikan opini yang memiliki panjangnya sekitar satu paragraf [11]. Fauzan Setya Ananto juga melakukan penelitian dengan pendekatan yang sama terhadap ulasan aplikasi MyPertamina dengan temuan tersebut, mendapatkan data bersih sebanyak 1289 dengan akurasi 77,42%, presisi 49,98%, dan recal 76,87% diperoleh dengan menggunakan klasifikasi *Naïve Bayes*, dengan menggunakan metode *k-fold cross*, teknik *cross validation* untuk klasifikasi data [12]. Penelitian lainnya juga yang dilakukan oleh Firman Noor Hasan Terhadap Layanan Grab Indonesia Menggunakan klasifikasi *Multinominal Naïve Bayes.* Hasil dari penelitian, didapatkan dataset sejumlah 1000 data, dengan 911 sentimen positif dan 89 data sentimen negatif. Yang menghasilkan tingkat *accuracy* sebesar 92,5%. Hal ini mengindikasikan bahwa mayoritas pelanggan merasakan kepuasan terhadap layanan dan fasilitas yang diberikan oleh Grab Indonesia [10].

Perlu dilakukan perbandingan dengan metode lainnya untuk mengevaluasi efektifitas metode *Naïve Bayes* dalam analisis sentimen. Salah satu perbandingan tersebut terdapat dalam penelitian yang dilakukan oleh Lutfi Budi Ilmawan antara klasifikasi *Support Vector Machine* (SVM) dan *Naïve Bayes* untuk analisis sentimen terhadap ulasan Tekstual di *Play Store*. Akurasi yang berhasil dicapai oleh metode *SVM* mencapai 81,46%, sementara metode *Naïve Bayes* hanya mencapai 75,41%. Ini menunjukkan bahwa *SVM* lebih efesien dalam melakukan analisis sentimen dibandingkan dengan metode *Naïve Bayes* [13]. Selain itu, merujuk pada penelitian sebelumnya mengeni analisis sentimen aplikasi Digital Korlantas POLRI, tidak banyak ditemukan penelitian yang membahas mengenai metoode *Naïve Bayes* sebagai algoritma pengklasifikasian. Berdasarkan penelitian [5] menerapkan metode *Support Vector Machine* untuk menganalisis sentimen apikasi Digital Korlantas POLRI, menghasilkan sentimen positif 598 data dan sentimen negatif 511 data. Dengan rasio data 90:10, mencapai tingkat *accuracy* 0,82.

penelitian ini bertujuan untuk mengkategorikan ulasan terhadap aplikasi Digital Korlantas POLRI di *Google Play Store* ke dalam dua kategori, yaitu sentimen positif dan negatif. serta mengevaluasi sejauh mana efektivitas metode *Naïve Bayes* dalam melakukan analisis sentimen pada penelitian ini.

2. METODOLOGI PENELITIAN

Dalam penelitian ini, analisis sentimen terhadap Digital Korlantas POLRI dilakukan dengan menerapkan metode *Naïve Bayes*. Alur penelitian tersebut diilustrasikan pada Gambar 1.



Gambar 1. Alur Penelitian

2.1 Pengambilan Data

Tahap awal dimulai dengan mengambil data menggunakan *Google Colab*, yang uumnya disebut sebagai proses *scraping* data. Berikut adalah proses pengambilan data yang ditunjukkan pada Gambar 2.

Gambar 2. Proses Pengambilan Data

Pada gambar 2, terlihat data yang diambil dari ulasan komentar pengguna terkait aplikasi Digital Korlantas POLRI di *Platform Google Play Store*. Pengambilan data dilakukan menggunkan *Google Colab* dengan total 1500 ulasan yang relevan dengan aplikasi. Proses pengumpulan data melibatkan teknik *web scraping*, dan setelahnya, dataset disimpan dalam format csv.

2.2 Pelabelan Dataset

Setelah berhasil mengambil data melalui teknik *web scraping*, langkah berikutnya adalah melakukan analisis dengan memberikan label pada setiap ulasan pengguna, mengelompokkannya dalam dua kategori, yaitu sentimen positif dan negatif. Label sentimen positif diberikan kepada pengguna yang merasakan manfaat dari aplikasi, sementara label sentimen negatif diberikan kepada pengguna yang merasa kurang terbantu oleh aplikasi tersebut [14].

2.3 *Preprocessing* Dataset

Preprocessing dataset adalah tahapan yang bertujuan untuk menghapus kata-kata yang tidak relevan dengan objek penelitian dan mengorganisir data yang sebelumnya berantakan menjadi data yang terstruktur [15][16]. Berikut terdapat beberapa tahapan *preprocessing* dataset, yaitu:

a. *Cleansing*, merupakan tahapan untuk menghilangkan noise atau simbol.
b. *Case Folding*, adalah tahapan mengubah semua huruf kapital menjadi kecil (*lower case*).
c. *Tokenizing*, adalah proses memanfaatkan spasi dan tanda baca untuk membagi kalimat menjadi bentuk token.
d. *Stopwords Removal*, merupakan tahapan yang melibatkan penghapusan istilah yang sering digunakan atau tidak memberikan kontribusi nilai pada sistem.
e. *Filter Tokens* (*by Length*), adalah tahapan untuk menghilangkan kata-kata berdasarkan panjangnya, dengan rentang antara 4 hingga 25 karakter.

2.4 Pembobotan Kata

Pemberian bobot pada setiap kata adalah teknik yang sering digunakan dalam pencarian informasi. Teknik ini juga disebut sebagai TF-IDF (*Term Frequency – Inverse Document Frequency*). Metode ini memberikan bobot pada sebuah istilah atau kata dalam dokumen dan dianggap sebagai metode yang lebih sederhana, tepat, dan efesien [17]. Persamaan (1) dibawah ini menyajikan rumus TF-IDF sebagai berikut:

$$W_{(t.d)} = Wtf_{(t.d)} \times idf_t \quad (1)$$

Keteranga:
$W_{(t.d)}$: Bobot TF-IDF
$Wtf_{(t.d)}$: Bobot kata dalam setiap dokumen
idf_t : Bobot *Inverse Document Frequency* dari nilai (log (N/df))
N : Jumlah seluruh dokumen
df : Jumlah dokumen yang mengandung *term*

2.5 Pembagian Data

Dalam penelitian ini, data diklasifikasikan dengan menggunakan operator *Split* Data dengan pembagian rasio 80:20, dimana 80% data latih dan 20% data uji.

2.6 Penerapan Algoritma *Naïve Bayes*

Naïve Bayes adalah suatu teknik klasifikasi objek yang memanfaatkan probabilitas dan statistika dalam memprediksi peluang suatu kejadian atau kondisi tertentu dengan mengandalkan asumsi kuat. *Naïve Bayes* mampu memberikan suatu hasil klasifikasi sebuah opini secara akurat, termasuk opini yang terdiri dari beberapa kalimat baik yang bersifat positif maupun negatif [10]. Dalam Teorema Bayes, Probabilitas bersyarat diungkapkan melalui suatu persamaan [18] seperti persamaan (2) berikut:

$$P\ (H \mid X\ ) = \frac{P\ (X|H)P(H)}{P(X)} \quad (2)$$

Keterangan:

X	: *Tuple* atau objek data
H	: Hipotesis atau *tuple X* adalah kelas C
$P\ (H \mid X\ )$	: Probabilitas bahwa hipotesis *H* benar untuk *tuple X*
$P\ (X\|H)$	: Probabilitas bahwa *tuple X* berada dalam kelas *C*
$P(H)$	: Probabilitas bahwa hipotesis *H* benar untuk setiap *tuple*
$P(X)$	: Probabilitas *prior* dari *tuple X*

2.7 Pengujian Evaluasi

Dalam tahap ini dilakukan evaluasi berdasarkan *Confusion matrix*, merupakan matriks yang menghitung jumlah data uji yang berhasil dikelompokkan dengan benar dan yang dikelompokkan dengan salah, untuk menilai seberapa baik inerja model klasifikasi. Penerapan matriks ini dapat membantu dalam menilai sejauh mana kualitas kinerja dari model klasifikasi tersebut [19]. Dalam tahap ini, dilakukan evaluasi berdasarkan *confusion matrix,* yang melibatkan perhitungan *accuracy*, *precision,* dan *recall* [20]. Berikut adalah Tabel 1 *Confusion Matrix:*

Tabel 1. *Confusion Matrix*

Prediction Class	***Actual Class***	
	Negative	***Positive***
Predict Negative	TN	FN
Predict Positive	FP	TP

Berikut rumus perhitungan *confusion matrix*, untuk mendapatkan nilai *accuracy*, *precision*, dan *recall* sebagai berikut. Dapat dijelaskan melalui persamaan (3), (4), dan (5).

$$Accuracy = \frac{(TP+TN)}{(TP+TN+FN+FP)} \quad (3)$$

$$Precision = \frac{TP}{(TP+FP)} \quad (4)$$

$$Recall = \frac{TP}{(TP+FN)} \quad (5)$$

3. HASIL DAN PEMBAHASAN

3.1 Pengambilan Data

Data diambil dengan *scraping*, menggunakan sebuah *library google-play-scraper* pada *Google Colab*. Metode ini secara otomatis mengambil data ulasan aplikasi Digital Korlantas POLRI yang terdapat pada *website Google Play Store*.

```
[ ]  #scrape jumlah ulasan yang diinginkan
     from google_play_scraper import Sort, reviews

     result, continuation_token = reviews(
         'id.qoin.korlantas.user',
         lang='id',
         country='id',
         sort=Sort.MOST_RELEVANT,
         count=1500,
         filter_score_with=None
     )
```

Gambar 3. *Scraping* Dataset

Pada Gambar 3 merupakan parameter dari proses *scraping*, dalam tahap pengambilan data, *library google-play-scraper* akan melakukan *scraping* berdasarkan parameter yang dimasukkan. Peneliti memasukan codigan seperti "id.qoin.korlantas.user" untuk mengambil data dari aplikasi Digitl Korlantas POLRI, dengan Bahasa yang diambil pada teks hanya Bahasa Indonesia, pengambilan dataset dilakukan berdasarkan ulasan yang relevan terhadap aplikasi Digital Korlantas POLRI, dengan jumlah sebanyak 1500 ulasan.

	userName	score	at	content
1350	Agastya Narendra	1	2023-12-23 10:48:18	Kecewa, ngeri bgt proses di satpas lama, 3 min...
198	Ijun Boy	5	2023-12-23 10:21:49	Saya kasih bintang 5 karena saya samgat setuju...
204	Kaharudin, SM. MM	2	2023-12-23 09:13:12	hari ini saya mau perpanjang SIM A online, unt...
864	ade saeful adhar	5	2023-12-22 12:51:27	Aplikasi bagus,mudah dan cepat pelayanan nya
737	Dede Imam	3	2023-12-22 11:38:34	Semua tahap sudah di lalui tapi pas terakhir g...

Gambar 4. Hasil *web scraping*

Pada Gambar 4 menampilkan proses *web scraping* yang menghasilkan sebanyak 1500 dataset dengan ulasan yang berasal dari tanggal terbaru. Kolom pada gambar tersebut mencakup informasi seperti *username*, *score*, *at*, *content*.

```
[ ]  my_df.to_csv("data-ulasan_DigitalKrlantasPOLRI.csv", index = False)
```

Gambar 5. Proses Penyimpanan Dataset ke File CSV

Setelah berhasil mengumpulkan dataset dengan jumlah ulasan yang diinginkan, data disimpan dalam format csv, seperti Gambar 5 diatas.

3.2 Pelabelan Dataset

Pada bagian ini, pelabelan akan diterapkan pada data yang sudah diperoleh dari langkah sebelumnya. Proses pelabelan data dilakukan secara manual oleh peneliti dengan menggunakan aplikasi *Microsoft Excel* 2016. Pelabelan dataset ini dibagi menjadi dua kategori, yaitu sentimen positif dan negatif. Dapat dilihat pada Tabel 2.

Tabel 2. Hasil Proses Pelabelan

Text	Sentimen
Sampe sekarang belum bisa, Laporannya TERJADI KESALAHAN saat registrasi identitas di perpanjangan sim (unggah dokumen)	Negatif
Perpanjangan sim online sehari jadi, bisa langsung diambil di polresta yang dipilih. Kalau diantar / dipaketkan mungkin lebih lama prosesnya	Positif
Aplikasi bagus, mudah dan cepat pelayanan nya	Positif
Aplikasinya eror terus mau buat SIM dengan cpt dan mudah asik gagal terus sistemnya	Negatif
Mantap, perpanjang SIM lebih mudah & murah	Positif

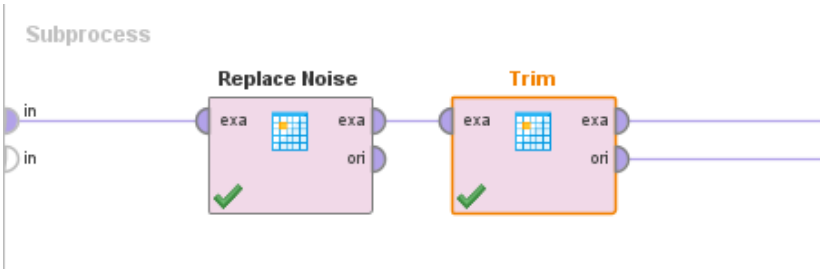
Pada Tabel 2 diatas memperlihatkan hasil dari proses pelabelan secara manual dengan total data sebanyak 1500, yang menghasilkan 536 data dengan sentimen positif dan 964 data dengan sentimen negatif.

3.3 *Preprocessing* Dataset

Setelah melalui tahap pengambilan data dan pelabelan, maka langkah berikutnya adalah *preprocessing* dataset menggunakan *software* RapidMiner. *Preprocessing* bertujuan untuk menyaring kata-kata yang tidak terstruktur menjadi data terstruktur. Tahapan ini terdiri atas lima langkah, yaitu *cleansing*, *case folding*, *tokenizing*, *stopwords removal*, dan *filter tokens by length.* Berikut adalah uraian dari tahapan *preprocessing* dataset sebagai berikut:

a. *Cleansing*

Proses *cleansing* dilakukan untuk menghilangkan noise atau simbol yang tidak diinginkan. Seperti icon [-!”#$%&^’()*+,./:;⇔?@\[\\\]_{|}~].



Gambar 6. Proses *Cleansing*

Pada Gambar 6, dalam tahapan ini, peneliti menggunakan operator *Replace* dan *Trim*, yang akan diintegrasikan ke dalam operator *Subprocess*. Operator *Replace Noise* berfungsi untuk menghilangkan noise atau simbol. Sedangkan operator *Trim* berfungsi untuk menghilangkan baris kosong.

Tabel 3. Hasil *Cleansing*

Text Sebelum	*Text Sesudah*
Sampe sekarang belum bisa, Laporannya TERJADI KESALAHAN saat registrasi identitas di perpanjangan sim (unggah dokumen)	Sampe sekarang belum bisa Laporannya TERJADI KESALAHAN saat registrasi identitas di perpanjangan sim unggah dokumen
Perpanjangan sim online sehari jadi, bisa langsung diambil di polresta yang dipilih. Kalau diantar / dipaketkan mungkin lebih lama prosesnya	Perpanjangan sim online sehari jadi bisa langsung diambil di polresta yang dipilih Kalau diantar dipaketkan mungkin lebih lama prosesnya

Pada Tabel 3 menunjukkan hasil dari proses *cleansing*. Perbedaan antara data sebelum diproses dan sesudah di proses terlihat jelas. Kolom kiri terdapat beberapa tanda baca seperti .,/(). Setelah melalui proses dengan operator *Replace* dan *Trim*, tanda baca tersebut berhasil dihilangkan, sebagaimana terlihat pada kolom kanan.

b. *Case Folding*

Setalah melakukan proses *cleansing*, langkah berikutnya adalah proses *case folding* guna mengubah semua huruf kapital menjadi kecil.

Tabel 4. Hasil *Case Folding*

Text Sebelum	*Text Sesudah*
Sampe sekarang belum bisa Laporannya TERJADI KESALAHAN saat registrasi identitas di perpanjangan sim unggah dokumen	sampe sekarang belum bisa laporannya terjadi kesalahan saat registrasi identitas di perpanjangan sim unggah dokumen
Perpanjangan sim online sehari jadi bisa langsung diambil di polresta yang dipilih Kalau diantar dipaketkan mungkin lebih lama prosesnya	perpanjangan sim online sehari jadi bisa langsung diambil di polresta yang dipilih kalau diantar dipaketkan mungkin lebih lama prosesnya

Tabel 4 menampilkan hasil dari tahap *case folding*. Dalam proses ini, operator *Transform Case* digunakan oleh peneliti. Pada kolom kiri, terdapat teks dengan huruf besar. Namun, setelah menjalani proses dengan operator *Transform Case*, teks yang sebelumnya berhuruf besar telah diubah menjadi huruf kecil, seperti yang terlihat pada kolom kanan.

c. *Tokenizing*

Setalah melakukan proses *case folding*, selanjutnya adalah prose *Tokenizing*, proses ini dilakukan untuk memecah satu kalimat menjadi bentuk token dengan menggunakan tanda baca dan spasi.

Tabel 5. Hasil *Tokenizing*

Text Sebelum	*Text Sesudah*
sampe sekarang belum bisa laporannya terjadi kesalahan saat registrasi identitas di perpanjangan sim unggah dokumen	sampe, sekarang, belum, bisa, laporannya, terjadi, kesalahan, saat, registrasi, identitas, di, perpanjangan, sim, unggah, dokumen
perpanjangan sim online sehari jadi bisa langsung diambil di polresta yang dipilih kalau diantar dipaketkan mungkin lebih lama prosesnya	perpanjangan, sim, online, sehari, jadi, bisa, langsung, diambil, di, polresta, yang, dipilih, kalau, diantar, dipaketkan, mungkin, lebih, lama, prosesnya

Tabel 5 menunjukkan hasil dari tahap *tokenizing*. Dalam proses ini, peneliti menggunakan operator *Tokenize*. Pada kolom kiri data masih berbentuk kalimat, tetapi setelah melalui proses *tokenizing*, kalimat tersebut dipecah menjadi kata-kata yang dipisahkan menggunakan tanda baca, sebagaimana terlihat pada kolom kanan.

d. *Stopwords Removal*

Setalah melakukan proses *tokenizing*, langkah berikutnya adalah prose *stopwords removal*, untuk menghapus istilah yang sering digunakan.

Tabel 6. Hasil *Stopwords Removal*

Text Sebelum	*Text Sesudah*
sampe, sekarang, belum, bisa, laporannya, terjadi, kesalahan, saat, registrasi, identitas, di, perpanjangan, sim, unggah, dokumen	sampe, laporannya, kesalahan, registrasi, identitas, perpanjangan, sim, unggah, dokumen
perpanjangan, sim, online, sehari, jadi, bisa, langsung, diambil, di, polresta, yang, dipilih, kalau, diantar, dipaketkan, mungkin, lebih, lama, prosesnya	perpanjangan, sim, online, sehari, langsung, diambil, polresta, dipilih, diantar, dipaketkan, prosesnya

Pada Tebel 6 menunjukkan hasil dari tahap *stopwords removal*. Dalam proses ini, peneliti menggunakan operator *Filter Stopwords (Dictionary)* dimana operator ini, menggunakan file *stopword* Bahasa Indonesia yang sudah dibuat. Perbedaan terlihat pada tabel tersebut, dimana kata-kata yang tidak memiliki makna telah dihilangkan.

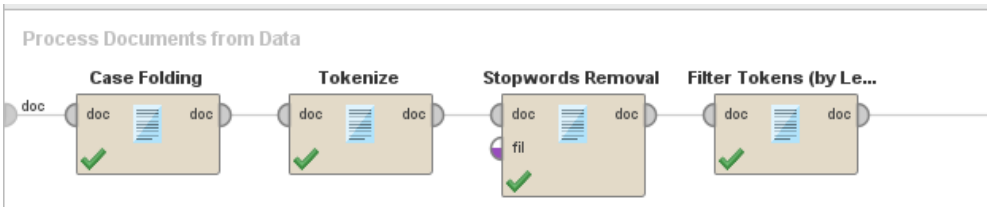
e. *Filter Tokens by Length*

Setalah melakukan proses *stopwords removal*, selanjutnya adalah prose *filter tokens by length*, proses dilakukan untuk menghilangkan kata-kata berdasarkan panjangnya.

Tabel 7. Hasil *Filter Tokens by Length*

Text Sebelum	*Text Sesudah*
sampe, laporannya, kesalahan, registrasi, identitas, perpanjangan, sim, unggah, dokumen	sampe, laporannya, kesalahan, registrasi, identitas, perpanjangan, unggah, dokumen
perpanjangan, sim, online, sehari, langsung, diambil, polresta, dipilih, diantar, dipaketkan, prosesnya	perpanjangan, online, sehari, langsung, diambil, polresta, dipilih, diantar, dipaketkan, prosesnya

Pada Tabel 7 merupakan hasil dari *filter tokens by length*. Dalaam proses ini, peneliti menggunakan operator *Filter Tokens (by Length)*. Yang akan menghapus kata-kata yang kurang dari minmal 4 karakter dan maksimal 25 karakter.

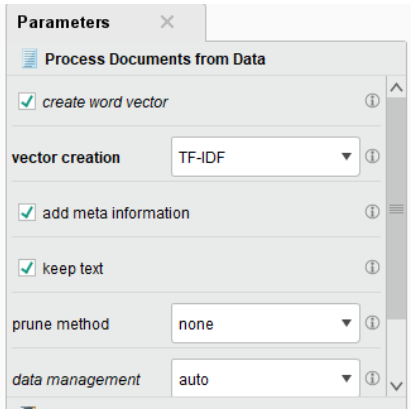


Gambar 7. Proses *Preprocessing*

Pada Gambar 7 merupakan proses *preprocessing*, di mana tahapan *case folding*, *tokenizing*, *stopwords removal*, dan *filter tokens by length* akan diintegrasikan ke dalam operator *Process Documents from Data*.

3.4 Pembobotan Kata

Setelah melewati tahap *preprocessing* dataset, langkah selanjutnya adalah melakukan pembobotan kata. Pada proses ini, setiap kata diberi bobot menggunakan metode TF-IDF. Dalam proses ini TF-IDF, tetap menggunakan operator *Process Documents from Data*.

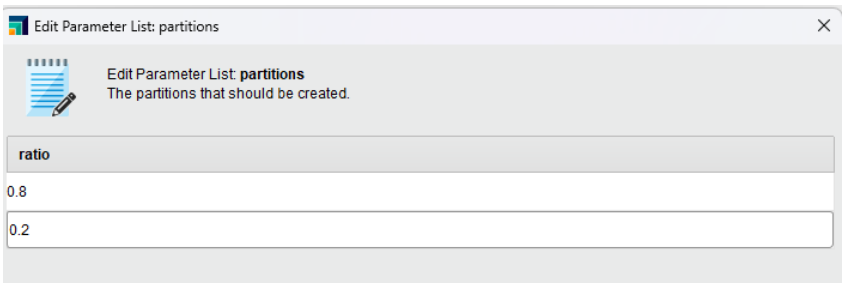


Gambar 8. Parameter TF-IDF

Gambar 8 diatas merupakan parameter dari operator *Process Documents from Data*. Terdapat opsi *create word vector* kemudian dicentang, dan untuk *vector creation* pilih TF-IDF.

3.5 Pembagian Data

Sebelum melanjutkan ke tahap penerapan algoritma *Naïve Bayes*, langkah selanjutnya adalah pembagian dataset ke dalam dua bagian, yakni data latih dan data uji. Proses ini, menggunakan operator *Split* Data.

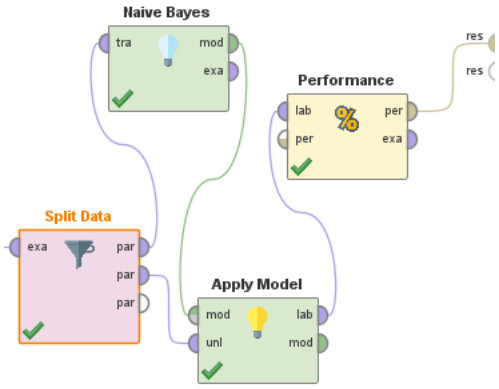


Gambar 9. Parameter *Split* Data

Gambar 9 merupakan parameter dari operator *Split* Data. Pada proses ini, pembagian rasio 80:20, mengahasilkan 80% data latih dan 20% data uji. Dengan pembagian tersebut, menghasilkan data latih 1200 dan data uji 300.

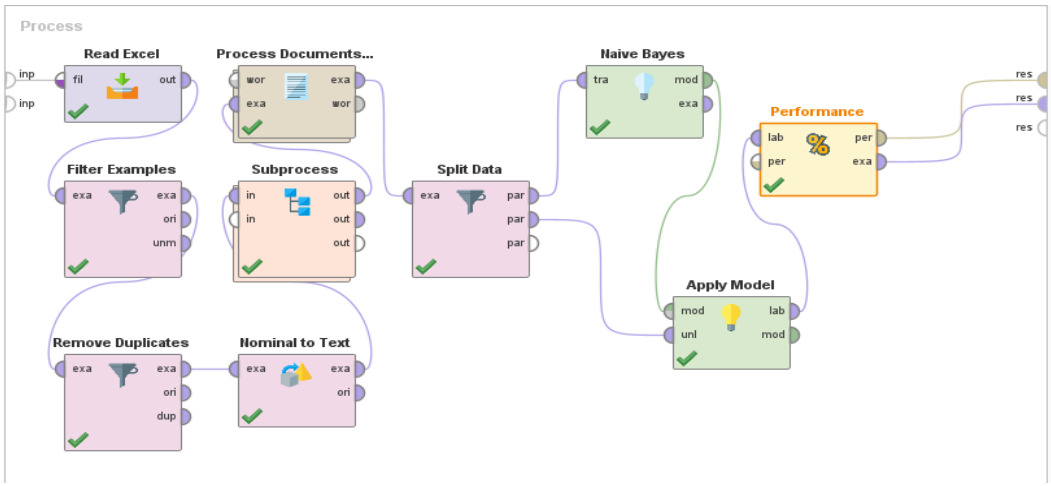
3.6 Penerapan Algoritma *Naïve Bayes*

Setelah melalui tahapan pembagian data. Langkah berikutnya adalah penerapan algoritma *Naïve Bayes*. Dalam proses ini, peneliti menggunakan operator *Naïve Bayes*, *Apply Model*, dan *Performance*.



Gambar 10. Penerapan Algoritma *Naïve Bayes*

Pada Gambar 10 merupakan penerapa algoritma *Naïve Bayes*. Dalam proses ini, data latih akan menjalani proses klasifikasi dengan menggunakan operator *Naïve Bayes*. Data latih yang telah diklasifikasikan akan diteruskan ke operator *Apply Model* untuk mendapatkan prediksi bersama dengan data uji. Operator *Performance* kemudian berfungsi untuk mengevaluasi model klasfikasi yang telah dibuat.



Gambar 11. Proses Keseluran dalam Analisis Sentimen

Gambar 11 menampilkan keseluruhan rangkaian proses analisis sentimen. Semua operator dihubungkan, dan setelah semua operator terrhubung, proses keseluruhan dijalankan dan menghasilkan *Output* seperti pada Gambar 12.

accuracy: 70.33%

	true Negatif	true Positif	class precision
pred. Negatif	131	27	82.91%
pred. Positif	62	80	56.34%
class recall	67.88%	74.77%	

Gambar 12. Hasil Penerapan Algoritma *Naïve Bayes*

3.7 Pengujian Evaluasi

Langkah terakhir dalam penelitian ini adalah tahap evaluasi, yang bertujuan untuk mengevaluasi seberapa baik kinerja algoritma *Naïve Bayes* dalam proses klasifikasi dan untuk mendapatkan nilai *accuracy*, *precision*, dan *recall*. Untuk memperoleh nilai-nilai tersebut, dilakukan menggunakan *confusion matrix*.

Tabel 8. Hasil *Performance Naïve Bayes*

Accuracy: 70.33%			
	True Negatif	***True Positif***	***Class Precision***
Pred. Negatif	131	27	82.91%
Pred. Positif	62	80	56.34%
Class Recall	67.88%	74.77%	

Pada proses pengujian evaluasi, digunakan sebanyak 300 data uji, terdiri dari 107 sentimen positif dan 193 sentimen negatif. Hasil evaluasi kinerja algoritma *Naïve Bayes* menunjukkan tingkat akurasi sebesar 70.33%, sebagaimana tercantum pada Tabel 8 diatas.

Berikut ini merupakan hasil perhitungan *confusion matrix* yang mendapatkan nilai *accuracy*, *precision*, dan *recall* sesuai dengan data yang diperoleh.

$$Accuracy = \frac{(TP+TN)}{(TP+TN+FN+FP)} = \frac{(80+131)}{(80+131+27+62)} = \frac{211}{300} = 0{,}70$$

$$Precision = \frac{TP}{(TP+FP)} = \frac{80}{(80+62)} = \frac{80}{142} = 0{,}56$$

$$Recall = \frac{TP}{(TP+FN)} = \frac{80}{(80+27)} = \frac{80}{107} = 0{,}74$$

Dari evaluasi dengan *confusion matrix*, didapatkan hasil *accuracy* sebesar 70.33%, *precision* sebesar 56.34%, dan *recall* sebesar 74.77%.

4. KESIMPULAN

Mengambil kesimpulan dari penelitian diatas, bahwa pengambilan data ulasan aplikasi Digital Korlantas POLRI dengan teknik *web scraping*, diperoleh dataset sebanyak 1500 dengan ulasan yang relevan. Setelah melakukan pelabelan secara manual pada dataset, menunjukkan adanya 536 data sentimen positif dan 964 data sentimen negatif. Sebelum penerapan algoritma *Naïve Bayes* dilakukan pembagian data terlebih dahulu, dengan rasio 80:20, yang menghasilkan 80% data latih dan 20% data uji. Dengan pembagian tersebut, dihasilkan data latih 1200 dan data uji 300. Dalam menggunakan algoritma *Naïve Bayes* untuk mengevaluasi kinerjanya, digunakan *confusion matrix* sebagai alat evaluasi. *Confusion matrix* digunakan untuk mendapatkan nilai *accuracy*, *precision*, dan *recall*. Tahap pengujian evaluasi, digunakan data sebanyak 300 data uji, terbagi atas 107 ulasan positif dan 193 ulasan negatif. Hasil dari evaluasi mendpatkan *accuracy* sebesar 70.33%, *precision* sebesar 56.34%, dan *recall* sebesar 74.77%. Dari hasil tersebut, dapat disimpukan bahwa banyak pengguna memberikan ulasan negatif, mencapai total 964 data, yang menunjukkan ketidakpuasan pengguna terhadap layanan aplikasi Digital Korlantas POLRI. Meskipun begitu, terdapat sejumlah pengguna yang merasakan manfaatnya dari aplikasi Digital Korlantas POLRI dan merasa belum sepenuhnya kecewa dengan layanan aplikasi tersebut. Disarankan untuk penelitian mendatang, dapat dipertimbangkan untuk melakukan pelabean secara otomatis guna meningkatkan efesien dalam penelitian. Selain itu, pengembangan penelitian dapat menggunakan metode klasifikasi lainnya seperti menggunakan metode *Deep Learning*, untuk perbandingan hasil nilai akurasinya.

REFERENCES

[1] W. Setiawan, "Era Digital dan Tantangannya," *Semin. Nas. Pendidik.*, pp. 1–9, 2017.

[2] Kominfo, "Pemerintah Kebut Digitalisasi Layanan Publik," *kominfo.go.id*, 2023. https://www.kominfo.go.id/content/detail/47280/pemerintah-kebut-digitalisasi-layanan-publik/0/artikel (accessed Nov. 14, 2023).

[3] Kepolisian, "KEPOLISIAN REPUBLIK INDONESIA," *kemenkeu.go.id*. https://jdih.kemenkeu.go.id/fulltext/2002/2TAHUN2002UU.htm (accessed Jan. 06, 2024).

[4] Digital Korlantas, "Digital Korlantas," *digitalkorlantas.id*, 2021. https://www.digitalkorlantas.id/ (accessed Nov. 14, 2023).

[5] N. R. Setiawan and E. R. Kaburuan, "Sentimen Analisis Review Aplikasi Digital Korlantas Pada Google Play Store Menggunakan Metode SVM," *J. Sisfokom (Sistem Inf. dan Komputer)*, vol. 12, no. 1, pp. 105–116, 2023, doi: 10.32736/sisfokom.v12i1.1614.

[6] H. Sibyan and N. Hasanah, "Analisis Sentimen Ulasan Pada Wisata Dieng Dengan Algoritma K-Nearest Neighbor (K-Nn)," *J. Penelit. dan Pengabdi. Kpd. Masy. UNSIQ*, vol. 9, no. 1, pp. 38–47, 2022, doi: 10.32699/ppkm.v9i1.2218.

[7] I. Saputra and D. A. Kristiyanti, "Machine Learning Untuk Pemula," *Inform. Bandung, Indones.*, 2022, Accessed: Nov. 14, 2023. [Online]. Available: https://elibrary.bsi.ac.id/readbook/220976/machine-learning-untuk-pemula

[8] A. P. Natasuwarna, "Seleksi Fitur Support Vector Machine pada Analisis Sentimen Keberlanjutan Pembelajaran Daring," *Techno.Com*, vol. 19, no. 4, pp. 437–448, 2020, doi: 10.33633/tc.v19i4.4044.

[9] M. P. Rahayu and Y. Farlina, "Penerapan Metode Naive Bayes Dalam Prediksi Penyebab Kecelakaan Kerja Cv. Deka Utama," *J. Larik Ldng. Artik. Ilmu Komput.*, vol. 1, no. 1, pp. 21–26, 2021, doi: 10.31294/larik.v1i1.472.

[10] F. N. Hasan and M. Dwijayanti, "Analisis Sentimen Ulasan Pelanggan Terhadap Layanan Grab Indonesia Menggunakan Multinominal Naïve Bayes Classifier," *J. Linguist. Komputasional*, vol. 4, no. 2, pp. 52–58, 2021, doi: https://doi.org/10.26418/jlk.v4i2.61.

[11] S. Nurwahyuni, "Online KRL access menggunakan metode Naive Bayes," *Swabumi*, vol. 7, no. 1, pp. 31–38, 2019.

[12] F. S. Ananto and F. N. Hasan, "Implementasi Algoritma Naïve Bayes Terhadap Analisis Sentimen Ulasan Aplikasi MyPertamina pada Google Play Store," *J. ICT Inf. Commun. Technol.*, vol. 23, no. 1, pp. 75–80, 2023, [Online]. Available: https://ejournal.ikmi.ac.id/index.php/jict-ikmi

[13] L. B. Ilmawan and M. A. Mude, "Perbandingan Metode Klasifikasi Support Vector Machine dan Naïve Bayes untuk Analisis Sentimen pada Ulasan Tekstual di Google Play Store," *Ilk. J. Ilm.*, vol. 12, no. 2, pp. 154–161, 2020, doi: 10.33096/ilkom.v12i2.597.154-161.

[14] E. P. Sutrisno and S. Amini, "IMPLEMENTASI ALGORITMA K-NEAREST NEIGHBOR PADA IMPLEMENTATION OF K-NEAREST NEIGHBOR ALGORITHM IN SENTIMENT ANALYSIS OF USER REVIEWS FOR DIGITAL," vol. 2, no. 2, pp. 687–695, 2023.

[15] F. Hashfi, D. Sugiarto, and I. Mardianto, "Sentiment Analysis of An Internet Provider Company Based on Twitter Using Support Vector Machine and Naïve Bayes Method," *Ultim. J. Tek. Inform.*, vol. 14, no. 1, pp. 1–6, 2022, doi: 10.31937/ti.v14i1.2384.

[16] Rahutomo;, I. Fahrur, and Haris;, "IMPLEMENTASI SUPPORT VECTOR MACHINE PADA ANALISA SENTIMEN TWITTER BERDASARKAN WAKTU," 2019. https://onesearch.id/Record/IOS4494.article-744/TOC (accessed Nov. 23, 2023).

[17] R. Melita, V. Amrizal, hendra bayu Suseno, and T. Dirjam, "Penerapan Metode Term Frequency Inverse Document Frequency (Tf-Idf) Dan Cosine Similarity Pada Sistem Temu Kembali Informasi Untuk Mengetahui Syarah Hadits Berbasis Web (Studi Kasus: Hadits Shahih Bukhari-Muslim)," *J. Tek. Inform.*, vol. 11, no. 2, pp. 149–164, 2018, doi: 10.15408/jti.v11i2.8623.

[18] D. Suyanto, "Data Mining untuk klasifikasi dan klasterisasi data," *Bandung Inform. Bandung*, 2017, Accessed: Nov. 14, 2023. [Online]. Available: https://opac.perpusnas.go.id/DetailOpac.aspx?id=1059041

[19] D. Normawati and S. A. Prayogi, "Implementasi Naïve Bayes Classifier Dan Confusion Matrix Pada Analisis Sentimen Berbasis Teks Pada Twitter," *J. Sains Komput. Inform. (J-SAKTI*, vol. 5, no. 2, pp. 697–711, 2021.

[20] R. Nadia, D. K. M. L, and F. Nhita, "Analisis Dan Implementasi Algoritma Naïve Bayes Classifier Terhadapa Pemilihan Gubernur Jawa Barat 2018 Pada Media Online," *e-Proceeding Eng.*, vol. 5, no. 1, pp. 1678–1700, 2018.