

ANALISIS SENTIMEN ULASAN APLIKASI BINAR PADA *GOOGLE PLAY STORE* MENGUNAKAN ALGORITMA *NAÏVE BAYES*

SENTIMENT ANALYSIS OF BINAR APPLICATION REVIEWS ON *GOOGLE PLAY STORE* USING *NAÏVE BAYES* ALGORITHM

Muhammad Raffi¹, Aries Suharso², Iqbal Maulana³

^{1,2,3}Universitas Singaperbangsa Karawang

1910631170032@student.unsika.ac.id

ABSTRACT

Binar is an online learning platform that provides courses and certifications in the digital field. The Binar app has been downloaded 500,000 times and has a *rating* of 3.6 on the *Google Play Store*. However, user *ratings* sometimes do not match their reviews. In application development, not only the number of *ratings* but also user opinions need to be considered. Therefore, developers must be able to interpret every opinion given, and sentiment analysis was conducted using the *Naïve Bayes Multinomial* and *Bernoulli* algorithms along with *Information Gain feature selection* to interpret user opinions. This study used the *Knowledge Discovery in Database* (KDD) method. The data used consisted of 713 reviews of the Binar app, including 518 positive and 195 negative reviews. The best results were obtained in the 9:1 data split scenario with the *Bernoulli Naïve Bayes* model achieving an *accuracy* of 93.06%, *precision* of 90.38%, *recall* of 100%, *f1-score* of 94.95%, and AUC of 0.996.

Keywords: *Sentiment Analysis, Multinomial Naïve Bayes, Bernoulli Naïve Bayes, Binar Academy*

ABSTRAK

Binar adalah platform pembelajaran *online* yang menyediakan kursus dan sertifikasi di bidang digital. Aplikasi Binar telah diunduh sebanyak 500 ribu kali dan memiliki *rating* 3,6 di *Google Play Store*. Dalam beberapa kasus terdapat *rating* yang tidak sesuai dengan isi ulasannya. Dalam pengembangan aplikasi, tidak hanya jumlah *rating* yang perlu diperhatikan, tetapi juga opini dari pengguna. Oleh karena itu, para pengembang harus dapat menginterpretasikan setiap opini yang diberikan, maka dilakukan analisis sentimen dengan menggunakan algoritma *Naïve Bayes Multinomial* dan *Bernoulli* serta *feature selection Information Gain* untuk menginterpretasikan opini pengguna. Metode yang digunakan dalam penelitian ini adalah *Knowledge Discovery in Database* (KDD). Data yang digunakan adalah 713 ulasan aplikasi Binar, dengan 518 ulasan positif dan 195 ulasan negatif. Pada skenario *split data* 9:1 dengan model *Bernoulli Naïve Bayes* menghasilkan performa terbaik dengan nilai *accuracy* 93,06%, *precision* 90,38%, *recall* 100%, *f1-score* 94,95%, dan AUC 0,996.

Kata Kunci: Analisis Sentimen, *Multinomial Naïve Bayes, Bernoulli Naïve Bayes, Binar Academy*

PENDAHULUAN

Penggunaan *smartphone* di Indonesia dipengaruhi oleh semakin pesatnya perkembangan teknologi. Menurut Pusat Data Ekonomi dan Bisnis Indonesia, Databoks-Katadata, persentase pengguna *smartphone* di Indonesia diperkirakan terus meningkat. Pada tahun 2015, hanya 28,6% populasi pengguna *smartphone* di Indonesia. Peningkatan jumlah pengguna *smartphone* mendorong pengembang aplikasi berinovasi dalam mengembangkan dan memasarkan aplikasi *mobile* sesuai kebutuhan, memperluas pasar, dan meningkatkan daya saing di industri teknologi. Hal ini menunjukkan pentingnya

adaptasi dan kreativitas dalam era digital. Pengguna dan pengembang mendapatkan kemudahan dalam mencari dan mendistribusikan aplikasi melalui berbagai platform distribusi aplikasi.

Salah satu aplikasi yang ada pada *Google Play Store* adalah aplikasi Binar. Aplikasi tersebut menawarkan berbagai konten pelatihan teknologi seperti kursus, video tutorial, artikel, dan sumber daya pembelajaran lainnya. Binar Academy, yang terkenal di Indonesia sebagai lembaga yang menyediakan kursus dan pelatihan di berbagai bidang teknologi seperti pemrograman, desain, dan *data science*. Sejak diluncurkan pada tanggal 18 Februari

2020 hingga bulan Maret 2023, aplikasi Binar telah berhasil meraih unduhan lebih dari 500 ribu kali di platform *Google Play Store* dan memperoleh *rating* sebesar 3,6.

Pada *Google Play Store* terdapat ulasan pada setiap aplikasi. Ulasan tersebut berisi opini pengguna mengenai kepuasan mereka terhadap aplikasi dan sering digunakan sebagai pertimbangan bagi calon pengguna dalam memilih aplikasi yang akan digunakan (Saputra dkk., 2019). Untuk mengetahui tingkat kepuasan pengguna, pihak pengembang menggunakan kolom ulasan yang disediakan pada *Google Play Store*. Dalam hal ini, kolom ulasan tersebut dimanfaatkan sebagai sumber informasi yang penting dan strategis bagi para pengembang dalam mengukur performa aplikasi yang dikembangkan dan memperbaiki layanan yang diberikan kepada pengguna.

Dalam beberapa kasus, terdapat beberapa *rating* yang tidak sesuai dengan isi ulasannya (Satria dkk., 2020). Dikarenakan banyaknya jumlah dan tidak terstrukturanya ulasan di *Google Play Store*, maka dibutuhkan suatu teknik untuk memahami sentimen pengguna aplikasi tersebut (Fransiska & Irham Gufroni, 2020). Dalam pengembangan aplikasi, tidak hanya jumlah *rating* yang perlu diperhatikan, tetapi juga isi komentar dari pengguna. Oleh karena itu, para pengembang harus dapat menginterpretasikan setiap komentar yang diberikan. Untuk membantu dalam menganalisis opini pengguna, analisis sentimen diperlukan sebagai teknik yang dapat digunakan untuk menganalisis sentimen. Analisis sentimen melibatkan evaluasi dan penilaian mengenai berbagai aspek, termasuk pendapat, sikap, emosi, dan penilaian terhadap entitas seperti produk, layanan, organisasi, individu, isu, peristiwa dan topik terkait (Liu, 2012). Analisis sentimen atau *opinion mining* merupakan jenis *Natural Language Processing* yang memproses teks untuk memahami perasaan atau opini masyarakat tentang suatu produk tertentu.

Beberapa metode klasifikasi telah dirancang untuk menganalisis sentimen dalam data berupa teks, salah satunya yaitu *Naïve Bayes*. Dalam pengolahan teks, terdapat dua jenis model klasifikasi *Naïve Bayes* yang dapat digunakan yaitu *Naïve Bayes Multinomial* dan *Bernoulli*. Model *Naïve Bayes Multinomial* menghitung frekuensi kemunculan setiap kata dalam dokumen, sementara model *Naïve Bayes Bernoulli* hanya memperhatikan apakah kata tersebut muncul pada dokumen atau tidak, tanpa memperhitungkan jumlah kemunculannya (Manning dkk., 2009). Penelitian terkait algoritma *Naïve Bayes* dilakukan oleh Wardani dkk. (2020) mengenai analisis sentimen menggunakan *Naïve Bayes Bernoulli* menghasilkan akurasi sebesar 90.19%, sementara untuk model *Multinomial* menghasilkan tingkat akurasi 93.45%. Selanjutnya, penelitian lainnya oleh Salsabila dkk. (2022) yang melakukan analisis sentimen ulasan aplikasi Tokopedia menggunakan algoritma *Naïve Bayes* menghasilkan akurasi sebesar 95,10%. Berdasarkan kedua penelitian tersebut metode *Naïve Bayes Bernoulli* dan *Multinomial* memiliki performa baik dalam klasifikasi sentimen. Secara umum, metode *Naïve Bayes* andal dalam klasifikasi teks dan dapat mencapai akurasi di atas 80%. Oleh karena itu, metode *Naïve Bayes* berpotensi menjadi metode yang andal dalam klasifikasi sentimen.

Dalam analisis sentimen, terdapat dua kelas yang dibedakan, yaitu positif dan negatif. Untuk melakukan hal tersebut, nilai setiap fitur dihitung. Namun, jumlah dimensi fitur yang tinggi dapat menyebabkan masalah dalam proses klasifikasi yang menjadi tidak efisien (Fanissa dkk., 2018). Oleh karena itu, *feature selection* diperlukan untuk mengurangi dimensi fitur dalam proses klasifikasi. Penggunaan *feature selection* juga dapat meningkatkan akurasi klasifikasi dari metode yang digunakan (Made, 2019). Dalam penelitian sebelumnya oleh Sari dkk. (2019) penggunaan *feature selection*

Information Gain dalam klasifikasi ulasan pengguna aplikasi dapat meningkatkan nilai akurasi model dan mendapatkan hasil akurasi sebesar 90,37%. Penelitian terkait *feature selection Information Gain* yang dikombinasikan dengan algoritma *Naïve Bayes* pada analisis sentimen oleh Isnanda dkk. (2021) performa *Naïve Bayes* dapat membantu meningkatkan performa sebesar 8% dengan hasil akhir akurasi sebesar 92%. Kedua penelitian tersebut menunjukkan bahwa *feature selection Information Gain* dapat meningkatkan akurasi model terutama ketika dikombinasikan dengan algoritma *Naïve Bayes*.

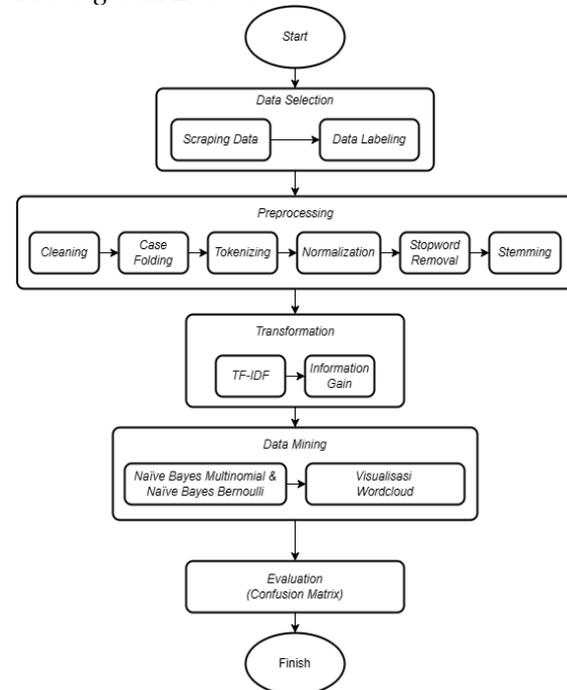
Penelitian ini dilakukan untuk membantu para pengembang dalam meningkatkan kualitas dan kepuasan pengguna dengan memanfaatkan data ulasan pengguna. Metode *Naïve Bayes Classifier* dipilih karena dianggap sederhana, cepat, dan memiliki performa yang tinggi dalam pengklasifikasian teks (Sari dkk., 2020). Metode *feature selection Information Gain* digunakan sebagai solusi untuk mengatasi masalah tingginya dimensi data. *Feature selection* merupakan suatu teknik untuk mengurangi tingginya dimensi atribut dalam data. Tujuan pengurangan dimensi data adalah untuk memilih atribut-atribut yang relevan dan mengurangi jumlah atribut sehingga dapat meningkatkan efisiensi proses klasifikasi dan dapat meningkatkan akurasi model (Aini dkk., 2018).

Penelitian ini akan menggunakan metode klasifikasi *Naïve Bayes Multinomial* dan *Bernoulli* untuk melakukan analisis sentimen pada ulasan pengguna aplikasi Binar di *Google Play Store*. PT. Lentera Bangsa Benderang (Binar Academy) memberikan dukungan yang positif terhadap penelitian ini dan berharap penelitian ini dapat berkontribusi dalam upaya meningkatkan kualitas dan layanan aplikasi Binar

METODE

Metode *Knowledge Discovery in Database* (KDD) digunakan dalam

penelitian ini karena memiliki keunggulan yaitu proses yang terstruktur untuk mengidentifikasi pola pada *dataset* yang kompleks sehingga memudahkan dalam pemahaman data (Ramos dkk., 2021). Penerapan metode KDD diperlukan karena data yang diambil memiliki unsur keragaman data serta tidak terstruktur data tersebut. Terdapat beberapa bahasa yang berbeda dan tidak baku dalam data ulasan tersebut, sehingga diperlukan tahapan *Preprocessing* data untuk mengubah data menjadi terstruktur dan mudah dipahami terutama data berbentuk teks yang perlu disesuaikan. Selain itu, penggunaan metode KDD juga dapat dikombinasikan dalam *feature selection* untuk membantu mempercepat proses analisis dan mengurangi jumlah data yang tidak relevan. Metode KDD terdiri dari 5 tahapan yaitu *Data Selection*, *Preprocessing*, *Transformation*, *Data Mining* dan *Evaluation*.



Gambar 1. Tahapan Penelitian

Penjelasan tahapan penelitian dengan metode KDD pada gambar 1 adalah sebagai berikut.

1. Data Selection

Pada tahapan ini akan dilakukan pengumpulan data, seleksi data dengan kriteria tertentu, dan memberi label pada setiap data ulasan oleh seorang ahli bahasa.

2. Preprocessing

Tahap ini merupakan awal pemrosesan dan dilakukan pembersihan data dari atribut yang tidak diperlukan untuk penelitian. Pembersihan data dapat mencakup deduplikasi data, pemeriksaan data, dan koreksi kesalahan dalam data. Berikut adalah 6 tahapan dalam *Preprocessing*:

- Cleaning*, merupakan tahapan pembersihan data dengan menghilangkan atribut yang kurang berpengaruh terhadap proses klasifikasi seperti huruf berulang lebih dari dua, *emoticon*, url, *hashtag*, dan simbol.
- Case Folding*, merupakan tahapan merubah data dengan mengubah huruf kapital menjadi huruf kecil yang bertujuan untuk menghindari case sensitive dari proses klasifikasi.
- Tokenizing*, merupakan tahapan memisahkan kalimat utuh menjadi pecahan kata per kata
- Normalization*, merupakan tahapan pengolahan atau perbaikan pada kata-kata yang tidak baku akibat dari kesalahan ketik, penyingkatan dan sebagainya menjadi kata baku.
- Stopword Removal*, merupakan tahapan menghilangkan kata-kata yang tidak bermakna dan tidak memiliki pengaruh signifikan dalam kalimat.
- Stemming*, merupakan tahapan mengubah suatu kata menjadi kata dasar dengan menghapus imbuhan dan akhiran tertentu.

3. Transformation

Tahap *Transformation* melibatkan perubahan bentuk data yang telah diproses sebelumnya menjadi format yang dapat digunakan dalam tahapan *Data Mining*. Dalam tahap ini, data akan dipecah antara data *training* dan data *testing* sesuai dengan skenario pengujian yang digunakan. Kemudian dilakukan pembobotan kata dengan *Term Frequency – Inverse Document Frequency* (TF-IDF). TF – IDF adalah sebuah algoritma transformasi data teks menjadi numerik.

TF-IDF adalah kombinasi dari dua teknik yaitu *Term Frequency* (TF) dan *Inverse Document Frequency* (IDF). Kedua teknik ini digunakan untuk mengatasi kelemahan dalam membuat prediksi yang relevan (Kulkarni & Shivananda, 2019). Persamaan untuk perhitungan TF-IDF menurut Verdikha dkk. (2018) sebagai berikut.

$$w_t = TF_{td} \times \log \left(\frac{N}{DF_t} \right)$$

Keterangan:

w_t = Bobot kata t terhadap dokumen d
 $TF_{t,d}$ = Banyaknya kata t pada dokumen d
 N = Jumlah dokumen keseluruhan
 DF_t = Jumlah dokumen yang memuat kata t

Selanjutnya dilakukan tahap *feature selection Information Gain* untuk memilih fitur paling relevan dalam mengklasifikasi data. Dengan menggunakan *Information Gain*, proses analisis data dapat dipercepat dengan mengurangi jumlah fitur yang tidak relevan dan meningkatkan akurasi model (Hasibuan & Marji, 2019). Persamaan untuk perhitungan *Information Gain* adalah sebagai berikut.

$$Entropy(S) = \sum_i^c -P_i \log_2 P_i$$

Keterangan:

c = Total nilai yang diperoleh dari kelas klasifikasi.

P_i = Total nilai sampel dari kelas i

Setelah memperoleh nilai *Entropy*, langkah perhitungan *Information Gain* dapat dilakukan dengan menggunakan persamaan sebagai berikut.

$$Gain(S,A) = Entropy(S) - \sum_{v \in Values(A)} \frac{|S_v|}{|S|} Entropy(S_v)$$

Keterangan:

$Gain(S,A)$: nilai *Gain* dari fitur

A : fitur

v : kemungkinan nilai fitur A

$Values(A)$: kemungkinan nilai

himpunan A

S_v : jumlah contoh nilai dari v

$Entropy(S_v)$: Entropy contoh nilai v

Setelah melakukan perhitungan, nilai *Gain* untuk setiap fitur akan diperoleh. Fitur-fitur yang memiliki nilai *Gain* tinggi dan relevan akan dipilih sebagai fitur data latih baru untuk diproses.

4. Data Mining

Pada tahapan *Data Mining*, tujuan penelitian akan ditentukan, seperti melakukan regresi, pengelompokan, klasifikasi, dan jenis analisis lainnya. Dalam penelitian ini, dilakukan klasifikasi dengan menggunakan algoritma *Naïve Bayes Multinomial* dan *Bernoulli*.

a. Multinomial Naïve Bayes

Multinomial Naïve Bayes merupakan sebuah metode klasifikasi pada pengolahan teks yang bekerja dengan cara menghitung frekuensi setiap *term* pada dokumen yang digunakan sebagai fitur untuk proses klasifikasi (Fanissa dkk., 2018). Jenis pengklasifikasi ini bekerja dengan baik pada data diskrit. Metode *Multinomial* dirancang untuk menentukan frekuensi dari kata (*term frequency*) yaitu berapa kali suatu kata muncul dalam suatu dokumen. Suatu istilah/kata mungkin sangat penting dalam menentukan sentimen, menjadikan metode ini sebagai salah satu pilihan klasifikasi pada analisis sentimen (Singh dkk., 2019). Dalam *Naïve Bayes Multinomial*, urutan kejadian kata dalam dokumen diabaikan, sehingga dokumen dianggap sebagai *bag of words*. Setiap kata diproses menggunakan distribusi *Multinomial*. Secara umum, persamaan untuk *Naïve Bayes Multinomial* adalah sebagai berikut (Fanissa dkk., 2018):

$$P(c|d) = P(c) \prod_{i=1}^n P(w_i|c)$$

Keterangan:

$P(c|d)$: peluang kelas c berdasarkan dokumen d

n : jumlah seluruh kata pada dokumen

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

Keterangan:

$P(c)$: peluang kelas c

c : kelas

N_c : jumlah dokumen kelas c

N : jumlah seluruh dokumen

$$P(w_i|c) = \frac{\text{count}(w_i|c) + 1}{\text{count}(c) + |V|}$$

Keterangan:

$P(w_i|c)$: peluang kata ke- i pada kelas c

$\text{count}(w_i|c)$: jumlah kata ke- i pada kelas c

$\text{count}(c)$: jumlah seluruh kata pada kelas c

$|V|$: jumlah kata unik pada semua kelas

b. Bernoulli Naïve Bayes

Bernoulli Naïve Bayes, juga dikenal sebagai *Multivariate Bernoulli Naïve Bayes* adalah teknik statistik yang menggunakan hitungan *biner*. Ini menghasilkan nilai *Boolean* sebagai keluarannya. Jenis pengklasifikasi ini bekerja dengan data diskrit (Chingmuankim, 2022). Model Klasifikasi *Bernoulli Naïve Bayes* adalah suatu metode yang menggunakan vektor atribut *biner* untuk merepresentasikan dokumen dan menunjukkan kemunculan serta ketidakhadiran kata dalam dokumen. Pada metode *Bernoulli Naïve Bayes*, kata (*term*) akan diberi nilai 1 apabila terdapat pada dokumen dan bernilai 0 apabila tidak terdapat pada dokumen seperti terdapat pada persamaan sebagai berikut (Wardani dkk., 2020).

$$d = \langle e_1, e_2, \dots, e_i, \dots, e_M \rangle \in \{0,1\}$$

Keterangan:

d : kumpulan nilai *biner* yang menunjukkan untuk setiap *term* itu muncul atau tidak dalam dokumen

M : banyaknya fitur

Peluang dokumen d terdapat dalam kelas c dihitung menggunakan persamaan sebagai berikut (Manning dkk., 2009).

$$P(c|d) \propto P(c) \prod_{\substack{1 \leq i \leq M \\ e_i=1}} P(t_i|c) \prod_{\substack{1 \leq i \leq M \\ e_i=0}} (1 - P(t_i|c))$$

Keterangan:

$P(t_i|c)$: peluang kemunculan kata (*term*) ke- i (t_i) terdapat dalam dokumen kelas c dalam posisi apapun dan kemungkinan berkali-kali.

$P(c)$: peluang sebuah dokumen muncul di kelas c .

Penduga dari $P(c)$ dihitung dengan menggunakan persamaan sebagai berikut (Manning dkk., 2009).

$$\hat{P}(c) = \frac{N_c}{N}$$

dan penduga dari $P(t_i|c)$ dihitung dengan menggunakan persamaan sebagai berikut (Manning dkk., 2009).

$$\hat{P}(t_i|c) = \frac{N_{ct}}{N_c}$$

Keterangan:

N_c : jumlah dokumen dengan kelas c pada data latih.

N_{ct} : banyaknya dokumen yang memuat *term* t dalam dokumen kelas c pada data latih.

N : jumlah total dokumen data latih.

Metode *Bernoulli* ini juga memerlukan *laplace smoothing* seperti metode *Multinomial* untuk mengatasi perhitungan peluang bernilai 0 yang diakibatkan oleh adanya kata pada data uji yang tidak terdapat pada data latih. Peluang tersebut dihitung menggunakan Persamaan (2.9) (Manning dkk., 2009).

$$\hat{P}(t_i|c) = \frac{N_{ct} + 1}{N_c + 2}$$

Kelas yang paling mungkin atau *maximum a posteriori* (MAP) untuk *Bernoulli Naïve Bayes* didefinisikan pada persamaan sebagai berikut dengan C adalah himpunan {positif, negatif} (Manning dkk., 2009).

$$c_{map} = \underset{c \in C}{\operatorname{argmax}} \hat{P}(c) \prod_{\substack{1 \leq i \leq M \\ e_i=1}} P(t_i|c) \prod_{\substack{1 \leq i \leq M \\ e_i=0}} (1 - P(t_i|c))$$

5. Evaluation

Pada tahapan *Evaluation* dilakukan evaluasi terhadap model yang telah dibuat untuk mengukur akurasi yang dihasilkan saat melakukan klasifikasi dengan algoritma *Naïve Bayes Multinomial* dan *Bernoulli*. Pada tahap evaluasi ini digunakan *Confusion Matrix* untuk menghitung nilai TP, FP, TN dan FN. Setelah dilakukan evaluasi menggunakan *Confusion Matrix*, maka dapat dihitung

beberapa metrik pengujian untuk mengukur performa model yang telah dibuat. Metrik-metrik tersebut meliputi *accuracy*, *precision*, *recall*, *f1-score*, dan AUC, yang memberikan informasi tentang seberapa baik model dapat melakukan klasifikasi pada data yang telah diberikan.

Confusion Matrix adalah suatu metode evaluasi untuk model klasifikasi berbentuk matriks yang digunakan untuk melakukan evaluasi terhadap suatu model klasifikasi. *Confusion Matrix* berisi informasi yang membandingkan hasil klasifikasi oleh model dengan hasil yang sebenarnya (Hadianto dkk., 2019). Terdapat 4 macam kejadian dalam *Confusion Matrix* yaitu:

1. TP (*True Positive*): data yang berhasil diklasifikasi ke dalam kelas positif.
2. TN (*True Negative*): data yang berhasil diklasifikasi ke dalam kelas negatif.
3. FP (*False Positive*): data yang tidak berhasil diklasifikasi ke dalam kelas positif.
4. FN (*False Negative*): data yang tidak berhasil diklasifikasi ke dalam kelas negatif.

Dengan nilai yang didapatkan pada *Confusion Matrix* maka dapat dihitung metrik-metrik pengujian sebagai berikut.

1. *Accuracy*, mengacu pada seberapa tepat model dalam melakukan klasifikasi data. *accuracy* dapat dihitung menggunakan persamaan sebagai berikut.

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

2. *Precision*, merupakan tingkat ketepatan hasil prediksi benar yang diinginkan oleh pengguna dengan hasil prediksi yang diberikan oleh suatu model. Perhitungan *precision* dapat dilakukan dengan persamaan sebagai berikut.

$$precision = \frac{TP}{TP + FP}$$

3. *Recall*, merupakan tingkat ketepatan suatu model dalam memprediksi data kelas positif berdasarkan keseluruhan data dengan nilai aktual positif. Perhitungan *recall* dapat dilakukan dengan persamaan sebagai berikut.

$$recall = \frac{TP}{TP + FN}$$

4. *F1-score*, merupakan sebuah nilai yang dihitung dari perbandingan rata-rata antara *precision* dan *recall*. Perhitungan *f1-score* dapat dilakukan dengan persamaan sebagai berikut.

$$f1 - score = 2 \times \frac{(Precision \times Recall)}{(Precision + Recall)}$$

5. AUC (*Area Under the Curve*) merupakan bagian dari luasan persegi yang dihasilkan dari grafik kurva ROC (*Receiver Operating Characteristic*), dengan nilai selalu berada di antara 0,0 dan 1,0. Kurva ROC didapat dari grafik yang menggunakan *False Positive (FP) Rate* pada sumbu x (horizontal) dan *True Positive (TP) Rate* pada sumbu y (vertikal). Rumus perhitungannya adalah sebagai berikut (Gorunescu, 2010):

$$TP Rate = \frac{TP}{TP + FN}$$

$$FP Rate = \frac{FP}{TP + FN}$$

HASIL DAN PEMBAHASAN

Hasil dari penelitian ini melibatkan analisis sentimen dengan mengklasifikasikan ulasan pengguna aplikasi Binar di *Google Play Store* menggunakan algoritma *Naïve Bayes Multinomial* dan *Bernoulli*, serta menerapkan *feature selection Information Gain*. Hasil klasifikasi dengan dua model *Naïve Bayes* tersebut dievaluasi dengan menggunakan *Confusion Matrix* dan beberapa metrik pengujian yang digunakan yaitu *accuracy*, *precision*, *recall*, *f1-score*, dan AUC untuk mengukur kualitas model *Naïve Bayes* mana yang lebih baik. Kemudian hasil klasifikasi tersebut akan divisualisasikan ke dalam *Wordcloud* sementara itu hasil akurasi perbandingan kedua model *Naïve Bayes* ditampilkan dalam bentuk tabel dan diagram perbandingan.

1. Data Selection

Tahapan pertama ini dilakukan pengumpulan data ulasan aplikasi Binar pada situs *Google Play Store* dengan

menggunakan teknik *Scraping*, dengan menggunakan bantuan *library google-play-scraper* dengan ketentuan data ulasan berbahasa Indonesia dan berasal dari negara Indonesia. Atribut yang diambil pada tahapan pengumpulan data ini yaitu *at* dan *content*. Data tersebut tersimpan dalam format *.csv*. Jumlah data ulasan pengguna aplikasi Binar yang diperoleh dari proses *Scraping* yang dilakukan pada tanggal 16 April 2023 adalah sebanyak 800 data.

Setelah data ulasan terkumpul maka dilakukan penyeleksian data yang akan digunakan. Atribut yang akan digunakan hanya *content* yang berisi data ulasan para pengguna terhadap aplikasi Binar. Data diseleksi dengan beberapa kriteria yang ditentukan yaitu data ulasan berbahasa Indonesia, tidak bersifat netral, dan memiliki relevansi dengan aplikasi yang diteliti. Jumlah data ulasan yang terseleksi sebesar 713 data.

Dalam membantu proses klasifikasi diperlukan sebuah label atau kelas yang dapat membantu algoritma dalam mengenali kelas tersebut. Dikarenakan *dataset* yang digunakan masih mentah dan tidak adanya label maka diperlukan ahli tata bahasa yang dapat mengenali kelas sentimen serta memberikan label ulasan tersebut. Beberapa ulasan dari hasil pelabelan data ditunjukkan pada tabel 1.

Tabel 1. Hasil Pelabelan Data

Ulasan	Label
Aplikasi yg oke banget	POSITIF
Jangan kasi iklan lagi ganggu	NEGATIF
Ui nya cakep!! Appnya jg gampang digunakan dan bermanfaat pol!	POSITIF
Ketika melihat gambar, lalu close atau back malah kembali ke awal materi. Tolong diperbaiki min. Terima kasih.	NEGATIF

Berdasarkan hasil pelabelan dan validasi data yang telah dilakukan oleh ahli tata bahasa Indonesia yaitu Aji Albaruddin, S.Pd. yang merupakan seorang lulusan

program studi Pendidikan Bahasa dan Sastra Indonesia dari Universitas Pasundan diperoleh jumlah ulasan hasil pelabelan yang ditunjukkan pada tabel 1.

Tabel 2. Hasil Data Selection

Label Ulasan	Jumlah Ulasan
Positif	518
Negatif	195
Total	713

2. Preprocessing

Tahapan ini mencakup 6 tahapan untuk mempersiapkan data untuk diproses pada tahapan *Data Mining*. *Preprocessing* dilakukan dengan kebutuhan pengolahan data untuk meningkatkan kualitas model dalam melakukan klasifikasi pada tahapan *Data Mining*. Dibawah ini merupakan penjelasan tahapan di dalam *Preprocessing* menggunakan bahasa pemrograman *Python* pada *Google Colab*. Hasil dari 6 tahapan *Preprocessing* dapat dilihat pada tabel 3.

Tabel 3. Hasil Preprocessing

Tahap	Hasil
Data Ulasan	Ui nya cakep!! Appnya jg gampang digunakan dan bermanfaat pol!
<i>Cleaning</i>	Ui nya cakep Appnya jg gampang digunakan dan bermanfaat pol
<i>Case Folding</i>	ui nya cakep appnya jg gampang digunakan dan bermanfaat pol
<i>Tokenizing</i>	['ui', 'nya', 'cakep', 'appnya', 'jg', 'gampang', 'digunakan', 'dan', 'bermanfaat', 'pol']
<i>Normalization</i>	['ui', 'nya', 'keren', 'aplikasi', 'juga', 'gampang', 'digunakan', 'dan', 'bermanfaat', 'pol']
<i>Stopword Removal</i>	['ui', 'nya', 'keren', 'aplikasi', 'gampang', 'digunakan', 'bermanfaat']
<i>Stemming</i>	['ui', 'nya', 'keren', 'aplikasi', 'gampang', 'guna', 'manfaat']

3. Transformation

Tahap *Transformation* dimulai dengan *split data* (pembagian data) ke dalam dua kelompok yaitu data *training* dan data *testing*. Untuk melakukan *split data* tersebut, ada tiga skenario pengujian

yang digunakan dengan rasio perbandingan yang berbeda antara data *training* dan data *testing*. Ketiga skenario tersebut adalah 9:1, 8:2 dan 7:3.

Dataset dibagi menjadi data *training* dan data *testing* menggunakan fungsi *train_test_split()* pada *library scikit-learn* dengan parameter *random_state=42*. Jumlah data ulasan pada tiga skenario pengujian dari hasil proses *split data* dapat dilihat pada tabel 4.

Tabel 4. Hasil Split data Ulasan

Skenario Rasio Perbandingan	Data Training	Data Testing
9:1	641	72
8:2	570	143
7:3	499	214

Setelah dilakukan *split data*, maka selanjutnya dilakukan proses ekstraksi fitur yaitu dengan melakukan pembobotan kata menggunakan TF-IDF dan *feature selection Information Gain* (IG). Langkah pertama yang dilakukan adalah dengan mengubah fitur pada data *training* dan data *testing* diubah menjadi representasi vektor menggunakan metode TF-IDF. Proses ekstraksi fitur ini dilakukan dengan menggunakan *library scikit-learn* pada fungsi *TfidfVectorizer()*. Selanjutnya, dilakukan *feature selection Information Gain* (IG) dengan menggunakan fungsi *SelectKBest()* pada *library scikit-learn*. Proses *feature selection* dilakukan dengan memilih fitur yang memiliki skor lebih besar dari *threshold* 0.001. Jumlah fitur pada hasil proses TF-IDF dan IG dapat dilihat pada tabel 5.

Tabel 5. Jumlah Fitur Hasil TF-IDF dan IG

Skenario Rasio Perbandingan	TF-IDF	IG
9:1	1114	613
8:2	1054	678
7:3	983	621

Dari hasil tersebut dapat disimpulkan bahwa semakin banyak data *training* yang digunakan, semakin banyak fitur yang dihasilkan menggunakan metode TF-IDF. Namun, dengan melakukan *feature selection* menggunakan metode *Information Gain*, jumlah fitur dapat ditekan sehingga tidak terlalu banyak. Salah satu hasil dari ekstraksi dan *feature*

selection pada skenario 1 (9:1) dapat dilihat pada gambar 1.

Gambar 2. Hasil TF-IDF dan IG

Gambar 1 menunjukkan hasil dari proses ekstraksi dan *feature selection* menggunakan metode TF-IDF dan *Information Gain*. Setelah itu, kata-kata tersebut diurutkan berdasarkan abjad dan dijadikan sebuah matriks yang berisi nilai kemungkinan suatu kata tersebut muncul dalam suatu dokumen. Matriks tersebut kemudian dapat digunakan sebagai fitur dalam proses klasifikasi dokumen.

4. Data Mining

Dalam tahap ini, dilakukan klasifikasi sentimen terhadap ulasan aplikasi Binar menggunakan dua model algoritma *Naïve Bayes*, yaitu model *Multinomial* dan *Bernoulli*. Proses implementasi algoritma *Naïve Bayes Multinomial* dan *Bernoulli* ini dilakukan dengan tiga skenario pengujian yang telah dijelaskan pada tahapan *Transformation*. Salah satu hasil dari implementasi algoritma *Naïve Bayes Multinomial* dan *Bernoulli* pada skenario 1 dapat dilihat pada gambar 3.

Gambar 3. Hasil Klasifikasi Ulasan Skenario 1

Gambar 3 menunjukkan hasil klasifikasi menggunakan algoritma *Naïve Bayes Multinomial* dan *Bernoulli*. Hasil klasifikasi *Naïve Bayes Multinomial* terletak di kolom *nb_Multinomial*, sedangkan hasil klasifikasi *Naïve Bayes Bernoulli* terletak di kolom *nb_Bernoulli*.

5. Evaluation

Tahap evaluasi merupakan tahap untuk melakukan evaluasi terhadap performa dari implementasi algoritma

Naïve Bayes Multinomial dan *Bernoulli* dalam melakukan klasifikasi dengan menggunakan *Confusion Matrix*. Metrik yang digunakan adalah TP (*True Positive*), TN (*True Negative*), FP (*False Positive*), FN (*False Negative*), *accuracy*, *precision*, *recall*, *f1-score* dan AUC. Hasil evaluasi dari tiga skenario pengujian dapat dilihat pada gambar 5, 6 dan 7.

Tabel 6. Hasil Pengujian Skenario 1 (9:1)

Metrik	Multinomial Naïve Bayes	Bernoulli Naïve Bayes
TP	47	47
FP	0	0
TN	18	20
FN	7	5
Accuracy	90,28%	93,06%
Precision	87,04%	90,38%
Recall	100%	100%
F1-score	93,07%	94,95%
AUC	0,988	0,996

Tabel 7. Hasil Pengujian Skenario 2 (8:2)

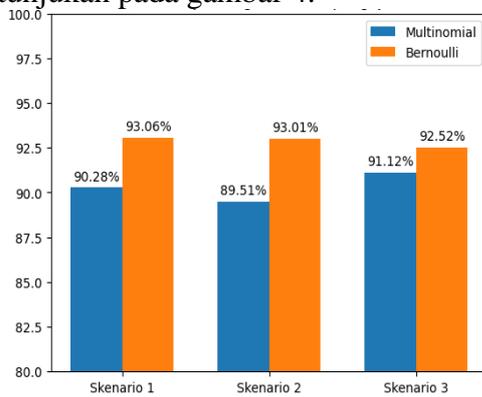
Metrik	Multinomial Naïve Bayes	Bernoulli Naïve Bayes
TP	106	106
FP	0	0
TN	22	27
FN	15	10
Accuracy	89,51%	93,01%
Precision	87,6%	91,38%
Recall	100%	100%
F1-score	93,39%	95,5%
AUC	0,995	0,995

Tabel 8. Hasil Pengujian Skenario 3 (7:3)

Metrik	Multinomial Naïve Bayes	Bernoulli Naïve Bayes
TP	162	161
FP	0	1
TN	33	37
FN	19	15
Accuracy	91,12%	92,52%
Precision	89,5%	91,48%
Recall	100%	99,38%
F1-score	94,46%	95,27%
AUC	0,99	0,983

Pada tabel 5, 6 dan 7 dapat dilihat bahwa *Naïve Bayes Multinomial* dan *Bernoulli* memberikan hasil yang cukup baik dalam melakukan klasifikasi pada ketiga skenario, dengan nilai *accuracy*, *precision*, *recall*, *f1-score* dan AUC yang cukup tinggi. Nilai AUC menunjukkan

bahwa kedua model masuk ke dalam klasifikasi “Sangat Baik” dalam membedakan kelas positif dan negatif. Berikut adalah grafik perbandingan dengan *library matplotlib.pyplot* antara *Naïve Bayes Multinomial* dan *Bernoulli* yang ditunjukkan pada gambar 4.



Gambar 4. Grafik Perbandingan Akurasi

Pada gambar 4 menunjukkan grafik perbandingan akurasi model pada tiga skenario pengujian. Dalam evaluasi model dengan menggunakan *Confusion Matrix* dan metrik evaluasi model pada ketiga skenario, disimpulkan bahwa model *Naïve Bayes Bernoulli* memiliki kinerja yang lebih baik dibandingkan dengan model *Naïve Bayes Multinomial*. model *Naïve Bayes Bernoulli* mencapai nilai akurasi tertinggi pada skenario pengujian pertama dengan rasio *split data* 8 banding 2 antara data *training* dan *testing*, dengan nilai *accuracy* sebesar 93,06%.

Hasil klasifikasi sentimen data ulasan aplikasi Binar divisualisasikan dalam bentuk *Wordcloud* dengan menggunakan *library Wordcloud* pada *Python* untuk mengetahui gambaran atau informasi mengenai data ulasan pengguna aplikasi Binar pada situs *Google Play Store*. Gambar 5 merupakan *Wordcloud* dari hasil visualisasi ulasan positif.



Gambar 5. Wordcloud Positif

Berdasarkan gambar 5 dapat dilihat bahwa semakin besar ukuran kata dalam *Wordcloud* maka semakin tinggi frekuensi kata tersebut muncul. Beberapa kata yang sering muncul dalam ulasan positif terhadap aplikasi Binar tersebut menunjukkan bahwa para pengguna merasa bahwa aplikasi Binar memberikan pembelajaran yang efektif (“ajar”), memiliki desain visual yang menarik dan User Interface yang atraktif (“keren”), memiliki materi yang berkualitas (“materi”), penempatan menu dan navigasi yang mudah dioperasikan (“mudah”) dan membantu dalam belajar menggunakan aplikasi tersebut (“aplikasi”). Sedangkan pada gambar 6 merupakan *Wordcloud* dari hasil visualisasi ulasan negatif.



Gambar 6. Wordcloud Negatif

Pada gambar 6 menunjukkan bahwa beberapa kata yang sering muncul dalam ulasan negatif terhadap aplikasi Binar tersebut menunjukkan bahwa pengguna aplikasi Binar merasa bahwa aplikasi tersebut lambat dalam pengoperasiannya (“lambat”), memerlukan waktu lama untuk memuat dan membuka halaman (“loading”), mengalami masalah teknik atau bug yang mengakibatkan malfungsi atau kesalahan dalam program yang

mengganggu kinerja dan dapat menyebabkan suatu aplikasi crash atau force close (“bug”) dan dirasa terlalu mahal atau tidak sebanding dengan fitur yang diberikan (“mahal”)

SIMPULAN

Berdasarkan hasil dari penelitian yang telah dilakukan, berikut adalah beberapa kesimpulan yang dapat diambil.

1. Hasil evaluasi dari tiga skenario pengujian yang telah dilakukan dengan menggunakan algoritma *Naïve Bayes Multinomial* dan *Bernoulli*, menunjukkan bahwa skenario pertama dengan perbandingan 9:1 antara data *training* dan *testing* menggunakan algoritma *Naïve Bayes Bernoulli* memberikan hasil terbaik. Model tersebut menghasilkan *accuracy* sebesar 93,06%, *precision* sebesar 87,04%, *Recall* sebesar 100%, *F1-score* sebesar 93,07%, dan AUC sebesar 0,988. Dengan nilai-nilai tersebut, klasifikasi model ini termasuk dalam kategori "Sangat Baik".
2. Kata yang sering muncul dalam ulasan positif terhadap aplikasi Binar tersebut menunjukkan bahwa para pengguna merasa bahwa aplikasi Binar memberikan pembelajaran yang efektif (“ajar”), menarik dan keren (“keren”), memiliki materi yang berkualitas (“materi”), mudah digunakan (“mudah”) dan membantu dalam belajar aplikasi tersebut (“aplikasi”). Sedangkan kata yang sering muncul dalam ulasan negatif menunjukkan bahwa pengguna aplikasi Binar merasa bahwa aplikasi tersebut lambat dalam pengoperasiannya (“lambat”), memerlukan waktu lama untuk memuat dan membuka halaman (“loading”), mengalami masalah teknik atau bug yang mengganggu (“bug”) dan dirasa terlalu mahal atau tidak sebanding dengan fitur yang diberikan (“mahal”).

DAFTAR PUSTAKA

- Aini, S. H. A., Sari, Y. A., & Arwan, A. (2018). *Feature selection Information Gain* untuk Klasifikasi Penyakit Jantung Menggunakan Kombinasi Metode K-Nearest Neighbor dan *Naïve Bayes*. *Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputer*, 2(9), 2546–2554. <http://j-ptiik.ub.ac.id>
- Chingmuankim, N. (2022). A comparative study of *Naïve Bayes Classifiers* with improved technique on Text Classification. *TechRxiv*, 1–8. <https://doi.org/10.36227/techrxiv.19918360.v1>
- Databoks. (2021). Pengguna Smartphone diperkirakan Mencapai 89% Populasi pada 2025 (15 September 2020). Diakses pada 10 Maret 2022, pukul 13.00 dari <https://databoks.katadata.co.id/datapublish/2020/09/15/pengguna-smartphone-diperkirakan-mencapai-89-populasi-pada-2025>
- Fanissa, S., Fauzi, M. A., & Adinugroho, S. (2018). Analisis Sentimen Pariwisata di Kota Malang Menggunakan Metode *Naïve Bayes* dan *Feature selection Query Expansion Ranking*. *Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputer*, 2(8), 2766–2770. <http://j-ptiik.ub.ac.id>
- Fransiska, S., & Irham Gufroni, A. (2020). Sentiment Analysis Provider by.U on *Google Play Store Reviews* with TF-IDF and Support Vector Machine (SVM) Method. *Scientific Journal of Informatics*, 7(2), 2407–7658. <http://journal.unnes.ac.id/nju/index.php/sji>
- Gorunescu, F. (2010). *Data Mining concepts, model, and techniques*. Chennai, India: Springer. <https://doi.org/10.1007/978-3-642-19721-5>
- Hadianto, N., Novitasari, H. B., & Rahmawati, A. (2019). Klasifikasi Peminjaman Nasabah Bank

- Menggunakan Metode Neural Network. *Jurnal Pilar Nusa Mandiri*, 15(2), 163–170. <https://doi.org/10.33480/pilar.v15i2.658>
- Hasibuan, M. R., & Marji. (2019). Pemilihan Fitur dengan *Information Gain* untuk Klasifikasi Penyakit Gagal Ginjal menggunakan Metode Modified K-Nearest Neighbor (MKNN). *Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputer*, 3(11), 10435–10443.
- Isnanda, A., Umaidah, Y., & Jaman, J. H. (2021). Implementasi *Naïve Bayes Classifier* Dan *Information Gain* Pada Analisis Sentimen Penggunaan E-Wallet Saat Pandemi. *Jurnal Teknologi Informatika Dan Komputer*, 7(2), 144–153. <https://doi.org/10.37012/jtik.v7i2.648>
- Kulkarni, A., & Shivananda, A. (2019). Natural Language Processing Recipes. In *Natural Language Processing Recipes*. Bangalore, India: Apress Berkeley, CA. <https://doi.org/10.1007/978-1-4842-4267-4>
- Liu, B. (2012) *Sentiment Analysis and Opinion Mining*. *Synthesis Lectures on Human Language Technologies*. California, Amerika Serikat: Morgan & Claypool Publishers, 5, 1-167. <https://doi.org/10.1007/978-3-031-02145-9>
- Made, I Adnyana, B. (2019). Penerapan *Feature selection* untuk Prediksi Lama Studi Mahasiswa. *Jurnal Sistem Dan Informatika*, 13, 72–76.
- Manning, C. D., Raghavan, P., & Schütze, H. (2009). *An Introduction to Information Retrieval*. England: Cambridge University Press Cambridge, 53(9), 462–463. <https://doi.org/10.1108/00242530410565256>
- Ramos, S., Soares, J., Cembranel, S. S., Tavares, I., Foroozandeh, Z., Vale, Z., & Fernandes, R. (2021). *Data Mining* techniques for electricity customer characterization. *Procedia Computer Science*, 186, 475–488. <https://doi.org/10.1016/j.procs.2021.04.168>
- Reza Satria, A., & Adinugroho, S. (2020). Analisis Sentimen Ulasan Aplikasi Mobile menggunakan Algoritma Gabungan *Naïve Bayes* dan C4.5 berbasis Normalisasi Kata Levenshtein Distance. *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, 4(11), 4154–4163. <http://j-ptiik.ub.ac.id>
- Salsabila, S. M., Alim Murtopo, A., & Fadhilah, N. (2022). Analisis Sentimen Pelanggan Tokopedia Menggunakan Metode *Naïve Bayes Classifier*. *Jurnal Minfo Polgan*, 11(2), 30–35. <https://doi.org/10.33395/jmp.v11i2.11640>
- Saputra, S. A., Rosiyadi, D., Gata, W., & Husain, S. M. (2019). Google Play E-Wallet Sentiment Analysis Using *Naïve Bayes* Algorithm Based on Particle Swarm Optimization. *Jurnal RESTI (Rekayasa Sistem Dan Teknologi Informasi)*, 3(3), 377–382.
- Sari, A. E., Widowati, S., & Lhaksmana, K. M. (2019). Klasifikasi Ulasan Pengguna Aplikasi Mandiri Online di *Google Play Store* dengan Menggunakan Metode *Information Gain* dan *Naïve Bayes Classifier*. *E-Proceeding of Engineering*, 6(2), 9143–9157.
- Singh, G., Kumar, B., Gaur, L., & Tyagi, A. (2019). Comparison between *Multinomial* and *Bernoulli Naïve Bayes* for Text Classification. *International Conference on Automation, Computational and Technology Management, ICACTM*, April, 593–596. <https://doi.org/10.1109/ICACTM.2019.8776800>
- Verdikha, N. A., Adjii, T. B., & Permasari, A. E. (2018). Komparasi Metode Oversampling

Untuk Klasifikasi Teks Ujaran Kebencian. Seminar Nasional Teknologi Informasi Dan Multimedia 2018, 85–90.

Wardani, N. S., Prahutama, A., & Kartikasari, P. (2020). Analisis Sentimen Pemandangan Ibu Kota Negara Dengan Klasifikasi *Naïve Bayes* Untuk Model *Bernoulli* Dan *Multinomial*. *Jurnal Gaussian*, 9(3), 237–246.
<https://doi.org/10.14710/j.gauss.v9i3.27963>.