ANALISIS SENTIMEN ULASAN PELANGGAN MIE GACOAN MENGGUNAKAN METODE NAÏVE BAYES CLASSIFIER

Ghina Shalihah, Rudi Kurniawan, Tati Suprapti
Teknik Informatika, STMIK IKMI Cirebon
Jalan Perjuangan No.10B, Karyamulya, Kec. Kesambi, Kota Cirebon, Jawa Barat
shalihahghina273@gmail.com

ABSTRAK

Indonesia terkenal dengan kekayaan kuliner yang beragam dan unik, salah satunya yaitu Mie Gacoan. Pada saat ini, bisnis kuliner Mie Gacoan sedang berkembang. Mie Gacoan menawarkan variasi tingkat kepedasan pada menu mie mereka, yang dapat disesuaikan dengan *preferensi* pelanggan. Jenis mie ini semakin populer, terutama di kalangan generasi muda, sehingga banyak restoran atau warung makan yang menyajikan hidangan serupa bermunculan. Memperoleh kepercayaan pelanggan merupakan tantangan untuk mereka yang terlibat dalam dunia bisnis, mendorong mereka untuk melakukan evaluasi menyeluruh terhadap produk dan layanan yang mereka sediakan. Permasalahan yang muncul dari ulasan atau pendapat dari pelanggan misalnya mengenai kurangnya pelayanan harus menjadi perhatian utama. Meskipun demikian, membaca dan mengelompokkan setiap ulasan yang diberikan oleh pelanggan memerlukan waktu yang cukup lama dan dianggap kurang *efisien*. Dalam *konteks* ini, analisis data pendapat pelanggan dari *platform* Twitter muncul sebagai solusi *alternatif*. Twitter, sebagai media sosial yang sering dimanfaatkan dalam pemasaran, memberikan kesempatan untuk *interaksi* pelanggan dengan pengusaha. Penelitian ini mengusulkan penerapan algoritma *Naïve Bayes Classifier* pada analisis sentimen ulasan pelanggan Mie Gacoan di Twitter pada Oktober 2023. Metode ini dipilih karena kebutuhan data *training* yang kecil dan sifatnya yang sederhana dan berhasil menghasilkan kinerja yang baik, dengan nilai akurasi sebesar 88,06%, *recall* sebesar 77,78%, dan *presision* sebesar 69,23%. Hasil dari penelitian ini dapat memberikan *kontribusi* positif dalam meningkatkan kualitas produk dan layanan, memperkuat relasi dengan pelanggan, serta mendukung keberlanjutan bisnis di tengah persaingan yang semakin ketat.

Kata kunci : *Analisis sentimen, Naïve Bayes Classifier, Mie Gacoan, Twitter, Text mining*

1. PENDAHULUAN

Di tengah Era *Globalisasi* ini, masyarakat terdorong dan termotivasi untuk mendirikan berbagai jenis usaha dan bisnis. Tujuan utamanya adalah memenuhi kebutuhan dan keinginan konsumen, sehingga dapat mencapai kepuasan konsumen dan meraih *profit seoptimal* mungkin [1]. Indonesia terkenal dengan kekayaan kuliner yang beragam dan unik [2], salah satunya yaitu Mie Gacoan. Pada saat ini, bisnis kuliner Mie Gacoan sedang berkembang. Mie Gacoan menawarkan variasi tingkat kepedasan pada menu mie mereka, yang dapat disesuaikan dengan preferensi pelanggan. Jenis mie ini semakin populer, terutama di kalangan generasi muda, sehingga banyak restoran atau warung makan yang menyajikan hidangan serupa bermunculan. Dalam situasi persaingan bisnis yang sengit, para pengusaha perlu mengutamakan kepuasan pelanggan agar mereka tetap setia dan tidak beralih ke produk pesaing. Keberhasilan ini hanya dapat dicapai jika kepuasan pelanggan terpenuhi [3].

Memperoleh kepercayaan pelanggan merupakan tantangan bagi pelaku bisnis, mendorong mereka untuk melakukan evaluasi menyeluruh terhadap produk dan layanan yang mereka tawarkan. *Identifikasi* masalah yang muncul dari ulasan pelanggan menjadi suatu keharusan. Meskipun demikian, membaca dan mengelompokkan setiap ulasan yang diberikan oleh pelanggan memerlukan waktu yang cukup lama dan dianggap kurang *efisien* [4]. Dari permasalahan yang dihadapi, salah satu solusi *alternatif* yang dapat diajukan adalah melalui analisis data pendapat pelanggan yang diperoleh dari *platform* Twitter [5]. Twitter merupakan salah satu *platform* media sosial yang sering dimanfaatkan sebagai sarana pemasaran produk dan juga digunakan untuk memperluas jangkauan bisnis. Pengguna Twitter dapat berinteraksi dengan pelaku bisnis lainnya, menjadi teman, atau mengikuti akun bisnis tersebut. Proses pengumpulan data dari Twitter dapat dilakukan dengan menghubungkan ke *Twitter-Harvest* dan menggunakan aplikasi seperti *RapidMiner*. Setelah data Tweet diperoleh, langkah selanjutnya adalah melakukan pemrosesan menggunakan teknik text mining untuk mengeliminasi data yang tidak diinginkan. Selanjutnya, data Tweet atau komentar akan dikelompokkan berdasarkan klasifikasi sentimen, yakni positif, negatif, dan netral. Metode perbandingan yang digunakan untuk pengelompokan ini adalah *Naïve Bayes* [6].

Penelitian ini menggunakan analisis sentiment dengan teknik *text mining* dan dengan menggunakan metode *Naïve Bayes.* Metode ini dipilih karena penggunaan metode *Naïve Bayes Classifier* hanya membutuhkan jumlah data *training* yang kecil sebagai *parameter* dalam proses *klasifikasi* [7].

Dengan menerapkan metode *Naive Bayes Classifier*, hasil analisis sentimen akan memberikan pemahaman yang lebih mendalam terhadap dinamika *interaksi* pelanggan, sekaligus membantu pemilik

bisnis, khususnya Mie Gacoan, dalam merespons dengan efektif terhadap umpan balik konsumen. Oleh karena itu, penelitian ini dapat memberikan *kontribusi* positif dalam meningkatkan kualitas produk dan layanan, memperkuat relasi dengan pelanggan, serta mendukung keberlanjutan bisnis di tengah persaingan yang semakin ketat.

2. TINJAUAN PUSTAKA

Terdapat beberapa penelitian yang dilakukan oleh penelitian sebelumnya yang menggunkan algoritma *Naïve Bayes Classifier* untuk penelitian klasifikasi, diantaranya yaitu penelitian yang dilakukan oleh [8], yang berjudul Analisis Sentimen Masyarakat terhadap Layanan Shopeefood Melalui Media Sosial Twitter dengan Algoritma *Naïve Bayes Classifier* dari hasil penelitian ini metode *Naive Bayes* mampu mengklasifikasikan sentimen positif dan negatif berupa data teks dengan baik dibuktikan dengan hasil performa yang didapatkan pada data *training* yaitu memiliki tingkat *accuracy* sebesar 98,17%, *precision* 97,68%, dan *recall* 98,69%. Sedangkan pada data *testing* memiliki *performa* dengan tingkat *accuracy* 90,62%, *precision* 88,23%, dan *recall* 93,75%. Hasil dari perhitungan AUC menghasilkan nilai sebesar 0.97 yang dimana nilai tersebut menunjukkan performa "Sangat Baik".

Berdasarkan hasil pengujian yang dilakukan oleh [7], pengklasifikasian sentimen dengan menggunakan algoritma *Naïve Bayes* memberikan hasil nilai *precision* sebesar 73,02%, *recall* sebesar 74%, serta akurasi sebesar 73,33%. Penelitian selanjutnya dilakukan oleh [9] Kehadiran aplikasi yang mempermudah untuk menelusuri ulasan mengenai suatu tempat atau hidangan memudahkan pembaca dalam memilih destinasi kuliner. Ulasan yang tersedia terbagi antara positif dan negatif. Penelitian ini menggunakan algoritma *Naive Bayes* berbasis *Particle Swarm Optimization* untuk mengevaluasi apakah terdapat peningkatan akurasi. Dataset yang digunakan terdiri dari ulasan restoran yang dibagi menjadi dua kelas, yaitu kelas positif dan kelas negatif. Pengujian data dilakukan dengan menggunakan metode *10 Fold Cross Validation*. Analisis sentimen terhadap ulasan restoran menggunakan Algoritma *Naive Bayes* berbasis *Particle Swarm Optimization* menghasilkan tingkat akurasi sebesar 82,45%. Hasil ini menunjukkan peningkatan dibandingkan dengan penggunaan algoritma *Naive Bayes saja*, yang mencapai tingkat akurasi sebesar 74,34%.

2.1. Data mining

Data mining merupakan proses yang menggunakan teknik *statistik*, matematika, kecerdasan buatan, dan *machine learning* untuk *mengekstraksi* serta *mengidentifikasi* informasi yang bermanfaat dan pengetahuan yang terkait dari berbagai *database* besar. Istilah *data mining* mencakup disiplin ilmu yang intinya adalah menemukan, menggali, atau menambang pengetahuan dari data atau informasi yang tersedia [10].

2.1.1. Analisis sentimen

Analisis *sentimen* merupakan proses *pengeksplorasi* teks yang mencakup tahapan analisis, pemahaman, pengolahan, dan *ekstraksi* data tekstual, termasuk opini, emosi, dan sikap terhadap suatu objek. Fokus utama penambangan data ini adalah teks yang berasal dari komentar, ulasan, atau testimoni yang terkait dengan produk yang sedang *diinvestigasi*. Tujuan dari analisis ini adalah untuk memperoleh pemahaman umum mengenai pandangan masyarakat terhadap objek yang sedang dianalisis [11].

2.2. Algoritma Naïve Bayes

Naïve Bayes merupakan metode klasifikasi yang sangat sederhana, mengasumsikan klasifikasi atribut, dan memanfaatkan prinsip *probabilitas* serta statistik yang diperkenalkan oleh ilmuwan Inggris, Thomas Bayes [12].

Teorema Bayes dapat diungkapkan secara matematis melalui persamaan berikut:

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

Dengan catatan bahwa $P(B) \neq 0$.

Pada dasarnya, teorema ini digunakan untuk menentukan peluang terjadinya kejadian A, bila diketahui bahwa kejadian B telah terjadi. Kejadian B sering disebut sebagai bukti atau informasi tambahan.

$P(A)$ merupakan *probabilitas apriori* dari kejadian A, yaitu *probabilitas* kejadian sebelum adanya bukti. Sementara itu, bukti atau kejadian B merupakan nilai atribut dari suatu *instansi* yang belum diketahui.

$P(A|B)$ adalah *probabilitas posteriori* dari kejadian A setelah adanya bukti B, dan dapat dihitung menggunakan teorema Bayes.

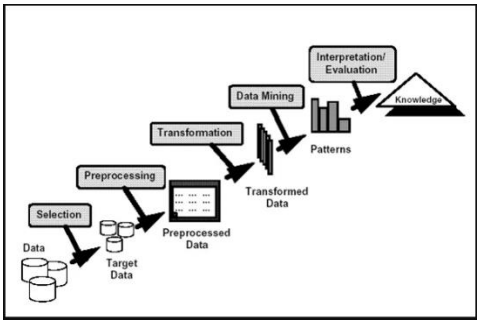
Karakteristik kunci dari algoritma Naive Bayes Classifier terletak pada asumsi yang sangat sederhana (*naif*), yaitu *independensi* yang dianggap kuat antar kondisi atau kejadian masing-masing.

2.3. RapidMiner

RapidMiner merupakan suatu *platform* perangkat lunak sumber terbuka yang difungsikan untuk menganalisis data, melakukan pemodelan prediktif, dan menggali informasi dari data. Desain platform ini ditujukan untuk memberikan dukungan kepada para profesional di bidang data, ilmuwan data, dan analis bisnis agar dapat menjelajahi, mengelola, serta melakukan analisis terhadap data dari berbagai sumber. RapidMiner dilengkapi dengan antarmuka grafis yang intuitif, memungkinkan pengguna untuk membuat alur kerja analisis data tanpa memerlukan penulisan kode pemrograman.

3. METODE PENELITIAN

Pada penelitian ini, metode KDD (*Knowledge Discovery in Databases*) digunakan sebagai landasan untuk *mengeksplorasi* dan *mengekstrak* pengetahuan dari basis data yang relevan. berikut adalah tahapan dari KDD terlihat pada Gambar 1 dibawah ini.



Gambar 1. Tahapan KDD (Sumber *https://dosbing.id/wp-content/uploads/2019/11/Annotation-2019-11-26-192902.png*)

1) Proses *Data Cleansing*
   melibatkan pengolahan data dan pemilihan data yang dianggap relevan.
2) Data *Integration*
   adalah proses menggabungkan data yang dianggap memiliki kesamaan, menjadikannya satu kesatuan.
3) Proses *Selection*
   melibatkan seleksi atau pemilihan data yang dianggap relevan untuk keperluan analisis.
4) Data *Transformation*
   adalah langkah di mana data terpilih diubah menjadi format yang sesuai dengan prosedur penggalian data.
5) Data *Mining*
   adalah tahapan di mana berbagai teknik diterapkan untuk mengekstrak pola-pola potensial yang menghasilkan informasi berharga.
6) Dalam *Pattern Evolution*, pola-pola yang teridentifikasi dievaluasi berdasarkan metrik yang ditetapkan.
7) *Knowledge Presentation*
   merupakan tahap akhir dalam proses KDD, di mana data yang telah diproses divisualisasikan untuk mempermudah pemahaman pengguna dan diharapkan mendorong tindakan berdasarkan analisis.

4. HASIL DAN PEMBAHASAN

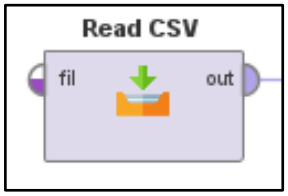
4.1. Hasil

4.1.1. *Selection*

Pada tahap ini dataset akan diseleksi, apakah atribut dapat digunakan atau tidak dengan menggunakan *rapidminer*.

Langkah-langkah yang dapat dilakukan adalah sebagai berikut :

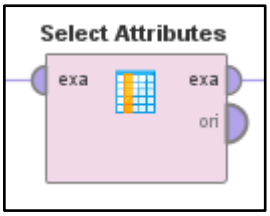
a) Masukkan dataset yang telah di ambil dari Tweet- Harvest tadi ke dalam read csv karena data dalam bentuk csv.



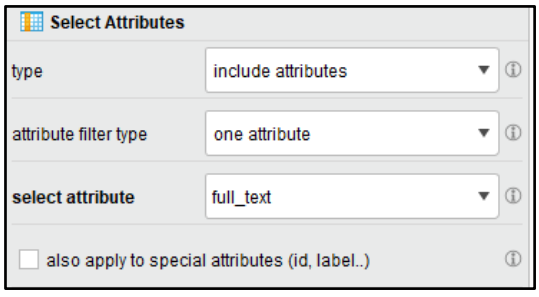
Gambar 2. Read CSV

Fungsi dari operasi "*Read CSV*" dalam *RapidMiner* adalah memfasilitasi *impor* data dari file *CSV* ke dalam lingkungan *RapidMiner*. Langkah ini memungkinkan pengguna untuk memulai analisis data dengan cepat dengan menggunakan sumber data *eksternal*. Dengan menggunakan "*Read CSV*," pengguna dapat memilih file *CSV* yang berisi data yang ingin diolah, dan *RapidMiner* akan membaca dan *mengonversi* data tersebut ke dalam *format* yang dapat digunakan di alur kerja analisis.

b) Kemudian masukkan operator *select attributes* dengan cara *drag* ke bagian halaman proses lalu atur *parameters* nya, *type* nya gunakan *include attributes*, *attribute filter type* nya pilih *one attribute*, kemudian *select attribute* nya pilih *full_text*



Gambar 3. Select atribute



Gambar 4. Parameter

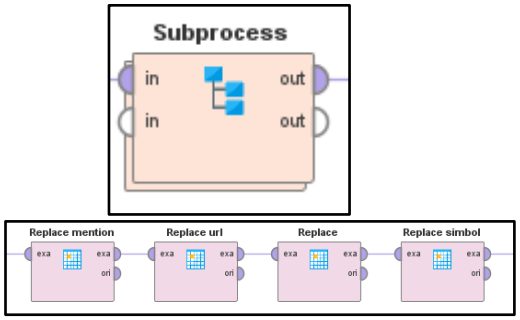
Operasi *Select Attributes* dalam *RapidMiner* berfungsi untuk memilih sebagian atribut atau kolom dari *dataset*. Fungsi ini memberikan kemampuan untuk mengontrol atribut mana yang akan dimasukkan atau dikecualikan dari proses analisis atau pemodelan. *Select Attributes* memungkinkan pengguna untuk mengelola *dimensi* data dengan memilih kolom-kolom tertentu sesuai kebutuhan analisis atau pemodelan yang diinginkan.

4.1.2. *Pre-Processing*

Pra-*pemrosesan* dalam *KDD* berperan dalam membersihkan, menggabungkan, dan mengubah data dari berbagai sumber sebelum melanjutkan ke tahap analisis. Proses-proses seperti membersihkan data, *mengintegrasikan* data, *mentransformasi* data, *mereduksi* dimensi, memilih data, *mendiskritisasi* data, melakukan *normalisasi*, memisahkan data, dan menangani ketidakseimbangan kelas dilakukan untuk memastikan bahwa data yang digunakan dalam proses penemuan pengetahuan dari basis data tersebut memiliki kualitas yang baik, *konsistensi*, dan siap untuk dijalankan analisis menggunakan berbagai metode di dalam *KDD*, termasuk penggunaan algoritma pembelajaran mesin.

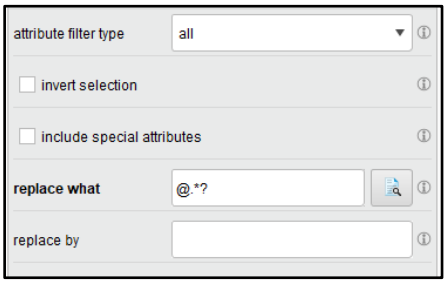
Langkah- langkah yang dapat dilakukan adalah sebagai berikut :

a) Tambahkan *operator select subprocess drag* ke dalam lembar proses lalu klik dua kali kemudian tambahkan *operator replace* di bagian lembar proses *select subprocess*.



Gambar 5. Tampilan operator subproses dan replace

Subproses dalam *RapidMiner* adalah sebuah proses yang diletakkan di dalam proses lain. Ketika operator *Subproses* dicapai selama *eksekusi* proses, terlebih dahulu seluruh *subproses* dijalankan. Setelah *eksekusi subproses* selesai, alur dikembalikan ke proses utama (induk). Setelah itu klik *replace* lalu kita atur *parameter* nya dengan cara klik *replace what* kemudian masukkan kata yang akan di hapus maka secara otomatis ketika di *run* kata yang terdapat *atribut text* akan terhapus.

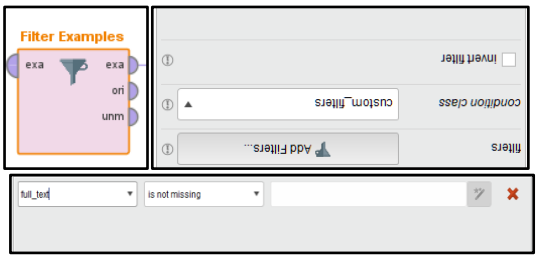


Gambar 6.Tampilan *parameter replace*

Operasi *Replace* dalam *RapidMiner* berfungsi untuk menggantikan nilai atau nilai-nilai tertentu dalam suatu *atribut* dengan nilai baru yang telah ditentukan. Fungsi ini berguna untuk melakukan pemrosesan data yang mencakup penggantian nilai-nilai yang tidak diinginkan atau perubahan tertentu dalam dataset.

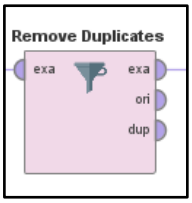
b) Langkah selanjutnya yaitu menambahkan *operator filter examples*, Caranya *drag operator filter example* ke lembar proses kemudian atur *parameter* nya dengan mengklik *add filter* dan pilih *full_text* serta *is not missing*.

Filter Examples dalam *RapidMiner* adalah suatu *operator* yang berfungsi untuk menyaring atau *memfilter* baris-baris data (*instances/examples*) dalam sebuah dataset berdasarkan suatu kriteria tertentu. Fungsinya memungkinkan pengguna untuk memilih *subset* data yang memenuhi kondisi atau aturan tertentu, sehingga memudahkan fokus pada bagian data yang *relevan* atau penting untuk analisis lebih lanjut.



Gambar 7. Tampilan filter examples

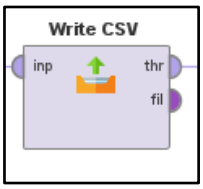
c) Setelah melakukan *filter example*, kemudian tambahkan *operator remove duplicate*. caranya pilih *remove duplicate drag* ke lembar proses lalu *run* atau jalankan maka *dataset* Twitter yang sama tersebut akan otomatis terhapus.



Gambar 8. *Operator remove duplicates*

Remove Duplicate dalam *RapidMiner* adalah suatu operasi yang berfungsi untuk menghilangkan baris-baris data yang memiliki nilai yang sama pada beberapa atau seluruh *atribut*, sehingga hanya menyisakan satu baris yang unik dari data tersebut.

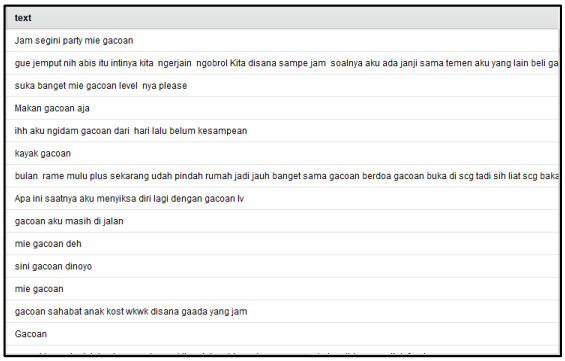
d) Langkah selanjutnya adalah menyimpan hasil data yang telah di proses tadi dengan menggunakan operator *write csv*. Caranya *darg* operator *write csv* ke lembar proses kemudian klik dua kali lalu simpan data di file *rapidminer* kemudian di dalam proses.



Gambar 9. Tampilan *write csv*

Write CSV dalam *RapidMiner* adalah suatu operasi yang berfungsi untuk menyimpan dataset atau hasil analisis ke dalam format file *CSV* (*Comma-Separated Values*). Tujuannya adalah untuk menyimpan data dalam bentuk teks yang sederhana dengan menggunakan koma sebagai pemisah antara nilai-nilai dalam setiap kolom.

e) Setelah melakukan pembersihan data, langkah selanjutnya adalah memberikan label pada data tersebut menggunakan kategori positif, negatif, dan netral. Proses pemberian label ini umumnya melibatkan analisis sentimen atau klasifikasi, di mana setiap *entitas* dalam data diberi label sesuai dengan kecenderungan atau karakteristik tertentu yang diinginkan.

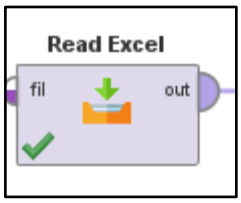


text
Jam segini party mie gacoan
gue jemput nih abis itu intinya kita ngerjain ngobrol Kita disana sampe jam soalnya aku ada janji sama temen aku yang lain beli ga
suka banget mie gacoan level nya please
Makan gacoan aja
ihh aku ngidam gacoan dari hari lalu belum kesampean
kayak gacoan
bulan rame mulu plus sekarang udah pindah rumah jadi jauh banget sama gacoan berdoa gacoan buka di scg tadi sih liat scg baka
Apa ini saatnya aku menyiksa diri lagi dengan gacoan lv
gacoan aku masih di jalan
mie gacoan deh
sini gacoan dinoyo
mie gacoan
gacoan sahabat anak kost wkwk disana gaada yang jam
Gacoan

Gambar 10. *Sampel* data yang sudah diberi label

Dalam analisis *sentimen* pada teks, data dapat diberi label positif jika mengandung *ekspresi* positif, label negatif jika mengandung *ekspresi* negatif, dan label netral jika tidak menunjukkan kecenderungan positif atau negatif yang *signifikan*. Pemberian label ini menjadi landasan untuk pengembangan model dan analisis lanjutan terkait sentimen atau klasifikasi pada data tersebut.

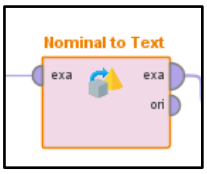
f) Setelah data diberi label, langkah selanjutnya adalah melakukan proses analisis menggunakan algoritma *Naïve Bayes*. Algoritma ini akan melakukan *evaluasi* dan klasifikasi data berlabel berdasarkan probabilitas serta dengan mengasumsikan kemandirian fitur-fitur yang terkandung dalam data tersebut. Langkah-langkah yang dapat dilakukan adalah sebagai berikut :
   a. Setelah data diberi label, langkah selanjutnya adalah menggabungkannya ke dalam proses analisis dengan menggunakan *operator Read Excel* pada *Rapidminer*.



Gambar 11. Tampilan *read excel*

Operator Read Excel ini berfungsi untuk membaca dan *mengimpor* data yang telah diberi label dari *file Excel* yang telah disimpan. Dengan demikian, *dataset* yang telah diberi label dapat diakses dan dimuat ke dalam lingkungan analisis untuk tahapan berikutnya, seperti pemrosesan lebih lanjut atau penerapan model algoritma.

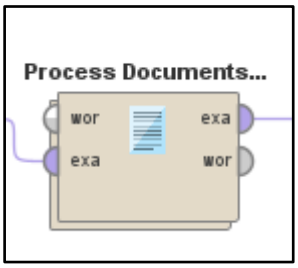
b. Selanjutnya, setelah data *diimpor* menggunakan *operator Read Excel*, langkah berikutnya adalah memberikan pemrosesan tambahan dengan menggunakan *operator Nominal to Text.*



Gambar 12. Tampilan nominal to text

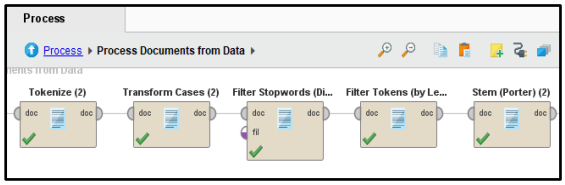
Operator Nominal to Text ini memiliki peran penting dalam *mengonversi variabel nominal* ke dalam *format* teks. Dengan menerapkan langkah ini, nilai-nilai *nominal* dalam *dataset* dapat diubah menjadi *representasi* teks yang lebih memudahkan *interpretasi* dan analisis selanjutnya. Proses ini merupakan bagian *integral* dari *pra-pemrosesan* data, mempersiapkan *dataset* untuk tahapan analisis lebih lanjut dengan memastikan *konsistensi* dan *interpretabilitas variabel* yang *relevan*.

c. Setelah data diberikan *operator Nominal to Text*, langkah berikutnya adalah melakukan proses dokumen yang melibatkan serangkaian operasi seperti *tokenisasi*, *transformasi* huruf, penyaringan kata penghenti (*stopword*), penyaringan *token*, dan *stemming*.



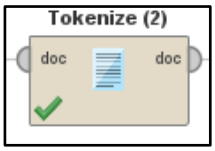
Gambar 13. Tampilan proses dokumen

Proses dokumen dalam *RapidMiner* berfungsi untuk memproses dan mempersiapkan data teks atau dokumen agar dapat diolah secara *efektif* dalam analisis data.



Gambar 14. Isi dari proses dokumen

d. Selanjutnya, dalam proses dokumen, *tokenisasi* berfungsi untuk memecah teks menjadi unit-unit kecil yang disebut token.



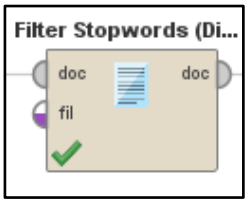
Gambar 15. *Tokenize*

Seperti kata atau frasa, sehingga *memfasilitasi* pemahaman struktur teks dan analisis lebih lanjut.

e. *Transform cases* berfungsi untuk menyamakan format huruf dalam teks, baik menjadi huruf kecil atau huruf besar, guna menghindari *ambiguitas* dalam analisis.

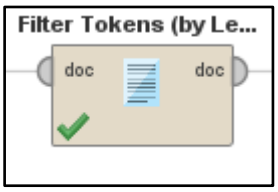
Gambar 16. *Transform cases*

f. *Filter stopword* berfungsi untuk menghilangkan kata-kata umum yang tidak memberikan informasi *signifikan*, seperti "dan," "atau," dan sejenisnya, sehingga memfokuskan analisis pada kata-kata kunci.



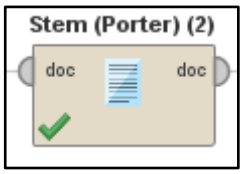
Gambar 17. *Filter stopwords*

g. *Filter tokenisasi* berfungsi untuk menyaring *token* berdasarkan kriteria tertentu, seperti membuang *token* yang terlalu pendek atau panjang, untuk membersihkan *dataset* dari informasi yang kurang *relevan*.



Gambar 18. *Filter tokens*

h. *Stemming* berfungsi untuk mengubah kata-kata dalam teks menjadi bentuk dasarnya dengan menghilangkan imbuhan atau akhiran kata, sehingga meningkatkan *konsistensi* dan *relevansi* analisis.



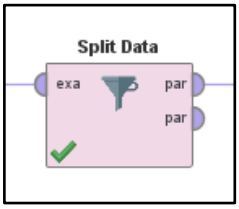
Gambar 19. *Stemming*

Dengan menerapkan langkah-langkah ini, proses dokumen mempersiapkan data teks dengan lebih baik untuk analisis lebih lanjut.

4.1.3. *Transformation*

Transformasi data ini dilakukan untuk memperbaiki kualitas data, mengurangi *kompleksitas*, atau mengungkap pola dan informasi tersembunyi yang tidak terlihat dalam bentuk awalnya. Langkah-langkah *transformasi* dapat mencakup *normalisasi* nilai-nilai data, penggabungan atau pengelompokan *fitur*, atau bahkan pembuatan *fitur* baru melalui *ekstraksi* informasi tambahan. Tujuan akhir dari proses *transform* adalah mempersiapkan data agar lebih sesuai untuk analisis lebih lanjut, *memfasilitasi identifikasi* pola pengetahuan, dan mendukung pengembangan model yang lebih akurat, langkah-langkahnya adalah sebagai berikut:

a) Setelah melalui proses dokumen menggunakan *operator* yang sesuai, langkah berikutnya adalah melakukan operasi pada data dengan menggunakan operator *Split Data*.

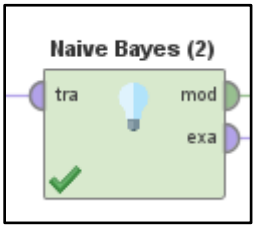


Gambar 20. *Split data*

Operator ini berfungsi untuk memisahkan *dataset* menjadi *subset* data pelatihan dan data pengujian, memungkinkan *evaluasi* kinerja model atau analisis lebih lanjut. Dengan membagi data ke dalam *subset* yang berbeda, kita dapat mengukur sejauh mana model atau analisis kita dapat diterapkan secara *efektif* pada data yang belum

dilihat sebelumnya. Proses ini merupakan langkah penting dalam menyiapkan data untuk penerapan model atau analisis lebih lanjut, bertujuan untuk meningkatkan kehandalan dan *generalisasi* hasil.

b) Setelah data diproses dan dibagi menjadi data latih dan data uji menggunakan *operator Split Data*, langkah selanjutnya adalah melanjutkan proses dengan mengklasifikasikan data menggunakan algoritma *Naive Bayes*.

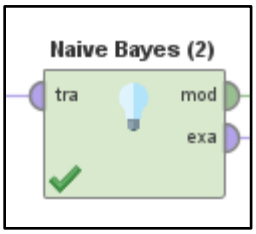


Gambar 21. *Algoritma Naive Bayes*

Algoritma *Naive Bayes* akan memanfaatkan informasi dari data latih untuk menghitung *probabilitas* kelas atau label pada data uji. Dengan demikian, proses klasifikasi ini akan memberikan *prediksi* terhadap kategori atau label yang paling mungkin untuk setiap *instan data* dalam *dataset* uji. *Evaluasi* hasil klasifikasi kemudian dapat memberikan pemahaman tentang *performa* algoritma *Naive Bayes* dalam *menggeneralisasi* dan *memprediksi* pada data yang belum pernah dilihat sebelumnya.

4.1.4. Data mining

Algoritma *Naive Bayes* adalah metode klasifikasi yang didasarkan pada *teorema Bayes* dengan *asumsi* bahwa setiap fitur dalam data bersifat *independen*.



Gambar 22. *Naive bayes*

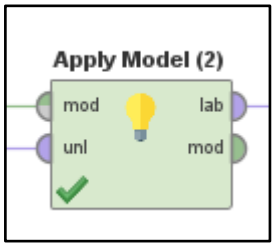
Algoritma ini digunakan untuk mengklasifikasikan data ke dalam kategori atau kelas tertentu dengan menghitung *probabilitas prior* dari setiap kategori dan *probabilitas likelihood* dari setiap fitur untuk setiap kategori. Beberapa keunggulan dari algoritma ini melibatkan tingkat akurasi yang tinggi, kecepatan dalam proses pengklasifikasi, kinerja yang baik dalam mengelola *variabel input kategoris*, serta kemampuannya menangani permasalahan multi-kelas. Penggunaan Algoritma *Naive Bayes* melibatkan berbagai bidang, termasuk *real-time prediction*, *multiclass prediction*, *text classification*, dan *rekomendasi*.

4.1.5. *Evaluasi*

Evaluasi merupakan proses penilaian dan pengukuran kualitas hasil yang ditemukan selama proses *ekstraksi* pengetahuan. Fungsinya adalah memastikan bahwa model, aturan, atau temuan pengetahuan yang dihasilkan dari analisis data memiliki *validitas*, akurasi, dan *relevansi* yang memadai. Langkah- langkah dalam *evaluasi* adalah sebagai berikut :

a) Apply Model

Apply model merujuk pada tahap dalam proses pemodelan di mana model yang telah dilatih pada data pelatihan diterapkan pada data baru untuk membuat *prediksi* atau klasifikasi.

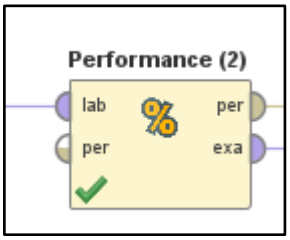


Gambar 23. *Apply Model*

Fungsinya adalah untuk menguji kemampuan model dalam *menggeneralisasi* dan memprediksi pada situasi yang belum pernah dilihat sebelumnya. Dengan menerapkan model pada data uji atau data *operasional*, kita dapat mengukur sejauh mana model dapat memberikan hasil yang akurat dan dapat diandalkan dalam memecahkan masalah atau memberikan *prediksi* yang bernilai. Hasil dari tahap ini juga digunakan untuk *mengevaluasi kinerja* model, *memvalidasi* ketepatan *prediksi*, dan memastikan bahwa model dapat digunakan secara *efektif* dalam *konteks* aplikasi yang diinginkan.

b) Performance

Setelah data *dievaluasi* melalui tahap *apply model*, langkah selanjutnya adalah menghitung nilai akurasi model menggunakan metode *performance* atau *evaluasi* kinerja.



Gambar 24. *Performance*

Proses ini melibatkan perbandingan antara hasil prediksi yang dihasilkan oleh model dengan nilai sebenarnya dari data uji. *Metrik performa* yang umum digunakan untuk mengukur akurasi termasuk akurasi (*accuracy*), presisi (*precision*), dan *recall*. Dengan mengukur kinerja model, kita dapat memahami seberapa baik model dapat

memprediksi dengan benar dan menilai kehandalan serta *relevansinya* dalam *konteks* aplikasi yang diinginkan. *Evaluasi* kinerja model sangat penting untuk memastikan bahwa hasil yang dihasilkan sesuai dengan tujuan analisis dan dapat diandalkan dalam pengambilan keputusan.

4.2. Pembahasan

Setelah melalui proses analisis, diperoleh nilai-nilai yang menggambarkan karakteristik dan pola dalam data. Hasil *evaluasi* menunjukkan sejumlah *metrik performa*, termasuk akurasi, *presisi*, dan *recall*, yang memberikan pemahaman mendalam tentang seberapa baik model dapat memprediksi dan mengklasifikasikan data.

accuracy: 88.06%

	true Positif	true Negatif	true Netral	class precision
pred. Positif	18	0	8	69.23%
pred. Negatif	0	13	0	100.00%
pred. Netral	0	0	28	100.00%
class recall	100.00%	100.00%	77.78%	

Gambar 25. Hasil data penelitian

4.2.1. Nilai Accuracy

Pada penelitian ini, diperoleh nilai akurasi sebesar 88,06%. Nilai ini mencerminkan sejauh mana model atau sistem klasifikasi dapat memprediksi dengan benar, diukur sebagai persentase dari jumlah *prediksi* yang benar terhadap total jumlah data yang dievaluasi. Akurasi sebesar 88,06% menunjukkan kinerja yang baik dalam mengklasifikasikan data karena tingkat akurasi tersebut di atas 80% atau 90%, [10], namun perlu dianalisis bersama dengan *metrik evaluasi* lainnya, seperti *presisi* dan *recall*, untuk mendapatkan pemahaman yang lebih mendalam tentang kemampuan model dalam mengelola instan positif dan negatif serta membuat penilaian yang lebih *holistik* terkait kualitas prediksi yang dihasilkan.

4.2.2. Nilai *Precision*

Dalam penelitian ini, nilai *presisi* sebesar 69,23%. Nilai ini mencerminkan tingkat ketepatan model dalam mengidentifikasi *instan* positif, dihitung sebagai *persentase* dari jumlah *prediksi* positif yang benar terhadap total *prediksi* positif yang dilakukan oleh model. Dengan nilai presisi sebesar 69,23%, dapat disimpulkan bahwa sekitar 69,23% dari *instan* yang diidentifikasi sebagai positif oleh model benar-benar termasuk dalam kategori positif. *Evaluasi* presisi ini memberikan gambaran lebih rinci tentang kemampuan model dalam mengelola hasil *prediksi* positif, sehingga dapat digunakan untuk mengoptimalkan kualitas *prediksi* dan memahami sejauh mana model dapat memberikan informasi yang *relevan* dan tepat.

4.2.3. Nilai *Recall*

Dan terakhir, nilai *recall* pada penelitian ini mencapai 77,78%. Dengan nilai *recall* sebesar 77,78%, dapat diartikan bahwa model mampu mendeteksi sekitar 77,78% dari total *instan* positif yang tersedia. Evaluasi *recall* ini memberikan gambaran tentang kemampuan model dalam mengelola hasil prediksi positif dan membantu dalam menilai sejauh mana model dapat mengidentifikasi secara *efektif instan* positif yang *relevan* dalam data.

5. KESIMPULAN DAN SARAN

Penelitian ini mengevaluasi efektivitas analisis sentimen terhadap ulasan pelanggan Mie Gacoan menggunakan *Naive Bayes Classifier*. Hasil evaluasi model menunjukkan tingkat akurasi sebesar 88,06%, *recall* sebesar 77,78%, dan presisi sebesar 69,23%. Tingkat akurasi yang tinggi mencerminkan kemampuan keseluruhan model, sementara tingkat recall yang tinggi menunjukkan kemampuan mendeteksi instan positif. Meskipun tingkat presisi menunjukkan ketepatan dalam mengklasifikasikan ulasan positif, keseimbangan antara presisi dan recall dianggap penting dalam penilaian holistik. Pendekatan ini memberikan gambaran positif terkait efektivitas analisis sentimen dengan *Naive Bayes Classifier* pada ulasan pelanggan Mie Gacoan, memberikan wawasan krusial bagi perusahaan. Algoritma *Naive Bayes* menggunakan probabilitas untuk mengklasifikasikan sentimen komentar, dengan *pra-pemrosesan* teks untuk mengidentifikasi fitur-fitur kunci. Dengan memahami sentimen pelanggan, perusahaan dapat merespons umpan balik secara lebih efektif. Penerapan *Naive Bayes Classifier* dalam analisis *sentimen* di *platform* seperti Twitter juga memberikan wawasan mendalam tentang *persepsi* pelanggan terhadap merek Mie Gacoan, yang dapat dimanfaatkan perusahaan untuk meningkatkan reputasi, memperkuat citra merek, dan membangun hubungan positif dengan pelanggan.

DAFTAR PUSTAKA

[1] J. Prasetyo, B. Ay, and I. A. Dpw, "KUALITAS PELAYANAN, STORE ATMOSPHERE DAN HARGA TERHADAP KEPUASAN KONSUMEN MIE GACOAN (Studi Kasus Pada Mie Gacoan Manahan)," 2021.

[2] Y. A. Singgalen, "Analisis Sentimen Konsumen terhadap Food, Services, and Value di Restoran dan Rumah Makan Populer Kota Makassar Berdasarkan Rekomendasi Tripadvisor Menggunakan Metode CRISP-DM dan SERVQUAL," *Build. Informatics, Technol. Sci.*, vol. 4, no. 4, pp. 1899–1914, 2023, doi: 10.47065/bits.v4i4.3231.

[3] C. A. B. S. Asma Wati, A. A. P. A. Suryawan Wiranatha, "ANALYSIS OF CUSTOMER SATISFACTION WITH PRODUCTS QUALITY AND SERVICES AT MIE GACOAN JIMBARAN REGENCY BALI ANALISIS KEPUASAN KONSUMEN TERHADAP KUALITAS PRODUK DAN LAYANAN PADA MIE GACOAN

JIMBARAN KABUPATEN BADUNG BALI," 2023.

[4] L. O. Sihombing, H. Hannie, and B. A. Dermawan, "Sentimen Analisis Customer Review Produk Shopee Indonesia Menggunakan Algortima Naïve Bayes Classifier," *Edumatic J. Pendidik. Inform.*, vol. 5, no. 2, pp. 233–242, 2021, doi: 10.29408/edumatic.v5i2.4089.

[5] W. Parasati, F. Abdurrachman Bachtiar, and N. Y. Setiawan, "Analisis Sentimen Berbasis Aspek pada Ulasan Pelanggan Restoran Bakso President Malang dengan Metode Naïve Bayes Classifier," 2020. [Online]. Available: http://j-ptiik.ub.ac.id

[6] H. Irsyad and M. R. Pribadi, "Opinion Classification on Indonesian Palm Oil Agriculture Using Naïve Bayes (Klasifikasi Opini Terhadap Pertanian Sawit Indonesia Menggunakan Naive Bayes)," *JATISI (Jurnal Tek. Inform. dan Sist. Informasi)*, vol. 6, no. 2, pp. 230–239, 2020.

[7] Vynska Amalia Permadi, "Analisis Sentimen Menggunakan Algoritma Naive Bayes Terhadap Review Restoran di Singapura," *J. Buana Inform.*, vol. 11, pp. 141–151, 2020.

[8] F. S. Mufidah, S. Winarno, F. Alzami, E. D. Udayanti, and R. R. Sani, "Analisis Sentimen Masyarakat Terhadap Layanan Shopeefood Melalui Media Sosial Twitter Dengan Algoritma Naïve Bayes Classifier," *JOINS (Journal Inf. Syst.*, vol. 7, no. 1, pp. 14–25, 2022, doi: 10.33633/joins.v7i1.5883.

[9] R. Sari, "Analisis Sentimen Review Restoran menggunakan Algoritma Naive Bayes berbasis Particle Swarm Optimization," *J. Inform.*, vol. 6, no. 1, pp. 23–28, 2019, doi: 10.31311/ji.v6i1.4695.

[10] M. Nasir and Verawaty, "Penerapan Algoritma Naive Bayes Classifier Untuk Evaluasi Kinerja Akademik Mahasiswa Universitas Bina Darma," *JIKO (Jurnal Inform. dan Komputer)*, vol. 5, no. 2, pp. 81–88, 2021, [Online]. Available: http://forlap.ristekdikti.go.id

[11] 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.

[12] R. Yulia Hayuningtyas, "Penerapan Algoritma Naïve Bayes untuk Rekomendasi Pakaian Wanita," *J. Inform.*, vol. 6, no. 1, pp. 18–22, 2019, [Online]. Available: http://ejournal.bsi.ac.id/ejurnal/index.php/ji/article/view/4685