**Jurnal Indonesia : Manajemen Informatika dan Komunikasi**
https://journal.stmiki.ac.id

E-ISSN: 2723-7079 | P-ISSN: 2776-8074
Vol. 5 No. 3 (2024) | September
OPEN ACCESS

https://doi.org/10.35870/jimik.v5i3.997

# Analisa Sentimen Drama Korea Melalui Media Sosial X dengan Menggunakan Algoritma *Naïve Bayes*

Putri Dwi Aprilia [1*], Sri Lestari [2]

[1*,2] Program Studi Sistem Informasi, Sekolah Tinggi Ilmu Komputer Cipta Karya Informatika, Kota Jakarta Timur, Daerah Khusus Ibukota Jakarta, Indonesia.

Email: putridwia19@gmail.com [1*], sri.lestari1203@gmail.com [2]

**Histori Artikel:**
*Dikirim* 25 Juli 2024; *Diterima dalam bentuk revisi* 8 Agustus 2024; *Diterima* 20 Agustus 2024; *Diterbitkan* 30 September 2024. 

## Abstrak

Di era digital ini, kemajuan teknologi memungkinkan akses mudah ke berbagai media sosial, termasuk Twitter atau X yang memiliki lebih dari 500 juta pengguna. Platform ini sering digunakan untuk berbagi opini melalui konsep microblogging. Di Indonesia, trend Korea seperti K-Pop dan K-Drama sangat populer, memicu berbagai diskusi di media sosial, khususnya di X, yang menghasilkan berbagai opini baik positif maupun negatif. Analisis sentimen adalah metode efektif untuk menggali opini publik dari data teks yang besar, seperti yang ditemukan di X, dengan mengelompokkan komentar menjadi kelas positif, , dan negatif menggunakan metode Naïve Bayes. Data tweet dikumpulkan dari X menggunakan hashtag terkait drama Korea. Data kemudian dibersihkan dan diolah untuk analisis sentimen menggunakan Naive Bayes. Algoritma Naïve Bayes didasarkan pada teorema Bayes dan sering digunakan dalam analisis sentimen karena akurasinya dan kecepatan dalam menghasilkan hasil. Penelitian ini menghasilkan precision dengan persentase ketepatan sebesar 88,44%. Untuk data bersentimen negatif, precision memiliki nilai 67,86%, recall (specificity) untuk sentimen negatif adalah 59,38%, menunjukkan bahwa model cukup baik dalam menemukan data yang benar-benar negatif. Sementara itu, recall untuk sentimen positif mencapai 91,70%. Nilai accuracy model adalah 84,33%, menunjukkan bahwa algoritma yang digunakan dapat mengklasifikasikan sentimen dengan baik menggunakan data yang ada. Hasil ini dapat bermanfaat bagi produsen konten, pemasar, dan penelitian budaya dan sosial.

**Kata Kunci:** Analisis Sentimen; Naive Bayes; K-Drama; X.

## Abstract

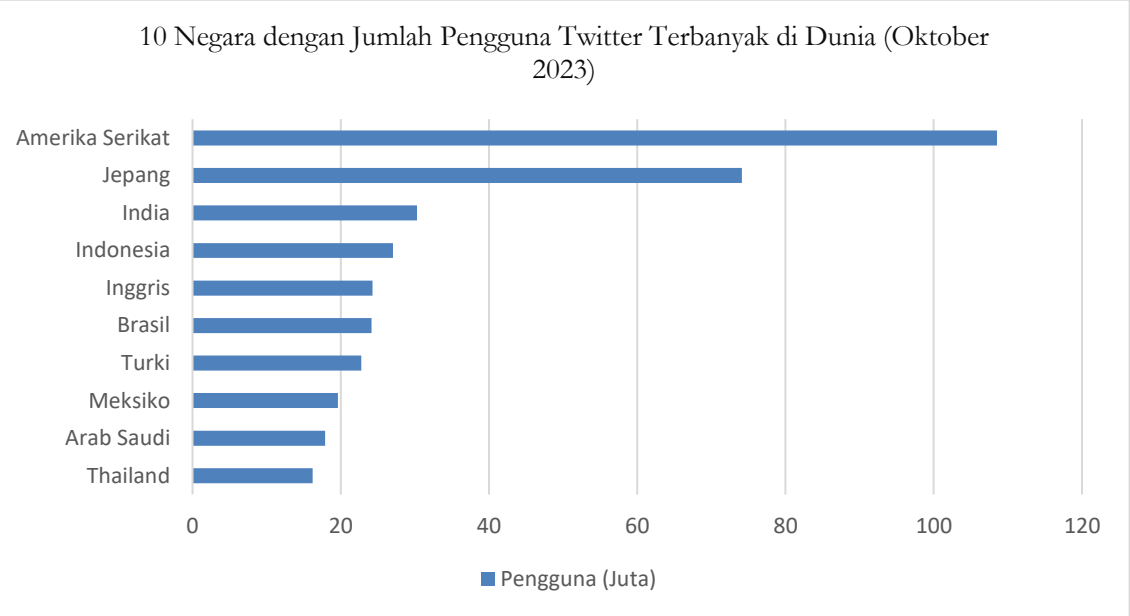
In this digital era, technological advancements allow easy access to various social media platforms, including Twitter or X, which has over 500 million users. This platform is often used to share opinions through the concept of microblogging. In Indonesia, Korean trends like K-Pop and K-Drama are highly popular, sparking various discussions on social media, especially on X, generating both positive and negative opinions. Sentiment analysis is an effective method for extracting public opinion from large text data, such as those found on X, by classifying comments into positive and negative classes using the Naïve Bayes method. Tweets were collected from X using hashtags related to Korean dramas. The data was then cleaned and processed for sentiment analysis using Naïve Bayes. The Naïve Bayes algorithm is based on Bayes' theorem and is often used in sentiment analysis due to its accuracy and speed in producing results. This study achieved a precision rate of 88.44%. For negative sentiment data, precision was 67.86%, and recall (specificity) for negative sentiment was 59.38%, indicating that the model is quite good at identifying truly negative data. Meanwhile, recall for positive sentiment reached 91.70%. The model's accuracy was 84.33%, showing that the algorithm used can classify sentiment well with the given data. These results can be useful for content creators, marketers, and cultural and social research.

**Keyword:** Sentiment Analysis; Naive Bayes; K-Drama; X.

Jurnal Indonesia : Manajemen Informatika dan Komunikasi
https://journal.stmiki.ac.id
E-ISSN: 2723-7079 | P-ISSN: 2776-8074
Vol. 5 No. 3 (2024) | September
OPEN ACCESS
https://doi.org/10.35870/jimik.v5i3.997

## 1. Pendahuluan

Di era digital ini, kemajuan teknologi dan informasi yang pesat memungkinkan kita untuk mengakses berbagai media sosial. Melalui platform media sosial ini, kita dapat dengan mudah mencari informasi dari berbagai belahan dunia. Salah satu platform media sosial yang banyak digunakan adalah *Twitter* atau *X*. Seperti halnya *X*, *microblogging* digunakan secara publik untuk menyampaikan pendapat, memberikan penilaian, dan membagikan berbagai opini melalui *tweet* (Mustofa & Novita, 2022).




Gambar 1. Jumlah Pengguna Twitter Terbanyak di Dunia

Menurut laporan *We Are Social*, Indonesia memiliki sekitar 27,5 juta pengguna *Twitter*/*X* pada Oktober 2023, menempatkan negara ini di peringkat keempat global. Amerika Serikat menempati posisi teratas dengan 108,55 juta pengguna *Twitter*/*X*, diikuti Jepang dan India dengan masing-masing 74,1 juta dan 30,3 juta pengguna. Inggris berada di bawah Indonesia dengan 24,3 juta pengguna, disusul Brasil dengan 24,15 juta, Turki 22,75 juta, Meksiko 19,6 juta, Arab Saudi 17,9 juta, dan Thailand 16,2 juta pengguna. Dalam laporan tersebut, *We Are Social* mencatat bahwa ada 666,2 juta pengguna *Twitter*/*X* di seluruh dunia pada Oktober 2023, menjadikan platform ini sebagai aplikasi media sosial dengan pengguna terbanyak ke-12 di dunia, di bawah *Snapchat*, *Douyin*, dan *Kuaishou*. Jumlah pengguna *Twitter*/*X* global pada Oktober 2023 meningkat 18,1% secara kuartalan dan 22,4% secara tahunan. Dalam hal gender, laki-laki mendominasi pengguna *Twitter*/*X* global dengan proporsi 61,2%, sedangkan perempuan 38,8%.

Tren Korea telah menjadi sangat populer di Indonesia dalam beberapa tahun terakhir, seperti dalam bidang mode, musik, yang biasa disebut *K-Pop*, dan drama Korea atau *K-Drama* yang telah menarik perhatian masyarakat Indonesia. Percakapan dan diskusi di media sosial, terutama di *X*, menghasilkan beragam kritik serta pro dan kontra. Banyak opini masyarakat tentang tren tersebut, salah satunya mengenai drama Korea, terdiri dari komentar positif dan negatif, yang menyebabkan keresahan tertentu (Amelia *et al.*, 2022).

Analisis sentimen adalah metode yang efektif untuk menggali opini publik dari data teks yang besar seperti yang ditemukan di *X* (Asri *et al.*, 2024). Analisis ini dapat dilakukan dengan mengelompokkan komentar menjadi kelas positif dan negatif menggunakan metode *Naïve Bayes*. Analisis sentimen merupakan proses untuk menemukan pola dalam kalimat, teks, atau data kata untuk mengidentifikasi isu sentimen, baik positif maupun negatif, yang umumnya ditulis di media sosial

Jurnal Indonesia : Manajemen Informatika dan Komunikasi
https://journal.stmiki.ac.id
E-ISSN: 2723-7079 | P-ISSN: 2776-8074
Vol. 5 No. 3 (2024) | September
OPEN ACCESS
https://doi.org/10.35870/jimik.v5i3.997

(Amrullah & Pane, 2023). Proses ini biasanya dilakukan secara otomatis menggunakan berbagai program atau platform yang mendukung analisis teks (Rahmawati *et al.*, 2017).

Algoritma *Naïve Bayes* didasarkan pada Teorema Bayes, yang mengasumsikan bahwa setiap kegiatan memberikan kontribusi yang sama penting atau bersifat independen dalam pemilihan kelas tertentu. Salah satu metode klasifikasi yang digunakan untuk memahami persepsi masyarakat dalam *text mining* adalah *Naïve Bayes*, yang sering disebut sebagai *Naïve Bayes Classifier* (Darwis *et al.*, 2021). Metode *Naïve Bayes* merupakan salah satu metode analisis sentimen yang paling populer dan sering digunakan dalam penelitian terkait karena akurasinya dan kemampuannya dalam menghasilkan analisis secara cepat dan efektif (Ghozali *et al.*, 2023).

Berdasarkan latar belakang di atas, penulis melakukan penelitian berjudul "Analisis Sentimen Ketertarikan Drama Korea Melalui Media Sosial *X* dengan Menggunakan Algoritma *Naïve Bayes*". Dengan menggunakan algoritma *Naïve Bayes* untuk memahami sentimen masyarakat Indonesia terhadap drama Korea melalui media sosial *X*, kita dapat memperoleh wawasan berharga mengenai tingkat ketertarikan masyarakat terhadap *K-Drama*, tren yang muncul, serta faktor-faktor yang memengaruhi persepsi mereka. Hal ini tidak hanya menguntungkan produsen konten dan pemasar, tetapi juga memberikan kontribusi signifikan bagi penelitian budaya dan sosial di Indonesia (Rina Noviana & Isram Rasal, 2023).

## 2. Metode Penelitian

### 2.1 Penerapan Metodologi

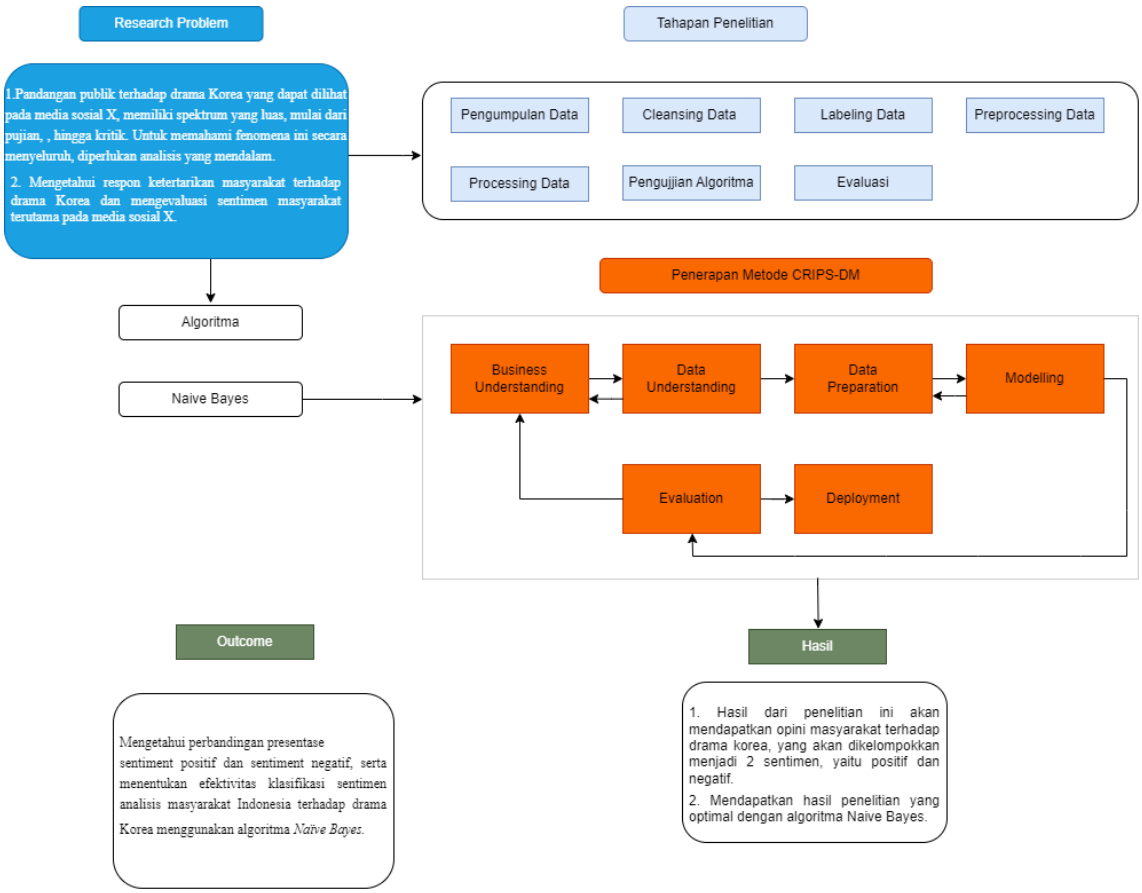
Dalam penelitian ini, analisis sentimen terhadap platform *X* mengenai drama Korea dilakukan dengan menerapkan algoritma *Naïve Bayes* yang diimplementasikan melalui perangkat lunak *RapidMiner*. *RapidMiner* merupakan sebuah platform yang dirancang khusus untuk mendukung berbagai proses pembelajaran mesin, baik yang sederhana maupun kompleks, termasuk di dalamnya pembelajaran mesin mendalam, penambangan teks, dan analisis prediktif. Platform ini banyak digunakan karena kemampuannya dalam mengolah data besar dengan cepat dan efisien (Slamet *et al.*, 2022).

Pengumpulan data dilakukan dengan metode *crawling* dari *tweet* yang diambil menggunakan *Google Colab* dan diproses menggunakan bahasa pemrograman *Python*. *Python* dikenal sebagai bahasa pemrograman yang dinamis, interpretatif, dan terstruktur, serta memiliki beragam pustaka yang sangat mendukung analisis data skala besar (Taufiqi & Nugroho, 2023). Setelah data dikumpulkan, langkah berikutnya adalah membersihkan data dari elemen-elemen yang tidak relevan, seperti *mention*, *hashtag*, dan simbol-simbol lainnya. Proses pembersihan atau *data cleansing* ini melibatkan beberapa tahap, di antaranya adalah tokenisasi, normalisasi teks, serta penghapusan *stop words*.

Setelah melalui proses pembersihan, data yang telah siap kemudian dianalisis menggunakan metode *CRISP-DM* (*Cross-Industry Standard Process for Data Mining*). Metode ini banyak digunakan karena menyediakan pendekatan terstruktur untuk analisis data besar, serta merupakan standar umum dalam data mining. *CRISP-DM* terdiri dari beberapa tahapan utama, seperti pemahaman bisnis, pemahaman data, persiapan data, pemodelan, evaluasi, dan penerapan model (Purnajaya *et al.*, 2022). Pendekatan ini membantu peneliti untuk memastikan bahwa setiap tahapan dalam proses analisis dilakukan secara sistematis dan konsisten, sesuai dengan standar yang diakui secara internasional (Sholeha *et al.*, 2022).

Penjelasan mengenai tahapan-tahapan dalam metodologi penelitian ini dapat dilihat pada gambar berikut.

Jurnal Indonesia : Manajemen Informatika dan Komunikasi
https://journal.stmiki.ac.id
E-ISSN: 2723-7079 | P-ISSN: 2776-8074
Vol. 5 No. 3 (2024) | September
OPEN ACCESS
https://doi.org/10.35870/jimik.v5i3.997




Gambar 2. Gambar Tahapan Penerapan Metodologi

Pada gambar di atas terdapat enam tahapan dalam metodologi *CRISP-DM* (Cross-Industry Standard Process for Data Mining), yang dijelaskan secara lebih rinci sebagai berikut:

1) Pemahaman Bisnis (*Business Understanding*)
   Tahap pertama ini berfokus pada pemahaman terhadap objek penelitian yang akan dilakukan. Tujuan utama dari penelitian ini adalah untuk mengidentifikasi dan menganalisis sentimen publik terhadap drama Korea yang disampaikan melalui media sosial *X*. Pada tahap ini, peneliti harus merumuskan dengan jelas tujuan penelitian, mendefinisikan pertanyaan bisnis, serta menetapkan target hasil yang ingin dicapai melalui analisis data yang tersedia.
2) Pemahaman Data (*Data Understanding*)
   Tahap pemahaman data bertujuan untuk mengenali struktur, sumber, dan karakteristik data yang akan digunakan dalam analisis. Data diambil dari media sosial *X* dalam bentuk *tweet* yang berisi opini tentang drama Korea. Pengambilan data dilakukan menggunakan API *X*, dengan kata kunci atau *hashtag* seperti "drama Korea." Pada tahap ini, peneliti juga perlu memverifikasi kualitas dan kelengkapan data, serta menilai relevansi data untuk menjawab tujuan penelitian.
3) Persiapan Data (*Data Preparation*)
   Pada tahap ini, data yang telah dikumpulkan akan melalui proses persiapan agar layak untuk dianalisis. Proses ini mencakup sejumlah aktivitas pengolahan data yang bertujuan untuk menghasilkan dataset yang bersih dan terstruktur. Tahap-tahap dalam persiapan data meliputi:
   a) Pengumpulan Data
      Data dikumpulkan dari media sosial *X* menggunakan teknik *crawling* berdasarkan kata kunci atau *hashtag* "drama Korea." Data yang diperoleh merupakan sekumpulan *tweet* yang relevan dengan penelitian.

Jurnal Indonesia : Manajemen Informatika dan Komunikasi
https://journal.stmiki.ac.id
E-ISSN: 2723-7079 | P-ISSN: 2776-8074
Vol. 5 No. 3 (2024) | September
OPEN ACCESS
https://doi.org/10.35870/jimik.v5i3.997

b) Pembersihan Data (*Data Cleansing*)
Pada tahap ini, data yang tidak relevan atau tidak valid dihapus untuk meningkatkan kualitas analisis. Beberapa langkah dalam proses pembersihan data meliputi:
1. Menghapus simbol, tautan, emotikon, tanda baca, serta karakter khusus lainnya yang tidak relevan dengan analisis.
2. Menormalisasi teks, misalnya dengan mengubah semua huruf menjadi huruf kecil agar lebih konsisten dalam analisis.
3. Menghapus data duplikat yang dapat memengaruhi hasil analisis.

c) Pelabelan Data (*Labelling*)
Setelah proses pembersihan, data dilabeli secara manual untuk membagi *tweet* ke dalam dua kelas sentimen, yaitu "positif" dan "negatif". Label ini digunakan untuk mengklasifikasikan opini publik berdasarkan sentimen yang disampaikan dalam *tweet*.

4) Pemodelan (*Modelling*)
Pada tahap pemodelan, data yang telah dipersiapkan kemudian diproses menggunakan algoritma *Naïve Bayes* untuk mengklasifikasikan sentimen. *Naïve Bayes* merupakan algoritma klasifikasi yang sering digunakan dalam analisis sentimen karena kecepatan dan keakuratannya dalam memprediksi hasil. Data dikelompokkan ke dalam dua kategori sentimen, yaitu "positif" dan "negatif", dengan menggunakan tiga atribut utama yang relevan. *RapidMiner* dipilih sebagai platform pemodelan karena memiliki kapabilitas yang kuat dalam analisis data dan pemodelan prediktif.

5) Evaluasi (*Evaluation*)
Tahap evaluasi bertujuan untuk mengukur performa dan keandalan model yang telah dibangun. Pada tahap ini, peneliti akan menilai apakah model tersebut telah sesuai dengan tujuan penelitian dan apakah model tersebut memberikan hasil yang akurat. Pengukuran akurasi, presisi, *recall*, dan nilai *F1* dilakukan untuk mengevaluasi efektivitas model dalam mengklasifikasikan sentimen dengan benar. Jika hasil evaluasi menunjukkan bahwa model tidak memenuhi standar yang diharapkan, penyesuaian atau pengulangan proses pemodelan dapat dilakukan (Alfandi Safira & Hasan, 2023).

6) Penerapan (*Deployment*)
Tahap terakhir dari metodologi *CRISP-DM* adalah penerapan model ke dalam lingkungan produksi. Pada tahap ini, model yang telah dievaluasi dan disetujui diterapkan untuk memberikan nilai bisnis, misalnya dalam konteks penelitian ini untuk memberikan wawasan kepada para produsen konten dan pemasar tentang opini publik terhadap drama Korea. Model yang diimplementasikan akan memberikan hasil yang dapat digunakan oleh pihak-pihak yang berkepentingan untuk mengambil keputusan yang berbasis data (Rachman & Oktavianti, 2021).

### 2.2 Proses Pengumpulan Data

Data tersebut diperoleh dari twitter dengan menggunakan skrip otomastis untuk X dengan kata kunci Drama Korea. Data yang berhasil diperoleh sebanyak 2.288 data. Dalam penelitian ini meliputi 5 proses yang dilakukan yaitu :

1) Crawling Data X
Objek pada deskripsi data penelitian ini adalah opini pengguna X terhadap drama Korea. Kata kunci atau hartag yang digunakan adalah drama Korea, dengan tweet yang menggunakan bahasa Indonesia.

2) Cleansing Data
Tweet yang belum diolah melalui beberapa proses, yaitu: cleaning, tokenize, transform cases, stopword dan filtering.

3) Labelling Data
Data yang sudah melewati tahap cleansing, beberapa data tweet dilabeli secara manual dengan MS. Excel dan untuk sisa data dilakukan pelabelan otomatis dengan menggunakan RapidMiner.

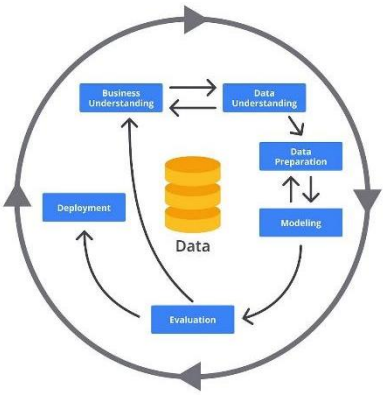
Jurnal Indonesia : Manajemen Informatika dan Komunikasi
https://journal.stmiki.ac.id
E-ISSN: 2723-7079 | P-ISSN: 2776-8074
Vol. 5 No. 3 (2024) | September
OPEN ACCESS
https://doi.org/10.35870/jimik.v5i3.997

4) Processing Data
Pada penelitian ini, pengujian akan dilakukan menggunakan pembobotan TF-IDF untuk mengevaluasi data yang diperoleh dari tahap pemodelan dengan Algoritma Naïve Bayes. Tujuan dari pengujian ini adalah untuk menentukan tingkat akurasi algoritma tersebut..
5) Evaluasi
Tahap terakhir adalah melihat hasil akurasi dan presentase sentimen positif dan negatif.

## 3. Hasil dan Pembahasan

### 3.1 Hasil

Pada Pada tahap ini, pendekatan yang digunakan adalah metode *Cross Industry Standard for Data Mining* (CRISP-DM). CRISP-DM adalah metode yang menggunakan model proses pengembangan data yang umum digunakan oleh para ahli untuk menyelesaikan masalah. Proses penelitian ini mengacu pada enam tahap yang ada dalam CRISP-DM, yaitu sebagai berikut:




Gambar 3. Metodologi CRISP-DM

3.1.1 *Business Understanding* (Pemahaman Bisnis)

Drama Korea atau *drakor* adalah serial televisi asal Korea Selatan yang sangat populer di berbagai belahan dunia, termasuk Indonesia. Drakor diminati oleh berbagai kalangan, khususnya remaja. Media sosial *X* sering digunakan untuk mengungkapkan opini masyarakat terkait dengan topik yang sedang populer, salah satunya adalah drama Korea. Tujuan utama dari penelitian ini adalah untuk menganalisis sentimen masyarakat Indonesia terhadap drama Korea dengan menggunakan algoritma *Naïve Bayes*, serta mencari nilai akurasi, presisi, dan *recall* untuk memahami opini publik secara keseluruhan.

3.1.2 *Data Understanding* (Pemahaman Data)

Langkah awal penelitian ini adalah pengumpulan data dengan teknik *crawling* dari media sosial *X*. Data yang dikumpulkan adalah *tweet* yang relevan dengan topik drama Korea. Pengumpulan data dilakukan dengan memanfaatkan API *X* dan diproses menggunakan *Google Colab*. Setelah data berhasil dikumpulkan, data disimpan dalam format CSV untuk mempermudah analisis lebih lanjut.

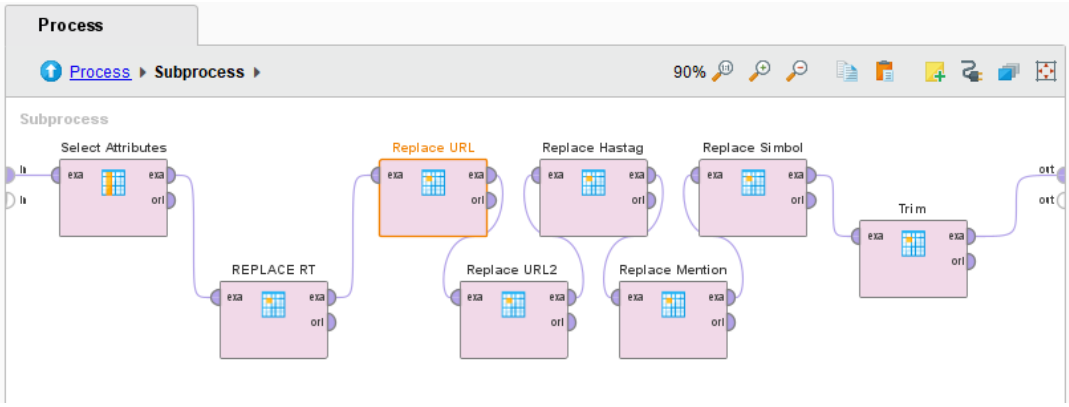
Jurnal Indonesia : Manajemen Informatika dan Komunikasi
https://journal.stmiki.ac.id
E-ISSN: 2723-7079 | P-ISSN: 2776-8074
Vol. 5 No. 3 (2024) | September
OPEN ACCESS
https://doi.org/10.35870/jimik.v5i3.997


Gambar 4. Proses Crawling Data

3.1.3 *Data Preparation* (Persiapan Data)

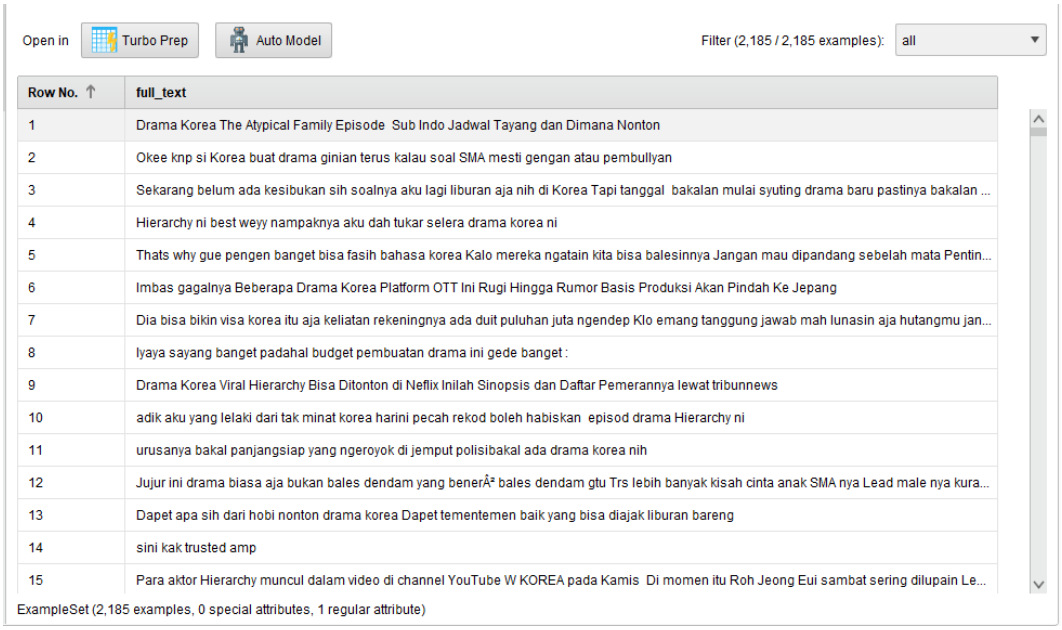
Persiapan data adalah salah satu tahapan kritis dalam analisis sentimen. Tahapan ini melibatkan beberapa langkah yaitu;

1) Pengumpulan Data
   Pengumpulan data atau *crawling data* dilakukan dengan menggunakan Google Colab untuk mengambil data tweet melalui media sosial X. Data yang berhasil dikumpulkan dari *crawling data* tersebut dengan menggunakan kata kunci drama korea adalah sebanyak 2.288 tweet.
2) Pembersihan Data (*Cleansing Data*)
   Tahap selanjutnya, data tweet yang telah dikumpulkan melalui *crawling data* akan melalui proses *cleansing data* dengan tujuan untuk membersihkan tweet dari kata – kata yang tidak diperlukan, kalimat duplikat, simbol – simbol atau karakter seperti *hastag* "#" , *mention* "@", link url seperti "https", dan angka yang tidak diperlukan dalam proses analisis sentimen. Dengan data awal 2.288 tweet, kemudian setelah melalui proses *cleansing* maka data yang didapatkan menjadi 2.185 tweet.




Gambar 5. Proses Pembersihan Data

Jurnal Indonesia : Manajemen Informatika dan Komunikasi
https://journal.stmiki.ac.id
E-ISSN: 2723-7079 | P-ISSN: 2776-8074
Vol. 5 No. 3 (2024) | September
OPEN ACCESS
https://doi.org/10.35870/jimik.v5i3.997



Open in Turbo Prep Auto Model Filter (2,185 / 2,185 examples): all

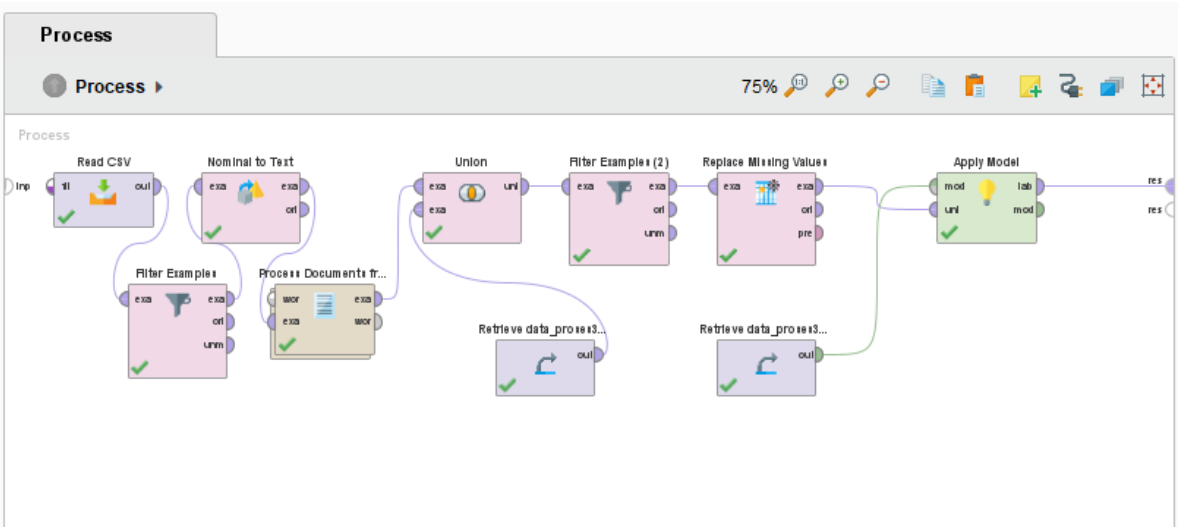
| Row No. | full_text |
|---|---|
| 1 | Drama Korea The Atypical Family Episode Sub Indo Jadwal Tayang dan Dimana Nonton |
| 2 | Okee knp si Korea buat drama ginian terus kalau soal SMA mesti gengan atau pembullyan |
| 3 | Sekarang belum ada kesibukan sih soalnya aku lagi liburan aja nih di Korea Tapi tanggal bakalan mulai syuting drama baru pastinya bakalan ... |
| 4 | Hierarchy ni best weyy nampaknya aku dah tukar selera drama korea ni |
| 5 | Thats why gue pengen banget bisa fasih bahasa korea Kalo mereka ngatain kita bisa balesinnya Jangan mau dipandang sebelah mata Pentin... |
| 6 | Imbas gagalnya Beberapa Drama Korea Platform OTT Ini Rugi Hingga Rumor Basis Produksi Akan Pindah Ke Jepang |
| 7 | Dia bisa bikin visa korea itu aja keliatan rekeningnya ada duit puluhan juta ngendep Klo emang tanggung jawab mah lunasin aja hutangmu jan... |
| 8 | Iyaya sayang banget padahal budget pembuatan drama ini gede banget : |
| 9 | Drama Korea Viral Hierarchy Bisa Ditonton di Neflix Inilah Sinopsis dan Daftar Pemerannya lewat tribunnews |
| 10 | adik aku yang lelaki dari tak minat korea harini pecah rekod boleh habiskan episod drama Hierarchy ni |
| 11 | urusanya bakal panjangsiap yang ngeroyok di jemput polisibakal ada drama korea nih |
| 12 | Jujur ini drama biasa aja bukan bales dendam yang benerÂ² bales dendam gtu Trs lebih banyak kisah cinta anak SMA nya Lead male nya kura... |
| 13 | Dapet apa sih dari hobi nonton drama korea Dapet tementemen baik yang bisa diajak liburan bareng |
| 14 | sini kak trusted amp |
| 15 | Para aktor Hierarchy muncul dalam video di channel YouTube W KOREA pada Kamis Di momen itu Roh Jeong Eui sambat sering dilupain Le... |

ExampleSet (2,185 examples, 0 special attributes, 1 regular attribute)

Gambar 6. Hasil Pembersihan Data

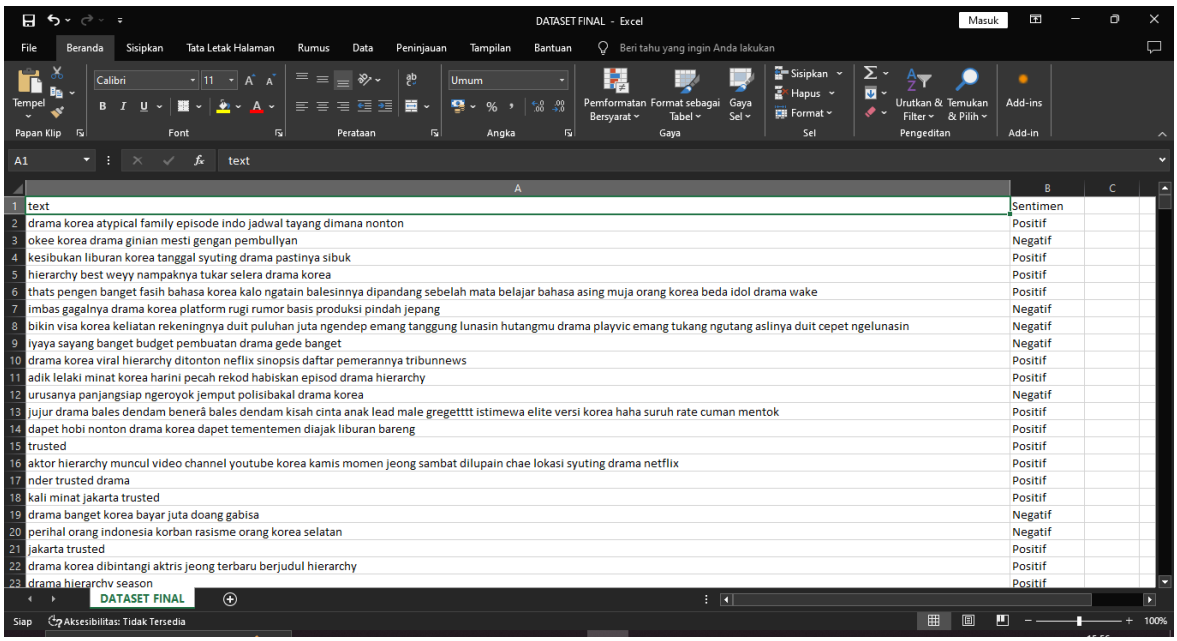
3) Labeling Data

Dalam tahap pelabelan data atau *labeling data*, 500 tweet dilabeli secara manual, sementara sisanya menggunakan model pelabelan otomatis. Data yang dilabeli secara manual ini digunakan sebagai data latih untuk memberi label sentimen dan untuk melatih model pelabelan otomatis. Hasil pelabelan ini mengelompokkan data ke dalam dua kategori, yaitu label "Positif" dan "Negatif".




Gambar 7. Proses Labeling

**Jurnal Indonesia : Manajemen Informatika dan Komunikasi**
https://journal.stmiki.ac.id
E-ISSN: 2723-7079 | P-ISSN: 2776-8074
Vol. 5 No. 3 (2024) | September
OPEN ACCESS
https://doi.org/10.35870/jimik.v5i3.997

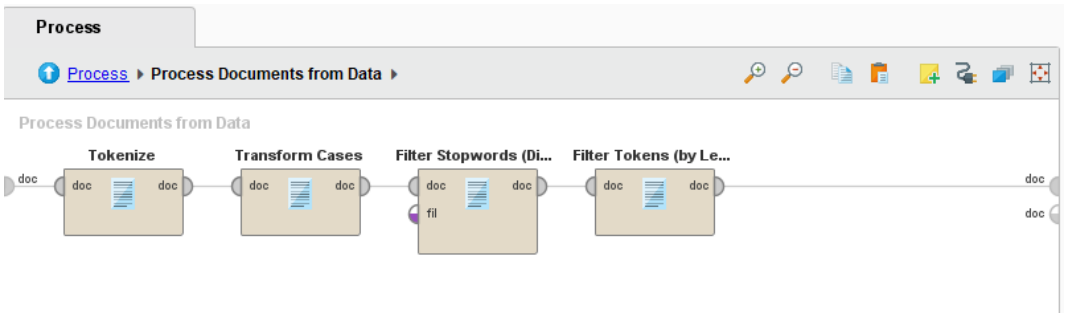



| | A | B |
|---|---|---|
| 1 | text | Sentimen |
| 2 | drama korea atypical family episode indo jadwal tayang dimana nonton | Positif |
| 3 | okee korea drama ginian mesti gengan pembullyan | Negatif |
| 4 | kesibukan liburan korea tanggal syuting drama pastinya sibuk | Positif |
| 5 | hierarchy best weyy nampaknya tukar selera drama korea | Positif |
| 6 | thats pengen banget fasih bahasa korea kalo ngatain balesinnya dipandang sebelah mata belajar bahasa asing muja orang korea beda idol drama wake | Positif |
| 7 | imbas gagalnya drama korea platform rugi rumor basis produksi pindah jepang | Negatif |
| 8 | bikin visa korea keliatan rekeningnya duit puluhan juta ngendep emang tanggung lunasin hutangmu drama playvic emang tukang ngutang aslinya duit cepet ngelunasin | Negatif |
| 9 | iyaya sayang banget budget pembuatan drama gede banget | Negatif |
| 10 | drama korea viral hierarchy ditonton neflix sinopsis daftar pemerannya tribunnews | Positif |
| 11 | adik lelaki minat korea harini pecah rekod habiskan episod drama hierarchy | Positif |
| 12 | urusanya panjangsiap ngeroyok jemput polisibakal drama korea | Negatif |
| 13 | jujur drama bales dendam benerâ bales dendam kisah cinta anak lead male gregetttt istimewa elite versi korea haha suruh rate cuman mentok | Positif |
| 14 | dapet hobi nonton drama korea dapet tementemen diajak liburan bareng | Positif |
| 15 | trusted | Positif |
| 16 | aktor hierarchy muncul video channel youtube korea kamis momen jeong sambat dilupain chae lokasi syuting drama netflix | Positif |
| 17 | nder trusted drama | Positif |
| 18 | kali minat jakarta trusted | Positif |
| 19 | drama banget korea bayar juta doang gabisa | Negatif |
| 20 | perihal orang indonesia korban rasisme orang korea selatan | Negatif |
| 21 | jakarta trusted | Positif |
| 22 | drama korea dibintangi aktris jeong terbaru berjudul hierarchy | Positif |
| 23 | drama hierarchy season | Positif |

Gambar 8. Hasil Labeling

3.1.4 Pemodelan (*Modelling*)

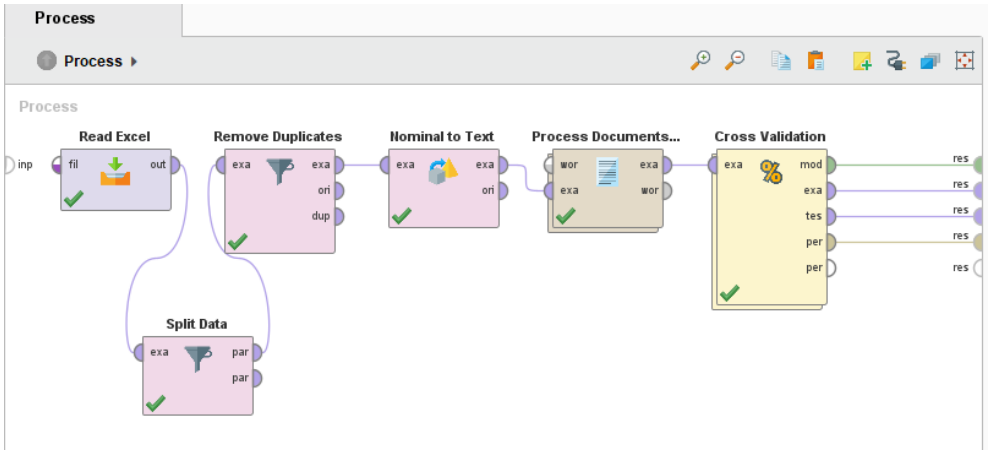
Modelling Pada tahap ini, peneliti akan melakukan preprocessing pada dataset untuk menyiapkan data yang bersih dan bebas dari noise, serta menghitung pembobotan kata yang akan digunakan dalam pemodelan. Proses text preprocessing ini dapat diperoleh 2.107 data tweet bersih yang siap digunakan. Pembobotan kata dilakukan menggunakan metode TF-IDF (*Term Frequency-Inverse Document Frequency*), yang berguna untuk menilai pentingnya sebuah kata dalam dokumen dibandingkan dengan seluruh corpus, sehingga menunjukkan relevansi kata dalam memahami isi dokumen tersebut..



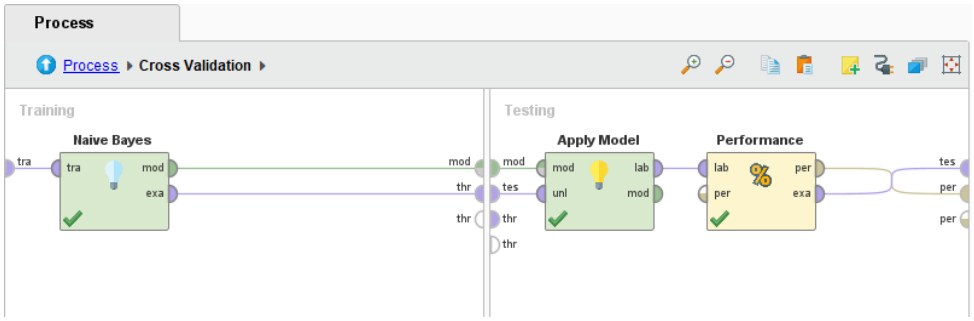

Gambar 9. Preprocessing Data

Pada tahap pemodelan, peneliti akan mengukur performa klasifikasi dengan menggunakan metode *split data*. Pada tahap ini, pengukuran performa klasifikasi akan dilakukan dengan menggunakan data yang telah dibagi. Proses pemodelan ini dilakukan menggunakan *software RapidMiner*.

Jurnal Indonesia : Manajemen Informatika dan Komunikasi
https://journal.stmiki.ac.id
E-ISSN: 2723-7079 | P-ISSN: 2776-8074
Vol. 5 No. 3 (2024) | September
OPEN ACCESS
https://doi.org/10.35870/jimik.v5i3.997




Gambar 10. Proses Modeling




Gambar 11. Proses Pengujian

3.1.5 Evaluasi (*Evaluation*)

Pada tahap ini akan dilakukan pengujian data dari tahapan modelling yang telah dibuat sebelumnya dengan menggunakan algoritma Naïve Bayes. Total dataset yang dipakai yaitu berjumlah 2.107 data. Berikut adalah hasil dari tahapan modeling yang dapat dilihat dari hasil perhitungan hasil RapidMiner yang ditunjukan pada gambar berikut.

PerformanceVector (Performance) | ExampleSet (Cross Validation)

Table View | Plot View

**accuracy: 84.33% +/- 2.30% (micro average: 84.33%)**

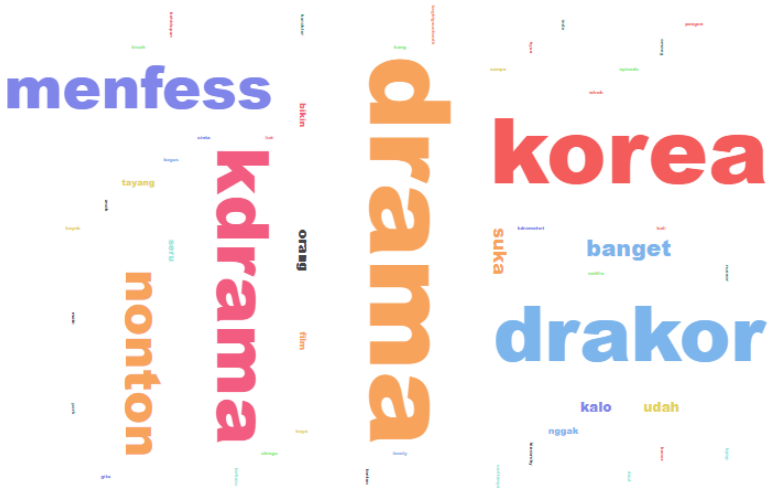
| | true Positif | true Negatif | class precision |
|---|---|---|---|
| pred. Positif | 1193 | 156 | 88.44% |
| pred. Negatif | 108 | 228 | 67.86% |
| class recall | 91.70% | 59.38% | |

Gambar 12. Hasil Proses Akurasi

Berdasarkan gambar di atas, dapat disimpulkan bahwa precision menunjukkan tingkat ketepatan data yang diprediksi positif terhadap jumlah data yang benar-benar positif, dengan persentase ketepatan sebesar 88,44%. Untuk data bersentimen negatif, precision memiliki nilai 67,86%. Recall (Specificity) untuk sentimen negatif adalah 59,38%, menunjukkan bahwa model cukup baik dalam menemukan

Jurnal Indonesia : Manajemen Informatika dan Komunikasi
https://journal.stmiki.ac.id
E-ISSN: 2723-7079 | P-ISSN: 2776-8074
Vol. 5 No. 3 (2024) | September
OPEN ACCESS
https://doi.org/10.35870/jimik.v5i3.997

data yang benar-benar negatif. Sementara itu, recall untuk sentimen positif mencapai 91,70%. Nilai accuracy model adalah 84,33%, menunjukkan bahwa algoritma yang digunakan dapat mengklasifikasikan sentimen dengan baik menggunakan data yang ada.




Gambar 13. Wordcloud Drama Korea

Wordcloud adalah visualisasi data yang digunakan untuk menampilkan frekuensi atau kepentingan kata-kata dalam sebuah teks.(Duei Putri *et al.*, 2022) Wordcloud ini menyusun kata-kata dengan ukuran yang berbeda, di mana kata yang lebih sering muncul atau lebih penting akan ditampilkan dalam ukuran yang lebih besar dibandingkan dengan kata-kata lainnya.

### 3.1.6 Penerapan (*Deployment*)

Deployment adalah tahap terakhir dalam pembuatan laporan hasil kegiatan dengan menggunakan metode CRIPS-DM. Pada tahap ini, laporan akhir disusun untuk menyajikan pengetahuan yang diperoleh atau pola yang ditemukan.

## 3.2 Pembahasan

Penelitian ini memanfaatkan pendekatan *Cross-Industry Standard Process for Data Mining* (CRISP-DM) untuk menganalisis sentimen masyarakat Indonesia terhadap drama Korea melalui platform media sosial *X* menggunakan algoritma *Naïve Bayes*. Penggunaan metode CRISP-DM terbukti efektif dalam memberikan struktur yang jelas dan sistematis untuk proses analisis data, terutama dalam konteks analisis sentimen (Alfandi Safira & Hasan, 2023). Pendekatan ini telah banyak digunakan dalam berbagai penelitian yang melibatkan pemrosesan data besar, khususnya yang berkaitan dengan opini publik di media sosial (Ghozali *et al.*, 2023; Mustofa & Novita, 2022).

Metode CRISP-DM memiliki tahapan yang memungkinkan penelitian ini untuk memetakan masalah dengan jelas dari awal hingga penerapan hasil. Tahapan pemahaman bisnis sangat krusial karena drama Korea atau *K-Drama* merupakan fenomena global yang sangat berpengaruh terhadap budaya populer di Indonesia. *K-Drama* tidak hanya menjadi tontonan hiburan, tetapi juga memicu diskusi publik yang intens di media sosial (Amelia *et al.*, 2022). Metode CRISP-DM membantu mengidentifikasi tujuan utama, yaitu memahami sentimen publik terhadap *K-Drama*, dengan fokus pada pengukuran akurasi, presisi, dan *recall* sentimen yang muncul.

Proses pengumpulan data dari *X* melalui teknik *crawling* dan pemanfaatan API terbukti berhasil dalam memperoleh data dalam jumlah besar (2.288 *tweet*). Ini sesuai dengan hasil penelitian lain yang juga menggunakan *crawling* data dari Twitter untuk analisis sentimen (Darwis *et al.*, 2021; Amrullah &

Jurnal Indonesia : Manajemen Informatika dan Komunikasi
https://journal.stmiki.ac.id
E-ISSN: 2723-7079 | P-ISSN: 2776-8074
Vol. 5 No. 3 (2024) | September
OPEN ACCESS
https://doi.org/10.35870/jimik.v5i3.997

Pane, 2023). Proses pembersihan data yang dilakukan dengan menghilangkan kata-kata yang tidak relevan, simbol, tautan, dan *hashtag* berperan penting dalam meningkatkan kualitas data untuk analisis sentimen. Langkah-langkah pembersihan ini penting untuk memastikan bahwa data yang dianalisis adalah data yang benar-benar representatif dan bebas dari *noise* yang dapat mengurangi akurasi hasil (Rina Noviana & Isram Rasal, 2023).

Algoritma *Naïve Bayes* digunakan dalam penelitian ini karena keakuratannya dalam melakukan klasifikasi sentimen secara cepat. Hasil penelitian menunjukkan bahwa algoritma ini memberikan hasil yang cukup baik dengan akurasi sebesar 84,33% dan *precision* untuk data positif mencapai 88,44%. Hasil ini sejalan dengan temuan dalam penelitian lain yang menggunakan *Naïve Bayes* untuk analisis sentimen, yang juga mencatat performa tinggi dalam mengklasifikasikan data (Amelia *et al.*, 2022; Alfandi Safira & Hasan, 2023). Namun, nilai *recall* untuk sentimen negatif lebih rendah, hanya 59,38%, yang menunjukkan bahwa algoritma ini memiliki keterbatasan dalam mendeteksi semua data yang benar-benar negatif, sebagaimana juga ditemukan dalam penelitian lain yang menggunakan metode serupa (Sholeha *et al.*, 2022).

Penggunaan *wordcloud* dalam penelitian ini membantu memvisualisasikan kata-kata yang paling sering muncul dalam diskusi terkait *K-Drama* di media sosial. *Wordcloud* menyusun kata-kata berdasarkan frekuensinya, memberikan wawasan cepat tentang topik yang paling banyak dibicarakan oleh pengguna *X* (Duei Putri *et al.*, 2022). Teknik visualisasi ini juga digunakan dalam penelitian lain untuk menyoroti tema-tema kunci dalam percakapan media sosial (Slamet *et al.*, 2022).

Hasil analisis sentimen dapat digunakan untuk memahami opini publik mengenai drama Korea, sehingga membantu produsen konten menciptakan program atau produk yang lebih sesuai dengan preferensi dan ekspektasi audiens (Purnajaya *et al.*, 2022). Dari sisi pemasaran, pemahaman yang lebih mendalam tentang sentimen publik terhadap *K-Drama* dapat membantu dalam strategi pemasaran produk yang menggunakan *K-Drama* sebagai elemen kampanye (Rahmawati *et al.*, 2017). Meskipun penelitian ini menghasilkan temuan yang relevan, ada beberapa keterbatasan yang harus diakui. Salah satu keterbatasan utama adalah jumlah data yang relatif kecil, meskipun 2.107 data yang telah dibersihkan cukup memadai, namun dataset yang lebih besar dapat memberikan hasil yang lebih baik (Asri *et al.*, 2024). Selain itu, *recall* yang lebih rendah pada sentimen negatif menunjukkan bahwa ada peluang untuk mengembangkan model yang lebih kuat dalam mendeteksi sentimen negatif secara lebih akurat. Penelitian selanjutnya dapat mengeksplorasi algoritma lain, seperti *Support Vector Machine* atau *Random Forest*, yang mungkin memberikan hasil yang lebih baik dalam analisis sentimen (Amrullah & Pane, 2023).

## 4. Kesimpulan

Penelitian ini berhasil menunjukkan bahwa masyarakat Indonesia memiliki pandangan yang beragam terhadap drama Korea (*K-Drama*), sebagaimana tercermin dalam diskusi di media sosial *X*. Melalui metode analisis sentimen dengan algoritma *Naïve Bayes*, penelitian ini mampu mengidentifikasi sentimen positif dan negatif terkait *K-Drama* dengan tingkat akurasi yang cukup tinggi. Penggunaan metode *Cross-Industry Standard Process for Data Mining* (CRISP-DM) dalam penelitian ini memberikan kerangka kerja yang sistematis dan terstruktur untuk pengumpulan, pengolahan, dan analisis data. Hasil pengujian model menunjukkan akurasi sebesar 84,33%, yang mengindikasikan bahwa algoritma *Naïve Bayes* efektif dalam mengklasifikasikan sentimen publik. Meskipun demikian, terdapat beberapa kelemahan dalam mendeteksi sentimen negatif, yang tercermin dari nilai *recall* untuk data negatif yang lebih rendah. Namun, secara keseluruhan, model yang dibangun telah terbukti handal dalam memberikan hasil yang relevan dan akurat terkait opini masyarakat Indonesia terhadap *K-Drama*.

E-ISSN: 2723-7079 | P-ISSN: 2776-8074
Vol. 5 No. 3 (2024) | September
OPEN ACCESS
https://doi.org/10.35870/jimik.v5i3.997

## 5. Ucapan Terima Kasih

Penulis mengucapkan terima kasih kepada semua pihak yang telah membantu dalam penelitian ini, baik secara langsung maupun tidak langsung. Penulis menyadari bahwa masih banyak kekurangan dalam penelitian ini. Oleh karena itu, penulis mengharapkan kritik dan saran yang membangun dari pembaca untuk menyempurnakan penelitian ini di masa depan.

## 6. Daftar Pustaka

Alfandi Safira, & Hasan, F. N. (2023). Analisis sentimen masyarakat terhadap Paylater menggunakan metode Naive Bayes Classifier. *ZONAsi: Jurnal Sistem Informasi, 5*(1), 59–70. https://doi.org/10.31849/zn.v5i1.12856

Amelia, R., Darmansah, D., Prastiwi, N. S., & Purbaya, M. E. (2022). Implementasi algoritma Naive Bayes terhadap analisis sentimen opini masyarakat Indonesia mengenai drama Korea pada Twitter. *JURIKOM (Jurnal Riset Komputer), 9*(2), 338. https://doi.org/10.30865/jurikom.v9i2.3895

Amrullah, M. S., & Pane, S. F. (2023). Analisis sentiment masyarakat terhadap kebijakan polisi tilang manual di Indonesia dengan SVM (M. N. Fauzan, Ed.). Penerbit Buku Pedia.

Asri, Y., Kuswardani, D., Horhoruw, L. F. M., & Ramadhana, S. A. (2024). *Machine Learning & Deep Learning: Analisis Sentimen Menggunakan Ulasan Aplikasi.* Uwais Inspirasi Indonesia.

Darwis, D., Siskawati, N., & Abidin, Z. (2021). Penerapan algoritma Naive Bayes untuk analisis sentimen review data Twitter BMKG Nasional. *Jurnal Tekno Kompak, 15*(1), 131. https://doi.org/10.33365/jtk.v15i1.744

Duei Putri, D., Nama, G. F., & Sulistiono, W. E. (2022). Analisis sentimen kinerja Dewan Perwakilan Rakyat (DPR) pada Twitter menggunakan metode Naive Bayes Classifier. *Jurnal Informatika Dan Teknik Elektro Terapan, 10*(1), 34–40. https://doi.org/10.23960/jitet.v10i1.2262

Ghozali, M. I., Sugiharto, W. H., & Iskandar, A. F. (2023). Analisis sentimen pinjaman online di media sosial Twitter menggunakan metode Naive Bayes. *KLIK: Kajian Ilmiah Informatika Dan Komputer, 3*(6), 1340–1348. https://doi.org/10.30865/klik.v3i6.936

Mustofa, A., & Novita, R. (2022). Klasifikasi sentimen masyarakat terhadap pemberlakuan pembatasan kegiatan masyarakat menggunakan *text mining* pada Twitter. *Building of Informatics, Technology and Science (BITS), 4*(1), 200–208. https://doi.org/10.47065/bits.v4i1.1628

Purnajaya, A. R., Lieputra, V., Tayanto, V., & Salim, J. G. (2022). Implementasi *text mining* untuk mengetahui opini masyarakat tentang climate change. *Journal of Information System and Technology, 3*(3), 36. https://doi.org/10.37253/joint.v3i3.7337

Rachman, R., & Oktavianti, R. (2021). Pengaruh kepercayaan konsumen terhadap loyalitas pelanggan dalam penggunaan sistem pembayaran online (Survei pengguna produk Unipin). *Prologia, 5*(1), 148. https://doi.org/10.24912/pr.v5i1.8200

E-ISSN: 2723-7079 | P-ISSN: 2776-8074
Vol. 5 No. 3 (2024) | September
OPEN ACCESS
https://doi.org/10.35870/jimik.v5i3.997

Rahmawati, A., Marjuni, A., & Zeniarja, J. (2017). Analisis sentimen publik pada media sosial Twitter terhadap pelaksanaan pilkada serentak menggunakan algoritma *Support Vector Machine*. *CCIT Journal, 10*(2), 197–206. https://doi.org/10.33050/ccit.v10i2.539

Rina Noviana, & Isram Rasal. (2023). Penerapan algoritma Naive Bayes dan SVM untuk analisis sentimen Boy Band BTS pada media sosial Twitter. *Jurnal Teknik Dan Science, 2*(2), 51–60. https://doi.org/10.56127/jts.v2i2.791

Sholeha, E. W., Yunita, S., Hammad, R., Hardita, V. C., & Kaharuddin, K. (2022). Analisis sentimen pada agen perjalanan online menggunakan Naïve Bayes dan K-Nearest Neighbor. *JTIM: Jurnal Teknologi Informasi Dan Multimedia, 3*(4), 203–208. https://doi.org/10.35746/jtim.v3i4.178

Slamet, R., Gata, W., Novtariany, A., Hilyati, K., & Jariyah, F. A. (2022). Analisis sentimen Twitter terhadap penggunaan artis Korea Selatan sebagai brand ambassador produk kecantikan lokal. *INTECOMS: Journal of Information Technology and Computer Science, 5*(1), 145–153. https://doi.org/10.31539/intecoms.v5i1.3933

Taufiqi, A. M., & Nugroho, A. (2023). Sentimen pengguna Twitter mengenai isu kebocoran data dengan algoritma Naïve Bayes. *Jurnal Nasional Ilmu Komputer, 4*(1), 1–11. https://doi.org/10.47747/jurnalnik.v4i1.1091.