

Analisis Sentimen Terhadap Opini Proyek Kereta Cepat Menggunakan Metode *Naïve Bayes Classifier*

Jabar Sanjaya¹, Tukino², Bayu Priyatna³, Shofa Shofia Hilabi⁴

¹²³⁴Sistem Informasi, Fakultas Ilmu Komputer, Universitas Buana Perjuangan Karawang

*¹si20.jabarsanjaya@mhs.ubpkarawang.ac.id, ²tukino@ubpkarawang.ac.id, ³bayu.priyatna@ubpkarawang.ac.id,

⁴shofa.hilabi@ubpkarawang.ac.id

Abstract

Twitter is one of the many social media that is used by most people. Twitter usage is generally used to post something, such as having an opinion about the country's infrastructure for example. As for one of the country's infrastructure that is being discussed is the fast train project, people have many views about this project. To find out people's opinions about it, Sentiment Analysis will be conducted. This study aims to determine the application of Sentiment Analysis using the *Naïve Naves Classifier* method on public opinion about the fast train project. The research starts from the data retrieval process using crawling techniques and getting approximately 2007 tweets that discuss the fast train project. Which then the data from the sentiment results in preprocessing, then each data is labeled. From the labeling results, the naïve bayes classifier method is applied. The results of the *Naïve Bayes* model evaluation are 81% Precision, 81% Recall, 81% F1-score. and then Accuracy of 81%. The final value produced is 57.23% of people comment or have a positive opinion about the fast train project and 42.77% have a negative opinion. From the results of this study it can be concluded that the community is more dominant in positive opinion about the fast train project than negative opinion, which shows that the majority of people can accept the fast train project.

Keywords: Twitter, Fast Train, Sentiment Analysis, Opinions, *Naïve Bayes Classifier*

Abstrak

Twitter merupakan satu dari banyaknya media sosial yang digunakan oleh hampir Sebagian besar masyarakat. Penggunaan twitter umumnya digunakan untuk memposting sesuatu, seperti beropini tentang infrastruktur negara misalnya. Adapun salah satu infrastruktur negara yang sedang ramai diperbincangkan adalah proyek kereta cepat, masyarakat memiliki banyak pandangan tentang proyek ini. Untuk mengetahui opini masyarakat, tentang hal itu maka akan dilakukan Sentimen Analisis. Penelitian ini bertujuan untuk mengetahui penerapan Sentimen Analisis menggunakan metode *Naïve Naves Classifier* pada opini masyarakat tentang proyek kereta cepat. Penelitian dimulai dari proses pengambilan data menggunakan teknik *crawling* dan mendapatkan kurang lebih 2007 tweet yang membahas proyek kereta cepat. Yang kemudian data dari hasil sentimen di *preprocessing*, lalu setiap data di labeling. Dari hasil labeling diterapkan metode *naïve bayes classifier*. Adapun hasil evaluasi model *Naïve Bayes* yaitu, Presisi 81%, Recall 81%, F1-score 81%. dan kemudian Akurasi sebesar 81%. Nilai akhir yang dihasilkan yaitu 57,23% masyarakat berkomentar atau beropini positif tentang proyek kereta cepat dan 42,77% beropini negatif. Dari hasil penelitian ini dapat disimpulkan bahwa masyarakat lebih dominan beropini positif tentang proyek kereta cepat daripada beropini negatif, yang mana ini menunjukkan bahwa mayoritas masyarakat dapat menerima adanya proyek kereta cepat.

Kata kunci: Twitter, Kereta Cepat, Sentimen Analisis, Opini, *Naïve Bayes Classifier*

©This work is licensed under a Creative Commons Attribution - ShareAlike 4.0 International License

1. Pendahuluan

Informasi dan data selalu hadir dalam langkah setiap orang. Penyebaran informasi bisa melalui banyak hal salah satunya melalui jejaring sosial. Dari banyaknya jenis informasi yang diterima oleh masyarakat, informasi mengenai pertumbuhan infrastruktur merupakan salah satu informasi yang paling banyak dinantikan karena berita dan Informasi ini penting untuk masyarakat demi mengetahui perkembangan sebuah negara.

Rata-rata masyarakat mengkonsumsi informasi dan berita dari Jejaring sosial media untuk pertama kalinya

di bandingkan dengan sumber tradisional seperti koran, majalah, atau media baca tradisional lainnya. Orang begitu mudah mempercayai apa yang mereka lihat di media sosial, yang membuat masyarakat mudah terpengaruh rumor dan berita *hoax*. Oleh sebab itu sangat penting untuk mendeteksi dan menyaringnya untuk menghindari penyebaran informasi yang tidak akurat [1]. Banyak informasi yang menyebar di jejaring social. Mulai dari informasi politik, bisnis, hiburan, dll. Begitu pula dengan informasi atau berita mengenai infrastruktur teknologi.

Kereta cepat merupakan sebuah proyek negara yang dimulai pada januari tahun 2016 dan selesai pada

oktober tahun 2023. Proyek ini diharapkan dapat menekan jumlah kemacetan rute tol dan jalan raya serta padatannya kereta regular. Proyek yang menghabiskan triliunan ini memiliki banyak tanggapan dari masyarakat, ada yang pro maupun kontra.

Kebanyakan orang lebih sering mengungkapkan pendapat serta opininya melalui media sosial dari pada dilakukan secara langsung. Media sosial yang biasa digunakan dalam mengungkapkan opini serta pendapat yang terkenal adalah *twitter* [2]. *Twitter* merupakan salah satu dari media sosial yang banyak digemari oleh masyarakat sebab kemudahan dan kesederhanaan dalam menggunakannya, dan pengguna dapat leluasa mengutarakan pandangan atau pendapatnya. Pada kasus-kasus sebelumnya banyak peneliti menggunakan data survei dari *twitter*, termasuk teks berbahasa Indonesia calon gubernur DKI Jakarta 2017, calon presiden 2014, dan penggunaan angkutan umum darat dalam dan luar kota [1]. Saat memberikan opini masyarakat cenderung akan menghasilkan 2 opini, yang pertama Positif, disini masyarakat memberikan suara dukungan atau pro terhadap suatu hal, terkadang mereka memberikan pujian terhadap hal tersebut, yang kedua Negatif, masyarakat disini memberikan suara penolakan atau kontra terhadap suatu hal, terkadang dalam opini yang disampaikan terdapat cacian dan makian yang dilontarkan. Apalagi dengan media sosial masyarakat dengan mudah menyampaikan pendapat mereka tanpa harus ketakutan identitas mereka akan diketahui oleh orang lain[3].

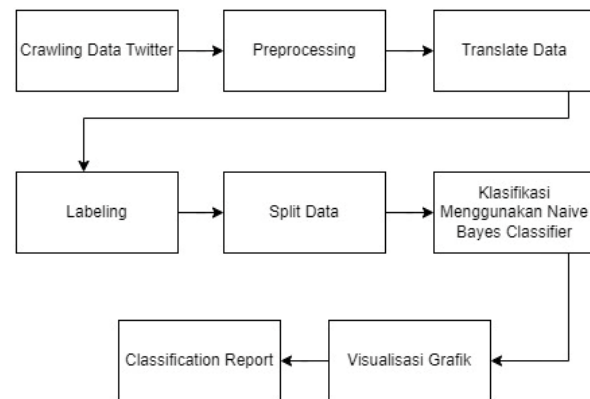
Penerapan Analisis Sentimen dengan metode *Naïve Bayes* ini telah dilakukan beberapa kali oleh penulis sebelumnya. Adapun penulis pertama Adittia Agustian yang menerapkan analisis sentimen menggunakan metode *naïve bayes* pada penggunaan kendaraan listrik. Penelitian ini dilakukan dengan mengumpulkan data tweet pengguna *twitter* yang membahas tentang penggunaan kendaraan listrik, dalam penelitiannya telah dihasilkan 1589 tweet yang membahas mengenai penggunaan kendaraan listrik. Informasi informasi setelah melakukan beberapa proses menghasilkan tingkat keakurasian sebesar 80%, presisi 82%, dan recall 80% [4]. Penelitian berikutnya yang dijadikan acuan adalah penelitian yang ditulis oleh Tri Sundara dengan judul penelitian “*Naïve Bayes Classifier* untuk Analisis Sentimen Isu Radikalisme”, yang mana tujuan dari penelitian ini adalah untuk mengklasifikasi opini publik berdasarkan data kelas positif dan negatif menggunakan metode *Naïve Bayes Classifier* dengan hasil akurasi 86% [5].

Di dalam penelitian ini, penulis akan mengimplementasikan analisis sentimen pada opini masyarakat tentang protek kereta cepat. Dalam melakukan sentimen tersebut data yang akan digunakan adalah data hasil *Crawling Data* menggunakan tools yang bernama *Twitter Harvest*, dan menghasilkan data mentah yang berjumlah 929 tweet. Lalu data tersebut akan melalui beberapa proses

sebelum di eksekusi dengan metode *Naïve Bayes Classifier* menggunakan Bahasa pemrograman python. Yang nantinya akan diketahui bagaimana hasil penerapan data sentimen menggunakan metode *naïve bayes classifier*, untuk klasifikasi data opini masyarakat tentang proyek kereta cepat.

2. Metode Penelitian

2.1 Alur Penelitian



Gambar 1. Alur Penelitian

Pada gambar di atas menggambarkan alur penelitian yang akan dilakukan, adapun alur yang akan dijelaskan mulai dari *crawling data*, *preprocessing*, *translating*, *labeling*, klasifikasi menggunakan *naïve bayes*, visualisasi data menggunakan *chart*, dan yang terakhir *classification report*.

2.2 Pengumpulan Data

Pada tahap ini dilakukan pengumpulan data yang bertujuan untuk menganalisa sekaligus mendapatkan data dan informasi yang berhubungan dengan penelitian kali ini [6]. Data yang diperoleh merupakan *tweet* pengguna mengenai proyek kereta cepat.

Pengumpulan data dilakukan untuk menganalisa dan memperoleh informasi yang berkaitan dengan penelitian. Data yang diperoleh adalah *tweet* masyarakat mengenai proyek kereta cepat pada media sosial *Twitter*. Data *tweet* didapat dengan cara *crawling*. Tool yang digunakan dalam proses *Crawling Data* adalah *Tweet Harvest*. Proses ini dilakukan dengan cara mengambil *tweet* yang berhubungan dengan proyek kereta cepat. Data yang didapat memiliki beberapa kolom, tetapi kolom yang digunakan hanya kolom yang berisi *tweet*, karena jika semua kolom digunakan, berpotensi terhambatnya proses penelitian [7], [8].

Pada penelitian kali ini data yang diproses menggunakan metode *naïve bayes classifier* terdapat 2 kategori, yaitu:

- a. Positif, merupakan *tweet* yang mengandung kata kata, dan arti yang positif, Adapun biasanya kata postifi di sini itu adalah kata kata yang mendukung dan

memuji suatu pihak atau setuju dengan sebuah pernyataan tertentu.

b. Negatif, merupakan *tweet* yang mengandung kata kata yang mengandung dan memiliki arti negatif, biasanya kata kata negatif ini digambarkan dengan sesuatu yang buruk, seperti umpatan atau kata kata yang mengandung penolakan terhadap suatu hal.

2.3 Preprocessing

Data yang akan diproses untuk analisis sentimen akan dibersihkan dari noise atau data-data yang tidak penting seperti duplikasi data, hastag(@), tautan atau link, tanda baca, angka, penggunaan bahasa slang, singkatan dan kata yang tidak baku [9]. Tahapan *preprocessing* terdiri dari beberapa tahap.

2.3.1 Text Clean

Text Clean merupakan proses yang bertujuan untuk menghapus data yang tidak dibutuhkan dari dokumen saat proses sentimen. Setelah data telah di proses melalui *text clean*, data siap diproses pada tahap selanjutnya. Adapun data yang tidak dibutuhkan bisa berupa angka, simbol, tanda baca, dll [10].

2.3.2 Case Folding

Case Folding merupakan salah satu tahap yang masih termasuk dalam *text preprocessing*. *Case folding* berkerja dengan merubah semua kata dan kalimat pada dokumen menjadi huruf kecil. Salah satu tujuan dari dilakukannya *case folding* ialah untuk mengatasi masalah penulisan yang tidak benar pada sebuah kalimat dalam menggunakan huruf besar dan huruf kecil yang tidak konsisten. Jika, semua dokumen diubah menjadi huruf kecil maka akan lebih mudah dalam melakukan proses analisis [11].

2.3.3 Stemming

Stemming ialah proses yang berguna untuk mengubah sebuah kata kembali ke bentuk dasar. Tujuannya ialah untuk memudahkan dalam mengelompokkan sekaligus menganalisis data tersebut [12]. Dalam python proses *stemming* biasanya menggunakan *library* Sastrawi, *library* ini hanya dikhususkan untuk bahasa indonesia. Jadi sebelum melakukan *stemming*, pastikan Bahasa yang akan diproses adalah bahasa indonesia [9].

2.3.4 Stopword

Pada proses *stopword* kata-kata yang tidak memiliki arti dan makna akan dihilangkan dari data yang akan diproses, karena akan hanya akan memperbesar ukuran data. Dalam melakukan *stopword* kata yang akan dihapus mengacu pada kamus yang telah disiapkan sebelumnya. Kamus *stopword* bisa dibuat sendiri atau bisa juga diunduh dari *website* tertentu [13].

2.4 Translate

Mengubah bahasa dari data yang akan diolah dimaksudkan memudahkan penulis dalam melakukan proses penelitian yang dilakukan. karena proses *labeling* akan

menggunakan *Textblob* yang berbasis *Lexicon Based* yang mana *library* ini hanya bisa memproses bahasa inggris.

2.5 Labeling

Tahap ini dilakukan untuk menentukan kategori yang akan diberikan kepada data. Data akan dipisah menjadi dua kategori yaitu positif dan negatif [14]. Pada tahap ini, *labeling* dilakukan menggunakan *Library Textblob* yang mana *library* ini menerapkan metode *Lexicon Based* untuk melakukan analisis sentiment pada data tertentu, dengan cara menganalisis pola kalimat untuk menghasilkan kategori sentimen.

2.6 Splitting Data

Split Data diperlukan untuk membagi keseluruhan data menjadi 2 bagian, yang pertama data latih, yang mana data ini dibutuhkan untuk melatih data model menggunakan metode tertentu. dan yang kedua data uji, yang mana data ini dibutuhkan untuk menguji kinerja data berdasarkan hasil dari model pelatihan sebelumnya. Pembagian data ini biasanya 80/20 atau 70/30. 80% untuk data latih atau *train* dan 20% untuk data uji atau *testing*. Dalam model klasifikasi, *Split Data* mempengaruhi hasil, oleh karena itu Data *Training* harus lebih besar dari Data *Testing*, karena hasil dipengaruhi oleh banyaknya data latih yang diberikan kepada model.

2.7 Naïve Bayes Classifier

Naïve Bayes Classifier adalah metode yang menggunakan distribusi kata kunci dalam membandingkan sebuah dokumen untuk menentukan nilai probabilitas dari suatu hipotesis setelah bukti yang tunjukan ada [15], [16]. *Naïve Bayes Classifier* didasarkan pada teorema bayes untuk menjelaskan bahwa setiap atribut adalah independen satu sama lain.

2.8 Visualisasi

Visualisasi disini yaitu menampilkan data dalam bentuk gambar. Data harus di *visualisasi* kan untuk kemudahan dalam mencerna informasi data yang akan disajikan dan memudahkan mengetahui kesalahan pada data yang disajikan. Banyak cara untuk mem *visualisasikan* data, salah satu caranya menggunakan *line chart*, *plot char*, *bar chart*, dan masih banyak lagi.

2.9 Classification Report

Laporan klasifikasi merupakan metrik evaluasi kerja pada model klasifikasi, yang mana di dalam *classification report* menunjukkan rincian tentang bagaimana performa model uji berjalan pada suatu metode [17]. Dalam *classification report* terdapat beberapa nilai penting yang akan ditampilkan, diantaranya :

a. *Precision*, merupakan nilai yang menjelaskan seberapa akurat sebuah model klasifikasi dalam

menentukan prediksi positif. Adapun rumus yang digunakan .

$$Precision = \frac{\text{True Positive}}{\text{True Positive} + \text{False Positive}} \quad (1)$$

b. *Recall*, merupakan nilai yang mengidentifikasi kemampuan model dalam menemukan nilai positif. Adapun rumus yang digunakan.

$$Recall = \frac{\text{True Positive}}{\text{True Positive} + \text{False Negative}} \quad (2)$$

c. *F1-Score*, merupakan nilai yang nilai nya merupakan penggabungan antara *precision* dan *recall*. Atau bisa disebut nilai rata-rata atau *mean* antara *precision* dan *recall*. Adapun rumus yang digunakan.

$$F1 \text{ Score} = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (3)$$

d. *Accuracy*, merupakan nilai yang berfungsi sebagai patokan seberapa baik model klasifikasi dalam memprediksi keseluruhan nilai. Adapun rumus yang digunakan [8].

$$Accuracy = \frac{TP+TN}{TP+FP+TN+FN} \quad (4)$$

3. Hasil dan Pembahasan

3.1 Pengumpulan Data

Di gunakan kata kunci “proyek kereta cepat” pada saat proses *crawling data* dan data yang dihasilkan dalam waktu 31 menit ialah 2007 tweet. Adapun data yang didapat meliputi *created_ad*, *id_str*, *full_text*, *_qoute_count*, *reply_count*, *favorite_count*, *lang*, *user_id*, *conversation_id*, *username*, dan *tweet_url*. Adapun data yang digunakan hanya tweet nya saja yang berada pada kolom *full_text*, hasil *crawling data* dapat dilihat pada tabel 1.

Tabel 1. Hasil Pengumpulan Data

Full_Text
@BoyRembaya @PartaiSocmed @jokowi Hambalang hanya secuil d banding proyek2 mangkrak era jokowi liatlah brp bandara d bangun ga berfungsi asal bangun2 tp hbs itu terbengkalai, itu baru bandara pelabuhan IKN kereta api cepat bandung itu hanya proyek egois gaya2an tp sgt membebaskan rakyat..
@lluvianeiva @NnAini17 @bangmenjie @Wicaksono_J @VIVAcoid @aniesbaswedan Iya dlu gw juga ga suka sama proyek tol trans jawa dan Kereta Cepat. Tapi makin kesini, makin kesini orang banyak pada paket itu.
@ahmahmNiam @addthoriq @AdyP93718291 @VIVAcoid Bukan tidak meyakini tp jokowi lebih pilih kerja sama dgn china, semua proyek pengennya dr china krn china sdh paham maunya pejabat kereta cepat jkt - bandung awalnya dgn jepang dibatalkan demi china, kasus BTS contoh perusahaan china menguntungkan pejabat RI
@KeretaCepatID Dalam beberapa tahun terakhir, proyek kereta api berkecepatan tinggi terus dikembangkan di seluruh dunia. Jika menilik sejarah, kereta ekspres terbatas pertama kali dikembangkan Jepang pada tahun 1964 dengan nama Shinkansen.

@ferrykoto @aniesbaswedan @andrinof_a_ch Masak sih belum kapok sama kereta cepat..? Proyek mercusuar gagal total ninggalin utang beberapa generasi.. Padahal dulu janjinya ANGARAN bukan dari uang rakyat, tapi lama2 pakai uang rakyat juga... Capek bos rakyat di bodohi terus dengan akal2an proyek strategis nasional...

@erasmus70 Prmbangunan yg sudah jalan ya pasti ada evaluasi berkala. Tiap2 proyek juga gitu. Skg tinggal lihat kualitas evaluatornya. Misal dulu evaluasi proses pembangunan kereta cepat. Jonan g setuju dilanjut, berarti ada mekanisme untuk stop proyek yg g feasible. Eeh dipecat

@KompasTV Import bahan pangan dilakukan setelah bendungan dibuat, untuk apa bendungan, embung dan irigasi itu ?? Proyek asal2an gak berguna !!! Banyak Jalan toll, pelabuhan laut dan udara, Kereta cepat dibangun tanpa study kelayakan lebih dulu dan akhirnya gak berguna buang2 duit negara !!

3.2 Preprocessing

Setelah melakukan *crawling data*, dilakukan proses *preprocessing*. Terdapat beberapa tahap, yang pertama *text clean*, *case folding*, *stemming*, *tokenizing* dan *stopword*. Hasil dari *preprocessing* dapat dilihat pada tabel di bawah ini.

Tabel 2. Hasil Text Clean

Full_Text	Text Clean
@BoyRembaya @PartaiSocmed @jokowi Hambalang hanya secuil di banding proyek2 mangkrak era jokowi liatlah brp bandara d bangun ga berfungsi asal bangun2 tp hbs itu terbengkalai itu baru bandara pelabuhan IKN kereta api cepat bandung itu hanya proyek egois gaya2an tp sgt membebaskan rakyat..	Hambalang hanya secuil d banding proyek2 mangkrak era jokowi liatlah brp bandara d bangun ga berfungsi asal bangun2 tp hbs itu terbengkalai itu baru bandara pelabuhan IKN kereta api cepat bandung itu hanya proyek egois gaya2an tp sgt membebaskan rakyat
@lluvianeiva @NnAini17 @bangmenjie @Wicaksono_J @VIVAcoid @aniesbaswedan Iya dlu gw juga ga suka sama proyek tol trans jawa dan Kereta Cepat. Tapi makin kesini, makin kesini orang banyak pada paket itu.	Iya dlu gw juga ga suka sama proyek tol trans jawa dan Kereta Cepat Tapi makin kesini makin kesini orang banyak pada paket itu
@ahmahmNiam @addthoriq @AdyP93718291 @VIVAcoid Bukan tidak meyakini tp jokowi lebih pilih kerja sama dgn china, semua proyek pengennya dr china krn china sdh paham maunya pejabat kereta cepat jkt - bandung awalnya dgn jepang dibatalkan demi china, kasus BTS contoh perusahaan china menguntungkan pejabat RI	Bukan tidak meyakini tp jokowi lebih pilih kerja sama dgn china semua proyek pengennya dr china krn china sdh paham maunya pejabat kereta cepat jkt bandung awalnya dgn jepang dibatalkan demi china kasus BTS contoh perusahaan china menguntungkan pejabat RI
@KeretaCepatID Dalam beberapa tahun terakhir, proyek kereta api berkecepatan tinggi terus dikembangkan di seluruh dunia. Jika menilik sejarah, kereta ekspres terbatas pertama kali dikembangkan Jepang pada tahun 1964 dengan nama Shinkansen.	Dalam beberapa tahun terakhir proyek kereta api berkecepatan tinggi terus dikembangkan di seluruh dunia Jika menilik sejarah kereta ekspres terbatas pertama kali dikembangkan Jepang pada tahun dengan nama Shinkansen
@ferrykoto @aniesbaswedan @andrinof_a_ch Masak sih belum kapok sama kereta cepat Proyek	Masak sih belum kapok sama kereta cepat Proyek

belum kapok sama kereta cepat..? Proyek mercusuar gagal total ninggalin utang beberapa generasi.. Padahal dulu janjinya ANGGARAN bukan dari uang rakyat, tapi lama2 pakai uang rakyat juga... Capek bos rakyat di bodohi terus dengan akal2an proyek strategis nasional...	mercusuar gagal total ninggalin utang beberapa generasi Padahal dulu janjinya ANGGARAN bukan dari uang rakyat tapi lama2 pakai uang rakyat juga Capek bos rakyat di bodohi terus dengan akal2an proyek strategis nasional
--	---

Text Clean pada preprocessing berfungsi untuk menghilangkan angka dan simbol tertentu yang tidak relevan digunakan untuk analisis sentimen.

Tabel 3. Hasil Case Folding	
Text Clean	Case Folding
Hambalang hanya secuil d banding proyek2 mangkrak era jokowi liatlah brp bandara d bangun ga berfungsi asal bangun2 tp hbs itu terbengkalai itu baru bandara pelabuhan IKN kereta api cepat bandung itu hanya proyek egois gaya2an tp sgt membebankan rakyat	hambalang hanya secuil d banding proyek2 mangkrak era jokowi liatlah brp bandara d bangun ga berfungsi asal bangun2 tp hbs itu terbengkalai itu baru bandara pelabuhan ikn kereta api cepat bandung itu hanya proyek egois gaya2an tp sgt membebankan rakyat
Iya dulu gw juga ga suka sama proyek tol trans jawa dan Kereta Cepat Tapi makin kesini makin kesini orang banyak pada paket itu	iya dulu gw juga ga suka sama proyek tol trans jawa dan kereta cepat tapi makin kesini makin kesini orang banyak pada paket itu
Bukan tidak meyakini tp jokowi lebih pilih kerja sama dgn china semua proyek pengennya dr china krn china sdh paham maunya pejabat kereta cepat jkt bandung awalnya dgn jepang dibatalkan demi china kasus BTS contoh perusahaan china menguntungkan pejabat RI	bukan tidak meyakini tp jokowi lebih pilih kerja sama dgn china semua proyek pengennya dr china krn china sdh paham maunya pejabat kereta cepat jkt bandung awalnya dgn jepang dibatalkan demi china kasus bts contoh perusahaan china menguntungkan pejabat ri
Dalam beberapa tahun terakhir proyek kereta api berkecepatan tinggi terus dikembangkan di seluruh dunia Jika menilik sejarah kereta ekspres terbatas pertama kali dikembangkan Jepang pada tahun dengan nama Shinkansen	dalam beberapa tahun terakhir proyek kereta api berkecepatan tinggi terus dikembangkan di seluruh dunia jika menilik sejarah kereta ekspres terbatas pertama kali dikembangkan jepang pada tahun dengan nama shinkansen
Masak sih belum kapok sama kereta cepat Proyek mercusuar gagal total ninggalin utang beberapa generasi Padahal dulu janjinya ANGGARAN bukan dari uang rakyat tapi lama2 pakai uang rakyat juga Capek bos rakyat di bodohi terus dengan akal2an proyek strategis nasional	masak sih belum kapok sama kereta cepat proyek mercusuar gagal total ninggalin utang beberapa generasi padahal dulu janjinya anggaran bukan dari uang rakyat tapi lama2 pakai uang rakyat juga capek bos rakyat di bodohi terus dengan akal2an proyek strategis nasional

Case Folding pada preprocessing berfungsi untuk mengubah semua huruf besar (uppercase) ke huruf kecil (lowercase) untuk memudahkan pemrosesan teks secara menyeluruh.

Tabel 4. Hasil Stemming dan Stopword	
Case Folding	Stemming dan Stopword

hambalang hanya secuil d banding proyek2 mangkrak era jokowi liatlah brp bandara d bangun ga berfungsi asal bangun2 tp hbs itu terbengkalai itu baru bandara pelabuhan ikn kereta api cepat bandung itu hanya proyek egois gaya2an tp sgt membebankan rakyat	hambalang hanya secuil banding proyek mangkrak era jokowi liat brp bandara bangun berfungsi bangun hbs bengkalai bandara labuh ikn kereta api cepat bandung proyek egois gaya sgt beban rakyat
iya dulu gw juga ga suka sama proyek tol trans jawa dan kereta cepat tapi makin kesini makin kesini orang banyak pada paket itu	iya dlu suka proyek tol trans jawa kereta cepat kesini kesini paket
bukan tidak meyakini tp jokowi lebih pilih kerja sama dgn china semua proyek pengennya dr china krn china sdh paham maunya pejabat kereta cepat jkt bandung awalnya dgn jepang dibatalkan demi china kasus bts contoh perusahaan china menguntungkan pejabat ri	jokowi pilih china proyek ken china krn china paham mau jabat kereta cepat jkt bandung jepang batal china bts contoh usaha china untung jabat ri
proyek kereta api cepat tinggi kembang dunia tahun sejarah kereta api cepat kembang jepang tahun nama shinkansen kereta peluru	proyek kereta api cepat kembang dunia tilik sejarah kereta ekspres batas kali kembang jepang nama shinkansen
masak sih belum kapok sama kereta cepat proyek mercusuar gagal total ninggalin utang beberapa generasi padahal dulu janjinya anggaran bukan dari uang rakyat tapi lama2 pakai uang rakyat juga capek bos rakyat di bodohi terus dengan akal2an proyek strategis nasional	masak sih belum kapok kereta cepat proyek mercusuar gagal total ninggalin utang generasi janji anggaran uang rakyat pakai uang rakyat capek bos rakyat bodohi akal proyek strategis nasional

Stemming digunakan untuk mengubah semua kata-kata Kembali ke bentuk dasarnya, seperti kata “bodohi” kembali ke bentuk dasarnya yaitu “bodoh”. Adapun stopword yaitu menghapus kata-kata yang tidak memiliki arti seperti kata “tp”, “gw”, “sgt”. Proses ini penting untuk meningkatkan akurasi saat melakukan pemodelan.

3.3 Translate

Mengubah teks menjadi bahasa inggris menggunakan library googletrans. Hasil dari translate dapat dilihat pada tabel di bawah ini.

Tabel 5. Hasil Translate	
Stemming dan Stopword	Translate
hambalang hanya secuil banding proyek mangkrak era jokowi liat brp bandara bangun berfungsi bangun hbs bengkalai bandara labuh ikn kereta api cepat bandung proyek egois gaya sgt beban rakyat	Hambalang is just a small part of the stalled projects of the Jokowi era. Look at how many airports were built that didn't function as long as they were built but then they were abandoned.
iya dlu gw juga ga suka sama proyek tol trans jawa dan kereta cepat tapi makin kesini	Yes, in the past I also didn't like the Trans Java toll road and high-speed train projects, but that's getting the case here

Bukan tidak meyakini tp jokowi lebih pilih kerja sama dgn china	It's not that he doesn't believe it, but Jokowi prefers cooperation with China
proyek kereta api cepat kembang dunia tilik sejarah kereta ekspres batas kali kembang jepang nama shinkansen	High-speed train projects have continued to be developed throughout the world in recent years. If you look back at their history
masak sih belum kapok kereta cepat proyek mercusuar gagal total ninggalin utang generasi janji anggaran uang rakyat pakai uang rakyat capek bos rakyat bodoh akal proyek strategis nasional	Cook still haven't given up on fast trains. The lighthouse project failed miserably, leaving several generations in debt, even though the promise was that the budget would not come from people's money.

3.4 Labeling

Dalam proses penglabelan penulis menggunakan *library textblob* untuk menentukan kategori yang akan diberikan kepada data. Adapun hasil *labeling* dapat dilihat pada tabel di bawah ini.

Tabel 6. Hasil Labeling	
Translate	Labeling
hambalang is just a small part of the stalled projects of the jokowi era. look at how many airports were built that didn't function as long as they were built but then they were abandoned.	Positif
yes, in the past i also didn't like the trans java toll road and high-speed train projects, but that's getting the case her.	Negatif
it's not that he doesn't believe it, but jokowi prefers cooperation with china.	Negatif
high-speed train projects have continued to be developed throughout the world in recent years. if you look back at their history.	Positif
cook still haven't given up on fast trains. the lighthouse project failed miserably, leaving several generations in debt, even though the promise was that the budget would not come from people's money.	Negatif

3.5 Klasifikasi Naïve Bayes

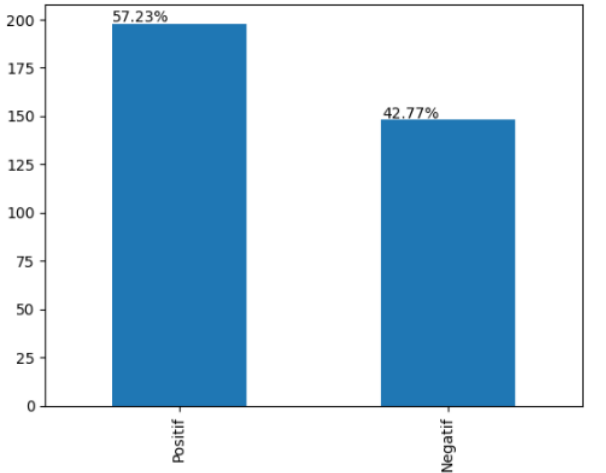
Di lakukan proses pemodelan menggunakan algoritma *Naïve Bayes Classifier*, dengan memproses data *train_set* menggunakan *Naïve Bayes Classifier Class*, yang diimport dari modul *textblob.classifiers*, adapun sintak yang digunakan dapat dilihat pada gambar 3.

```
from textblob.classifiers import NaiveBayesClassifier
cl = NaiveBayesClassifier(train_set)
print('Akurasi Test:', cl.accuracy(dataset))
```

Gambar 2. Klasifikasi Naïve Bayes

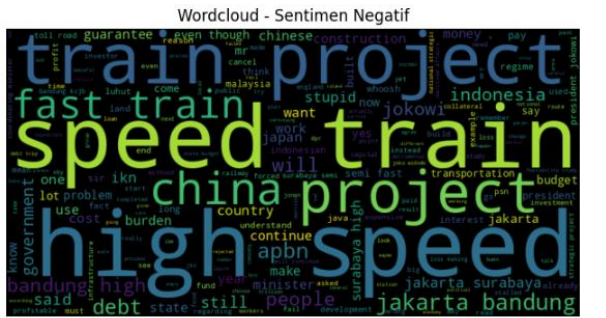
Dan dihasilkan akurasi tes menggunakan *Naïve Bayes Classifier*, sebesar 81%.

Setelah mengetahui nilai akurasi pemodelan menggunakan *Naïve Bayes Classifier*, Langkah berikutnya adalah menampilkan *visualisasi* data hasil klasifikasi. Dalam menampilkan *visualisasi* data penulis disini menggunakan *plot bar* untuk menampilkan hasil data dan *wordcloud* untuk menampilkan data yang sering muncul dalam teks analisis, Adapun gambar *plot bar* dapat dilihat pada gambar 3.



Gambar 3. Visualisasi data hasil klasifikasi

Selanjutnya akan ditampilkan *wordcloud* untuk mengetahui kata apa saja yang sering muncul saat pemodelan menggunakan *Naïve Bayes Classifier*. *Wordcloud* dapat dilihat pada gambar 5 dan 6.



Gambar 4. Wordcloud Negatif

Pada *wordcloud* negatif akan ditampilkan kata-kata yang sering keluar dengan label sentimen negatif

-
- [6] S. Nanda, D. Muallfah, and D. A. Fitri, "Analisis Sentimen Kepuasan Pengguna Terhadap Layanan Streaming Mola Menggunakan Algoritma Random Forest," 2022.
- [7] I. Afdhal *et al.*, "Penerapan Algoritma Random Forest Untuk Analisis Sentimen Komentar Di YouTube Tentang Islamofobia," *Jurnal Nasional Komputasi dan Teknologi Informasi*, vol. 5, no. 1, 2022.
- [8] A. Putri Nardilasari, S. Shofia Hilabi, and B. Priyatna, "Analisis Sentimen Calon Presiden 2024 Menggunakan Algoritma SVM Pada Media Sosial Twitter," 2026.
- [9] Y. Nooryuda Prasetya and D. Winarso, "Penerapan Lexicon Based Untuk Analisis Sentimen Pada Twiter Terhadap Isu Covid-19," vol. 11, no. 2, 2021.
- [10] R. Rahman Salam, M. Fajri Jamil, Y. Ibrahim, Rahnadden, Soni, and Herianto, "Indonesian Journal of Machine Learning and Computer Science Sentiment Analysis of Cash Direct Assistance Distribution for Fuel Oil Using Support Vector Machine Analisis Sentimen Terhadap Bantuan Langsung Tunai (BLT) Bahan Bakar Minyak (BBM) Menggunakan Support Vector Machine," vol. 3, pp. 27–35, 2023.
- [11] G. Yunanda, D. Nurjanah, and S. Meliana, "Recommendation System from Microsoft News Data using TF-IDF and Cosine Similarity Methods," *Building of Informatics, Technology and Science (BITS)*, vol. 4, no. 1, Jun. 2022, doi: 10.47065/bits.v4i1.1670.
- [12] M. Rahardi, A. Aminuddin, F. F. Abdulloh, and R. A. Nugroho, "Sentiment Analysis of Covid-19 Vaccination using Support Vector Machine in Indonesia." [Online]. Available: <https://t.co/h5x41UO3tF>
- [13] V. Mohan, "Preprocessing Techniques for Text Mining-An Overview." [Online]. Available: <https://www.researchgate.net/publication/339529230>
- [14] D. Muallfah, Ramadhoni, R. Gunawan, and D. M. Suratno, "Analisis Sentimen Komentar YouTube TvOne Tentang Ustadz Abdul Somad Dideportasi Dari Singapura Menggunakan Algoritma SVM," 2023.
- [15] J. Mantik *et al.*, "Analisis Sentimen Masyarakat Terhadap Tweet Ruang Guru Menggunakan Algoritma Naive Bayes Classifier (NBC)," Online, 2022.
- [16] Syahril Dwi Prasetyo, Shofa Shofiah Hilabi, and Fitri Nurapriani, "Analisis Sentimen Relokasi Ibukota Nusantara Menggunakan Algoritma Naive Bayes dan KNN," *Jurnal KomtekInfo*, pp. 1–7, Jan. 2023, doi: 10.35134/komtekinfo.v10i1.330.
- [17] D. Duei Putri, G. F. Nama, and W. E. Sulistiono, "Analisis Sentimen Kinerja Dewan Perwakilan Rakyat (DPR) Pada Twitter Menggunakan Metode Naive Bayes Classifier," *Jurnal Informatika dan Teknik Elektro Terapan*, vol. 10, no. 1, Jan. 2022, doi: 10.23960/jitet.v10i1.2262.