

Jurnal final turnitin

by adit.ihzar@gmail.com 1

Submission date: 11-Aug-2022 12:26AM (UTC-0400)

Submission ID: 1881244527

File name: 1811500584_jurnal_final.docx (116.36K)

Word count: 4224

Character count: 26381

PENERAPAN ALGORITME NAÏVE BAYES UNTUK ANALISIS PANDANGAN MASYARAKAT TERHADAP PEMINDAHAN IBU KOTA INDONESIA

Aditiya Ihzar Eka Prayogo¹, Painem²

² Fakultas Teknologi Informasi, Teknik Informatika, Universitas Budi Luhur, Jakarta Selatan, Indonesia
Email: ¹1811500584@student.budiluhur.ac.id, ²painem@budiluhur.ac.id

⁷**Abstrak-** Pemerintah sudah mencanangkan pemindahan ibu kota. Namun dibalik hal tersebut terdapat kontroversi dari masyarakat yang merasa pro (Positif) dan kontra (Negatif), masing-masing memiliki sudut pandang yang berbeda. Reaksi-reaksi yang berbeda tersebut perlu dilihat dengan gambaran umum, bagaimana pendapat dan respon masyarakat, apakah menyetujui rencana pemindahan ibu kota negara atau tidak. Pada era dengan teknologi yang sangat canggih seperti sekarang ini, penilaian sentiment publik terhadap berbagai isu untuk mendapatkan gambaran secara umum cukup mudah dilakukan dan dapat dilakukan secara cepat. Hal ini terjadi karena maraknya penggunaan media sosial oleh masyarakat. Media sosial saat ini dapat digunakan sebagai media hiburan, tetapi juga menjadi media untuk menyatakan pendapat serta pandangan terkait berbagai topik. Begitu juga dengan topik rencana pemindahan ibu kota negara, dimana reaksi masyarakat banyak dituangkan dalam berbagai media, baik berupa diskusi maupun media-media sosial seperti twitter. Salah satu media sosial banyak yang digunakan oleh masyarakat Indonesia untuk menyatakan pendapat adalah twitter. Twitter adalah media sosial yang dapat menghubungkan antara pengguna satu dengan pengguna lain. Pendapat pemindahan ibu kota yang didapatkan dari twitter perlu dilakukan analisis. Salah satu metode analisis sentiment yang digunakan untuk menganalisa topik pemindahan ibu kota dan media sosial twitter adalah metode Naive Bayes Classifier (NBC). Hasil analisis berdasarkan 1.272 tweet menunjukkan bahwa sentimen positif sebesar 86.95 % dan sentimen negatif sebesar 13.05 % pada periode 10 May 2022 hingga 1 Juni 2022, sementara hasil pengujian yang diperoleh adalah nilai akurasi sebesar 93 %, presisi sebesar 87 % dan recall sebesar 100 %.

Kata Kunci: Analisis Sentimen, Twitter, Klasifikasi, Naive Bayes

APPLICATION OF NAVE BAYES ALGORITHM FOR ANALYSIS OF COMMUNITY VIEW ON INDOONESIAN CAPITAL MOVEMENT

³⁴**Abstract-** The government has announced the relocation of the capital city. But behind this, there is controversy among people who feel pro (Positive) and con (Negative), each of which has a different point of view. These different reactions need to be seen with a general picture, what are the opinions and responses of the people, whether they agree with the plan to move the national capital or not. In today's era of highly sophisticated technology, assessing public sentiment on various issues to get an overview is relatively easy and can be done quickly. This is due to the widespread use of social media by the public. Today's social media can be used as a medium of entertainment, but also as a medium for expressing opinions and views on various topics. Likewise with the planned relocation of the national capital, where public reactions are widely described in various media, both in the form of discussions and on social media such as Twitter. One of the social media that is widely used by Indonesian people to express their opinions is Twitter. Twitter is a social media that can connect one user to another. Opinions about the relocation of the capital obtained from Twitter need to be analyzed. One of the sentiment analysis methods used to analyze the topic of moving the capital city from social media Twitter is the Naive Bayes Classifier (NBC) method. The results of the analysis based on 1,272 tweets showed positive sentiment of 86.95% and negative sentiment of 13.05% in the period 10 May 2022 to 1 June 2022, while the test results obtained were 93% accuracy, 87% precision, and 100% recall.

Keywords: Sentiment Analysis, Twitter, Classification, Naive Bayes

1. PENDAHULUAN

Rencana pemindahan ibu kota ke Kalimantan Timur secara resmi di deklarasikan pada bulan agustus 2019 oleh Presiden Joko Widodo. Namun dibalik hal tersebut terdapat kontroversi dari masyarakat yang merasa pro dan kontra terhadap pemindahan ibu kota. Saat ini, masyarakat dalam mengemukakan pendapat (komentar) atau dalam melakukan penilaian bisa dilakukan melalui media Sosial. Salah satu media sosial yang digunakan dalam memberikan komentar atau penilaian adalah Twitter. Twitter adalah komunitas informal dengan kontribusi mini ke blog yang memiliki jumlah klien dan basis informasi yang besar dan berkembang. Twitter dalam banyak kasus digunakan sebagai tempat untuk menawarkan pendapat karena memungkinkan pengguna untuk berbagi pesan

menggunakan pesan singkat yang disebut tweets. Ada sebuah ruang untuk mengarahkan eksplorasi dengan memecah perasaan dan pendapat secara komputasi [1].

Berdasarkan data dari website statista, kemajuan klien Twitter pada tahun 2017 mencapai 109 juta klien dinamis dan pada tahun 2022 mencapai 229 juta klien dinamis di seluruh dunia. Klien web menyusun sentimen dan penilaian tentang berbagai subjek di administrasi Twitter. Saat klien mengekspresikan tema yang berbeda, misalnya, masalah pemerintahan, ini menjadikan Twitter sumber data yang potensial dan mahir. Penelitian ini menggunakan Twitter karena memiliki beberapa manfaat, khususnya digunakan oleh berbagai kelompok klien, memiliki pesan singkat yang disebut tweet yang berisi berbagai kesimpulan umum, terus berkembang, dan menyebarkan berita dengan cepat. Berbagai topik dibicarakan secara luas di Twitter mulai dari masalah keuangan, masalah legislatif, sosial, budaya, regulasi, dan lainnya. Salah satu topik yang pernah trending adalah pemindahan ibu kota ke Kalimantan Timur. Pada periode 10 Mei 2022 hingga 1 Juni 2022 didapatkan 1.272 tweet yang berhubungan dengan tema pemindahan ibu kota. Berdasarkan 1.272 tweet yang didapatkan maka perlu dilakukan analisa untuk mengetahui pendapat masyarakat tentang pemindahan ibu kota. Baik pendapat masyarakat yang positif maupun yang negatif. Salah satu metode analisis sentimen yang digunakan adalah metode Naive Bayes Classifier (NBC).

Pada penelitian sebelumnya mengenai analisis sentimen yang dilakukan oleh [5] metode Naive Bayes Classifier untuk Warga Jakarta Terhadap Kehadiran Mass Rapid Transit dari penelitian tersebut dapat memprediksi sentimen dari tweet yang sudah dikumpulkan terkait animo masyarakat terhadap MRT Jakarta dengan akurasi sebesar 75%. Sedangkan pada penelitian yang sudah dilakukan [10] metode Analisis Sentimen Masyarakat Terhadap Pembelajaran Daring di Era Pandemi Covid-19 pada Media Sosial Twitter Menggunakan Ekstraksi Fitur Countvectorizer dan Algoritma K-Nearest Neighbor dari penelitian tersebut pandangan (sentimen) masyarakat Indonesia terhadap pembelajaran daring cenderung ke arah sentimen yang baik (positif) sebesar 76,56%

Algoritme *Naive Bayes* untuk mengkonsolidasikan probabilitas sebelumnya dan probabilitas konsistensi dalam persamaan soliter, yang dapat digunakan untuk menghitung probabilitas dari setiap urutan probabilitas yang mungkin. Setelah selesai kami memilih pengaturan dengan nilai terbesar. Dalam penelitian yang diarahkan oleh [3], Peneliti tersebut berhasil mendapatkan akurasi yang sangat akurat dalam melakukan klasifikasi tweet menggunakan Algoritme *Naive Bayes* yaitu sebesar 99,62%, *Naive Bayes* juga digunakan oleh [4], untuk mengklasifikasi data tweet yang berjumlah 2000 dan mendapatkan akurasi yang cukup baik yaitu 74%. Dengan permasalahan diatas peneliti bermaksud untuk membuat sebuah sistem untuk melakukan analisis sentimen dengan Algoritme klasifikasi *Naive Bayes*. Dataset yang digunakan berupa teks kicauan (Tweet) yang bersumber pada media sosial Twitter dengan menggunakan Teknik *Crawling*. Proses pengumpulan data yang dilakukan menggunakan pustaka Tweepy dengan kata kunci “pemindahan ibu kota baru”, “ibu kota baru”, “ibu kota pindah”, “pindah ibu kota”, dan “ibukota pindah” yang dilakukan di rentan waktu 10 Mei 2022 hingga 1 Juni 2022. Selama proses pengumpulan data, peneliti mengambil attribute id, nama user, tanggal dan isi text. Setelah dilakukan pengumpulan data, tweet tersebut dilakukan labeling data agar memberikan label pada sentiment, kemudian tahap *pre-processing* dengan dilakukannya penghilangan simbol dan tanda baca yang tidak diperlukan, *normalisasi* data dilakukan dengan konversi kata yang semula tergolong ke dalam tidak baku menjadi kata baku, menghilangkan angka, dan penghapusan *stopword*. Kemudian masuk di tahap prediksi sentimen menggunakan metode *Naive Bayes*. Berdasarkan uraian tersebut penulis mengangkat judul topik penerapan algoritme *Naive Bayes* untuk analisis pandangan masyarakat terhadap pemindahan ibu kota Indonesia.

Penelitian ini dilakukan dengan tujuan untuk melakukan pengklasifikasian tweet berbahasa Indonesia terhadap pemindahan ibu kota ke Kalimantan melalui media sosial Twitter dengan menggunakan Algoritme *Naive Bayes* dan untuk menguji tingkat akurasi dari Algoritme *Naive Bayes* dalam klasifikasi berbasis text.

2. METODE PENELITIAN

2.1 Data Penelitian

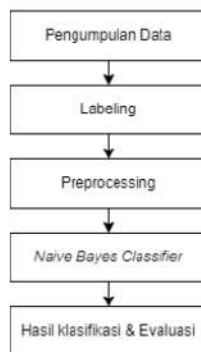
Twitter menyediakan API (*Application Programming Interface*). Twitter API adalah untuk pengembang. Twitter API memungkinkan pengguna untuk membaca, menulis, dan mengambil data dari Twitter. Twitter API juga memungkinkan pengembang untuk mengambil informasi dan data pengguna tentang topik di Twitter atau lokasi tertentu. Penulis memperoleh Data dengan menggunakan teknik *crawling* dengan bantuan pustaka Tweepy. Pengumpulan Data dilakukan pada rentan waktu 10 Mei 2022 hingga 1 Juni 2022 dengan memasukan keyword “pemindahan ibu kota baru”, “ibu kota baru”, “ibu kota pindah”, “pindah ibu kota”, dan “ibukota pindah” sehingga diperoleh data dengan total 1.796.

Tabel 1. Sample Data Tweet

Full Name	Tweet	Date
yurikanamnam	@dutaSherliana ¹⁷ Pemandahan ibu kota akan menurunkan kesenjangan antar kelompok pendapatan dan indikasi ketimpangan akan menyempit.	2022-05-11 02:36:42
²⁷ Dewan Mahasiswa Justicia FH UGM	Wajah-wajah itu tidak datang dari ruang hampa. Melainkan kebijakan pemindahan ibu kota, yang pada akhirnya melatarbelakangi terukirnya ragam wajah dalam karya-karya yang kami sajikan. https://t.co/w1SNuiSaYI	2022-05-10 11:38:11
Dewan Mahasiswa Justicia FH UGM	yaitu mereka-mereka yang diuntungkan melalui kebijakan pemindahan ibu kota. Dengan penyajian yang sederhana, barangkali potret wajah yang terpampang merupakan suatu yang mahal dan tak ternilai. Sebab, ialah kekuasaan menjadi harga yang harus dibayarkan.	2022-05-10 11:38:23
Margareth Saalam	Segala macam hoak diciptakan kelompok sakit hati terhadap Ibu Kota baru Negara. Tetap wasapda. IKN Untuk Negeri https://t.co/01E314G51O	2022-05-23 02:05:25
Ramdani	@OposisiCerdas Yang jadi masalah Luhut mau bangun ibu kota baru...danannya dr hutang...yang bayar rakyat...dia dapat komisi dr investor...kan sialan dangkalan.....	2022-05-23 04:28:17
Anwar jahid Pena	Rakyat ga butuh Ibu kota baru..	2022-05-22 14:32:24

2.2 Penerapan Metode

Dalam membangun kerangka kerja pemeriksaan perasaan Twitter yang menggunakan teknik *Naive Bayes*, diperlukan penggambaran keseluruhan yang menunjukkan perkembangan eksplorasi yang sedang diselesaikan sehingga siklus eksplorasi berjalan seperti yang diharapkan. Proses kerja ini membahas tahapan yang mendasari hingga batas terjauh dari kerangka kerja yang berjalan, yang terdapat pada Gambar 1 berikut :



Gambar 1. Tahapan Metode Penelitian

Berdasarkan Gambar 1, setidaknya ada 5 tahapan dalam penelitian ini dimulai dengan tahap pengumpulan data dengan memanfaatkan proses *crawling* data Twitter menjadi dataset mentah, kemudian tweet yang sudah berbentuk file .csv dilakukan proses *import* ke dalam program lalu kemudian dimasukkan kedalam database untuk dilakukan pelabelan kelas sentimen dan *preprocessing*. Pada tahap *preprocessing* dilakukan pembersihan, pemilihan dan pemilihan kata yang layak untuk diproses ke dalam tahap selanjutnya, Pada tahap *Naive Bayes Classifier* dilakukan tahap pembagian data dengan ratio (90:10) untuk data latih dan data uji. Setelah itu dilakukan

proses perhitungan *Document Frequency* (DF). Setelah perhitungan DF, kemudian dilakukan proses klasifikasi yang dilanjutkan dengan tahap selanjutnya yaitu menampilkan proses hasil dari klasifikasi dan evaluasi berupa hasil validasi akurasi atau perhitungan untuk menentukan nilai akurasi dan akurasi, dan hasil pencetakan metode divisualisasikan dalam bentuk persentase.

2.2.1 Pengumpulan Data

Penelitian ini untuk mengumpulkan data yang digunakan dengan memanfaatkan teknik *crawling* Twitter. Teknik tersebut dilakukan pada rentan waktu 10 Mei 2022 hingga 1 Juni 2022 dengan menggunakan *library* yang telah disediakan yakni Tweepy dan Twitter API. Twitter API dibutuhkan untuk menjalankan Tweepy sehingga data tweet dapat diakses melalui akun developer Twitter. *Crawling* data dilakukan berdasarkan *keyword* "pemindahan ibu kota baru", "ibu kota baru", "ibu kota pindah", "pindah ibu kota", dan "ibukota pindah". Setelah itu dataset tweet yang masih mentah disimpan dalam bentuk *csv* file lebih lanjut.

2.2.2 Labeling Data

Proses pelabelan kelas sentimen pada setiap tweet dikategorikan dengan melihat karakter, pesan atau informasi yang dimuat pada sebuah tweet. Sistem penamaan (*Labeling*) harus dimungkinkan secara fisik oleh pencipta dengan penandaan kelas langsung mengingat subjektivitas kalimat secara individual pada tweet tanpa melalui sistem *preprocessing* terlebih dahulu. Penandaan perasaan dibedakan menjadi dua kelas, yaitu baik yang spesifik dan pesimis. Kelas positif ditujukan untuk tweet yang berisi perasaan puas, ceria, dan hebat dengan hijrahnya Ibu Kota Negara ke Kalimantan Timur, sedangkan kelas negatif ditujukan pada tweet yang cenderung kecewa, tidak senang serta tidak puas dengan adanya pemindahan Ibu Kota Negara ke Kalimantan Timur. Setelah dilakukan proses pelabelan maka data yang akan disimpan di database akan dilakukan penyaringan sehingga akan menyisakan tweet yang akan dilakukan proses selanjutnya yaitu *preprocessing*.

2.2.3 Preprocessing

Informasi tweet yang didapat dari sistem *crawling* tidak dapat dimanfaatkan karena sebenarnya memiliki pesan atau struktur kalimat yang sporadis atau tidak terstruktur karena memiliki banyak sekali keributan. Oleh karena itu, pada tahap *prapemrosesan* ini, penting untuk mengubah jenis informasi dari informasi tidak terstruktur menjadi informasi terorganisir, bertekad untuk membuat tweet bersih dengan tujuan agar mereka dapat ditangani untuk mengeluarkan data dari laporan teks dan bekerja dengan fase siklus berikutnya. Tahapan *preprocessing* dalam penelitian ini meliputi:

a. Case folding

Di dalam tahapan ini karakter-karakter yang mengandung huruf besar akan dikonversi menjadi huruf kecil (*lowercase*) seperti kata "Indonesia" diubah menjadi "indonesia", dan lain sebagainya.

b. Cleansing

Proses *cleansing* digunakan untuk menghilangkan atau menghapus gambar-gambar umum Twitter, misalnya username, @ (pemberitahuan), dan karakter yang berbeda seakan huruf utama (a-z) tertinggal. Sistem pemurnian ini memiliki beberapa fase, khususnya: menghilangkan karakter selain huruf (a-z), teks dengan 1 orang, URL, dan spasi berlebihan atau teks kosong sebagai spasi.

c. Mengganti Slang word

Proses merubah *slang word* setiap kata kekinian atau yang tidak dianggap baku akan dikonversi menjadi bentuk bakunya, seperti kata "dri" menjadi "dari" dan lain sebagainya. Proses ini membutuhkan kamus slang word yang telah tersedia dalam basis data (database).

d. Menghapus Stopword

Proses *stopword* dilakukan dengan menggunakan pustaka dari Nltk.corpus. Pada tahap ini, kata-kata tidak penting atau kata-kata umum tidak berdampak pada peristiwa kata-kata yang berkelanjutan. Pembatalan kata menggunakan *stopword* bahasa Indonesia yang telah disusun dan ditambahkan secara fisik mengingat masih ada beberapa kata yang belum diingat untuk *stopword* tersebut. Metode yang digunakan untuk menghapus kata-kata menyerupai model yang menyertainya, di mana kata-kata "ini", dan "sekali lagi" ada di daftar kata berhenti, dan kata itu dihapus.

e. Stemming

Tahapan ini dilakukan dengan memanfaatkan pustaka dari Sastrawi, pada setiap kata yang memiliki imbuhan (berimbuhan) akan dikonversikan menjadi kata dasar seperti “memakan” dihapus imbuhan “me-” sehingga menjadi kata “makan” dan lain sebagainya.

2.2.4 Naïve Bayes Classifier

Dari proses yang telah dilakukan sebelumnya telah didapatkan data dari hasil proses *preprocessing* yang kemudian akan diproses ke dalam klasifikasi. Hasil klasifikasi sentimen yang di dapatkan dengan menerapkan metode *Naïve Bayes Classifier*, terdapat 3 tahapan utama, yaitu pembagian dataset training dan testing, perhitungan DF (*Document Frequency*) dan klasifikasi.

a. Pembagian Data

Pada tahap ini, setelah dilakukannya *preprocessing*. Dataset dibagi dalam proses pembagian data ini dengan rasio 90 untuk data *training* berbanding 10 untuk data *testing*.

b. Perhitungan *Document Frequency*

Tahap ini adalah cara paling umum untuk menurunkan bobot istilah (kata-kata) ke catatan atau tweet untuk mengetahui relevansinya dalam korpus (berbagai laporan). Tahap pembobotan diingat untuk siklus vektorisasi yang diharapkan dapat mengubah teks dalam tweet menjadi angka-angka yang dapat ditangani dan dicari desainnya selama proses penataan arsip. Resep perhitungan pada kondisi (1) adalah sebagai berikut.

$$w_{ij} = \frac{1}{\sum_{i=1}^n w_{ij}} \log\left(\frac{w_{ij}}{w_i}\right) \quad (1)$$

c. Klasifikasi *Naïve Bayes Classifier*

Tahap karakterisasi dengan memanfaatkan *Naïve Bayes* merupakan tahap persiapan laporan yang telah diproses sebelumnya dan diberi tanda untuk mendapatkan hasil opini. Sistem pengelompokan ini membutuhkan informasi penandaan manual mulai dari penyiapan informasi hingga pembuatan model. Ada dua tahap dalam siklus karakterisasi, yaitu

1. Model dibangun berdasarkan analisa pada data *training*
2. Kemudian melakukan klasifikasi dari hasil model yang didapatkan akan dimanfaatkan kembali untuk dilakukannya klasifikasi terhadap label yang belum diketahui

Pada Algoritme *Naïve Bayes* setiap dokumen dipresentasikan dengan memasukkan “a1, a2, a3, ..., an” di mana a1 adalah kata pertama dan berikutnya sampai, sedangkan V yaitu label kategori. Selanjutnya yaitu mencari nilai tertinggi dari kategori teks yang diujikan (V_{MAP}) persamaan V_{MAP} yaitu pada rumus (2) sebagai berikut:

$$V_{MAP} = \underset{v_i \in V}{\operatorname{argmax}} P(v_j) \cap_i P(a_i|v_j) \quad (2)$$

Nilai $P(v_j)$ dihitung pada saat data latih, dengan rumus (2) sebagai berikut :

$$P(v_j) = \frac{|Docs_j|}{|training|} \quad (3)$$

$|Docs_j|$ adalah jumlah dokumen yang memiliki kategori j pada dokumen latih dan $|training|$ adalah jumlah dokumen latih.

$$P(a_i|v_j) = \frac{|n_i + 1|}{|n + kosakata|} \quad (4)$$

Keterangan :

- n_i : Jumlah kemunculan kata a_i pada dokumen latih yang berkategori v_j
- N : jumlah seluruh kata pada dokumen latih yang berkategori v_j
- kosakata : Jumlah kata unik pada seluruh dokumen latih

3. HASIL DAN PEMBAHASAN

Implementasi metode dalam penelitian ini dilakukan dengan empat (4) tahapan utama. Tahapan utama tersebut diproses secara berurutan, tahapan utama yang dimaksud antara lain: Tahap pengumpulan data, tahap labeling data, tahap preprocessing, tahap klasifikasi dengan Algoritme *Naïve Bayes Classifier*

3.1 Tahap Pengumpulan Data

Berdasarkan sub bab (2.2.1), Data yang dikumpulkan untuk penelitian ini menggunakan teknik *crawling* pada Twitter. Adapun teknik *crawling* dilakukan pada rentan waktu 10 Mei 2022 hingga 1 Juni 2022 mengimplementasikan library yang telah tersedia yakni Tweepy dan Twitter API. *Crawling* data dilakukan berdasarkan keyword “pemindahan ibu kota baru”, “ibu kota baru”, “ibu kota pindah”, “pindah ibu kota”, dan “ibukota pindah”. Setelah itu dataset tweet yang masih mentah disimpan dalam bentuk csv file.

Dataset yang dikumpulkan melalui proses *crawling* dengan informasi antara lain : *tweet id*, *full_text* (tweet), *created_at* (tanggal), *tweet.user.name* (username), yang akan disimpan ke dalam file excel (.csv) yang kemudian akan dimasukkan kedalam basis data (Database) secara *Import* Dataset didalam program berjumlah sebanyak 1.272 data tweet.

3.2 Tahap Labeling

Berdasarkan sub bab (2.2.2) Proses pelabelan (*labeling*) dilakukan dengan penamaan kelas langsung mengingat subjektivitas kalimat secara individual pada tweet tanpa melalui sistem pra-pemrosesan terlebih dahulu. Penandaan opini dipisahkan menjadi dua kelas, khususnya pasti dan negatif. Kelas positif ditujukan untuk tweet yang berisi perasaan terpenuhi, bahagia, dan hebat dengan migrasi Ibukota Negara ke Kalimantan Timur, sedangkan kelas negatif ditujukan pada tweet yang cenderung kecewa, tidak senang serta tidak puas dengan adanya pemindahan Ibu Kota Negara ke Kalimantan Timur. Berikut ini contoh dataset yang telah dilakukan proses labeling.

Tabel 2. Dataset Labeling

<i>Tweet</i>	<i>Label</i>
pindah kota perata bangun ekonomi indonesia infrastruktur bukti nyata	Positif
heboh tolak pindah ibukota	Negatif

Dari 1.272 dataset yang sudah dilakukan proses pelabelan (*labeling*), sebanyak 272 yang akan dilakukan tahap *Preprocessing* dan tahap *Classification*. Proses ini bertujuan untuk menyeimbangkan total tweet yang negatif dan positif agar mencapai nilai akurasi maksimal pada tahap *Classification* Menggunakan Algoritme *Naïve Bayes*.

3.3 Tahap Preprocessing

Setelah dataset dilakukan proses pelabelan dan proses maka selanjutnya dataset diproses dengan proses seperti *case folding*, menghapus teks dengan 1 karakter, menghapus URL, mengganti *slang word*, menghapus *stop word*, dan *stemming*.

a. Case Folding

Pada Tabel 3 merupakan proses untuk mengubah untuk mengubah isi teks dokumen menjadi huruf kecil (*lower case*), melalui proses *case folding*, sehingga isi teks menjadi huruf kecil semua. Berikut contohnya:

Tabel 3. Case Folding

<i>Tweet</i> sbelum	<i>Tweet</i> sesudah
b'Rakyat...??? Masyarakat kaliii.... Pajak yg ditentukan juga biasa Ada di daerah lain.... Yang mau pindah silahkan... Yg gak mau pindah silahkan.... Gak Ada elu juga gak papa Tapi pemindahan ibu kota sangat2 tepat.... Agar pemerintah fokus melayani	b'rakyat...??? masyarakat kaliii.... pajak yg ditentukan juga biasa ada di daerah lain.... yang mau pindah silahkan... yg gak mau pindah silahkan.... gak ada elu juga gak papa tapi pemindahan ibu kota sangat2 tepat.... agar pemerintah fokus melayani

b. *Cleansing*

Pada Tabel 4 merupakan proses untuk menghapus karakter selain a-z. Sehingga isi teks yang bukan huruf a-z akan dihapus, seperti angka, simbol, dan seterusnya. Berikut contohnya:

Tabel 4. Cleansing

<i>Tweet</i> sebelum	<i>Tweet</i> sesudah
b'rakyat...??? masyarakat kaliii.... pajak yg ditentukan juga biasa ada di daerah lain.... yang mau pindah silahkan... yg gak mau pindah silahkan.... gak ada elu juga gak papa tapi pemindahan ibu kota sangat2 tepat.... agar pemerintah fokus melayani	rakyat masyarakat kaliii pajak yg ditentukan juga biasa ada di daerah lain yang mau pindah silahkan yg gak mau pindah silahkan gak ada elu juga gak papa tapi pemindahan ibu kota sangat tepat agar pemerintah fokus melayani

c. *Mengganti Slang Word*

Pada Tabel 5 merupakan proses untuk mengganti kata yang tidak baku (slang) menjadi kata baku, seperti kata “yg” dan “gak” diubah menjadi “yang” dan “tidak”. Berikut contohnya:

Tabel 5. Slang Word

<i>Tweet</i> sebelum	<i>Tweet</i> sesudah
rakyat masyarakat kaliii pajak yg ditentukan juga biasa ada di daerah lain yang mau pindah silahkan yg gak mau pindah silahkan gak ada elu juga gak papa tapi pemindahan ibu kota sangat tepat agar pemerintah fokus melayani	rakyat masyarakat kaliii pajak yang ditentukan juga biasa ada di daerah lain yang mau pindah silahkan yang tidak mau pindah silahkan tidak ada lu juga tidak papa tapi pemindahan ibu kota sangat tepat agar pemerintah fokus melayani

d. *Menghapus Stop Word*

Pada Tabel 6 menggambarkan terjadinya proses penghapusan kata yang dikategorikan sebagai kata tidak terlalu penting dengan mengacu kepada kamus *stop word*, seperti kata dan, untuk, bagi, dan seterusnya. Berikut contohnya:

Tabel 6. Stop Word

<i>Tweet</i> sebelum	<i>Tweet</i> sesudah
rakyat masyarakat kaliii pajak yang ditentukan juga biasa ada di daerah lain yang mau pindah silahkan yang tidak mau pindah silahkan tidak ada lu juga tidak papa tapi pemindahan ibu kota sangat tepat agar pemerintah fokus melayani	rakyat masyarakat kaliii pajak ditentukan daerah pindah silahkan pindah silahkan lu papa pemindahan kota pemerintah fokus melayani

e. *Stemming*

Pada Tabel 7 menggambarkan penghapusan imbuhan dari tweet untuk menghasilkan cukup kata dasarnya saja. Berikut contohnya:

Tabel 7. Stemming

<i>Tweet</i> sebelum	<i>Tweet</i> sesudah
rakyat masyarakat kaliii pajak ditentukan daerah pindah silahkan pindah silahkan lu papa pemindahan kota pemerintah fokus melayani	rakyat masyarakat kaliii pajak tentu daerah pindah silah pindah silah lu papa pindah kota perintah fokus layan

3.4 Tahap Klasifikasi *Naïve Bayes Classifier*

a. Tahap Pembagian Data

Setelah proses *preprocessing* dilanjutkan dengan pemisahan data menjadi data latih dan data uji, maka proses pemisahan data dilakukan secara manual di dalam program. Untuk proses pemisahan data dengan perbandingan rasio 90:10, 90% untuk data latih dan 10% untuk data uji.

b. Perhitungan *Document Frequency*

Tahap perhitungan untuk mengubah teks dalam tweet menjadi angka yang dapat ditangani dan dicari desainnya selama proses pemesanan arsip. Langkah awal pembobotan adalah memecah kalimat menjadi berbagai kata (token), kemudian pada saat itu mengambil kata khusus untuk melakukan pembobotan. Seperti pada contoh Tabel 8 berikut:

Tabel 8. Proses Tokenisasi

Tweet	Token	Unik Token
pindah kota perata bangun ekonomi indonesia infrastruktur bukti nyata	pindah	pindah
	kota	kota
	perata	perata
	bangun	bangun
	ekonomi	ekonomi
	indonesia	indonesia
	infrastruktur	infrastruktur
heboh tolak pindah ibukota	bukti	bukti
	nyata	nyata
	heboh	heboh
	tolak	tolak
	pindah ibukota	- ibukota
Total	13	12

Setelah mendapatkan kata yang unik dari setiap token, dilanjutkan dengan perhitungan TF. Setelah itu dilanjutkan dengan proses DF untuk kemudian diolah menjadi probabilitas yang akan digunakan untuk perhitungan prediksi sentimen. Proses perhitungan seperti pada Tabel 9 berikut:

Tabel 9. Perhitungan DF

Token	TF		DF
	d1	d2	
pindah	1	1	2
kota	1	0	1
perata	1	0	1
bangun	1	0	1
ekonomi	1	0	1
indonesia	1	0	1
infrastruktur	1	0	1
bukti	1	0	1
nyata	1	0	1
heboh	0	1	1
tolak	0	1	1
ibukota	0	1	1

c. Tahap *Naïve Bayes Classifier*

Dalam tahapan ini dilakukan proses *Naïve Bayes Classifier*, langkah pertama yaitu membuat model klasifikasinya dengan menghitung jumlah probabilitas kata untuk kategori positif dan negatif dengan dataset

training sebanyak 244 dari 1272 data. Proses perhitungan jumlah probabilitas kata dengan membagi total kata dengan banyaknya kata, dan didapatkan hasil perhitungan sebagai berikut.

Tabel 10. Perhitungan Probabilitas

Contoh Token	Probabilitas
pindah	0.1277139208173691
kota	0.3864406779661017
perata	0.023728813559322035
bangun	0.019157088122605363
ekonomi	0.03389830508474576
indonesia	0.14915254237288136
infrastruktur	0.008939974457215836
bukti	0.006385696040868455
nyata	0.006385696040868455
heboh	0.007662835249042145
tolak	0.010217113665389528
ibukota	0.05874840357598978

Setelah didapatkan nilai probabilitas pada setiap tokennya, langkah berikutnya yakni mengalikan keseluruhan nilai probabilitas masing-masing kategori untuk didapatkan batas nilai akhir dari setiap kategori nya, dengan hasil sebagai berikut :

$$P_{\text{kategori}} = \text{nilai probabilitas kategori} * (\text{total data kategori} / \text{total data training})$$

$$P_{\text{positif}} = 9,740540823468385e^{-18} * (120/244) = 7,295815054e^{-8}$$

$$P_{\text{negatif}} = 2,5475699800127307e^{-15} * (124/244) = 3,960415512e^{-7}$$

Setelah dilakukan perhitungan probabilitas maka didapatkan batas nilai akhir dari setiap kategori seperti pada hasil diatas. Apabila nilai probabilitas nya melebihi dari batas akhir nilai positif maka kata atau kalimat tersebut dinyatakan sebagai kata atau kalimat negatif.

4. KESIMPULAN

Hasil dari penelitian ini dengan menyelesaikan tahapan perancangan, pembuatan sistem dan pengujian aplikasi, disertai dengan dataset juga algoritme yang diusulkan. Berdasarkan 1.272 tweet, arah pandangan (sentimen) masyarakat indonesia terhadap pemindahan ibu kota indonesia cenderung kearah sentimen positif sebesar 86.95 % pada periode Mei – Juni 2022. Tahapan utama dalam penelitian ini adalah *Crawling*, *Labeling*, *Preprocessing* dan *Classification* menggunakan *Naïve Bayes Classifier*. Tahap *Preprocessing* yang baik menjadi penentu dalam terbentuknya hasil yang optimal untuk tahap selanjutnya. Penggunaan algoritme *Naïve Bayes Classifier* dalam melakukan analisis sentimen dapat berjalan dengan baik, dengan nilai pengujian dan evaluasi tertinggi yang diperoleh sebesar: akurasi 93%, presisi 87%, dan recall 100%.

DAFTAR PUSTAKA

- 15
- [1] Hidayat, A.M. and Syafrullah, M., "ALGORITMA NAÏVE BAYES DALAM ANALISIS SENTIMEN UNTUK KLASIFIKASI PADA LAYANAN INTERNET PT . XYZ", 2017, pp. 91–95.
 - [2] Bappenas, Dampak Ekonomi dan Skema Pembiayaan Pemindahan Ibu Kota Negara. 2019.
 - [3] Purwiantono, F.E. and Aditya, A., "KLASIFIKASI SENTIMEN SARA, HOAKS DAN RADIKAL PADA POSTINGAN MEDIA SOSIAL MENGGUNAKAN ALGORITMA NAIVE BAYES MULTINOMIAL TEXT", 2020, pp. 68–73.
 - [4] Yuyun, Hidayah, N. and Sahibu, S., "Algoritma Multinomial Naïve Bayes Untuk Klasifikasi Sentimen Pemerintah Terhadap Penanganan Covid-19, 2021.
 - [5] Afrizal, S., "Implementasi Metode Naïve Bayes untuk Analisis Sentimen Warga Jakarta Terhadap Kehadiran Mass Rapid Transit," JURNAL INFORMATIK, 2019
 - [6] Afrizal, S., Irmanda, H.N. and Falih, N., "Implementasi Metode Naïve Bayes untuk Analisis Sentimen Warga Jakarta Terhadap Kehadiran Mass Rapid Transit", 2019, pp. 157–168.

- [7] Antinasari, P., Perdana, R.S. and Fauzi, M.A., "Analisis Sentimen Tentang Opini Film Pada Dokumen Twitter Berbahasa Indonesia Menggunakan Naive Bayes Dengan Perbaikan Kata Tidak Baku", 2017, pp. 1733–1741.
- [8] Arsi, P. and Waluyo, R., "ANALISIS SENTIMEN WACANA PEMINDAHAN IBU KOTA INDONESIA MENGGUNAKAN ALGORITMA SUPPORT VECTOR MACHINE (SVM)", *Jurnal Teknologi Informasi dan Ilmu Komputer (JTIIK)*, 2021, pp. 147–156.
- [9] Isnain, A.R., Margo, N.S. and Alita, D., "Sentiment Analysis Of Government Policy On Corona Case Using Naive Bayes Algorithm," *IJCCS (Indonesian Journal of Computing and Cybernetics Systems)*, 2021, p. 55.
- [10] Priandi, M. and Painem, "Analisis Sentimen Masyarakat Terhadap Pembelajaran Daring di Era Pandemi Covid-19 pada Media Sosial Twitter Menggunakan Ekstraksi Fitur Countvectorizer dan Algoritma K-Nearest Neighbor," *Seminar Nasional Mahasiswa Ilmu Komputer dan Aplikasinya (SENAMIKA) Jakarta-Indonesia*, 2021.
- [11] Sadianoro, A.V. and Zuliarso, E., "Analisis Sentimen Twitter Menggunakan Text Mining Dengan Algoritma Naive Bayes Classifier ANALISIS SENTIMEN TWITTER MENGGUNAKAN TEXT MINING DENGAN ALGORITMA NAIVE BAYES CLASSIFIER", 2018, pp. 69–73.

Jurnal final turnitin

ORIGINALITY REPORT

28%

SIMILARITY INDEX

24%

INTERNET SOURCES

7%

PUBLICATIONS

11%

STUDENT PAPERS

PRIMARY SOURCES

1	conference.upnvj.ac.id Internet Source	4%
2	Submitted to Universitas Budi Luhur Student Paper	4%
3	e-journal.uajy.ac.id Internet Source	3%
4	123dok.com Internet Source	2%
5	repository.uin-suska.ac.id Internet Source	2%
6	ejournal.upnvj.ac.id Internet Source	2%
7	jtiik.ub.ac.id Internet Source	1%
8	id.123dok.com Internet Source	1%
9	economy.okezone.com Internet Source	1%

10	jurnal.teknokrat.ac.id Internet Source	1 %
11	Submitted to University of Melbourne Student Paper	1 %
12	ejurnal.stmik-budidarma.ac.id Internet Source	1 %
13	ieomsociety.org Internet Source	<1 %
14	Eko Budi Santoso, Aryo Nugroho. "Analisis Sentimen Calon Presiden Indonesia 2019 Berdasarkan Komentar Publik Di Facebook", Eksplora Informatika, 2019 Publication	<1 %
15	repository.upnvj.ac.id Internet Source	<1 %
16	proceeding.unived.ac.id Internet Source	<1 %
17	indonesiabaik.id Internet Source	<1 %
18	Submitted to Sriwijaya University Student Paper	<1 %
19	sistemasi.ftik.unisi.ac.id Internet Source	<1 %
20	www.liputan6.com Internet Source	<1 %

21	www.researchgate.net Internet Source	<1 %
22	Delpima Suhita, Melinedri Melinedri, Tika Meldina, Yulya Muharmi, Nella Novita. "Gadget dan Perkembangan Sosial Anak Usia Dini", <i>Zuriah : Jurnal Pendidikan Anak Usia Dini</i> , 2020 Publication	<1 %
23	Submitted to Universitas Brawijaya Student Paper	<1 %
24	dprexternal3.dpr.go.id Internet Source	<1 %
25	ejournal.uksw.edu Internet Source	<1 %
26	etd.repository.ugm.ac.id Internet Source	<1 %
27	familylaw2018.com.au Internet Source	<1 %
28	jurnal.unsil.ac.id Internet Source	<1 %
29	Meylan Wongkar, Apriandy Angdresey. "Sentiment Analysis Using Naive Bayes Algorithm Of The Data Crawler: Twitter", 2019 Fourth International Conference on Informatics and Computing (ICIC), 2019 Publication	<1 %

30	Yudhy Lady Joane, Alicia Sinsuw, Agustinus Jacobus. "Rancang Bangun Aplikasi Deteksi Kemiripan Dokumen Teks Menggunakan Algoritma Ratcliff/Obershelp", Jurnal Teknik Informatika, 2017 Publication	<1 %
31	adoc.tips Internet Source	<1 %
32	doku.pub Internet Source	<1 %
33	download.garuda.ristekdikti.go.id Internet Source	<1 %
34	journal.itltrisakti.ac.id Internet Source	<1 %
35	repository.uinjkt.ac.id Internet Source	<1 %
36	tel.archives-ouvertes.fr Internet Source	<1 %
37	www.gesuri.id Internet Source	<1 %
38	www.pantau.com Internet Source	<1 %
39	Bayu Yanuargi. "Analisis Sentemen Terhadap Aplikasi Bukalapak Sebelum IPO dan Sesudah	<1 %

IPO Menggunakan Algoritma Naive Bayes", JNANALOKA, 2022

Publication

40

Suprianto, Muhammad Fadlan, Muhammad, Yusni Amaliah, Mussallimah. "Retrieval Information Using Generalized Vector Space Models And Sentiment Analysis Using Naive Bayes Classifier For Evaluation Of Lecturers By Students", 2020 Fifth International Conference on Informatics and Computing (ICIC), 2020

Publication

<1 %

Exclude quotes Off

Exclude matches Off

Exclude bibliography Off

Jurnal final turnitin

PAGE 1

PAGE 2

PAGE 3

PAGE 4

PAGE 5

PAGE 6

PAGE 7

PAGE 8

PAGE 9

PAGE 10
