



Analisis Sentimen Ibukota Negara Baru Menggunakan Metode Naïve Bayes Classifier

Dewi Aryanti*

Sistem Informasi, STMIK Borneo Internasional Balikpapan, Balikpapan
Jl. A.W. Syahrani No.04, RT.32, Batu Ampar, Kec. Balikpapan Utara, Kota Balikpapan, Kalimantan Timur, Indonesia
Email: dewi.19@stmik-borneo.ac.id

Email Penulis Korespondensi: dewi.19@stmik-borneo.ac.id

Submitted: 23/07/2022; Accepted: 30/07/2022; Published: 31/07/2022

Abstrak—Twitter adalah salah satu situs web media sosial yang paling populer di internet saat ini. Situs ini mencakup berbagai topik, termasuk hukum, politik, budaya, ekonomi dan banyak lagi. Di Twitter, banyak orang membicarakan ide pemindahan ibukota Indonesia. Namun ada ketidaksepakatan antara orang-orang terkait wacana tersebut, karena masing-masing pihak memiliki pendapat dan pandangan yang berbeda. Untuk memahami proses data yang diekstraksi secara otomatis dari teks tidak terstruktur, maka analisis sentimen dikembangkan. Proses ini digunakan untuk menentukan opini keseluruhan dari suatu populasi tentang suatu subjek dalam wacana publik, seperti yang dapat dilihat dari beragam pendapat di Twitter. Dalam penerapan analisis sentimen ini penelitian menggunakan metode naïve bayes untuk diterapkan pada tweets topik ibukota baru untuk tujuan klasifikasi kelas sentimen pada media sosial twitter. Teknik klasifikasi dilakukan dengan cara mengklasifikasikan menjadi 2 kelas yaitu positif dan negatif. Berdasarkan hasil pengujian yang dilakukan terhadap tweets sentimen ibukota baru dari media sosial twitter sebanyak 1.065 cuitan yang diuji sentimen mengenai mengenai ibukota baru dari media sosial twitter 289 menyatakan positif dan 776 menyatakan negatif (Metode Naïve Bayes diperoleh akurasi = 94,18%,)

Kata Kunci: Twitter; Ibukota; Analisis Sentimen; Naïve Bayes; Klasifikasi

Abstract—Twitter is one of the most popular social media websites on the internet today. The site covers a wide range of topics, including law, politics, culture, economics and more. On Twitter, many people talked about the idea of moving the capital of Indonesia. But there is disagreement between people regarding the discourse, because each side has different opinions and views. In order to understand the process of data automatically extracted from unstructured texts, then sentiment analysts were developed. This process is used to determine the overall opinion of a population about a subject in public discourse, as can be derived from diverse opinions on Twitter. In the application of sentiment analysis, the study used the naïve bayes method to be applied to tweets of new capital topics for the purpose of class classification of sentiment on social media tweeters. Technical classification is carried out by classifying into 2 classes, namely positive and negative. Based on the results of testing conducted on the new capital's sentiment tweets from social media twitter as many as 1,065 tweets tested sentiment about the new capital from social media twitter 289 stated positive and 776 stated negative Naïve Bayes method obtained accuracy = 94.18%,)

Keywords: Twitter; Capital; Sentiment Analysis; Naïve Bayes; Classification

1. PENDAHULUAN

Saat ini, media sosial berkembang pesat di Internet, dan salah satu yang paling populer adalah Twitter (Zulfa dan Winarko, 2017). Seperti media sosial lainnya, Twitter adalah platform media sosial yang memungkinkan pengguna untuk berinteraksi secara pribadi atau publik. Berbagai topik dibicarakan di Twitter, antara lain politik, ekonomi, sosial, budaya dan hukum (Jurnal Teran 148 Teknologi Informasi dan Ilmu Komputer (JTIK), Vol. 8, No. 1, Februari 2021, hlm. 147-156 dan Mancera, 2019) (Kusen dan Stembeck, 2018) (Öztürk dan Ayvaz, 2017). Fitur hashtag memungkinkan pengguna Twitter untuk melihat apa yang sedang dibahas secara real time. Selain itu, kata kunci di Twitter juga bisa menjadi sumber percakapan bagi pengguna. Tidak jarang item berita muncul sebagai berita utama di media online, dengan diskusi yang sudah berlangsung di Twitter. Bahkan diskusi di Twitter dapat memicu berita (Priyatno dkk., 2019). Saat ini pengguna twitter khususnya Indonesia telah memberikan dampak besar pada sudut pandang topik tertentu (Pandey Rajpoot dan Saraswat, 2017).

Salah satu trending topik di Twitter adalah Hal itu terkait dengan isu ibukota baru di Indonesia. Wacana Berita tentang ibukota baru terus bermunculan, tetapi wacana tentang tersebut kembali trending setelah pemerintahan Presiden Jokowi membahas hal tersebut (Septiana, 2018). Wacana Ibukota Baru adalah bagian dari kepentingan nasional sebagai upaya menyongsong Persaingan global (Toun, 2018). Tapi di balik hal tersebut Para pihak percaya bahwa ada perselisihan Pro dan kontra, masing-masing memiliki kelebihannya sendiri dengan Perspektif yang berbeda. Hal ini menyebabkan Munculnya fenomena perdebatan, khususnya di twitter yang menunjukkan kepedulian terhadap wacana publik tersebut (Andriani, 2018). Kecendrungan pengguna twitter dalam memposting konten dapat diketahui dengan cara analisis sentimen (Pritama, Andrian, dan Nugroho, 2019). Dengan memisahkan pendapat yang berbeda berdasarkan kelas positif dan negatif.

Analisis sentimen adalah proses ekstraksi, Secara otomatis memproses dan memahami data dalam bentuk teks tidak terstruktur untuk mengambil informasi sentimen yang terkandung dalam kalimat opini atau opini (Brah, Touahria, & Tari, 2019). Analisis sentimen untuk menilai opini dan kecendrungan terhadap suatu topik memiliki kecenderungan negatif dan positif (Rozi, Pramono & Dahlan, 2012). Analisis sentimen dapat diterapkan pada perspektif di semua bidang termasuk ekonomi, politik, masyarakat, dan ekonomi hukum. Media sosial Twitter ini membuka jendela bagi peneliti untuk mempelajari sentimen, sentimen, dan opini publik melalui analisis sentimen

(Qiu, Lin, dan Shuai, 2019). Dalam penerapan Analisis Sentimen Menggunakan metode pembelajaran mesin memiliki Beberapa metode yang umum digunakan seperti KNN, Naive Bayes, Random Forest dan support Vectore Machine (SVM) (Zuhdi, Utami, dan Raharjo, 2019) (Wang, Liu, dan Ma, 2019) (Fauzi, 2018) (Thakur dan Deshpande, 2019). Banyak teknik analisis yang digunakan dalam analisis sentimen, karena sebagian besar pertanyaan adalah kalimat non-standar dari pengguna media sosial. Keakuratannya akan diklasifikasikan sebagai positif atau negatif. Saat ini, analisis sentiment Algoritma seperti Naive Bayes, Maximum Entropy, dan Support Vector Machines biasanya digunakan. Kehadiran Twitter dan penggunaannya menawarkan kesempatan bagi peneliti untuk mempelajari lebih lanjut. Membahas Opini melalui Analisis Sentimen Beberapa metode pembelajaran mesin yang sering digunakan dalam analisis sentimen antara lain K-Nearest Neighbors (K-NN), Naive Bayes Classifiers (NBC), Random Forests, dan Decision Trees. Metode NBC adalah salah satu metode klasifikasi dalam teknik data mining yang menerapkan teori bayes. NBC sendiri merupakan metode klasifikasi yang sangat efektif jika dibandingkan dengan metode lainnya yang menunjukkan bahwa NBC lebih unggul dalam hal akurasi, presisi, recall dan f-measure.

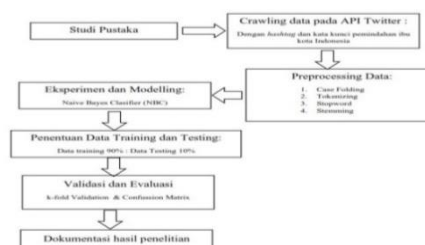
Pada penelitian sebelumnya Analisis sentimen juga telah dipelajari di bidang ekonomi. Fransiska (2019) menganalisis sentimen konsumen pada toko online JD.com. Opini konsumen ini diklasifikasikan menjadi sentimen positif, negatif, dan netral menggunakan algoritma pengklasifikasi Naive Bayes. Hasil analisis menunjukkan akurasi 96,44% [32]. Manuel J.Sanchez-Franco et al., 2018 juga melakukan analisis sentimen di bidang ekonomi, yaitu mengidentifikasi layanan di industri perhotelan. Dalam penelitian Menggunakan data berupa ulasan pelanggan, total 47.172 ulasan berasal dari 33 hotel di Las Vegas. Data Dari media sosial Yelp. Klasifikasi dilakukan dengan menggunakan algoritma classifier Naive Bayes. Tak hanya dalam bidang ekonomi, Mehdi (2019) melakukan penelitian dalam bidang politik mengenai Analisa Sentimen menggunakan algoritma naïve bayes untuk mengidentifikasi kecenderungan suatu opini publik di twitter tentang pemilihan presiden di Indonesia pada tahun 2019. Hasil analisa dalam penelitian tersebut menyebutkan hasil positif yaitu 79.5% untuk sentimen Jokowi i-Ma'rif dan 64% untuk Prabowo-Sandi. Budi Haryanto dkk (2019) juga melakukan penelitian pada subjek yang sama, yaitu analisis sentimen subjek pemilihan presiden Indonesia di tahun yang sama. Studi ini mengkategorikan sentimen di beberapa komunitas di media sosial Facebook. Komunitas tersebut adalah @detikcom, @tribunews dan liputan6online. Dalam penelitian ini, algoritma pengklasifikasi Naive Bayes juga digunakan dalam proses analisis sentimen. Hasil analisis menunjukkan bahwa Jokowi-Ma'rif memiliki sentimen positif sebesar 56,76% dan sentimen negatif sebesar 43,24%. Sementara itu, pasangan mata uang Prabowo-Sandi menunjukkan sentimen sebesar 24,21% sentimen positif dan 75,79% sentiment negatif. Dari penelitian tersebut didapat kesimpulan bahwa algoritma yang digunakan cepat, terukur dan akurat.

Berdasarkan tinjauan pustaka, penulis mengusulkan untuk menerapkan metode naïve Bayes untuk diterapkan pada tweet topik ibu kota baru Indonesia dan untuk mengklasifikasikan kategori sentimen di media sosial Twitter. Naïve Bayes adalah metode untuk mengklasifikasi dengan menggunakan metode statistic dan probabilitas yang dikemukakan oleh ilmuwan Inggris yaitu Thomas Bayes (Rifai, Jatnika, & Valentino, 2019) (Syarli, 2016). Naïve Bayes adalah penghitungan secara statistik untuk memprediksi peluang yang akan datang berdasarkan pengalaman atau permasalahan yang dihadapi sebelumnya sehingga dikenal sebagai Teorema Bayes (Pratiwi & Nugroho, 2016). Dengan penerapan metode naïve bayes, diharapkan hasil akurasi yang baik, jika melihat dari kecenderungan opini publik yang terkandung dalam kontroversi media sosial Twitter tentang ibu kota baru. Dengan adanya penelitian ini yang menggunakan metode Naive Bayes Classifier (NBC) untuk menganalisis sentimen wacana pemerintah tentang topik di media massa online Twitter yaitu ibukota baru yang nantinya akan diklasifikasikan menjadi dua kategori yaitu positif dan negatif. Melalui analisis sentimen ini diharapkan dapat mengidentifikasi isu-isu yang terkandung dalam kontroversi tersebut dengan topik ibu kota Indonesia agar dapat dijadikan bahan evaluasi untuk penggunaan selanjutnya

2. METODOLOGI PENELITIAN

2.1 Tahapan Penelitian

Penelitian ini berjenis eksperimen, Adapun tahapan yang akan dilakukan ditunjukkan pada Gambar 1. Gambar merupakan tahap penelitian, dimulai dengan tahap penelitian kepustakaan. Pada tahap ini, beberapa referensi dengan Topik penelitian dan penyaringan literatur yang relevan saat ibu kota baru.



Gambar 1. Tahap Penelitian



Tabel 1. Data Hasil Crawling

No	Text
1	Fraksi PKS @FPKSDPRRI makin bulat menolak rencana pemindahan Ibukota Negara (IKN) menyusul banjir yang terjadi di Kabupaten Penajam Paser Utara (PPU), Kalimantan Timur . https://t.co/mPxtTweJh5
2	Pengamat Ekonomi: Cara Berpikir Pemerintah Soal Pemindahan Ibukota Negara Masih Kuno. Di zaman serba digital sekarang ini semestinya Pemerintah berpikir ke depan dan lebih mengutamakan digitalisasi yang masif. https://t.co/VlzEzTuvuq
3	Menurut gw Jakarta tu udah masuk kategori Trianapolis. Udah susah banget disembuhin dan dijadikan tempat layak bagi semua orang. Makanya gw dukung sih pemindahan ibukota. Malah kaya, “Kok ga dari dulu sih?”.
4	RT @Kaca678910: Pemindahan Ibukota Negara SANGAT TIDAK ADA URGENSINYA semua karena Dendam dan AMBISI penguasaan Lahan ² di Pulau Kalimantan,...
5	sch! ikut cerelisasi itu dimana dan gimana? ????
6	? Trivia: Stasiun KA Bandung dirancang untuk persiapan pemindahan ibukota dari Jakarta ke Bandung RT @alextham878: Slamet Maarif : 'Setop proyek ibukota baru krn mubazir & gk ada manfaatnya.' Pemindahan ibu kota melalui kajian dr seluru...
7	@alextham878 Jika PA 212 ngomong begitu berarti pemindahan ibukota sudah benar.
8	Pemindahan ibukota kabupaten yang dari zaman Belanda... sampai skg ada yg blm sepenuhnya rampung, misalnya Kediri blm sepenuhnya pindah ke Ngasem krn masih digantung pindah tidaknya ke Pare
9	RT @m_sael: Proyek pemindahan ibukota kabupaten pertama dalam sejarah Indonesia ada di Bandung, dulu dipindahkan ke Baleendah tp hanya +- 1...
10	Proyek pemindahan ibukota kabupaten pertama dalam sejarah Indonesia ada di Bandung, dulu dipindahkan ke Baleendah tp hanya +- 10 tahun. Lalu pindah ke Soreang pd 1987. Yg di era Orde Baru lainnya... Kab Semarang dipindahin ke Ungaran, Magelang dipindah ke Mungkid
11	@SoundOfYogi Kemarin2 @m_sael QRT postingan dia soal pemindahan ibukota malah di QRT balik jg sm si Mus https://t.co/5fIs2v691i
12	RT @alextham878: Slamet Maarif : 'Setop proyek ibukota baru krn mubazir & gk ada manfaatnya.' Pemindahan ibu kota melalui kajian dr seluru...

Tahap kedua adalah melakukan Crawling data media sosial. Berbagai media sosial telah mengubah cara untuk berekspresi dan mengeluarkan pendapat. Setiap pengguna bisa mengunggahnya melalui berbagai situs, salah satunya Twitter. Berbagai informasi yang ada di dalam Twitter tersebut tentunya akan dapat dianalisis untuk mempermudah dalam pengambilan keputusan tidak terkecuali pada kasus pemindahan ibu kota yang menjadi topik penelitian ini. Data primer dikoleksi dari Twitter menggunakan API Twitter. Adapun topik data teks yang diambil dari Twitter merupakan percakapan terkait ibukota baru dari November 2021 sampai dengan Desember 2021, seperti tampak pada Tabel 1. Pada tahapan ini cara untuk filter topik yakni dengan menginputkan hastag atau kata kunci “ibukota baru, pindah ibukota, ikn”. Jumlah data yang berhasil diCrawling menggunakan library tweepy adalah sebanyak 3528 tweets merupakan data non-labelling. Kemudian selanjutnya adalah memeriksa setiap data dan melabelkannya ke dalam dua kelas yaitu positif dan negatif. Tahapan ketiga yakni preprocessing. Pada tahapan ini dilakukan beberapa operasi yaitu case folding, tokenizing, stopword, filter example (by length). Tahap keempat merupakan tahap eksperimen dan pemodelan. Pada tahap ini model jaringan klasifikasi dibentuk menggunakan algoritma Naive Bayes. Setelah penentuan model klasifikasi, selanjutnya adalah tahap kelima yaitu melakukan splitting data dengan perbandingan data training dan data testing adalah 90:10 yang kemudian dikomputasi menggunakan metode Naive Bayes sehingga menghasilkan luaran. Dari data luaran testing kemudian langkah keenam dilakukan evaluasi performansi menggunakan metode k-fold validation dan confusion matrix. Pada tahap ini dilakukan tuning parameter untuk mengoptimalkan performansi algoritma. Adapun perangkat lunak yang digunakan adalah Rapidminer. Tahap ketujuh atau terakhir merupakan tahap menyusun dokumentasi hasil penelitian yang nantinya dapat digunakan untuk berbagai keperluan.

Metode Preprocessing Dataset yang dikumpulkan atau dicrawling dari twitter belum terstruktur dengan baik. Oleh karena itu dibutuhkan tahapan pra-proses termasuk beberapa langkah berikut yaitu case folding, tokenizing, stopword, dan stemming. Namun sebelumnya teks akan dilakukan handling duplicate, yaitu memeriksa dataset apakah ada data yang ganda atau tidak. Maka dari 3528 data diseleksi kembali menjadi 1.065 data yang

terbebas dari data ganda. Case folding merupakan langkah yang bertujuan untuk mengkonversi upper-case menjadi lower-case yang diperlukan untuk tahap pemrosesan berikutnya. Sedangkan tokenizing adalah proses memenggal kalimat menjadi kata per kata. Proses selanjutnya adalah stopwords yaitu merupakan filter terhadap kata-kata yang tidak diperlukan dalam pemrosesan. Kemudian yang terakhir adalah filter example by length, adalah suatu proses untuk menghilangkan kata yang tidak penting. Setelah seluruh preprocessing dilakukan, maka dataset siap untuk diklasifikasi. Pada tahap ini juga dilakukan splitting data training dan testing. Adapun pembagiannya adalah sebanyak 915 data training dan data testing sebanyak 150 data, total keseluruhan adalah 1.065 data.

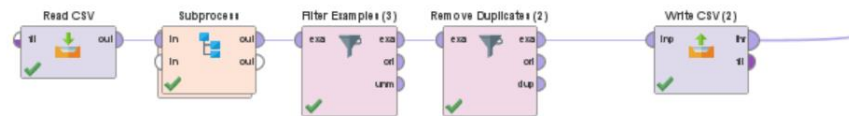
Analisis sentiment adalah proses penentuan polaritas, pengkategorian sentiment dan pengelompokan berdasarkan jumlah yang telah ditargetkan. Metode naïve bayes classifier adalah metode klasifikasi berdasarkan teorema bayes. Berikut algoritma pengklasifikasian naïve bayes,

Algoritma

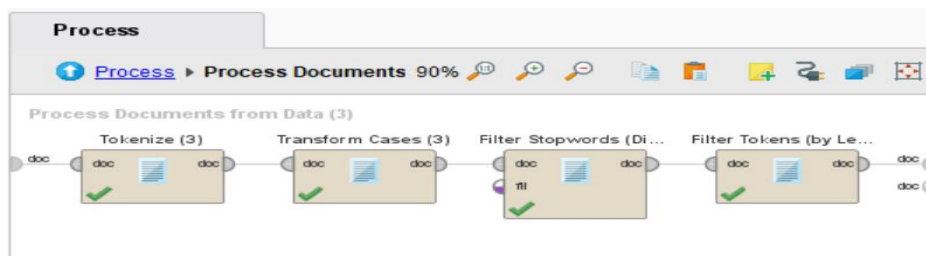
1. Mempersiapkan data latih dari dataset beserta dengan target kelas, katakanlah kelas A dan B.
2. Probabilitas Prior dari kedua kelas A dan B dikalkulasi seperti berikut :
 - Kelas A = jumlah objek dari kelas A dibagi jumlah total dari keseluruhan objek
 - Kelas B = jumlah objek dari kelas B dibagi jumlah total dari keseluruhan objek
3. Mengkalkulasi jumlah total dari kemunculan kata kedua kelas A dan B misalnya n_i
 - n_a = jumlah total dari frekuensi kata dari kelas A
 - n_b = jumlah total dari frekuensi kata dari kelas B
4. Mengkalkulasi probabilitas kondisional dari kata kunci kejadian untuk setiap kelas yang diberikan.
 - $P(\text{kata1} / \text{kelas A}) = \text{JumlahKata} / n_i(A)$
 - $P(\text{kata1} / \text{Kelas B}) = \text{JumlahKata} / n_i(B)$
 - $P(\text{kata2} / \text{Kelas A}) = \text{JumlahKata} / n_i(A)$
 - $P(\text{kata2} / \text{Kelas B}) = \text{JumlahKata} / n_i(B)$
 -
 - $P(\text{kataN} / \text{kelas B}) = \text{JumlahKata} / n_i(B)$
5. Distribusi harus dilakukan secara berurutan untuk menghindari masalah frekuensi nol (zero frequency).
6. Sekarang suatu dokumen baru dari M diklasifikasikan berdasarkan perhitungan distribusi dari kedua kelas A dan B. $P(M/W)$.
 - Mencari $P(A / W) = P(A) * P(\text{kata1} / \text{kelas A}) * P(\text{kata2} / \text{kelas A}) \dots \dots (P(\text{kataN} / \text{kelas A}))$
 - Mencari $P(B / W) = P(B) * P(\text{kata1} / \text{kelas B}) * P(\text{kata2} / \text{kelas B}) \dots \dots (P(\text{kataN} / \text{kelas B}))$
7. Setelah menghitung probabilitas untuk kedua kelas A dan B, kelas dengan probabilitas tertinggi adalah kelas baru untuk dokumen M.

3. HASIL DAN PEMBAHASAN

Dalam proses analisis sentimen, terdapat tiga jenis cakupan, yaitu level dokumen, level kalimat, dan level aspek. Analisis sentimen pada level dokumen akan direpresentasikan dalam bentuk kelas positif, negatif atau opini. Pada level kalimat, analisis sentimen akan diekspresikan dalam setiap kalimat, dimulai dengan mengidentifikasi subjek atau objek dalam kalimat dan kemudian mengidentifikasinya sebagai positif atau negatif. Karena kalimat adalah bentuk dokumen pendek, tingkat aspek dan tingkat kalimat dapat diabaikan. Namun, mungkin juga ada beberapa celah di tingkat dokumen yang memerlukan lebih banyak masukan. Sedangkan tingkat aspek bisa berarti memiliki entitas di atas tingkat kalimat, misalnya dalam bentuk paragraf. Adapun di Penelitian ini dilakukan pada tataran dokumen dan tidak melihat subjek atau objek kalimat.



Gambar 2. Desain Preprocessing



Gambar 3. Tahapan tokenizing, transform cases dan stopwords

ExampleSet (Remove Duplicates (2))

Open in Turbo Prep Auto Model Filter (1,065 / 1,065 examples)

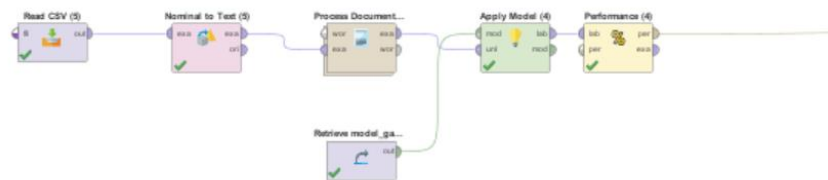
Row No.	Text
1	Fraksi PKS makin bulat menolak rencana pemindahan Ibukota Negara IKN menyusul banjir yang terjadi di Kabupaten Penajam Paser Utara ...
2	Pengamat Ekonomi: Cara Berpikir Pemerintah Soal Pemindahan Ibukota Negara Masih Kuno Di zaman serba digital sekarang ini semestiny...
3	Menurut gw Jakarta tu udah masuk kategori Trianapolis Udah susah banget disembuhin dan dijadikan tempat layak bagi semua orang Maka...
4	Pemindahan ibukota Negara SANGAT TIDAK ADA URGENSINYA semua karena Dendam dan AMBISI penguasaan Lahan* di Pulau Kalimantan...
5	sch ikut cerelisasi itu dimana dan gimana
6	Trivia: Stasiun KA Bandung dirancang untuk persiapan pemindahan ibukota dari Jakarta ke Bandung
7	Pemindahan ibu kota melalui kajian dr seluruh ...
8	Jika PA 212 ngomong begitu berarti pemindahan ibukota sudah benar
9	Pemindahan ibukota kabupaten yang dari zaman Belanda sampai skg ada yg bim sepenuhnya rampung
10	Proyek pemindahan ibukota kabupaten pertama dalam sejarah Indonesia ada di Bandung
11	Yg di era Orde Baru lainnya Kab Semarang dipindahin ke Ungaran
12	Kemarin2 Qpostingan dia soal pemindahan ibukota malah di Qbalik jg sm si Mus
13	Pemerintah dalam hal terkail pemindahan ibukota Baru harus bisa melakukan pemetaan agar terbebas dari rawan bencana seh...

Gambar 4. Hasil dari tahapan preprocessing

Dataset berjumlah 1065 dilakukan preprocessing dengan urutan case folding, tokenizing, transform cases dan stopword, seperti pada Gambar 3. Setelah itu dataset dilakukan pembagian data 120 dan 30 secara manual, kemudian mengacak susunan data training set. Tujuannya adalah untuk mengatasi permasalahan kombinasi data yang tidak tersebar pada dataset. Metode ini akan melakukan training dan testing sebanyak pada posisi data yang berbeda. Hasil akurasi dari kedua pembagian data testing akan dirata-ratakan, sehingga dapat memvalidasi seluruh bagian dataset, ilustrasi tampak seperti pada Gambar 4.



Gambar 5. Desain Pengujian Metode Naïve Bayes



Gambar 6. Desain Pengujian Data Latih 120

Result History PerformanceVector (Performance (4))

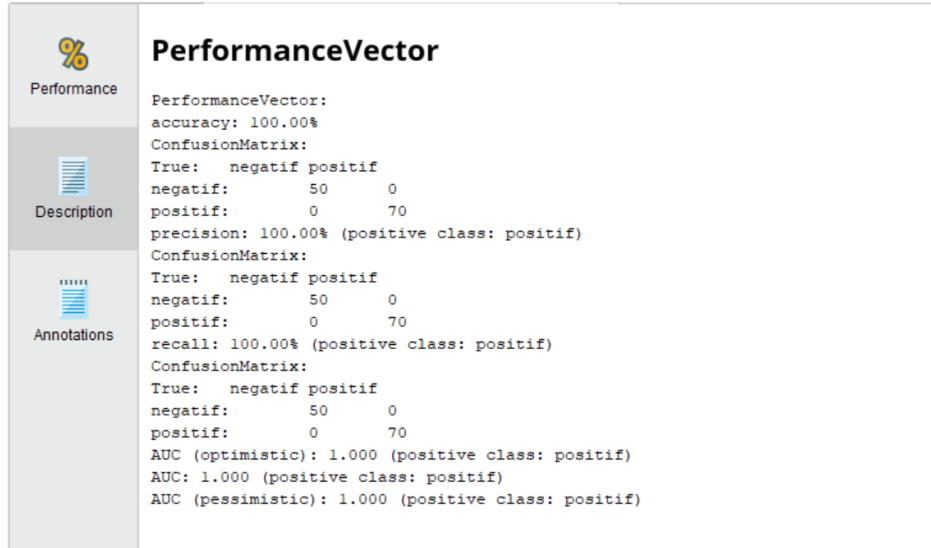
Criterion: accuracy, precision, recall, AUC (optimistic), AUC, AUC (pessimistic)

Table View Plot View

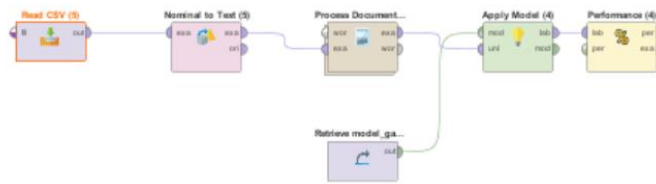
accuracy: 100.00%

	true negatif	true positif	class precision
pred. negatif	50	0	100.00%
pred. positif	0	70	100.00%
class recall	100.00%	100.00%	

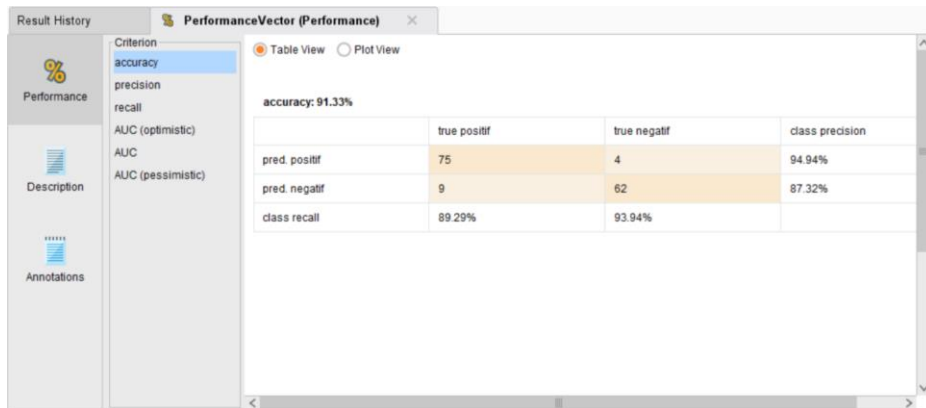
Gambar 7. Hasil Prediksi dan Akurasi Metode Naïve Bayes dengan Data Latih 120



Gambar 8. Description hasil data latih 120



Gambar 9. Desain Pengujian Data Latih 30

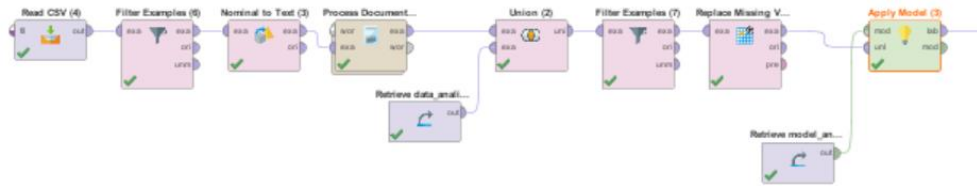


Gambar 10. Hasil Prediksi dan Akurasi Metode Naïve Bayes dengan Data Latih 30

Pengujian dilakukan dengan menggunakan software Rapidminer versi 9.9. Pada proses dari pembacaan file dataset hingga proses pembagian data latih 150 yang dibagi secara manual menjadi 120 dan 30 dapat dilihat melalui Gambar 6 dan 8. Pada Gambar tersebut terlihat bahwa terdapat proses pembacaan dari file dataset sebanyak dua buah file masukan data latih yang telah dilakukan tahapan preprocessing sebelumnya. Pada tahap ini dilakukan penentuan fungsi Map yaitu menspesifikkan nilai-nilai dari atribut yang dipilih untuk nilai-nilai yang baru, dan operator ini dapat diterapkan untuk jenis data baik numerik maupun nominal. Sedangkan set role adalah atribut yang digunakan pada dataset yang dikonversi terlebih dahulu dari nominal ke teks untuk kemudian dataset diproses. Proses inti yaitu klasifikasi menggunakan metode naïve bayes ditunjukkan seperti gambar 5. Pada tahapan ini merupakan implementasi pemodelan naïve bayes, dimana pada Langkah ini dilakukan proses training dan testing data set.

Setelah mengaplikasikan model Naïve Bayes, maka tahap selanjutnya adalah menguji performansi dataset seperti terlihat pada Gambar 10. Hasil pengujian dari dataset yang telah dilatih menggunakan menunjukkan bahwa akurasi yang dicapai sebesar 94,18%. Artinya system sudah cukup baik dalam mendeteksi kelas sentimen. Adapun uraian terperinci data confusion matrix terlihat pada Gambar 11. Nilai TN=771, TP=345, FP=55 dan FN=14. Sehingga dapat diketahui dataset training berlabel negatif sebanyak 826 dan data berlabel positif sebanyak 359. Untuk class recall untuk label aktual negatif adalah 93,34% dan class recall untuk label aktual positif yakni

96,10%. Sedangkan untuk nilai class precision pada prediksi negatif adalah 98,22% dan class precision untuk semua prediksi positif adalah 86,25%, seperti tampak pada gambar 11.



Gambar 11. Desain Pengujian Data Uji

PerformanceVector (Performance (2))			
Table View <input checked="" type="radio"/> Plot View <input type="radio"/>			
Criterion	accuracy	accuracy: 94.18%	
	precision		
	recall		
	AUC (optimistic)		
	AUC		
	AUC (pessimistic)		
		true negatif	true positif
			class precision
	pred. negatif	771	14
	pred. positif	55	345
	class recall	93.34%	96.10%

Gambar 12. Hasil Prediksi dan Akurasi Metode Naïve Bayes Data Uji

PerformanceVector

```

PerformanceVector:
accuracy: 94.18%
ConfusionMatrix:
True: negatif positif
negatif: 771 14
positif: 55 345
precision: 86.25% (positive class: positif)
ConfusionMatrix:
True: negatif positif
negatif: 771 14
positif: 55 345
recall: 96.10% (positive class: positif)
ConfusionMatrix:
True: negatif positif
negatif: 771 14
positif: 55 345
AUC (optimistic): 0.981 (positive class: positif)
AUC: 0.914 (positive class: positif)
AUC (pessimistic): 0.910 (positive class: positif)
    
```

Gambar 13. Description hasil Data Uji

Dari hasil pengujian dapat terlihat bahwa terdapat 55 data yang seharusnya memiliki nilai aktual sentimen negatif dideteksi sebagai sentimen positif. Jika dihitung dalam persen maka nilai FP ini adalah sebesar 5,23% yang berarti masih dapat ditoleransi. Terlebih kesalahan positif dalam kasus sentimen politik lebih baik terjadi dibandingkan dengan kesalahan negatif, dikarenakan sentimen politik merupakan hal yang sangat sensitif dan berpeluang untuk menimbulkan keresahan publik. Selain itu untuk mengukur seberapa baik performansi harmonik precision dan recall dari suatu model klasifikasi juga dapat menggunakan F1 Score, yang dapat diukur dengan menggunakan persamaan (1). Berdasarkan persamaan (1) dan nilai precision=0,87 dan recall=0,96, maka didapatkan nilai F1 Score sebesar 0,91 yang mengindikasikan bahwa model klasifikasi sudah memiliki precision dan recall yang baik

4. KESIMPULAN

Penelitian ini bertujuan untuk mengklasifikasikan sentimen yang terkait dengan ibukota baru dari media sosial Twitter menggunakan classifier Naive Bayes. Hasil dari dataset sebanyak 1065 data didapatkan nilai akurasi 94,18%, presisi 0,87, dan reccal 0,96. Harmonized Recall and Precision atau yang juga dikenal dengan F1-Score memiliki nilai 0,92 yang artinya sistem ini bagus dalam mendeteksi sentimen. Nilai recall yang tinggi juga bekerja dengan baik untuk sistem ini, karena politik adalah isu yang sensitif. Penelitian selanjutnya dapat direkomendasikan untuk menambahkan dataset dan menambahkan perbandingan dengan model classifier lainnya.

REFERENCES

[1] M. Safiullah, P. Pathak, S. Singh, and A. Anshul, "Social media as an upcoming tool for political marketing effectiveness," *Asia Pacific Manag. Rev.*, vol. 22, pp. 10–15, 2017.
 [2] K. Sailunaz and R. Alhadj, "Emotion and Sentiment Analysis from Twitter Text," *J. Comput. Sci.*, 2019.



- [3] D. Neu, G. Saxton, A. Rahaman, and J. Everett, “Twitter and social accountability : Reactions to the Panama Papers,” *Crit. Perspect. Account.*, vol. 61, pp. 38–53, 2019.
- [4] I. Zulfa and E. Winarko, “Sentimen Analisis Tweet Berbahasa Indonesia dengan Deep Belief Network,” *IJCCS*, vol. 11, no. 2, 2017.
- [5] F. K. Ningrum and M. M. Syarah, “INSTAGRAM DAN TWITTER SEBAGAI STRATEGI HUMAS PT JASA MARGA (PERSERO) TBK DALAM MENINGKATKAN PELAYANAN,” *J. Commun. Sci. Islam. Da’wah*, vol. 2, no. 2, pp. 122–133, 2018.
- [6] L. Terán and J. Mancera, “Dynamic profiles using sentiment analysis and twitter data for voting advice applications,” *Gov. Inf. Q.*, no. February 2018, pp. 1–16, 2019.
- [7] E. Kušen and M. Strembeck, “Politics , sentiments , and misinformation : An analysis of the Twitter discussion on the 2016 Austrian Presidential Elections,” *Online Soc. Networks Media*, vol. 5, pp. 37–50, 2018.
- [8] N. Öztürk and S. Ayvaz, “Sentiment Analysis on Twitter : A Text Mining Approach to the Syrian Refugee Crisis,” *Telemat. Informatics*, no. October, 2017.
- [9] A. M. Priyatno, M. M. Muttaqi, F. Syuhada, and A. Z. Arifin, “Deteksi Bot Spammer Twitter Berbasis Time Interval Entropy dan Global Vectors for Word Representations Tweet ’ s Hashtag,” *J. Ilm. Teknol. Sist. Inf.*, no. June, 2019.
- [10] A. M. Zuhdi, E. Utami, and S. Raharjo, “ANALISIS SENTIMEN TWITTER TERHADAP CAPRES INDONESIA 2019 DENGAN METODE K-NN,” *J. Inf. Politek. Indonusa Surakarta*, vol. 5, pp. 1–7, 2019.
- [11] Q. Wang, K. Liu, and K. Ma, “Emotional Analysis of Public Opinions in Colleges and Universities : Based on Naive Bayesian Classification Method,” *J. Phys.*, 2019.
- [12] M. A. Fauzi, “Random Forest Approach fo Sentiment Analysis in Indonesian Language,” *Indones. J. Electr. Eng. Comput. Sci.*, vol. 12, no. 1, pp. 46–50, 2018.
- [13] R. K. Thakur and M. V. Deshpande, “Kernel Optimized-Support Vector Machine and Mapreduce framework for sentiment classification of train reviews,” *Indian Acad. Sci.*, vol. 44, no. 1, pp. 1–14, 2019. [25] L. Oktasari, Y. H. Chrisnanto, and Yuniarti, “Text Mining Dalam Analisis Sentimen Asuransi Menggunakan Metode Naive Bayes Classifier,” in *Prosiding SNST*, 2016, pp. 37–42.
- [14] D. Xhemali, C. J. Hinde, and R. G. Stone, “Naïve Bayes vs . Decision Trees vs . Neural Networks in the Classification of Training Web Pages,” *Int. J. Comput. Sci. Issues*, vol. 4, no. 1, pp. 16–23, 2009.
- [15] M. Bilal, H. Israr, M. Shahid, and A. Khan, “Sentiment classification of Roman-Urdu opinions using Naive Bayesian , Decision Tree and KNN classification techniques,” *J. King Saud Univ. - Comput. Inf. Sci.*, vol. 28, no. 3, pp. 330–344, 2015.
- [16] V. A. Fitri, R. Andreswari, and M. A. Hasibuan, “Sentiment Analysis of Social Media Twitter with Case of Anti- LGBT Campaign in Indonesia using Naïve Bayes , Decision Tree , and Random Forest Algorithm Algorithm,” *Procedia Comput. Sci.*, vol. 161, pp. 765–772, 2019.
- [17] M. M. Ismail and K. M. Lhaksamana, “Sentimen Analisis Pada Media Online Mengenai Pemilihan Presiden 2019 Dengan Menggunakan Metode Naive Bayes,” in *e-proceeding of engineering*, 2019, vol. 6, no. 2, pp. 9997–10004.
- [18] B. Haryanto et al., “Facebook Analysis of Community Sentiment on 2019 Indonesian Facebook Analysis of Community Sentiment on 2019 Indonesian Presidential Candidates from Facebook Opinion Data Presidential Candidates from Facebook Opinion Data,” in *Procedia Computer Science*, 2019, vol. 161, pp. 715– 722.
- [19] G. A. Buntoro, “ANALISIS SENTIMEN CALON GUBERNUR JAWA TIMUR 2018 DENGAN METODE NAIVE BAYES CLASSIFIER,” *J. Informatics Peita Nusant.*, vol. 4, no. 1, 2019.
- [20] F. V. Sari and A. Wibowo, “ANALISIS SENTIMEN PELANGGAN TOKO ONLINE JD.ID MENGGUNAKAN METODE NAÏVE BAYES CLASSIFIER BERBASIS KONVERSI IKON EMOSI,” *J. SIMETRIS*, vol. 10, no. 2, pp. 681–686, 2019.