

# Analisis Sentimen Pindah Ibu Kota Berbasis *Naive Bayes Classifier*

P. Arsi<sup>1</sup>, B.A. Kusuma<sup>2</sup>, A. Nurhakim<sup>3</sup>

<sup>1,2,3</sup>*Prodi Informatika, Universitas Amikom Purwokerto*

*Jl. Let. Jend. Pol. Soemarto, Watumas, Purwanegara, Purwokerto*

E-mail : ukhti.prima@amikompurwokerto.ac.id<sup>1</sup>, bagus@amikompurwokerto.ac.id<sup>2</sup>,  
azizan.nurhakim@gmail.com<sup>3</sup>

**Abstract**—The development of social media makes it easier for users to accelerate access to information on the internet. Access to information that was previously difficult to obtain was so easy nowadays. Social media allows its users not only to consume but also to participate in, create, comment on and distribute a variety of content in various formats. There are many social media that have developed on the internet, one of the most popular is Twitter. Twitter is a social media that allows its users to interact personally or openly. Through the hashtag feature, Twitter users can find out what topics are being discussed in real-time. In addition, keywords on Twitter can also be a source of conversation by users. One of the hot topics discussed on Twitter is the issue of relocating Indonesia's capital city. However, behind this there is controversy from those who feel the pros and cons, each having their own point of view. This has led to the emergence of a debate phenomenon, especially on Twitter, which actually shows collective concern about public discourse. The tendency of Twitter users to post content can be determined by means of sentiment analysis. The results of this study indicate that the accuracy value obtained is 94,33%. In this study, the Naive Bayes Classifier (NBC) method is proposed to analyze the sentiment towards government discourse in the online mass media Twitter on the topic of moving the Indonesian capital by classifying it into positive and negative. By conducting this sentiment analysis, it is hoped that the problems contained in the controversy over the topic of relocating the capital can be identified, so that it can be used as evaluation material for further purposes.

**Keywords**—Twitter, naive bayes, sentiment analysis, polling

**Abstrak**—Perkembangan media sosial memudahkan pengguna dalam percepatan akses informasi di internet. Akses informasi yang awalnya sulit diperoleh begitu mudah sekarang ini. Media sosial memungkinkan penggunaannya tidak hanya mengonsumsi tapi juga berpartisipasi, membuat, mengomentari dan menyebarkan beragam konten dalam berbagai format. Banyak media sosial yang berkembang di internet, salah satu yang banyak digemari adalah Twitter. Twitter merupakan media sosial yang memungkinkan para pengguna untuk berinteraksi secara personal ataupun terbuka. Melalui fitur *hashtag* para pengguna Twitter dapat mengetahui topik yang sedang dibahas secara *real-time*. Selain itu kata kunci pada Twitter dapat pula menjadi sumber perbincangan oleh pengguna. Salah satu topik yang ramai diperbincangkan di Twitter adalah terkait *issue* pemindahan ibu kota Indonesia. Namun dibalik hal tersebut terdapat kontroversi dari pihak yang merasa pro dan kontra, masing-masing memiliki sudut pandang sendiri. Hal ini menyebabkan munculnya fenomena perdebatan khususnya di Twitter yang sebenarnya menunjukkan perhatian kolektif mengenai wacana publik. Kecenderungan pengguna Twitter dalam memposting konten dapat diketahui dengan cara analisa sentiment. Pada penelitian ini diusulkan metode Naive Bayes Classifier (NBC) untuk menganalisa sentimen terhadap wacana pemerintah di media massa online Twitter pada topik pemindahan ibukota Indonesia dengan cara mengklasifikasikan menjadi positif, dan negatif. Hasil penelitian ini menunjukkan bahwa nilai akurasi yang diperoleh sebesar 94,33%. Dengan dilakukannya analisa sentimen ini diharapkan dapat diketahui permasalahan yang terdapat pada kontroversi topik pemindahan ibukota, sehingga dapat dijadikan sebagai bahan evaluasi untuk kepentingan lebih lanjut.

**Kata Kunci**—Twitter, naive bayes, analisis sentimen, polling

## I. PENDAHULUAN

Dewasa ini, media sosial telah menjadi akses dari segala bentuk komunikasi [1]. Media sosial ini membuat penggunaannya dapat melakukan banyak hal. Tidak hanya sekedar mendapatkan informasi di dalamnya, tetapi juga menjadi sarana untuk berpartisipasi, yaitu membuat konten, memberikan komentar dan menyebarkan berbagai informasi dalam berbagai format, seperti teks, suara dan gambar [2]. Tak hanya itu, media sosial juga dapat digunakan sebagai tempat untuk menjalin komunikasi, membentuk komunitas, mengelola usaha dan juga organisasi masyarakat, serta memudahkan lembaga pemerintah untuk bisa terhubung dengan public [3]. Salah satu media sosial yang memberikan kebebasan bagi penggunaannya untuk bisa berinteraksi secara terbuka adalah twitter [4][5].

Banyak hal yang dibahas oleh para pengguna media

sosial twitter, di antaranya ekonomi, politik, sosial, budaya, dan hukum [6][7][8]. Twitter sendiri memiliki sebuah menu yang disebut *hashtag*. Dengan menggunakan fitur *hashtag* tersebut, pengguna bisa lebih mudah untuk mengetahui topik apa yang saat ini banyak diperbincangkan di twitter. Bahkan seringkali suatu topik sudah menjadi tren, atau yang kerap disebut *trending topic* di Twitter sebelum berita tersebut muncul sebagai berita utama di media masa. Tak jarang pula jika pembahasan suatu topik di Twitter justru menjadi sumber munculnya sebuah berita [9]. Melihat hal tersebut, tak heran jika di Indonesia, media sosial Twitter dikatakan bisa memberikan dampak besar dalam topik tertentu [10].

Belakangan ini, salah satu topik yang menjadi sorotan dan kerap menjadi *trending topic* di Twitter adalah mengenai pemindahan ibu kota negara Indonesia. Seperti yang telah

kita ketahui, di setiap era presiden memang selalu muncul pembahasan mengenai hal tersebut. Namun di Era Presiden Joko Widodo saat ini, wacana pemindahan ibu kota kembali santer setelah presiden mengemukakan kembali hal tersebut [11]. Masalah pemindahan ibu kota ini tentunya bukan hal yang mudah dan merupakan kepentingan yang besar bagi negara kita terutama dalam upaya menyongsong kompetisi global [12]. Meskipun demikian, dalam sebuah topik pasti menimbulkan pro kontra, tak terkecuali soal wacana pemindahan ibu kota. Topik ini pun tak lepas dari perhatian pengguna Twitter dan menjadi sumber perdebatan [13]. Kecenderungan dari pengguna Twitter dalam mengemukakan pendapatnya dapat diketahui melalui analisis sentiment [14]. Hal itu dilakukan dengan cara memisahkan opini yang disampaikan oleh pengguna Twitter berdasarkan kelasnya yaitu positif dan negatif. Setelah mengklasifikasikan opini, dapat diambil pula kesimpulan mengenai faktor apa yang paling sering dibahas di dalam opini-opini tersebut [15].

Analisis sentimen memiliki kemampuan yang sangat baik untuk mengekstraksi pengetahuan dari pengguna sosial media ke beberapa kelas target. Analisis sentimen juga bisa digunakan untuk menganalisis perasaan dan suasana hati seseorang terhadap suatu hal. Analisis Sentiment bisa menghasilkan apa yang ingin kita ketahui, misalnya tentang ulasan dari sebuah kalimat yang diutarakan [16].

Banyak teknik Analisis yang digunakan dalam analisis sentimen karena kebanyakan permasalahannya adalah kalimat yang tidak baku dari pengguna sosial media. Akurasi akan diklasifikasikan dalam bentuk evaluasi positif, negatif atau netral. Saat ini, analisis sentimen biasanya menggunakan algoritma seperti *Naïve Bayes*, *Maximum Entropy*, dan *Support vector machine* [17]–[19].

Kebudayaan Twitter dan penggunaannya membuka peluang bagi peneliti untuk bisa mempelajari lebih lanjut mengenai sebuah opini melalui Analisa sentiment [20]. Beberapa metode pembelajaran mesin pada Analisa Sentimen yang sering digunakan antara lain adalah *K-Nearest Neighbor* (K-NN), *Naive Bayes Clasifier* (NBC), *Random Forest*, dan *Decision Tree* [21]–[24]. Metode NBC adalah salah satu metode klasifikasi dalam Teknik data mining yang menerapkan Teori Bayes [25]. NBC ini sendiri merupakan metode klasifikasi yang bekerja sangat baik jika dibandingkan dengan metode yang lainnya [26]. Hasil perbandingan performa antara NBC, *Decision Tree*, *Random Forest* dan KNN menunjukan bahwa NBC unggul dalam hal akurasi, presisi, recall, dan f-measure [27][28].

Pada penelitian sebelumnya mengenai Analisa Sentimen yang dilakukan oleh Mehdi (2019), algoritma *Naive Bayes Clasifier* diterapkan 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-Ma'ruf dan 64% untuk Prabowo-Sandi [29]. Budi Haryanto dkk (2019) juga melakukan penelitian terkait topik yang sama yaitu Analisa Sentimen pada topik pemilihan presiden di Indonesia di tahun yang sama.

Penelitian ini mengklasifikasi sentiment di beberapa komunitas pada media sosial Facebook. Komunitas tersebut adalah @detikcom, @tribunews dan @liputan6online. Dalam penelitian tersebut juga digunakan algoritma *Naïve Bayes Clasifier* untuk proses Analisa Sentimen. Hasil analisa menunjukkan 56,76% sentimen positif dan 43,24% sentimen negatif untuk Jokowi-Ma'ruf. Sementara itu untuk pasangan Prabowo-Sandi menunjukkan 24,21% sentimen positif dan 75,79% sentiment negative [30]. Di tahun 2019, Ghulam juga melakukan penelitian analisa sentimen di bidang politik yaitu mengenai calon gubernur Jawa Timur. Dalam penelitian ini dilakukan klasifikasi opini di Twitter. Hasil analisa menunjukkan 77% untuk Khofifah dan 76% untuk Gus Pul [31].

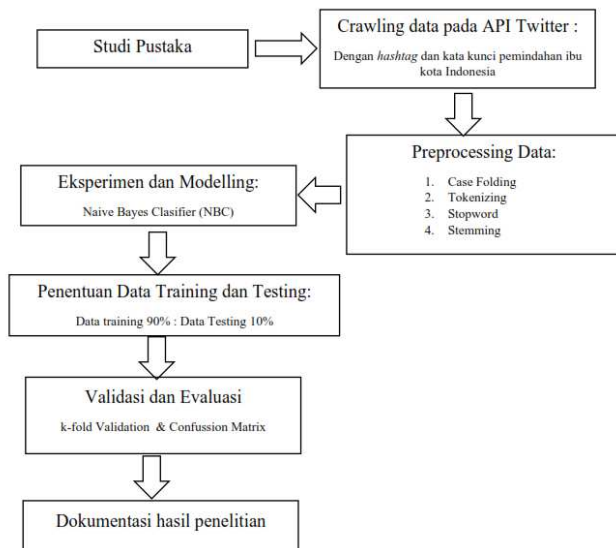
Tak hanya di bidang politik, penelitian mengenai Analisa Sentimen juga pernah dilakukan di bidang ekonomi. Fransiska (2019) menganalisa sentiment konsumen toko online JD.ID. Opini-opini dari konsumen tersebut diklasifikasikan menggunakan algoritma *Naïve Bayes Clasifier* ke dalam sentiment positif, negatif dan netral. Hasil analisa menunjukkan akurasi 96,44% [32]. Manuel J. Sanchez-Franco dkk, 2018 juga melakukan analisa sentiment di bidang ekonomi, yaitu mengidentifikasi layanan di bidang perhotelan. Dalam penelitian ini digunakan data berupa ulasan pelanggan berjumlah 47.172 ulasan yang berasal dari 33 hotel di Las Vegas. Data tersebut didapat dari media sosial Yelp. Klasifikasi dilakukan dengan menggunakan algoritma *Naïve Bayes Clasifier*. Dari penelitian tersebut didapat kesimpulan bahwa algoritma yang digunakan tersebut cepat, terukur dan akurat [33].

Pada penelitian ini diusulkan metode *Naive Bayes Clasifier* (NBC) untuk menganalisa sentimen terhadap wacana pemerintah di media masa *online* Twitter pada topik pemindahan ibukota Indonesia dengan cara mengklasifikasikan menjadi positif dan negatif. Dengan dilakukannya analisa sentimen ini diharapkan dapat diketahui permasalahan yang terdapat pada kontroversi topik pemindahan ibukota Indonesia, sehingga dapat dijadikan sebagai bahan evaluasi untuk kepentingan lebih lanjut.

## II. METODE PENELITIAN

### A. Tahapan Penelitian

Penelitian ini berjenis eksperimen, adapun tahapan yang dilakukan seperti tampak pada Gambar 1. Pada Gambar merupakan tahapan penelitian yang diawali dengan melakukan tahapan studi pustaka. Pada tahapan ini dikumpulkan beberapa referensi yang berkaitan dengan topik penelitian serta melakukan *filter* terhadap pustaka yang relevan dengan pemindahan ibu kota.



Gambar 1. Tahapan penelitian

Tabel 1. Dataset penelitian (*crawling data*)

	<i>Tweet</i>
1	Pede banget udah mau main pindahkan ibukota sj. Padahal mayoritas rakyat gak mau pindah. Pak @aniesbaswedan mohon jangan mendukung proyek pencitraan ini sebelum ada hasil survey resmi dari lembaga asing yg independen apakah rakyat mmg menginginkan pemindahan ibukota. cc @ILCtv1 <a href="https://t.co/kJhaFeiFmw">https://t.co/kJhaFeiFmw</a>
2	"RT @KangFathanNow: Anggaran Pemindahan Ibukota Negara harus terukur.. #SuaraParlemen <a href="https://t.co/xAg6sVNC36">https://t.co/xAg6sVNC36</a> "
3	Presiden Jokowi Gelar Ratas Lanjutan Rencana pemindahan Ibukota <a href="https://t.co/Jb0GSYXV0n">https://t.co/Jb0GSYXV0n</a>
4	@asumsico BPIP tu kerjanya ngapain sih.. mending bubarin aja, terus anggarannya dialihkan aja ke pemindahan ibukota.
5	Konsep pemindahan ibukota RI sdh Final, seandainya Presiden @jokowi lengser sekalipun pengantinya harus meneruskan , karena kita bukan ingin membangun ibukota parpol tapi sebuah ibukota Indonesia
.....	.....
1.420	@erwannusantara Kalah pamor dengan Preaiden @aniesbaswedan dalam mengatasi permasalahan Jakarta kok terus pindah Ibukota, oalah jook...jook...!

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 pemindahan ibu kota dari 10 Januari 2020 sampai dengan 17 Maret 2020, seperti tampak pada Tabel 1. Pada tahapan ini cara untuk *filter* topik yakni dengan menginputkan *hashtag* atau kata kunci “pemindahan ibu kota Indonesia”. Jumlah data yang berhasil di *Crawling* menggunakan *library tweepy Python* adalah sebanyak 1.420 *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*, dan *stemming*.

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 keluaran. 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.

### B. 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 1.420 data diseleksi kembali menjadi 1.007 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 *stopword* yaitu merupakan *filter* terhadap kata-kata yang tidak diperlukan dalam pemrosesan. Terakhir adalah *stemming*, adalah suatu proses untuk mengubah suatu kata berimbuhan menjadi bentuk dasarnya. Setelah seluruh preprocessing dilakukan, maka dataset siap untuk diklasifikasi. Pada tahap ini juga dilakukan *splitting* data *training* dan *testing*. Adapun pembagiannya adalah sebanyak 917 data *training* dan data *testing* sebanyak 90 data, total keseluruhan adalah 1.007 data.

### C. Metode Klasifikasi Naive Bayes

Analisis sentimen adalah proses penentuan polaritas, klasifikasi sentiment dan pengelompokan dalam suatu jumlah yang ditargetkan. Salah satu metode yang dapat digunakan adalah Naive Bayes Classifier (NBC). NBC merupakan sebuah metode klasifikasi yang berlandaskan pada teorema Bayes. Berikut merupakan algoritme Naive Bayes Classifier.

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 katakanlah  $n_i$ .
    - $n_a$  = 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 seragam 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  $P(A / W) = P(A) * P(\text{kata1} / \text{kelas A}) * P(\text{kata2} / \text{kelas A}) * \dots * P(\text{kataN} / \text{kelas A})$
    - Mencari  $P(B / W) = P(B) * P(\text{kata1} / \text{kelas B}) * P(\text{kata2} / \text{kelas B}) * \dots * P(\text{kataN} / \text{kelas B})$
  7. Setelah menghitung probabilitas untuk kedua kelas A dan B, kelas dengan probabilitas tertinggi adalah kelas baru untuk dokumen M.

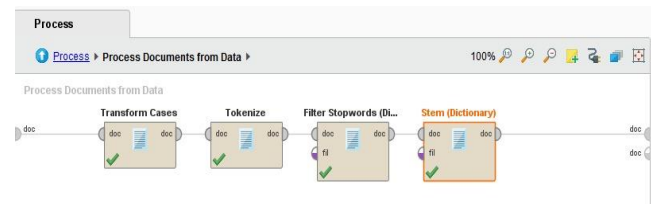
#### D. Validasi dan Evaluasi

Validasi dan evaluasi diperlukan dalam melihat sejauh mana performansi dari sistem yang telah dibangun. Pada penelitian ini digunakan metode *k-fold validation* dengan parameter  $k=5$ . Teknik ini melakukan iterasi dari proses *splitting* data *training* dan *testing*. Dalam setiap iterasi data dibagi ke dalam lima bagian yang sama besar Sedangkan untuk evaluasi digunakan metode *confusion matrix*.

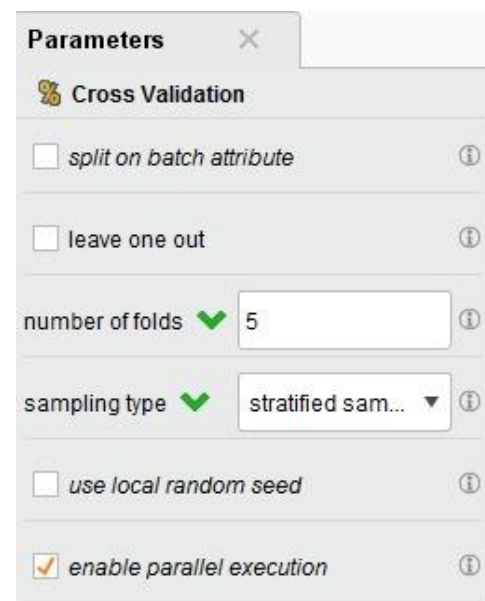
### III. HASIL DAN PEMBAHASAN

Dalam pemrosesan analisis sentimen akan ada tiga cakupan kategori yaitu level dokumen, level kalimat dan level aspek. Pada level dokumen dalam analisis sentimen akan diekspresikan dalam bentuk positif, negataif atau sebuah kelas pendapat. Pada level kalimat dalam analisis sentimen akan diekspresikan dalam setiap kalimat, di mulai dengan mengidentifikasi subjek atau objek dalam kalimat lalu mengidentifikasikannya menjadi positif atau negatif. Level aspek dan level kalimat bisa sedikit diabaikan karena kalimat merupakan bentuk dokumen pendek. Meskipun demikian, level dokumen juga bisa terdapat sedikit celah yang membutuhkan lebih banyak opini. Sedangkan level aspek dapat berarti memiliki entitas di atas level kalimat, misalkan dalam bentuk suatu paragraf. Adapun dalam penelitian ini dilakukan pada level dokumen dengan tidak melihat subjek atau objek dari suatu kalimat.

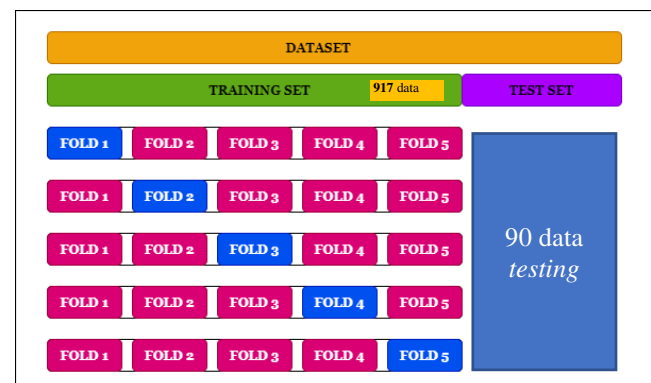
Dataset berjumlah 1.007 dilakukan *preprocessing* dengan urutan *case folding*, *tokenizing*, *stopword*, dan *stemming*, seperti pada Gambar 2. Setelah itu data dilakukan *splitting* dengan teknik *k-fold validation* dengan parameter  $k=5$ , seperti tampak pada Gambar 3. *K-fold validation* akan membagi dataset menjadi lima bagian yang sama besar, kemudian mengacak susunan data *trainingset*. Tujuannya adalah untuk mengatasi permasalahan kombinasi data yang tidak tersebar pada *dataset*. Metode ini akan melakukan *training* dan *testing* sebanyak lima kali pada posisi data yang berbeda. Hasil akurasi dari kelima luaran akan dirata-ratakan, sehingga dapat memvalidasi seluruh bagian *dataset*, ilustrasi tampak seperti pada Gambar 4.



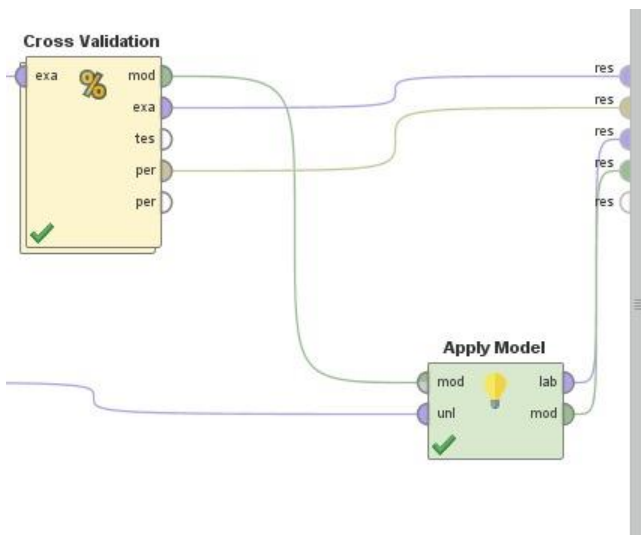
Gambar 2. Tahapan *tokenizing*, *stopword* dan *stemming*



Gambar 3. Setting parameter *k-fold*



Gambar 4. Ilustrasi *k-fold validation* dengan  $k=5$

Gambar 5. Diagram sistem menggunakan *k-fold validation*Gambar 6. Implementasi model *Naive Bayes*

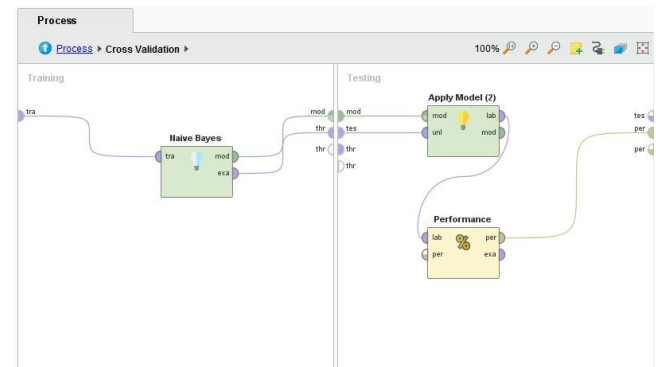
Pengujian dilakukan dengan menggunakan *software* Rapidminer versi 9.0. Pada proses dari pembacaan file dataset hingga proses *k-fold validation* dapat dilihat melalui Gambar 5. Pada Gambar 5 terlihat bahwa terdapat proses pembacaan dari file *dataset* sebanyak dua buah file masukan (*training* dan *testing*) 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 menggunakan metode *k-fold validation*.

Proses inti yaitu klasifikasi menggunakan metode *Naive Bayes* ditunjukkan seperti pada Gambar 6. Pada tahapan ini merupakan implementasi pemodelan *Naive Bayes*, dimana pada langkah ini dilakukan proses *training* dan *testing dataset*.

Setelah mengaplikasikan model *Naive Bayes*, maka tahap selanjutnya adalah menguji performansi seperti terlihat pada Gambar 7. Hasil pengujian dari *dataset* yang telah dilatih menggunakan *k-fold* menunjukkan bahwa akurasi yang dicapai sebesar 94,33%  $\pm$  1,28%. Artinya

sistem sudah cukup baik dalam mendeteksi kelas sentimen Adapun uraian terperinci data *confusion matrix* terlihat pada Tabel 2. Nilai TN=550, TP=315, FP=49 dan FN=3. Sehingga dapat diketahui *dataset training* berlabel negatif sebanyak 599 dan data berlabel positif sebanyak 318. Untuk *class recall* untuk label aktual negatif adalah 91,82% dan *class recall* untuk label aktual positif yakni 99,06%. Sedangkan untuk nilai *class precision* pada prediksi negatif adalah 99,46% dan *class precision* untuk semua prediksi positif adalah 86,54%, seperti tampak pada Tabel 2.

$$F1\ Score = 2 * \frac{precision * recall}{precision + recall} \quad (1)$$



Gambar 7. Proses Evaluasi

Tabel 2. Tabel *confusion matrix* NBC

	Akurasi : 94,33% $\pm$ 1,28%		
	True negative	True positive	Class precision
Prediction Negative	550	3	99,46%
Prediction Positive	49	315	86,54%
Class Recall	91,82%	99,06%	

Dari hasil pengujian dapat terlihat bahwa terdapat 49 data yang seharusnya memiliki nilai aktual sentimen negatif dideteksi sebagai sentimen positif. Jika dihitung dalam persen maka nilai FP ini adalah sebesar 5,34% 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*=0,99, maka didapatkan nilai *F1 Score* sebesar 0,92, yang mengindikasikan bahwa model klasifikasi sudah memiliki *precision* dan *recall* yang baik.

#### IV. KESIMPULAN

Penelitian ini bertujuan untuk mengklasifikasikan sentimen yang berasal dari media sosial twitter terkait pemindahan ibu kota dengan menggunakan *Naive Bayes Classifier*. Hasil yang didapat dari dataset sejumlah 1007



data adalah nilai akurasi sebesar 94,33%, *precision* sebesar 0,87 dan *recall* 0,99. Nilai harmonik *recall* dan *precision* atau yang disebut juga sebagai *F1-Score* adalah 0,92 yang berarti sistem ini sudah baik dalam mendeteksi sentimen. Nilai *recall* yang tinggi juga sangat baik diterapkan pada sistem ini, dikarenakan politik merupakan hal yang sensitif. Penelitian ke depan dapat direkomendasikan untuk menambahkan dataset dan menambahkan perbandingan model classifier yang lainnya.

### UCAPAN TERIMA KASIH

Ucapan terima kasih disampaikan kepada LPPM Universitas Amikom Purwokerto yang telah mendanai penelitian ini melalui program Penelitian Dosen Muda Amikom.

### DAFTAR PUSTAKA

- [1] M. Safiullah, P. Pathak, S. Singh, and A. Anshul, "Social media as an upcoming tool for political marketing effectiveness," *Asia Paci f i c Manag. Rev.*, vol. 22, pp. 10–15, 2017.
- [2] K. Sailunaz and R. Alhaji, "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. C. Pandey, D. S. Rajpoot, and M. Saraswat, "Twitter sentiment analysis using hybrid cuckoo search method," *Inf. Process. Manag.*, vol. 53, pp. 764–779, 2017.
- [11] D. Septiana, "Palangka Raya the Capital City of Indonesia : Critical Discourse Analysis on News about Moving the Capital City from Jakarta," *Adv. Soc. Sci. Educ. Humanit.*, vol. 280, no. Basa, pp. 190–202, 2018.
- [12] N. R. Toun, "Analisis Kesiapan Pemerintah Provinsi Kalimantan Tengah dalam Wacana Pemindahan Ibu Kota Negara Republik Indonesia ke Kota Palangkaraya," *J. Acad. Praja*, vol. 1, pp. 129–148, 2018.
- [13] F. Andriani, "FENOMENA SOCIAL CIMBER MELALUI TWITWAR," *J. Pustaka Komun.*, vol. 1, 2018.
- [14] S. F. Pratama, R. Andrean, and A. Nugroho, "Analisis Sentimen Twitter Debat Calon Presiden Indonesia Menggunakan Metode Fined-Grained Sentiment Analysis," *J. Inf. Technol. Comput. Sci.*, vol. 4, no. 2, 2019.
- [15] A. Novantirani, M. K. Sabariah, and V. Effendy, "Analisis Sentimen pada Twitter untuk Mengenai Penggunaan Transportasi Umum Darat Dalam Kota dengan Metode Support Vector Machine," in *Proceeding of Engineering*, 2015, vol. 2, no. 1, pp. 1177–1183.
- [16] I. F. Rozi, H. S. Pramono, and E. A. Dahlan, "Implementasi Opinion Mining ( Analisis Sentimen ) untuk Ekstraksi Data Opini Publik pada Perguruan Tinggi," *J. EECCIS*, vol. 6, no. 1, pp. 37–43, 2012.
- [17] D. T. Lukmana, S. Subanti, and Y. Susanti, "ANALISIS SENTIMEN TERHADAP CALON PRESIDEN 2019 DENGAN SUPPORT VECTOR MACHINE DI TWITTER," in *SEMINAR NASIONAL PENELITIAN PENDIDIKAN MATEMATIKA*, 2019, pp. 154–160.
- [18] F. Tang, L. Fu, B. Yao, and W. Xu, "Aspect based fine-grained sentiment analysis for online reviews," *Inf. Sci. (Ny)*, 2019.
- [19] F. Ali *et al.*, "Transportation Sentiment Analysis Using Word Embedding and Ontology-based Topic Modeling," *Knowledge-Based Syst.*, 2019.
- [20] J. Qiu, Z. Lin, and Q. Shuai, "Investigating the Opinions Distribution in the Controversy on Social Media," *Inf. Sci. (Ny)*, 2019.
- [21] 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.
- [22] 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.
- [23] 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.
- [24] 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.
- [26] 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.
- [27] 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.
- [28] 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.
- [29] 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.
- [30] 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.
- [31] G. A. Buntoro, "ANALISIS SENTIMEN CALON GUBERNUR JAWA TIMUR 2018 DENGAN METODE NAIVE BAYES CLASSIFIER," *J. Informatics Peita Nusantara*, vol. 4, no. 1, 2019.
- [32] 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.
- [33] M. J. Sánchez-franco, A. Navarro-garcía, and F. J. Rondán-cataluña, "A naive Bayes strategy for classifying customer satisfaction : A study based on online reviews of hospitality services," *J. Bus. Res.*, no. June, pp. 0–1, 2018.