

Klasifikasi Interaksi Kampanye di Media Sosial Menggunakan *Naïve Bayes Kernel Estimator*

Aryo Nugroho^{1,5}, Rumaisah Hidayatillah², Surya Sumpeno^{3,4}, Mauridhi Hery Purnomo^{3,4}

Abstract—The development of technology also influences changes in campaign patterns. Campaign activities are part of the process of Election of Regional Heads. The aim of the campaign is to mobilize public participation, which is carried out directly or through social media. Social media becomes a channel for interaction between candidates and their supporters. Interactions that occur during the campaign period can be one indicator of the success of the closeness between voters and candidates. This study aims to get the pattern of campaign interactions that occur on Twitter social media channels. This interaction pattern is classified as a model in measuring the success of campaigns on social media. The research begins with obtaining data through the data retrieval process using the API feature provided by Twitter. Furthermore, pre-processing is carried out before data can be processed in an algorithmic method. This stage is done to improve data quality so as to improve accuracy. Naive Bayes Classifier was chosen because of a simple procedure, then Kernel Estimator (KE) was used to improve performance. The use of naive Bayes Kernel Estimator can improve model performance from 76.74% to 80.14%. Testing models with split percentage methods on several combinations get satisfactory results.

Intisari— Kegiatan kampanye adalah bagian dari proses Pemilihan Kepala Daerah. Penggunaan teknologi memengaruhi perubahan pola kampanye di masa kini. Kampanye bertujuan untuk menggalang partisipasi publik dan dapat dilakukan secara langsung maupun melalui media sosial. Media sosial menjadi sebuah saluran untuk berinteraksi antara calon dengan para pendukungnya. Interaksi yang terjadi dalam masa kampanye menjadi salah satu indikator keberhasilan sosialisasi program

para calon. Makalah ini bertujuan untuk mendapatkan pola interaksi kampanye yang terjadi pada saluran media sosial Twitter. Selanjutnya model dari penelitian dalam makalah ini dapat digunakan dalam mengukur keberhasilan kampanye di media sosial. Penelitian diawali dengan mendapatkan data melalui proses *data crawling* menggunakan API yang ditentukan oleh media sosial. Selanjutnya dilakukan praproses sebelum data dapat diolah dalam sebuah metode algoritme. Perbaikan kualitas data saat praproses dilakukan untuk meningkatkan akurasi. *Naive Bayes Classifier* dipilih karena prosedurnya yang sederhana, kemudian ditingkatkan unjuk kerjanya dengan menggunakan *Kernel Estimator* (KE). Matriks unjuk kerja (*confusion matrix*) yang didapatkan menunjukkan, penggunaan KE pada *naive Bayes Classifier* dapat meningkatkan unjuk kerja model dari 76,74% menjadi 80,14%. Sedangkan pengujian model dengan metode *percentage split* pada beberapa kombinasi memberikan hasil yang memuaskan.

Kata Kunci— Pola Interaksi, Klasifikasi, *Naive Bayes*, *Kernel Estimator*.

I. PENDAHULUAN

Pemilihan Kepala Daerah (pilkada) merupakan sebuah kegiatan rutin yang dilakukan di berbagai daerah di Indonesia. Di tahun 2018 ada 269 pilkada serentak di 224 kabupaten, 36 kota, dan 9 provinsi. Selanjutnya pada tahun 2018 ada 171 pilkada dengan rincian 115 kabupaten, 39 kota, dan 17 propinsi [1]. Kegiatan ini tentu memerlukan sosialisasi hingga dapat menarik simpati calon pemilih untuk memilih pemimpin di wilayahnya. Kegiatan yang dilakukan dapat berupa kampanye terbuka hingga kampanye tertutup serta memanfaatkan berbagai alat peraga kampanye yang dipasang di berbagai sudut kota. Namun, pola ini bertambah dengan adanya media sosial yang banyak digunakan penduduk Indonesia serta kemajuan teknologi internet. Peranan media sosial untuk menggalang partisipasi masyarakat menjadi semakin besar dalam kegiatan kampanye melalui penggunaan berbagai aplikasi.

Analisis interaksi di media sosial dalam kegiatan kampanye dapat membuka ruang penelitian berdasarkan materi (*content*) yang ditulis (*post*) dan/atau dikomentari, maupun interaksi jejaring sosial antar pengguna di dalamnya [2]. Makalah ini bertujuan mendapatkan pola interaksi jejaring sosial dalam kegiatan kampanye suatu pasangan calon. Pengolahan data untuk mencari pola dalam media sosial sudah sering dilakukan. Tujuan dari pencarian pola ini beragam, ada yang dipakai dalam mendeteksi stres, penyebaran rumor, dan lainnya [3], [4].

Untuk mencari pola interaksi dalam media sosial dibutuhkan sebuah metode, baik itu pengenalan jejaring sosial (*social network*), analisis sentimen (*sentiment analysis*), maupun interaksi dari respons para pengguna terhadap suatu pernyataan yang dipublikasikan (*posting*). Hasil dari pengenalan pola

¹ Mahasiswa, Departemen Teknik Elektro, Fakultas Teknologi Elektro, Institut Teknologi Sepuluh Nopember, Jln. Teknik Mesin, Gedung B, C, dan AJ, Kampus ITS, Sukolilo, Surabaya, Jawa Timur, 60111 INDONESIA (telp: 031-5994251; fax: 031-5931237; e-mail: aryo.nugroho14@mhs.ee.its.ac.id)

² Alumnus, Program Studi Informatika, Fakultas Ilmu Komputer, Universitas Narotama, Jln Arief Rahman Hakim 51, Surabaya, Jawa Timur, INDONESIA, (telp: 031-5946404; fax: 031-5931213; email: rumaisahhidayatillah.14@fasilkom.narotama.ac.id)

³ Dosen, Departemen Teknik Elektro, Fakultas Teknologi Elektro, Institut Teknologi Sepuluh Nopember, Gedung B & C, Jln. Teknik Mesin, Kampus ITS, Sukolilo, Surabaya, Jawa Timur, 60111 INDONESIA (telp: 031-5922936; e-mail: surya@ee.its.ac.id, hery@ee.its.ac.id)

⁴ Dosen, Departemen Teknik Komputer, Fakultas Teknologi Elektro, Institut Teknologi Sepuluh Nopember, Gedung B & C, Jln. Teknik Mesin, Kampus ITS, Sukolilo, Surabaya, Jawa Timur, 60111 INDONESIA (telp: 031-5922936; e-mail: surya@ee.its.ac.id, hery@ee.its.ac.id)

⁵ Dosen, Fakultas Ilmu Komputer, Universitas Narotama, Jln Arief Rahman Hakim 51, Surabaya, Jawa Timur, INDONESIA (telp: 031-5946404; fax: 031-5931213; e-mail: aryo.nugroho@narotama.ac.id)

interaksi dapat disajikan dengan metode prediksi seperti *naive Bayes*, *C4.5*, *fuzzy*, dan jaringan syaraf tiruan (*artificial neural network*).

Penelitian ini bertujuan untuk menganalisis pola dan mengelompokkan atribut berdasarkan kesamaan tertentu. Pada penelitian ini digunakan dua varian dari metode *naive Bayes*, yaitu *naive Bayes* dan *naive Bayes Kernel Estimator* (KE) dengan berbagai kombinasi uji *percentage split*.

Konten interaksi dalam media sosial memiliki tiga tipe, yaitu tanggapan positif, tanggapan negatif, dan kontroversi. Tanggapan positif mewakili sekelompok individu yang menyatakan persetujuan, mendukung, percaya dan sepakat. Tanggapan negatif menunjukkan ketidaksetujuan, ketidakpercayaan atau tidak sepakat atas sebuah tulisan dalam media sosial [5]. Sedangkan kontroversi terjadi jika sebuah tulisan memicu perdebatan antar kelompok yang pro dan kontra dalam sebuah tulisan. Ketiga tipe ini dapat memengaruhi interaksi dalam media sosial. Dalam penelitian ini algoritme *naive Bayes* akan dipakai untuk menemukan model interaksi kampanye yang terbaik dalam membangun kedekatan di media sosial.

Penelitian ini menggunakan studi kasus partisipasi politik publik terhadap pemilihan Kepala Daerah di Provinsi Jawa Tengah. Data diambil dari media sosial Twitter dengan memasukkan kata kunci berupa nama calon ke dalam program penggalian data.

II. INTERAKSI KAMPANYE DI MEDIA SOSIAL

Saat ini manusia hampir tidak pernah lepas dari telepon pintar (*smartphone*) di genggaman tangannya [6]. Di saat melakukan aktivasi sebuah telepon pintar, secara otomatis akan dimiliki sebuah akun khususnya Google Account bagi pengguna Android [7]. Akun ini otomatis mengaktifkan aplikasi *Online Social Network* (OSN) yaitu Google+ dan memungkinkan pengguna menggunakan akun ini untuk mendaftarkan diri pada beragam aplikasi yang lain, sehingga banyak orang menggunakan beberapa OSN secara bersamaan untuk beragam kegiatan [8]. Manusia berinteraksi menggunakan Facebook dengan temannya. Di saat yang sama juga mengaktifkan aplikasi Twitter untuk mendapatkan berita terkini. Kemudian, aplikasi Instagram digunakan untuk melihat dan mengunggah foto kegiatannya. Secara otomatis dapat pula dilakukan pengaturan agar tulisan atau unggahan pada suatu aplikasi dapat muncul pula pada aplikasi yang lain. Integrasi beberapa aplikasi dapat membuat unggahan pada Instagram otomatis muncul pula di Facebook dari akun pengguna. Motivasi mengaktifkan pengaturan secara simultan ini selain untuk alasan kepraktisan juga menambah kedekatan interaksi dengan relasi dan teman.

Fenomena kedekatan manusia dengan telepon pintarnya ini telah menjadi gaya hidup. Berbagai kegiatan dibagikan melalui OSN dan dapat dibaca melalui layar telepon pintar [9]. Beragam merk dan jenis ditawarkan untuk memudahkan penggunaannya, sehingga berbagai kalangan pun memanfaatkan fenomena ini dalam mempromosikan kegiatannya seperti kalangan bisnis hingga meraih partisipasi publik pada kampanye pemilihan umum [10].

Sebuah riset telah dilakukan di India untuk melakukan optimasi penyebaran informasi dalam sebuah OSN. Penyebaran informasi dilakukan dengan menurunkan proses epidemik, yaitu mengatur kontrol dengan menyesuaikan anggaran yang ada. Beberapa parameter diujicobakan untuk mendapatkan strategi terbaik. Hasil yang didapatkan bervariasi dari waktu ke waktu, sesuai dengan tingkat ketertarikan komunitas dan materi yang disampaikan. Penelitian lain dilakukan pada 53 partisipasi publik dalam IdeaScale. Hasil penelitian ini merekomendasikan beberapa teknik yang dapat dilakukan untuk mempromosikan ide kampanye dalam OSN melalui penggunaan fitur pada Twitter [11]. Penelitian ini menjadi dasar pengambilan data dilakukan pada aplikasi Twitter. Namun, penelitian ini hanya menunjukkan penggunaan Twitter tanpa mengeksplorasi pola kejadian dalam kampanye hariannya.

Pada tahun 2013 di Turki terjadi sebuah perlawanan kepada pemerintah yang diinisiasi oleh sebuah gerakan kecil di Gezi Park, salah satu distrik di Istanbul. Dalam beberapa hari saja gerakan ini menyebar lebih luas melalui peran OSN seperti Facebook, Twitter dan YouTube. Kejadian ini menunjukkan bahwa tanpa diduga penyebaran informasi melalui internet dapat menggalang mobilisasi partisipasi politik, khususnya dalam melakukan protes pada kepemimpinan Perdana Menteri Recep Tayyip Erdogan [12]. Penelitian terkait membangun partisipasi politik pun telah menunjukkan adanya perubahan dari masa ke masa. Pada umumnya, konsep yang dipakai berupa "*local involvement*" dan "*empowering people*", yang telah berubah dengan adanya perkembangan teknologi informasi melalui penggunaan OSN. Secara tradisional, semula hanya ditawarkan sebuah program kerja dan penokohan seseorang. Di masa kini, membangun partisipasi politik dilakukan dengan membangun relasi dan keterlibatan, mulai dari mendesain program bersama publik, menguji desain program, hingga mengusung tema secara bersama. Melalui pola ini, maka partisipasi politik adalah gerakan bekerja dengan imajinasi bersama dalam meraih tujuan [13].

Sebuah penelitian yang dilakukan pada tahun 2015 juga menunjukkan adanya sinergi antara politik dan media sosial. Sinergi ini menunjukkan adanya kecenderungan pemanfaatan media sosial di berbagai negara dalam kegiatan penggalangan partisipasi publik di kampanye politik. Tim kampanye menggunakan media sosial untuk mengumumkan pendaftaran calon kandidat, membangun *personal image* yang dekat dengan masyarakat, menggalang para relawan dan pendukung, penggalangan dana, menyampaikan visi dan misi serta program kandidat, hingga mobilisasi peserta kampanye. Beberapa kegiatan politik yang diteliti berlangsung di Moldova, Iran, Hongkong, dan revolusi yang terjadi di Mesir serta Tunisia yang dikenal dengan *social media movement*. Sedangkan peran media sosial sebagai katalis terjadi dalam fenomena Arab Spring, yaitu sebuah gerakan perubahan yang terjadi di beberapa negara benua Arab [14].

Perpaduan antara kampanye di televisi dan pemanfaatan OSN juga telah dilakukan dengan mengamati sebuah acara debat kandidat di televisi. Acara ini juga dipadukan dengan interaksi para penonton melalui OSN selama debat kandidat

berlangsung. Studi dilakukan pada the *Scottish parliamentary election and EU referendum* pada tahun 2016 untuk melihat adanya hubungan antara jumlah penonton televisi dalam acara debat publik dengan penggunaan perangkat telepon pintar. Hasil dari penelitian ini menunjukkan adanya korelasi positif dalam konteks acara televisi bertema politik dengan penggunaan media sosial [15]. Studi literatur terkait media sosial menunjukkan sebuah tulisan atau unggahan yang memiliki interaksi yang baik tidak hanya dihasilkan dari sebuah tulisan yang memicu komentar positif atau negatif saja, tetapi akan lebih banyak terjadi interaksi pada tulisan yang memicu kontroversi pada suatu waktu yang tepat. Interaksi yang tinggi dapat memicu kunjungan atau permintaan pertemanan yang berguna dalam membangun kedekatan antara pemilih dan calonnya [16], [17]. Beberapa penelitian tersebut menunjukkan adanya peran yang signifikan dari penggunaan OSN dalam kegiatan kampanye politik, tetapi belum menunjukkan pola interaksi yang dapat digunakan untuk mengukur keberhasilan sebuah kampanye.

Mengukur keberhasilan sebuah kampanye dapat dilakukan dengan melihat pola yang terjadi pada suatu waktu. Kemudian dilakukan klasifikasi untuk menggolongkan tingkat interaksi yang terjadi. Banyak penelitian telah dilakukan dengan menggunakan berbagai algoritme klasifikasi. Namun, pada penelitian ini digunakan *naive Bayes* dengan pertimbangan algoritmenya memiliki proses yang sederhana tetapi tetap diharapkan mendapatkan hasil terbaik. *Naive Bayes* dikembangkan dari teorema Bayes, yang selanjutnya terus disempurnakan dengan berbagai perbaikan untuk mendapatkan unjuk kerja yang lebih baik [5]. *Naive Bayes* KE diuji untuk mendapatkan unjuk kerja yang lebih baik. Makalah ini akan memberikan kontribusi pada pola interaksi yang dapat menunjukkan tingkat tinggi atau rendahnya interaksi dalam sebuah kampanye di media sosial. Hasilnya diharapkan akan melengkapi penelitian sebelumnya yang membahas tentang pengaruh tulisan atau unggahan positif dan negatif dalam media sosial [17].

III. NAÏVE BAYES KERNEL ESTIMATOR

Tahapan penelitian ini dilakukan dengan langkah-langkah sesuai Gbr. 1. Secara umum, penelitian dilakukan dengan tahap yang sama dengan penerapan algoritme *naive Bayes*. Namun, pada penelitian ini dilakukan pengembangan dengan menggunakan algoritme *naive Bayes* KE, sebagai kebaruan dari penelitian sejenis yang pernah dilakukan.

A. Pengumpulan Data

Pengumpulan data adalah proses awal dari *data mining*, yaitu kegiatan mengunduh data berskala besar dari *website* dengan melibatkan sebuah *agent* atau *crawling tool*. Tujuannya adalah mendapatkan data untuk penelitian maupun kebutuhan lainnya yang bersifat pribadi. Sumber data dapat berasal dari web ataupun media sosial, selain sumber yang lain. Pengambilan data ini memerlukan sumber daya dan waktu yang memadai [18]. Beberapa aplikasi media sosial memberikan persyaratan tertentu untuk melakukan *data crawling* menggunakan API, seperti pada Facebook dan Twitter [19], [20]. Penelitian ini diawali dari membangun sebuah aplikasi berbahasa Python yang terkoneksi dengan fitur *Search Engine*

dari Twitter. Data yang diambil berupa percakapan (*tweet*) dengan kata kunci (*keyword*) yang diperlukan, sehingga interaksi pengguna dan peristiwa yang terjadi dapat diperoleh [21].



Gbr. 1 Tahapan penelitian.

B. Pemrosesan Awal

Konsep utama dari data adalah sebagai bukti valid suatu peristiwa, sehingga integritas dan akurasi dalam pengelolaannya harus diutamakan. Ketika dua komponen tersebut sudah terpenuhi, maka data dapat digunakan untuk menentukan hasil. Data dapat digunakan sebagai sumber informasi untuk membantu pengambilan sebuah keputusan. Namun, sangat jarang data mentah langsung digunakan. Data harus diproses dan dipresentasikan sedemikian rupa sehingga menjadi data yang dapat digunakan secara tepat. Kualitas data sering kali dipengaruhi oleh jumlah data. Penelitian ini sendiri melibatkan data dalam jumlah yang cukup banyak. Terdapat ribuan baris data yang dapat diperoleh. Guna meningkatkan kualitas data yang diteliti, langkah awal yang harus dilakukan adalah melakukan pengecekan *error* pada data.

Kegiatan pada tahapan ini meliputi pengecekan data, pembersihan data (*data cleansing*), transformasi data (*data transformation*), hingga penentuan label dan pengurutan data (*sorting*). Pembersihan data yang bersifat heterogen bertujuan untuk membuang data yang tidak diperlukan [22]. Proses yang dilakukan adalah memfilter, memodifikasi, serta membuang data yang tidak relevan untuk memenuhi kriteria *dataset*. Parameter yang berupa angka (numerik) akan meningkatkan akurasi yang lebih baik dalam penggunaan *naive Bayes* [5].

Transformasi data [23] dilakukan untuk meningkatkan kemampuan interpretasi dan tampilan hasilnya. Proses yang dilakukan adalah mengubah bentuk data menjadi struktur atau format lain. Proses ini bertujuan untuk mengurangi rentang nilai antar data, karena selisih dari data yang terlalu jauh dapat mengakibatkan penurunan akurasi hasil pelatihan (*training*). Beberapa metode yang dapat digunakan antara lain transformasi akar (*square root transformation*), transformasi logaritma (*log*), transformasi *ArcSin*, transformasi kuadrat (*cubic*), dan transformasi *inverse* atau *reverse*. Penelitian ini menggunakan transformasi logaritma dengan rumus $\log(x+1)$, dengan x = nilai asli data, sedangkan penambahan angka 1 dimaksudkan untuk memastikan data asli memiliki nilai selain 0 [24].

Pengurutan data dilakukan setelah data ditransformasikan dan diberi label guna keperluan klasifikasi. Pelabelan dilakukan menggunakan sebuah aturan tertentu berdasarkan karakteristik data.

C. Naive Bayes Classifier

Klasifikasi data menggunakan algoritme *naive Bayes* sangat efektif dalam proses klasifikasi. Metode *naive Bayes* didasarkan pada probabilitas dan teorema Bayes yang diperkenalkan oleh penemunya, Thomas Bayes [25]. Pada metode ini, diasumsikan bahwa setiap variabel bersifat bebas (independen) dan sebuah fitur (variabel) diasumsikan tidak saling berkaitan. Perumusan umum teorema Bayes adalah seperti pada (1).

$$P(h|e) = \frac{P(e|h) P(h)}{P(e)} \tag{1}$$

dengan

$P(h|e)$: Probabilitas akhir dimana hipotesis h terjadi jika diberikan bukti e .

$P(e)$: Probabilitas awal e terjadi tanpa memandang bukti lain.

$P(e|h)$: Probabilitas bukti e terjadi akan mempengaruhi hipotesis h .

$P(h)$: Probabilitas awal h tanpa memandang bukti.

Prinsip dasar pada Bayes adalah bahwa hasil dari hipotesis (h) dapat diperkirakan dari beberapa bukti e yang diamati, sehingga pada klasifikasi ini perumusan akan menjadi (2).

$$P(\text{existing}|\text{data}) = \frac{P(\text{existing}|\text{class}) P(\text{class})}{P(\text{existing})} \tag{2}$$

dengan *class* adalah level interaksi yang terjadi sedangkan *existing* adalah banyaknya *reply*, *retweet*, dan *like*.

Perumusan *naive Bayes* untuk klasifikasi dapat dituliskan menjadi

$$P(Y|X) = \frac{P(Y) \prod_{i=1}^q P(X_i|Y)}{P(X)} \tag{3}$$

dengan

$P(Y|X)$: probabilitas data dengan vektor X pada kelas Y .

$P(Y)$: probabilitas awal kelas Y .

$\prod_{i=1}^q P(X_i|Y)$: probabilitas bebas kelas Y dari semua fitur dalam vektor X .

$P(X)$: probabilitas awal kelas X .

D. Naive Bayes Kernel Estimator (KE)

Penelitian ini mengusulkan sebuah metode perbaikan dengan menambahkan KE pada perhitungan klasifikasi *naive Bayes* [26]. Diharapkan metode ini dapat meningkatkan unjuk kinerja sistem. KE adalah perkiraan dari *probability density function*. Untuk setiap nilai x , maka KE ditentukan dengan perumusan sebagai berikut.

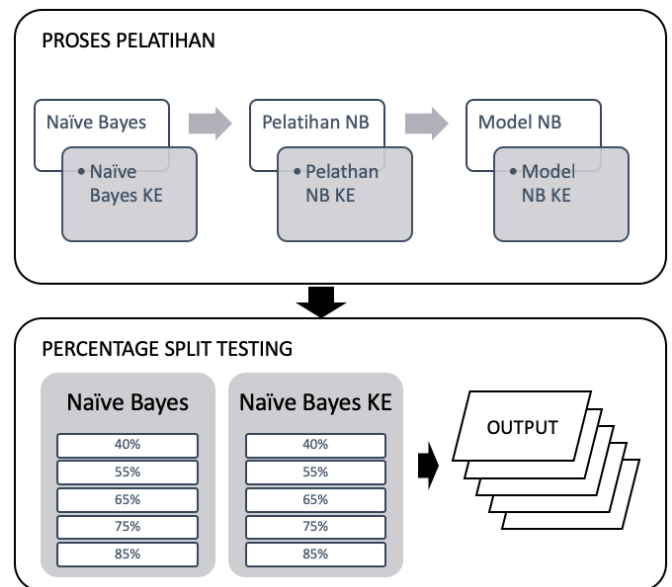
$$f_h(x) = \frac{1}{nh} \sum_{i=1}^n K \left(\frac{x - x_i}{h} \right) \tag{4}$$

dengan x_1, x_2, \dots, x_n adalah sampel acak dari distribusi, n adalah ukuran sampel, $K(\cdot)$ adalah *kernel smoothing function*, dan h adalah *bandwidth*. *Kernel smoothing function* mendefinisikan

bentuk kurva yang digunakan untuk menghasilkan *probability density function*. Mirip dengan histogram, distribusi *kernel* membangun fungsi untuk merepresentasikan distribusi probabilitas menggunakan data sampel. Perbedaannya adalah pada histogram nilai-nilai ditempatkan ke dalam kotak diskret. Sedangkan distribusi *kernel* menjumlahkan fungsi-fungsi penghalusan komponen untuk setiap nilai data untuk menghasilkan kurva probabilitas yang mulus dan berkesinambungan. Hasil yang diharapkan akan memberikan unjuk kerja yang semakin bagus dan menjadi kontribusi utama penelitian sejenis dalam penggunaan algoritme *naive Bayes*.

E. Pelatihan dan Pengujian Data

Naive Bayes dan *naive Bayes KE* digunakan dalam proses pelatihan dan pengujian untuk mendapatkan model klasifikasi yang memiliki unjuk kerja paling baik. Langkah awal adalah melakukan pelatihan pada *dataset*, dilanjutkan dengan pengujian menggunakan metode *percentage split* untuk beberapa kombinasi. Alur prosesnya ditunjukkan pada Gbr. 2.



Gbr. 2 Pelatihan dan pengujian data.

Pengujian dengan metode *percentage split* dilakukan dengan kombinasi 40% hingga 85%. Hasil yang didapat dibandingkan unjuk kerjanya guna mendapatkan model klasifikasi yang terbaik.

IV. HASIL EKSPERIMEN DAN PEMBAHASAN

A. Twitter Crawling

Berdasarkan kata kunci berupa "Ganjar Pranowo" dapat dikumpulkan sejumlah data dengan spesifikasi: *tweets ID*, tanggal (*date*), waktu (*timestamp*), zona waktu, *tweets*, *reply*, *retweet*, dan *like*, seperti ditunjukkan pada Tabel I. Proses ini berlangsung pada rentang waktu 29 Maret hingga 11 April 2018 sampai mendapatkan data yang cukup banyak untuk dapat dilakukan analisis.

Selama rentang waktu empat belas hari, didapatkan sebanyak 9.972 *tweets* data, 6.028 *reply*, 18.776 *retweet*, dan 25.646 *like* yang selanjutnya dipilah dalam tahapan berikutnya.

TABEL I
PENGUMPULAN DATA

Keywords	"Ganjar Pranowo"
Rentang waktu	29 Maret 2018 – 11 April 2018
Jumlah tweets	9.972
Jumlah reply	6.028
Jumlah retweet	18.776
Jumlah like	25.646

B. Pemrosesan Awal Data

Hasil pembersihan data mendapatkan beberapa parameter berupa *reply*, *retweet*, dan *like* yang akan digunakan sebagai *dataset*. Nilai parameter ini berupa angka sehingga akan memudahkan dalam proses selanjutnya. Data yang semua atributnya memiliki nilai 0 dibuang sehingga jumlah jumlah data turun dari 9.972 menjadi 4.507 data.

Transformasi data dilakukan dengan mengubah nilai semula menggunakan rumus sesuai yang telah dibahas dalam bagian sebelumnya. Penggunaan transformasi data telah menghasilkan data yang lebih baik, yaitu perbedaan selisih angka menjadi tidak terlalu jauh. Diharapkan tingkat akurasinya akan baik ketika dilakukan pelatihan data.

Pemberian label *dataset* adalah sebagai berikut. Jika hanya ada satu interaksi, maka label akan bernilai "low", jika ada dua interaksi maka label akan bernilai "medium", dan jika terdapat tiga interaksi (*reply*, *retweet*, dan *like*) maka label akan bernilai "high". Tabel II menunjukkan ilustrasi sederhananya.

TABEL II
CONTOH PENENTUAN LABEL

<i>Reply</i>	<i>Retweet</i>	<i>Like</i>	Reaksi
0	0	1	"Low"
0	1	1	"Medium"
1	1	1	"High"

Selanjutnya dilakukan proses pengurutan data berdasarkan label (reaksi) yang telah ditetapkan. Persentase reaksi pada *dataset* disajikan pada Tabel III.

TABEL III
PERSENTASE REAKSI PADA *TWEET*

Reaksi	Persentase (%)
"Low"	55
"Medium"	30
"High"	15

Dari Tabel III terlihat bahwa Calon Gubernur Ganjar Pranowo yang juga merupakan calon petahana telah sukses mendapatkan respons dari para pengikutnya. Hal ini dapat ditandai dari 55% reaksi yang tergolong kategori rendah (*low*) dan 30% reaksi yang tergolong sedang (*medium*). Sedangkan reaksi yang tinggi (*high*) hanya 15% dari keseluruhan interaksi yang berlangsung, sehingga kegiatan penggalangan partisipasi publik di media sosial khususnya Twitter tergolong sukses bagi calon ini.

C. Pelatihan dengan Naive Bayes

Berdasarkan sampel (*instance*) sejumlah 4.507, diperoleh unjuk kerja dengan klasifikasi *naive Bayes* yang dapat diklasifikasikan dengan benar adalah 76,74%.

TABEL IV
CONFUSION MATRIX NAIVE BAYES

N =4.507		<i>Predicted</i>		
		"High"	"Medium"	"Low"
<i>Actual</i>	"High"	394	289	0
	"Medium"	70	838	419
	"Low"	3	267	2.227

Sedangkan uraian unjuk kerja dari klasifikasi dapat dijelaskan dari *confusion matrix* pada Tabel IV, yaitu sebagai berikut.

- Terdapat tiga *predicted* dan *actual class* yaitu "high", "medium", dan "low".
- Hasil prediksi "high" terjadi sebanyak 467, prediksi "medium" sebanyak 1.394 kali, dan "low" sebanyak 2.646 kali.
- Hasil *actual* menunjukkan adanya "high" sebanyak 683, "medium" sebanyak 1.327, dan "low" sebanyak 2.497.
- Jumlah sampel yang diklasifikasikan dengan benar (*match*) adalah 3.459.

Maka, akurasinya dapat dihitung sebagai berikut.

$[(TL+TM+TH)/n] \times 100\% = [(2227+838+394)/4507] \times 100\% = 76,74\%$, dengan TL = *True Low*, TM = *True Medium*, TH = *True High*.

D. Pelatihan Naive Bayes dengan Kernel Estimator

Setelah diuji coba dengan metode *naive Bayes*, selanjutnya diupayakan perbaikan unjuk kerja dengan menggunakan *naive Bayes KE*. Perbaikan metode ini mendapatkan kenaikan unjuk kerja dengan hasil perolehan sebesar 80,14%.

TABEL V
CONFUSION MATRIX NAIVE BAYES KDE

N =4.507		<i>Predicted</i>		
		"High"	"Medium"	"Low"
<i>Actual</i>	"High"	417	190	76
	"Medium"	0	704	623
	"Low"	0	6	2491

Uraian unjuk kerja dari klasifikasi ini dapat dijelaskan dari *confusion matrix* pada Tabel V sebagai berikut.

- Terdapat tiga *predicted* dan *actual class* yaitu "high", "medium", dan "low".
- Hasil prediksi "high" terjadi sebanyak 417, prediksi "medium" sebanyak 894 kali, dan "low" sebanyak 3.190 kali.
- Hasil *actual* menunjukkan adanya "high" sebanyak 683, "medium" sebanyak 1.327, dan "low" sebanyak 2.497.
- Jumlah sampel yang diklasifikasikan dengan benar (*match*) adalah 3.612.

Maka, akurasinya dapat dihitung sebagai berikut.

$[(TL+TM+TH)/n] \times 100\% = [(417+704+2491)/4507] \times 100\% = 80,14\%$.

E. Pengujian

Tahap ini dilakukan menggunakan metode *percentage split* dengan beberapa kombinasi, yaitu 40%, 55%, 65%, 75%, dan 85%. Kombinasi yang dilakukan adalah memasukkan sejumlah data untuk pelatihan dan sisanya sebagai data uji. Cara ini

dilakukan terhadap dua algoritme, yaitu *naive Bayes* dan *naive Bayes KE*.

Ringkasan hasil *confusion matrix* dari proses uji untuk kedua algoritme disajikan pada Tabel VI dan Tabel VII.

TABEL VI
HASIL UJI NAIVE BAYES

Percentage Split (%)	Waktu proses (s)	Hasil (%)
40	0,41	77,36
55	0,24	76,53
65	0,25	77,17
75	0,16	80,12
85	0,07	79,44

TABEL VII
HASIL UJI NAIVE BAYES KERNEL ESTIMATOR

Percentage Split (%)	Waktu proses (s)	Hasil (%)
40	0,36	80,92
55	0,34	80,28
65	0,29	80,02
75	0,17	80,30
85	0,09	79,59

Hasil pengujian dengan kedua algoritme menunjukkan bahwa semakin kecil nilai *percentage split* membutuhkan waktu pemrosesan yang semakin besar. Hasil uji *naive Bayes* dengan *percentage split* makin besar akan mendapatkan persentase hasil lebih baik, sedangkan hasil uji *naive Bayes KE* menunjukkan persentase hasil yang lebih baik pada *percentage split* yang lebih kecil.

Hasil yang didapat dari proses uji ditampilkan pula dalam bentuk grafik batang, dalam Gbr. 3. Dari hasil beberapa kombinasi uji menggunakan metode *percentage split* dapat diketahui bahwa penggunaan algoritme *naive Bayes KE* dapat meningkatkan persentase hasil uji. Nilai tertinggi *naive Bayes* adalah pada *percentage split* sebesar 75% dengan hasil 80,12%, sedangkan nilai tertinggi *naive Bayes KE* pada *percentage split* sebesar 40% dengan hasil 80,92%.

Hasil pengukuran *confusion matrix* untuk uji terbaik dengan menggunakan algoritme *naive Bayes* dan *naive Bayes KE* disajikan pada Tabel VIII dan Tabel IX.

TABEL VIII
CONFUSION MATRIX NAIVE BAYES UJI

75%		Predicted		
		“High”	“Medium”	“Low”
Actual	“High”	99	72	0
	“Medium”	21	187	108
	“Low”	0	23	617

TABEL IX
CONFUSION MATRIX NAIVE BAYES KERNEL ESTIMATOR UJI

40%		Predicted		
		“High”	“Medium”	“Low”
Actual	“High”	262	113	28
	“Medium”	3	395	363
	“Low”	1	8	1.531

Uji *naive Bayes* terbaik diperoleh pada *percentage split* sebesar 75%, dengan data yang diprediksi secara tepat adalah “high” sebanyak 99, “medium” sebanyak 187, dan “low” sebanyak 61, sehingga yang diprediksi tepat sesuai dengan *actual* adalah 903 data atau akurasi 80,12%.

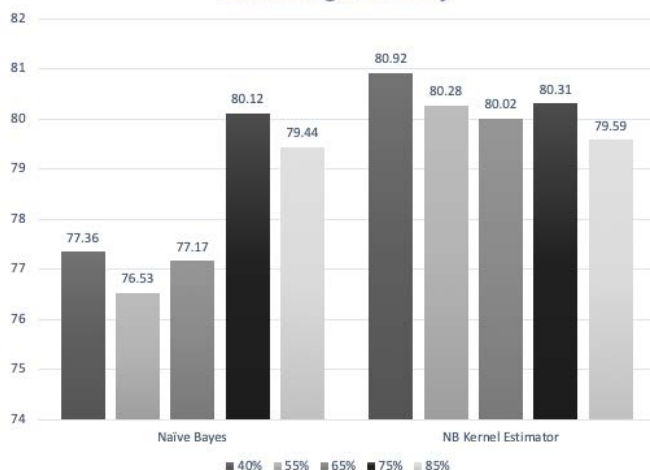
Uji *naive Bayes KE* terbaik diperoleh pada *percentage split* 40%. Data yang diprediksi tepat sebagai “high” adalah 262, “medium” adalah 395, dan “low” sebanyak 1.531, sehingga jumlah data yang diprediksi secara tepat adalah 2.188, dengan akurasi 80,92%.

Hasil penelitian model prediksi dengan *naive Bayes KE* ini dapat melengkapi penelitian sebelumnya tentang tinggi rendahnya pengaruh tulisan/unggahan di media sosial [17], sehingga ukuran tingkat interaksi, khususnya dalam partisipasi politik di media sosial, dapat diprediksikan.

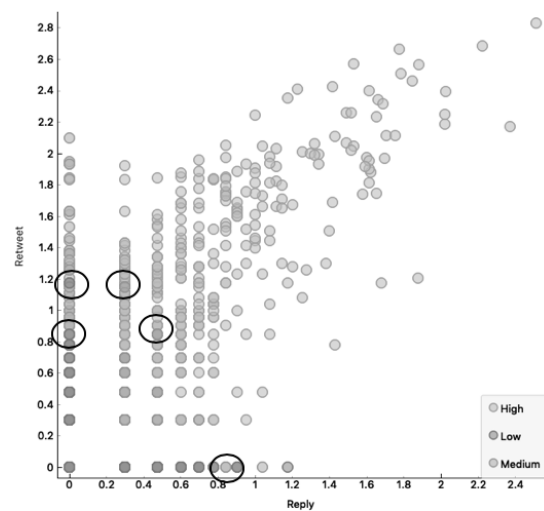
F. Analisis Hasil

Berbagai faktor dapat memengaruhi unjuk kerja dari klasifikasi. Jumlah data yang dikumpulkan dan rentang nilai yang diperoleh dapat memengaruhi akurasi. Diagram tebar (*scatter plot*) pada Gbr. 4 menunjukkan sebaran data dari masing-masing label interaksi yaitu “high”, “medium”, dan “low”. Hasil yang diperoleh menunjukkan bahwa terjadi tumpang tindih beberapa kelompok data atau bernilai sama, yang ditunjukkan dengan lingkaran hitam pada gambar.

Perbandingan Hasil Uji



Gbr. 3 Perbandingan hasil uji.



Gbr. 4 Diagram tebar.

Pada sisi dekat sumbu y terdapat dua kelompok data yang berimpitan antara “low” dan “medium”, sedangkan pada posisi dekat sumbu x terdapat kelompok data yang berimpitan antara “low” dan “medium”. Di sekitar nilai *reply* antara 0,3-0,5 juga didapatkan dua kelompok data yang tumpang tindih antara “medium” dan “high”. Tumpang tindihnya data ini terlihat juga pada data yang akan diolah. Hal ini dapat mengakibatkan tidak tercapainya akurasi melebihi 90% setelah melalui proses klasifikasi menggunakan metode yang diusulkan.

V. KESIMPULAN

Hasil pelatihan terhadap data dengan menggunakan metode klasifikasi *naive Bayes* dan perbaikan metode menggunakan *naive Bayes Kernel Estimator* (KE) menunjukkan hasil yang cukup memuaskan. *Naive Bayes* KE dapat meningkatkan akurasi unjuk kerja dari nilai awal 76,74% naik menjadi 80,14%.

Sejalan dengan itu, hasil uji juga menunjukkan dengan peningkatan metode menggunakan *naive Bayes* KE menghasilkan akurasi sebesar 80,92%, atau 2.188 data yang diprediksi secara tepat. Sedangkan penggunaan *naive Bayes* hanya menghasilkan 80,12% atau hanya memprediksi 1.531 data secara tepat.

Penggunaan *naive Bayes* KE dapat menghasilkan akurasi yang lebih baik dikarenakan pada proses perhitungannya terjadi penghalusan komponen untuk setiap nilai data sehingga menghasilkan kurva probabilitas yang lebih baik. Hal ini dimungkinkan karena nilai-nilai pembentuk kurva dihaluskan melalui penggunaan *kernel smoothing function*.

Hasil klasifikasi menunjukkan bahwa penggunaan algoritme *naive Bayes* KE dapat menghasilkan pemodelan prediksi yang lebih baik. Prediksi ini dapat membantu mengukur pola interaksi kampanye di media sosial. Tingkat interaksi “low”, “medium” hingga “high” dapat diperoleh dari tulisan atau unggahan yang memicu banyaknya tanggapan dengan kombinasi *reply*, *retweet*, dan *like*.

UCAPAN TERIMA KASIH

Terima kasih dihaturkan kepada Kementerian Riset Teknologi dan Pendidikan Tinggi yang telah memberikan dana untuk terselenggaranya penelitian ini. Tak lupa kepada Ferdi Masyaril Widjanarko, Dessy Sagitarini, Arsi Raniah Faza dan Ardi Ilham Falah yang telah menyediakan sarana laboratorium yang menunjang dalam menyiapkan data dan pengolahannya.

REFERENSI

- [1] W. Budiharto dan M. Meiliana, “Prediction and Analysis of Indonesia Presidential Election from Twitter Using Sentiment Analysis,” *J. Big Data*, Vol. 5, No. 1, hal. 51-60, Dec. 2018.
- [2] P. Wang, W. He, dan J. Zhao, “A Tale of Three Social Networks: User Activity Comparisons across Facebook, Twitter, and Foursquare,” *IEEE Internet Comput.*, Vol. 18, No. 2, hal. 10–15, Mar. 2014.
- [3] (2017) “Detecting Stress Based on Social Interactions in Social Networks - IEEE Journals & Magazine,” [Online]. <https://ieeexplore.ieee.org/document/7885098>, tanggal akses: 22-Nov-2018.
- [4] (2016) “An Agent Based Model of Spread of Competing Rumors Through Online Interactions on Social Media - IEEE Conference Publication.” [Online]. <https://ieeexplore.ieee.org/document/7408413>, tanggal akses: 22-Nov-2018.
- [5] S.-J. Bang dan W. Wu, “Naive Bayes Ensemble: A New Approach to Classifying Unlabeled Multi-Class Asthma Subjects,” *2016 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, 2016, hal. 460–465.
- [6] M.N. Al-Azam, M.M. Achlaq, A. Nugroho, A.G. Sooi, dan A. Winaya, “Broadcasting the Status of Plant Growth Chamber using Bluetooth Low Energy,” *MATEC Web Conf.*, Vol. 164, hal. 1-11, 2018.
- [7] A.G. Sooi, A. Nugroho, M.N.A. Azam, S. Sumpeno, dan M.H. Purnomo, “Virtual Artifact: Enhancing Museum Exhibit Using 3D Virtual Reality,” *2017 TRON Symposium (TRONSHOW)*, 2017, hal. 1–5.
- [8] C. Kartiko dan G.B. Hertantyo, “Peningkatan Kualitas Aplikasi Pemantau Media Sosial dan Media Daring Menggunakan Metode WebQEM,” *J. Nas. Tek. Elektro Dan Teknol. Inf. JNTETI*, Vol. 7, No. 2, hal. 144-149, 2018.
- [9] K. Michael, “Are You Addicted to Your Smartphone, Social Media, dan More?: The New AntiSocial App Could Help,” *IEEE Consum. Electron. Mag.*, Vol. 6, No. 4, hal. 116–121, Oct. 2017.
- [10] K. Kandhway dan J. Kuri, “Campaigning in Heterogeneous Social Networks: Optimal Control of SI Information Epidemics,” *IEEEACM Trans. Netw.*, Vol. 24, No. 1, hal. 383–396, Feb. 2016.
- [11] J. Saldivar, C. Rodríguez, F. Daniel, F. Casati, dan L. Cernuzzi, “On the (In)effectiveness of the Share/Tweet Button: Idea Management for Civic Participation,” *IEEE Internet Comput.*, Vol. 21, No. 5, hal. 38–47, 2017.
- [12] I. I. G. şen, “Internet Freedom and Political Participation in Turkey: Legal Framework and Practice,” *2014 First International Conference on eDemocracy eGovernment (ICEDEG)*, 2014, hal. 105–108.
- [13] D. Opazo, M. Wolff, dan M.J. Araya, “Imagination and the Political in Design Participation,” *Des. Issues*, Vol. 33, No. 4, hal. 73–82, Oct. 2017.
- [14] D. Gayo-Avello, “Social Media, Democracy, and Democratization,” *IEEE Multimed.*, Vol. 22, No. 2, hal. 10–16, Apr. 2015.
- [15] K. Gorkovenko, N. Taylor, dan J. Rogers, “Social Printers: A Physical Social Network for Political Debates,” *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, 2017, hal. 2269–2281.
- [16] K. Lee, J. Mahmud, dan J. Chen, “Who Will Retweet This? Detecting Strangers from Twitter to Retweet Information,” *ACM Trans. Intell. Syst. Technol. TIST - Surv. Pap. Regul. Pap. Spec. Issue Soc. Media Process.*, Vol. 8, No. 6, hal. 1-10, Sep. 2017.
- [17] M. Glenski dan T. Weninger, “Rating Effects on Social News Posts and Comments,” *ArXiv160606140 [cs.SI]*, Jun. 2016.
- [18] H. Hao, R. Padman, B. Sun, dan R. Telang, “Examining the Social Influence on Information Technology Sustained Use in a Community Health System: A Hierarchical Bayesian Learning Method Analysis,” *2014 47th Hawaii International Conference on System Sciences (HICSS)*, 2014, hal. 2751–2758.
- [19] Y. Chen, J. Hu, H. Zhao, Y. Xiao, dan P. Hui, “Measurement and Analysis of the Swarm Social Network With Tens of Millions of Nodes,” *IEEE Access*, Vol. 6, hal. 4547–4559, 2018.
- [20] S. Servia-Rodríguez, A. Fernández-Vilas, R.P. Díaz-Redondo, dan J.J. Pazos-Arias, “Inferring Contexts From Facebook Interactions: A Social Publicity Scenario,” *IEEE Trans. Multimed.*, Vol. 15, No. 6, hal. 1296–1303, Oct. 2013.
- [21] M.Y. Nejad, M. Hosseinzadeh, dan M. Mohammadi, “Hijab in Twitter: Advocates and Critics: A Content Analysis of Hijab-Related Tweets,” *IEEE Technol. Soc. Mag.*, Vol. 37, No. 2, hal. 46–51, Jun. 2018.
- [22] S. Ahmed dan M.M. Skoric, “My Name Is Khan: The Use of Twitter in the Campaign for 2013 Pakistan General Election,” *2014 47th Hawaii International Conference on System Sciences (HICSS)*, 2014, hal. 2242–2251.
- [23] H. Nurrahmi, R. Wijayanti, A.F. Rozie, dan A. Arisal, “Twitter Data Transformation for Network Visualization Based Context Analysis,” *2018 International Conference on Information and Communications Technology (ICOIAC)*, 2018, hal. 525–530.
- [24] T. Busjahn, R. Bednarik, dan C. Schulte, “What Influences Dwell Time During Source Code Reading?: Analysis of Element Type and Frequency As Factors,” *Proceedings of the Symposium on Eye Tracking Research and Applications*, 2014, hal. 335–338.
- [25] Z. Sari, M. Sarosa, dan S. Suhari, “‘Si Tole’ Chatterbot untuk Melatih

- Rasa Percaya Diri Menggunakan Naive Bayes Classification,” *J. Nas. Tek. Elektro Dan Teknol. Inf. JNTETI*, Vol. 7, No. 1, hal. 64-71, 2018.
- [26] Z.L. Xiang, X.R. Yu, A.W.M. Hui, dan D.K. Kang, “Novel Naive Bayes based on Attribute Weighting in Kernel Density Estimation,” *2014 Joint 7th International Conference on Soft Computing and Intelligent Systems (SCIS) and 15th International Symposium on Advanced Intelligent Systems (ISIS)*, 2014, hal. 1439–1442.