

Terbit online pada laman web jurnal: <http://jurnal.iaii.or.id>**JURNAL RESTI****(Rekayasa Sistem dan Teknologi Informasi)**

Vol. 4 No. 3 (2020) 504 - 512

ISSN Media Elektronik: 2580-0760

Analisis Sentimen Pemindahan Ibu Kota Negara dengan *Feature Selection* Algoritma *Naive Bayes* dan *Support Vector Machine*Faried Zamachsari¹, Gabriel Vangeran Saragih², Susafa'ati³, Windu Gata⁴^{1,2,3,4}Magister Ilmu Komputer, Fakultas Ilmu Komputer, STMIK Nusa Mandiri Jakarta¹14002335@nusamandiri.ac.id, ²14002327@nusamandiri.ac.id, ³susafa.suf@nusamandiri.ac.id, ⁴windu@nusamandiri.ac.id**Abstract**

The decision to move Indonesia's capital city to East Kalimantan received mixed responses on social media. When the poverty rate is still high and the country's finances are difficult to be a factor in disapproval of the relocation of the national capital. Equitable development and an increasingly vibrant business environment are factors that support the relocation of the national capital. Twitter as one of the popular social media, is used by the public to express these opinions. How is the tendency of community responses related to the move of the National Capital and how to do public opinion sentiment analysis related to the move of the National Capital with *Feature Selection Naive Bayes Algorithm* and *Support Vector Machine* to get the highest accuracy value is the goal in this study. Sentiment analysis data will take from public opinion using Indonesian from Twitter social media tweets in a crawling manner. Search words used are #IbuKotaBaru and #PindahIbuKota. The amount of data taken in the period 23-April-2020 until 2-May-2020 gets the results of 849 data tweets. The stages of the research consisted of collecting data through social media Twitter, polarity, preprocessing consisting of the process of transform case, cleansing, tokenizing, filtering and stemming. The use of feature selection to increase the accuracy value will then enter the ratio that has been determined to be used by data testing and training. The next step is the comparison between the *Support Vector Machine* and *Naive Bayes* methods to determine which method is more accurate. In the data period above it was found 24.26% positive sentiment 75.74% negative sentiment related to the move of a new capital city. Accuracy results using *Rapid Miner* software, the best accuracy value of *Naive Bayes* with *Feature Selection* is at a ratio of 9:1 with an accuracy of 88.24% while the best accuracy results *Support Vector Machine* with *Feature Selection* is at a ratio of 5:5 with an accuracy of 78.77%.

Keywords: naive bayes, support vector machine, feature selection, sentiment analysis, new capital city

Abstrak

Keputusan pemindahan ibu kota negara Indonesia ke Kalimantan Timur mendapat tanggapan beragam pada media sosial. Disaat tingkat kemiskinan masih tinggi dan keuangan negara sedang sulit menjadi faktor ketidaksetujuan pemindahan ibu kota negara. Pemerataan pembangunan dan kondisi dunia usaha yang semakin bergairah menjadi faktor yang mendukung pemindahan ibu kota negara. *Twitter* sebagai salah satu media sosial yang populer, digunakan masyarakat untuk mengungkapkan opini-opini tersebut. Bagaimana kecenderungan tanggapan masyarakat terkait pemindahan ibu kota negara serta bagaimana melakukan analisis sentimen opini masyarakat terkait pemindahan ibu kota negara dengan *Feature Selection Algoritma Naive Bayes* dan *Support Vector Machine* untuk mendapatkan nilai akurasi paling tinggi menjadi tujuan dalam penelitian ini. Data analisis sentimen akan mengambil dari opini masyarakat yang menggunakan bahasa Indonesia dari *tweet* media sosial *Twitter* secara *crawling*. Pencarian kata yang digunakan adalah #IbuKotaBaru dan #PindahIbuKota. Jumlah data yang diambil pada periode 23-April-2020 s.d 2-Mei-2020 mendapatkan hasil sebanyak 849 data *tweet*. Tahapan penelitian terdiri dari pengumpulan data melalui media sosial *Twitter*, polaritas, *preprocessing* terdiri dari proses *transform case*, *cleansing*, *tokenizing*, *filtering* dan *stemming*. Penggunaan *feature selection* untuk meningkatkan nilai akurasi yang selanjutnya akan memasukan rasio yang telah ditentukan untuk digunakan oleh data *testing* dan *training*. Tahap selanjutnya, yaitu perbandingan antara algoritma *Support Vector Machine* dan *Naive Bayes* untuk menentukan algoritma mana yang lebih akurat. Dalam periode data diatas didapatkan 24,26% sentimen positif dan 75,74% sentimen negatif terkait pemindahan ibu kota baru. Hasil akurasi dengan memakai perangkat lunak *Rapid Miner* nilai akurasi terbaik algoritma *Naive Bayes* dengan *Feature Selection* ada pada rasio 9:1 dengan hasil akurasi 88,24% sedangkan hasil akurasi terbaik algoritma *Support Vector Machine* dengan *Feature Selection* ada pada rasio 5:5 dengan hasil akurasi 78,77%.

Kata kunci: naive bayes, support vector machine, feature selection, analisis sentimen, ibu kota baru

1. Pendahuluan

Jakarta - Presiden Joko Widodo (Jokowi) akhirnya resmi memilih Provinsi Kalimantan Timur (Kaltim) sebagai ibu kota negara Indonesia yang baru. Hal itu diumumkan langsung di Istana Negara, Jakarta Pusat. "Lokasi ibu kota baru yang paling ideal adalah di sebagian Kabupaten Penajam Paser Utara dan sebagian di Kabupaten Kutai Kartanegara Provinsi Kalimantan Timur" [1].

Beragam reaksi pro dan kontra terhadap rencana pemindahan ibu kota muncul dalam masyarakat. Ibu kota pindah pada saat angka kemiskinan masih tinggi dan pemindahan ibu kota dikritik karena keuangan negara tengah sulit adalah beberapa alasan yang menolak pemindahan ibu kota negara untuk saat ini. Menteri Pekerjaan Umum dan Perumahan Rakyat (PUPR), Basuki Hadimuljono, menuturkan ada banyak pertimbangan kenapa ibu kota harus dipindahkan dari Jakarta. Salah satunya terkait faktor daya dukung Jakarta seperti kemacetan atau banjir. Ketua Umum Asosiasi Pengusaha Indonesia (Apindo), Hariyadi Sukamdani, menyebut bahwa pemindahan ibu kota dari Jakarta ke luar Jawa akan membawa dampak positif, khususnya bagi dunia usaha. Kepala Badan Koordinasi Penanaman Modal (BKPM) Thomas Lembong menyebut rencana pemindahan ibukota dapat menjadi angin segar dan membawa sentimen positif bagi para investor jika sudah terealisasi. Faktor-faktor diatas adalah faktor yang mendukung pemindahan ibu kota negara[2].

Keputusan pemindahan ibu kota ini sangat menarik karena telah menimbulkan opini beragam baik yang setuju ataupun tidak setuju dari masyarakat dalam jejaring sosial dan mikroblog daring *Twitter*. *Twitter* memungkinkan pengguna untuk mengirim dan membaca pesan berbasis teks hingga 280 karakter yang dikenal dengan sebutan kicauan (*tweet*). *Twitter* adalah media sosial paling populer dalam menyampaikan opini [3]. *Tweet* yang dilakukan oleh masyarakat merupakan sumber data yang valid untuk dilakukan analisis sentimen [4]. Penyampaian opini di *Twitter* cenderung menggunakan bahasa informal dan istilah yang dipersingkat [5]. Beberapa tagar pada media sosial *Twitter* bermunculan diantaranya adalah #IbuKotaBaru dan #PindahIbuKota.

Analisis sentimen merupakan bidang studi yang menganalisis pendapat, sentimen, penilaian, evaluasi, sikap, dan emosi seseorang terkait suatu topik, layanan, produk, individu, organisasi, atau kegiatan tertentu [6]. Analisis sentimen adalah salah satu cabang penelitian dari *text mining* yang berguna untuk mengklasifikasi dokumen teks berupa opini berdasarkan sentimen. *Text mining* memiliki potensi nilai komersial, *sentiment analysis* atau *opinion mining* adalah studi komputasional dari opini-opini orang, sentimen dan emosi melalui entitas dan atribut yang dimiliki yang diekspresikan dalam bentuk teks [7].

Analisis sentimen dapat mengklasifikasikan opini-opini tersebut diantaranya menggunakan algoritma data mining seperti klasifikasi *Naive Bayes* [8]. *Naive Bayes* merupakan sebuah pengklasifikasian probalistik sederhana yang menghitung sekumpulan probabilitas dengan menjumlahkan frekuensi dan kombinasi nilai dari dataset yang diberikan. Algoritma menggunakan teorema *bayes* dan mengansumsikan semua atribut independen atau tidak saling ketergantungan yang diberikan oleh nilai pada variabel kelas. *Naive Bayes* juga didefinisikan sebagai pengklasifikasian dengan metode probabilitas dan statistik yang dikemukakan oleh ilmuwan inggis Thomas Bayes, yaitu memprediksi peluang di masa depan berdasarkan pengalaman di masa sebelumnya[9]. Kelebihan lain, akurasi metode ini cukup baik pada data besar seperti analisis sentimen [10]. Problem mendasar pada *opinion mining* adalah mengekstrak opini berdasarkan klasifikasi [10]. Besarnya manfaat bagi pengetahuan dan industri, membuat analisis sentimen berkembang pesat [11].

Analisis sentimen dapat juga menggunakan algoritma algoritma *Support Vector Machine* (SVM). Konsep SVM pertama kali dipresentasikan pada tahun 1992 di *Annual Workshop on Computational Learning Theory*. Konsep SVM sebenarnya adalah kombinasi yang harmonis dari teori-teori komputasi yang sudah ada puluhan tahun sebelumnya, seperti kernel, margin *hyperplane*, dan konsep-konsep pendukung lainnya, namun hingga tahun 1992, belum pernah ada upaya untuk merangkai komponen-komponen tersebut [12]. Teknik SVM digunakan untuk mendapatkan fungsi pemisah (*hyperplane*) yang optimal untuk memisahkan observasi yang memiliki nilai variabel target yang berbeda *hyperplane* ini dapat berupa *line* pada *two dimension* dan dapat berupa *flat plane* pada *multiple dimension* [13].

Menurut Basari bahwa *Particle Swarm Optimization* (PSO) merupakan teknik optimasi yang sangat sederhana untuk menerapkan dan memodifikasi beberapa parameter. Dalam PSO terdapat beberapa teknik untuk melakukan pengoptimasian diantaranya meningkatkan bobot atribut (*attribute weight*) terhadap semua atribut atau variabel yang dipakai, menseleksi atribut (*attribute selection*) dan *feature selection* [14]. Salah satu *feature selection* adalah *Weight by Corellation* yang artinya pembobotan pada *attribute* dengan cara menghubungkan (korelasi) satu *attribute* dengan atribut lainnya [15].

Penelitian dengan menggunakan analisis sentimen berdasarkan opini masyarakat sehubungan dengan keputusan pemerintah dengan menggunakan algoritma Klasifikasi *Naive Bayes* diantaranya adalah analisis sentimen terhadap pelaksanaan kurikulum 2013 [16], analisis sentimen terhadap ujian nasional secara online [17], dan analisis sentimen penerapan e-KTP [18]. Merujuk kepada penelitian sebelumnya, penelitian ini berkaitan dengan keputusan pemerintah dengan menggunakan algoritma klasifikasi *Naive Bayes*. Tujuan

penelitian ini adalah melakukan analisis sentimen terhadap *tweet* berbahasa Indonesia pada *Twitter* berupa opini masyarakat terhadap keputusan pemerintah Indonesia pindah ibukota dengan menggunakan algoritma klasifikasi *Naive Bayes*.

Analisis sentimen dan klasifikasi respon masyarakat terhadap pemindahan ibu kota negara ke Kalimantan Timur pada *twitter* [19]. Hasil pengujian akurasi dengan memakai perangkat lunak *Rapid Miner* dengan metode *Naive Bayes* menghasilkan nilai akurasi terendah sebesar 57,06% dari rasio 1:9 dan menghasilkan nilai akurasi tertinggi sebesar 62,09% dari rasio 7:3 pada proses pengujian dengan memakai sampel sebesar 606 data. Hasil pengujian akurasi dengan memakai perangkat lunak *Rapid Miner* dengan metode *Support Vector Machine* menghasilkan nilai akurasi terendah sebesar 54,10% dari rasio 9:1 dan menghasilkan nilai akurasi tertinggi sebesar 63,19% dari rasio 7:3 pada proses pengujian dengan memakai sampel sebesar 606 data. Analisis sentimen keputusan pemindahan ibu kota negara menggunakan klasifikasi *Naive Bayes* [20]. Kesimpulan ketika dilakukan proses manual pada *preprocessing* diperlukan pemahaman penggunaan bahasa Indonesia terutama penggunaan kalimat dan kata dasar yang termasuk opini positif atau opini negatif. Nilai akurasi tertinggi diperoleh 92,00% pada rasio 90:10 dan akurasi rata-rata dari lima rasio yang diukur adalah 89,86%.

Penggunaan *feature selection* di Algoritma *Support Vector Machine* untuk Sentimen Analisis Komisi Pemilihan Umum [21]. Penggunaan *feature selection* meingg penggunaan *feature selection* meningkatkan nilai untuk *accuracy* dan nilai AUC. Analisis Sentimen *E-Wallet* Pada *Google Play* Menggunakan Algoritma *Naive Bayes* Berbasis *Particle Swarm Optimization* [22]. Penelitian dengan model algoritma *Naive Bayes Classifier* tanpa *feature selection* dan dibandingkan dengan model *Naive Bayes Classifier* dengan *feature selection* dapat disimpulkan bahwa penggunaan *feature selection Particle Swarm Optimization (PSO)* dapat meningkatkan nilai dari *Accuracy* dan AUC. Peningkatan yang didapatkan sangat signifikan yang sebelumnya model *Naive Bayes Classifier* tanpa *feature selection* hanya menghasilkan 82.30% dan nilai AUC 0.780 setelah ditambahkan penggunaan *feature selection* menjadi 83.60% untuk *accuracy* dan nilai AUC 0.801.

Setelah melakukan penelitian dengan model algoritma *Naive Bayes Classifier* tanpa *feature selection* dan dibandingkan dengan model *Naive Bayes Classifier* dengan *feature selection* dapat disimpulkan bahwa penggunaan *feature selection Particle Swarm Optimization (PSO)* dapat meningkatkan nilai dari *Accuracy* dan AUC.

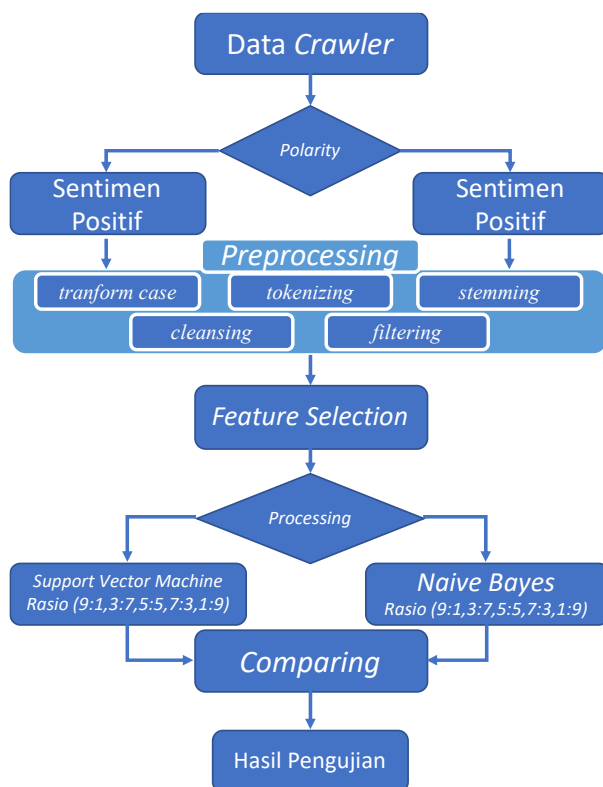
Berdasarkan penelitian yang telah dilakukan sebelumnya penulis tertarik untuk melakukan penelitian analisis sentimen dengan menerapkan perbandingan algoritma dan perbandingan rasio. Perbandingan

algoritma yang digunakan adalah algoritma *Support Vector Machine* dengan *feature selection* dan algoritma *Naive Bayes* dengan *feature selection* dengan perbandingan rasio 9:1; 7:3; 5:5; 3:7; 1:9.

Tujuan yang ingin dicapai dari penelitian ini untuk mengetahui respon mengenai setuju atau tidaknya masyarakat terhadap pemindahan ibu kota negara melalui media sosial *Twitter* di Indonesia dan Mengetahui rasio terbaik antara data *training* dan *testing* yang akan dimasukkan pada algoritma *Support Vector Machine* dengan *feature selection* dan *Naive Bayes* dengan *feature selection* dalam menentukan nilai akurasi menentukan respon masyarakat terhadap pemindahan ibu kota negara.

2. Metode Penelitian

Metode penelitian dimulai dari data *crawler* atau proses pengumpulan data dilanjutkan dengan *polarity* dimana data akan diberikan label menjadi sentimen positif dan sentimen negatif. Tahapan selanjutnya adalah *preprocessing* yang terdiri dari *transform case*, *cleansing*, *tokenizing*, *filtering* dan *stemming*. Untuk mengoptimalkan akurasi dilakukan proses *feature selection*. Proses selanjutnya adalah memasukan rasio yang telah ditentukan untuk digunakan dalam data *testing* dan *training* dan diakhiri dengan membandingkan antara algoritma *Support Vector Machine* dan *Naive Bayes* untuk menentukan algoritma yang lebih akurat dalam melakukan analisis sentimen. Proses diatas digambarkan dalam Gambar 1.



Gambar 1. Metode Penelitian

2.1. Data Crawler

Pengumpulan data menggunakan media sosial *Twitter*. Data yang di gunakan adalah data *tweet* dengan kata kunci #IbuKotaBaru dan #PindahIbuKota. Dari masing-masing respon, diambil dengan menggunakan beberapa percobaan pengujian data latih dan data uji dengan menggunakan rasio 9:1; 7:3; 5:5; 3:7; dan 1:9 pada setiap algoritma. Data yang diambil adalah data pada *Twitter* pada bulan tanggal 23-April-2020 s.d 2-Mei-2020.

2.2. Polarity

Polarity akan dilakukan secara manual untuk memberikan label sentimen positif atau sentimen negatif terhadap data yang dipilih.

2.3. Preprocessing

Tahap *preprocessing* meliputi tahap *transform case*, *cleansing*, *tokenizing*, *filtering* dan *stemming*. *Transform case* akan mengubah semua huruf menjadi huruf kecil. *Cleansing* menghapus *character link* url https, http, @ dan #. *Tokenizing* akan memecah kalimat menjadi kata. *Filtering* digunakan untuk mengambil kata-kata yang mempunyai arti penting. *Stemming* digunakan untuk mengambil kata dasar dari setiap kalimat yang telah dipecah perkata pada tahap *tokenizing*.

2.4. Feature Selection (Weight by Correlation)

Untuk pengoptimasian diantaranya meningkatkan bobot atribut (*attribute weight*) terhadap semua atribut atau variabel yang dipakai sehingga dapat menaikkan nilai akurasi dari model *machine learning* yang dikembangkan.

2.5. Processing

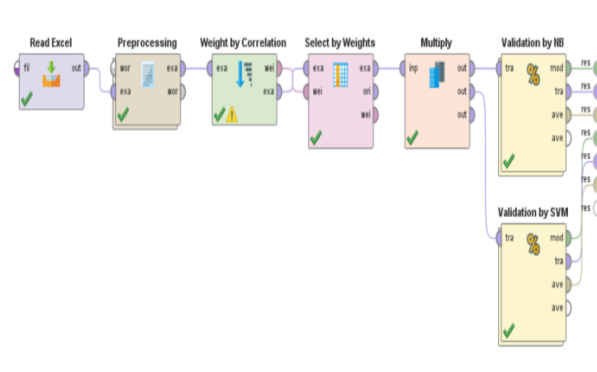
Data akan dipilih secara acak dalam melakukan proses pengecekan akurasi dari data *training* dan *testing* dan akan melakukan pengujian terhadap rasio yang telah ditentukan untuk setiap algoritma baik algoritma *Support Vector Machine* atau *Naive Bayes* sehingga dari data yang dihasilkan dapat diketahui hasil keakuratan data serta hasil rasio terbaik yang dapat digunakan dalam melakukan pengujian data.

2.6. Comparing

Membandingkan algoritma *Support Vector Machine* dan *Naive Bayes* pada hasil akurasi yang didapatkan dari *Confusion Matriks* berupa akurasi data, *class recall*, dan *class precision* pada setiap rasio antara data latih dan data uji.

3. Hasil dan Pembahasan

Tahap ini penulis akan melakukan pembahasan proses analisis dan klasifikasi dengan menggunakan *Rapid Miner* untuk mengumpulkan data, membandingkan algoritma klasifikasi, menguji akurasi algoritma klasifikasi serta melakukan pembahasan hasil klasifikasi respon masyarakat, perbandingan algoritma dan hasil akurasi terbaik.

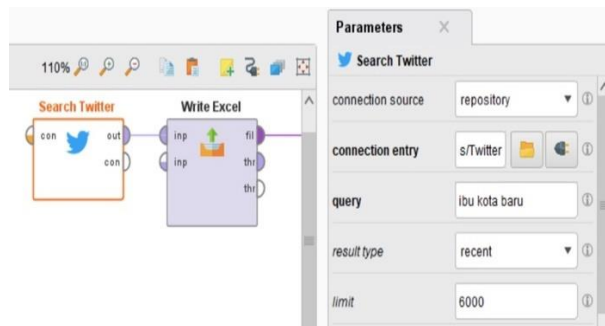


Gambar 2. Pemodelan Penelitian

Gambar 2 diatas menunjukkan hasil implementasi proses pemodelan dalam *Rapid Miner*. Pada proses *aply* model akan dilaksanakan sebanyak 5 kali sesuai jumlah rasio antara data *training* dan data *testing* yang telah ditentukan yaitu dengan perbandingan rasio 9:1; 7:3; 5:5; 3:7; 1:9.

3.1. Data Crawler

Data Crawler adalah pengumpulan data menggunakan media sosial *Twitter*. Hasil dari proses *crawling* dengan kata kunci #IbuKotaBaru dan #PindahIbuKota mulai tanggal 23-April-2020 s.d 2-Mei-2020 untuk #IbuKotaBaru mendapatkan 5.687 *tweet* sedangkan #PindahIbuKota mendapatkan 465 *tweet*.



Gambar 3. Pengambilan Data *Twitter*

Gambar 3 merupakan proses pemodelan untuk pengambilan data melalui media sosial *Twitter*. Dari total data yang didapat sebanyak 6.152 dilakukan proses penghapusan *tweet* duplikat dan didapatkan data 849 untuk diolah.

3.2. Polarity

Hasil data yang telah dilakukan penambangan akan di lakukan proses pemisahan pada kelompok sentimen positif atau sentimen negatif. Proses pemisahan sentimen dilakukan secara manual.

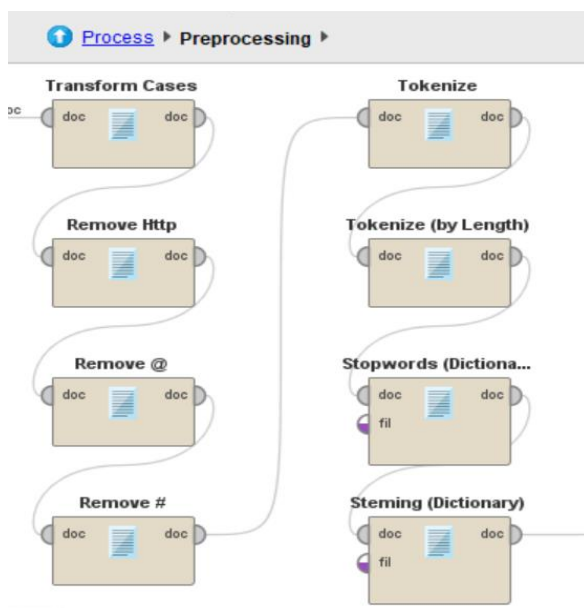
Tabel 1 menunjukkan contoh *tweet* yang telah dilakukan *polarity*. Dari total 849 *tweet* yang didapatkan sebanyak 206 (24,26%) dengan sentimen positif atau mendukung pemindahan ibu kota dan 643 (75,74%) dengan sentimen negatif atau tidak mendukung pemindahan Ibu Kota.

Tabel 1. Contoh Hasil *Polarity*

No.	Tweet	Sentimen
1	Lanjutkan untuk Indonesia Maju Ibu kota Baru.	Positif
2	Semoga ibu kota baru tetap berjalan, orang yg ga setuju ibu kota baru pindah adalah pokitikus2 yg takut kehilangan pamornya karna pasti otomatis orang lokal diibukota baru lah nantinya the new politikus.	Positif
3	Tapi kesaksian pak PS, pak Jokowi berjuang adalah demi kepentingan bangsa dan rakyat. Ibu kota baru tentu untuk itu. Mengapa harus distop pak FJ?	Positif
4	Kenapa pak presiden tdk pakai lahan calon ibu kota negara sbg lokasi peretakan sawah baru. Mumpung blm jelas kapan mau E20, mending buat bikin sawah sj dulu.	Negatif
5	Peneliti Inggris: Calon Ibu Kota Baru RI Rawan Tsunami.	Negatif
6	Ratusan juta Rakyat tak butuh Ibu Kota Negara yg Baru. Rakyat hanya butuh Rezim baru.	Negatif

3.3. Preprocessing

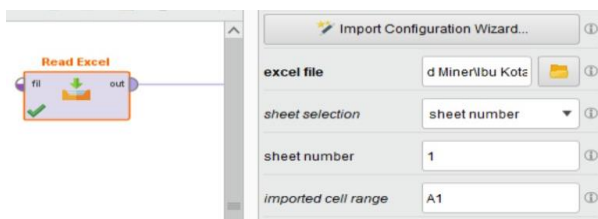
Data hasil *Polarity* akan dimasukkan dalam tahap *preprocessing* yang terdiri dari proses *Transform Cases*, *Remove Http/Https*, *Remove @*, *Remove #*, *Tokenize*, *Filter Tokens by Length* dan *Filter Stopword*. Tahapan tersebut digambarkan dalam gambar 4.



Gambar 4. Alur *Preprocessing*

3.3.1. Read Data Excel

Proses ini akan membaca sebuah data *ExampleSet* dari file *excel* yang ditentukan.



Gambar 5. *Read Data Excel*

Pada gambar 5 menunjukkan proses pembacaan data text yang berisi *tweet* yang telah dilakukan proses *polarity* berupa sentimen positif dan negatif.

3.3.2. Transform Cases

Proses ini akan mengubah text yang masih mengandung huruf kapital menjadi huruf kecil semua. Hal ini dilakukan agar ketika dilakukan proses ke dalam model

klasifikasi terdapat keseragaman huruf agar tidak terjadi kesalahan dalam proses *tokenize*.

3.3.3. Removal

Proses ini akan menghilangkan *https*, *http*, *@* dan *#* pada text *tweet*. Hal ini dilakukan karena data yang diambil berasal dari *Twitter* yang dalam text *tweet* masih mengandung text *https*, *http*, *@* dan *#*.

3.3.4. Tokenizing

Proses tokenisasi pada data teks adalah pemecahan kalimat *tweet* menjadi potongan karakter atau satuan kata sesuai kebutuhan yang sering disebut token. Pada proses *tokenizing* dilakukan proses *filtering* data untuk mengambil kata-kata yang memiliki minimal 3 character dan maksimal 25 character . Untuk kata yang terdiri kurang dari 4 character atau lebih dari 25 character akan dihapus.

3.3.5. Filtering

Tahap *filtering* adalah tahap mengambil kata-kata penting dari hasil token dengan cara menggunakan algoritma *stoplist* (membuang kata kurang penting). *Stoplist/stopword* adalah kata-kata yang tidak deskriptif yang dapat dibuang dalam pendekatan Indonesian *Stopword Removal*. Contoh *stopword* adalah “yang”, “dan”, “di”, “dari”.

3.3.6. Stemming

Hasil dari *filtering* akan dilanjutkan dengan proses *stemming* yaitu menghilangkan imbuhan yang terdapat pada masing-masing kata sehingga menjadi kata dasar dengan menggunakan indonesian *stemming* untuk *review* berbahasa Indonesia.

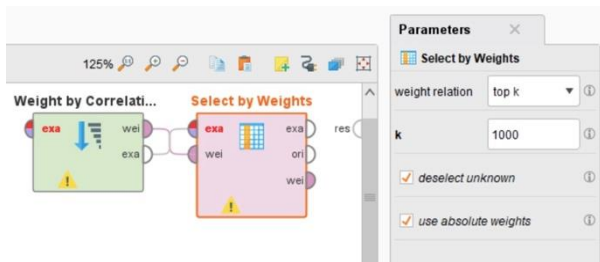
3.4. Feature Selection (Weight by Correlation)

Feature Selection (Weight by Correlation) berfungsi untuk menghitung bobot berdasarkan korelasi antara satu fitur dengan fitur yang ada dalam dataset dengan mempertimbangkan nilai label. Setelah fitur dilakukan penghitungan bobot maka dilakukan *sort ascending* fitur berdasarkan nilai bobot itu sendiri. Terdapat 2.383 *regular attributes* yang dihasilkan dari 849 data contoh yang diproses. Pada proses *feature selection (Weight by Correlation)* akan mengambil 1.000 data pertama yang mempunyai urutan 1 sampai dengan 1.000 agar tidak semua fitur masuk dalam model. Hal ini dilakukan dengan menggunakan operator *select by weight* dengan parameter *weight relation* = top k. Dengan nilai k=1.000. Penggunaan kedua operator ini dapat dilihat pada gambar 6.

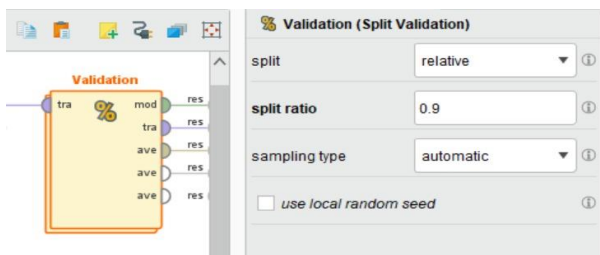
3.5. Processing

Data akan dibagi menjadi 2 yaitu data *training* dan data *testing*. Data *training* dan *testing* akan diambil secara acak dengan jumlah 849 data yang telah diambil dari *preprocessing* data. Perhitungan performansi menggunakan operator *split validation*, yang digunakan untuk menilai atau memvalidasi keakuratan sebuah

model yang dibangun berdasarkan rasio tertentu. Pemisahan menggunakan *relative* dengan menggunakan rasio pemisah (*split ratio*) yang akan diisikan sesuai rasio yang ditetapkan dan *sampling type automatic*.



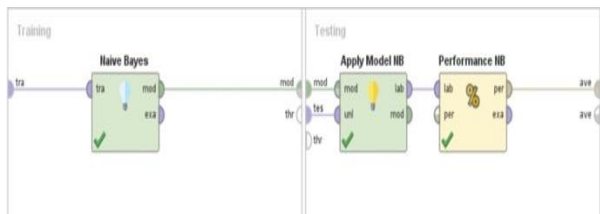
Gambar 6. Feature Selection



Gambar 7. Split Validation

Gambar 7 menunjukkan *split validation* dengan *split ratio* 0.9. *Split ratio* 0.9 mempunyai arti bahwa data yang digunakan untuk *training* adalah 90% dan data yang digunakan untuk *testing* adalah 10%. Ada 5 *variable split ratio* yang akan di gunakan untuk memproses data yaitu 0.9, 0.7, 0.5, 0.3 dan 0.1.

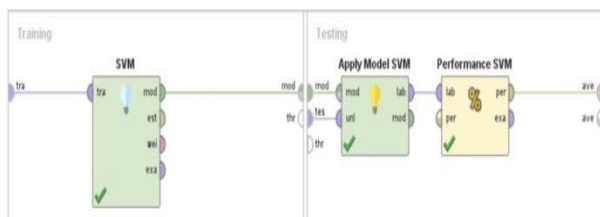
3.5.1. Algoritma Naive Bayes



Gambar 8. Implementasi Algoritma Naive Bayes

Gambar 8 menunjukkan proses klasifikasi dengan *Rapid Miner* dengan algoritma *Naive Bayes* untuk mengklasifikasikan data penelitian dilanjutkan dengan proses *Apply Model* agar dapat membaca data. Untuk melihat nilai akurasi dilakukan dengan memasang *performance* sehingga data akan dilakukan evaluasi.

3.5.2 Algoritma Support Vector Machine



Gambar 9. Implementasi Algoritma Support Vector Machine

Gambar 9 menunjukkan proses klasifikasi dengan *Rapid Miner* dengan algoritma *Support Vector Machine* untuk mengklasifikasikan data penelitian dilanjutkan dengan proses *Apply Model* agar dapat membaca data. Untuk melihat nilai akurasi dilakukan dengan memasang *performance* sehingga data akan dilakukan evaluasi.

3.6 Comparing

Dalam melakukan pengujian ini, data dibagi menjadi 2 yaitu data *training* dan data *testing*. Data yang digunakan sebagai data *training* dan data *testing* diambil secara acak dari 849 data yang akan dianalisis. Dari 606 data yang ada, data akan dibagi menjadi beberapa bagian untuk proses *training* dan *testing*. Dengan pembagian, 90% data *training* dengan 10% data *testing* yang ditulis dengan rasio 9:1, 70% data *training* dengan 30% data *testing* yang ditulis dengan rasio 7:3, 50% data *training* dengan 50% data *testing* yang ditulis dengan rasio 5:5, 30% data *training* dengan 70% data *testing* yang ditulis dengan rasio 3:7 dan 10% data *training* dengan 90% data *testing* yang ditulis dengan rasio 1:9. Pada *Rapid Miner* ratio pembagian dimasukkan dengan nilai persentase yang diubah kedalam nilai desimal seperti 0.9, 0.7, 0.5, 0.3 dan 0.1.

Pada proses pembagian data apabila hasilnya desimal pembulatan akan disesuaikan dengan angka di sebelah kiri koma, jika angka tersebut adalah 4, 3, 2 atau 1 maka akan dilakukan pembulatan kebawah atau mengikuti nilai di sebelah kanan koma. Pembulatan ke atas akan dilakukan apabila angka di sebelah kiri koma adalah 5, 6, 7, 8, atau 9. Hasil pengujian ini akan menunjukkan nilai *accuracy*, *class precision*, *class recall* dan *AUC* pada setiap rasio yang telah ditentukan pada algoritma *Naive Bayes* dan *Support Vector Machine*.

Tabel 2. Hasil Implementasi Algoritma Naive Bayes dengan Feature Selection

Rasio	Data Training	Data Testing	Accuracy	Precision	Recall	AUC
9:1	764	85	88.24%	87.50%	98.44%	0.657
7:3	594	255	85.49%	87.86%	93.78%	0.611
5:5	425	424	74.76%	87.15%	78.19%	0.601
3:7	255	594	76.43%	89.95%	77.56%	0.667
1:9	85	764	74.08%	84.57%	80.48%	0.503
Rata-Rata			79.80%	87.41%	85.69%	0.608

Tabel 2 menunjukkan rekap hasil dari klasifikasi menggunakan algoritma *Naive Bayes*. Dapat dilihat hasil akurasi terbaik *Naive Bayes* dengan *polarity* manual ada pada rasio 9:1 dengan hasil akurasi 88.24% dan hasil rata-rata *accuracy* 79.80% , rata-rata *precision* 87.41%, hasil rata-rata *recall* 85.69% dan rata-rata *AUC* 0.608.

Tabel 3. Hasil Accuracy Algoritma Naive Bayes dengan Feature Selection Rasio 9:1

accuracy: 88.24%

	true POSITIF	true NEGATIF	class precision
pred. POSITIF	12	1	92.31%
pred. NEGATIF	9	63	87.50%
class recall	57.14%	98.44%	

Tabel 3 adalah hasil pengujian dari rasio data latih dan data uji 9:1 yang merupakan nilai akurasi paling tinggi pada algoritma Naive Bayes dengan Feature Selection. Nilai akurasi adalah hasil dari gabungan nilai class recall dan class precision yang didapat yaitu 88.24%. Nilai class precision 87.50% dan nilai class recall adalah 98.44%. Karena nilai recall lebih tinggi dari nilai precision berarti tingkat ketepatan sistem dalam membaca polarity lebih tinggi.



Gambar 10. Grafik AUC Algoritma Naive Bayes dengan Feature Selection Rasio 9:1

Gambar 10 adalah Kurva Receiver Operating Characteristic (ROC) hasil pengujian dari rasio data latih dan data uji 9:1 pada algoritma Naive Bayes dengan Feature Selection. Hasil pengujian diatas menghasilkan Area Under Curve (AUC) dengan nilai 0.657. Rasio 9:1 bukanlah rasio tertinggi untuk mendapatkan nilai AUC karena nilai AUC tertinggi ada pada rasio 3:7 dengan nilai 0.667.

Tabel 4. Hasil Implementasi Algoritma Support Vector Machine dengan Feature Selection

Rasio	Data Training	Data Testing	Accuracy	Precision	Recall	AUC
9:1	764	85	76.74%	76.83%	98.44%	0.769
7:3	594	255	78.43%	78.05%	99.48%	0.796
5:5	425	424	78.77%	78.38%	99.38%	0.770
3:7	255	594	77.78%	77.32%	100.00%	0.786
1:9	85	764	78.01%	77.58%	99.83%	0.747
Rata-Rata			77.95%	77.63%	99.43%	0.774

Tabel 4 menunjukkan rekap hasil dari klasifikasi menggunakan algoritma Support Vector Machine. Dapat dilihat hasil akurasi terbaik Support Vector Machine

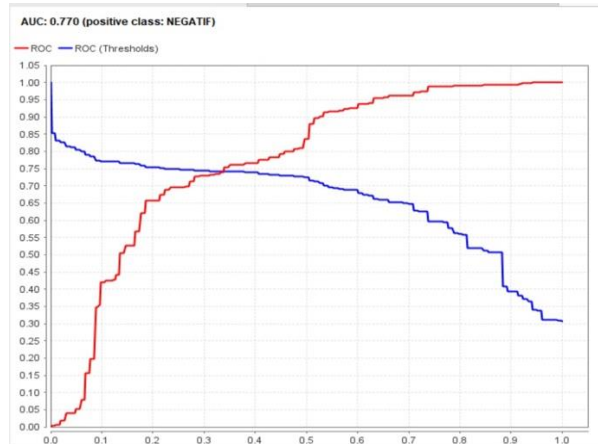
dengan polarity manual ada pada rasio 5:5 dengan hasil akurasi 78.77% dan hasil rata-rata accuracy 77.95% , rata-rata precision 77.63%, hasil rata-rata recall 99.43% dan rata-rata AUC 0.774.

Tabel 5. Hasil Accuracy Algoritma Support Vector Machine dengan Feature Selection Rasio 5:5

accuracy: 78.77%

	true POSITIF	true NEGATIF	class precision
pred. POSITIF	15	2	88.24%
pred. NEGATIF	88	319	78.38%
class recall	14.56%	99.38%	

Tabel 5 adalah hasil pengujian dari rasio data latih dan data uji 5:5 yang merupakan nilai akurasi paling tinggi pada algoritma Support Vector Machine dengan Feature Selection. Nilai akurasi adalah hasil dari gabungan nilai class recall dan class precision yang didapat yaitu 78.77%. Nilai class precision 78.38% dan nilai class recall adalah 99.38%. Karena nilai recall lebih tinggi dari nilai precision berarti tingkat ketepatan sistem dalam membaca polarity lebih tinggi.

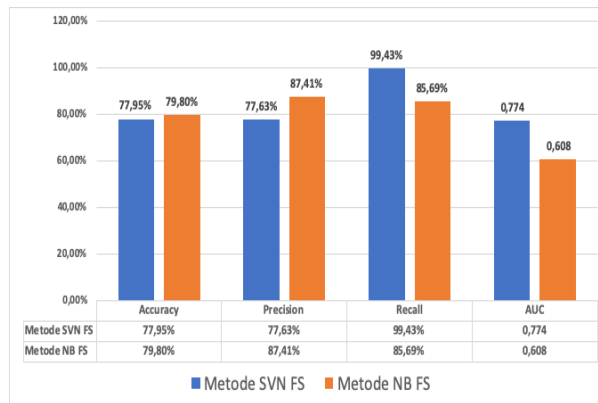


Gambar 11. Grafik AUC Algoritma Support Vector Machine dengan Feature Selection Rasio 5:5

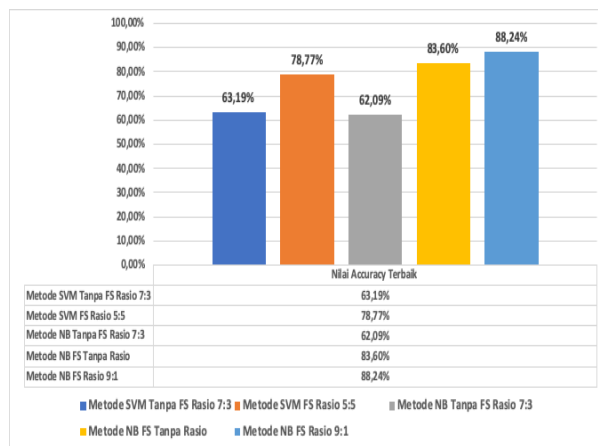
Gambar 11 adalah Kurva Receiver Operating Characteristic (ROC) hasil pengujian dari rasio data latih dan data uji 5:5 pada algoritma Support Vector Machine dengan Feature Selection. Hasil pengujian diatas menghasilkan Area Under Curve (AUC) dengan nilai 0.770. Rasio 5:5 bukanlah rasio tertinggi untuk mendapatkan nilai AUC karena nilai AUC tertinggi ada pada rasio 7:3 dengan nilai 0.796.

Gambar 12 adalah perbandingan rata-rata nilai algoritma Support Vector Machine dan Algoritma Naive Bayes. Setelah melakukan analisis menggunakan algoritma Naive Bayes dengan Feature Selection dan Support Vector Machine dengan Feature Selection pada hasil perbandingan nilai rata-rata dari algoritma, diperoleh nilai akurasi dan class precision tertinggi dimiliki oleh algoritma Naive Bayes dengan total rata-rata akurasi adalah 79.80% dan class precision 87.41% dan pada

recall dan AUC tertinggi dimiliki oleh algoritma *Support Vector Machine* dengan *recall* 99.43% dan AUC 0.774.



Gambar 12. Perbandingan Rata-Rata Nilai Algoritma *Support Vector Machine* dan Algoritma *Naive Bayes*



Gambar 13. Perbandingan Nilai *Accuracy* Terbaik pada Algoritma *Support Vector Machine* dan Algoritma *Naive Bayes*

Gambar 13 adalah perbandingan nilai *accuracy* terbaik pada algoritma *Support Vector Machine* dan algoritma *Naive Bayes* dengan penggunaan *feature selection* pada dan penggunaan perbandingan rasio antara data *training* dan data *testing*. Pada penelitian sebelumnya tentang analisis sentimen dan klasifikasi respon masyarakat terhadap pemindahan ibu kota negara ke Kalimantan Timur pada *twitter* tanpa menggunakan *Feature Selection* hasil pengujian akurasi dengan memakai perangkat lunak *Rapid Miner* dengan metode *Naive Bayes* menghasilkan nilai akurasi tertinggi sebesar 62,09% dari rasio 7:3. Hasil pengujian akurasi dengan memakai perangkat lunak *Rapid Miner* dengan metode *Support Vector Machine* menghasilkan nilai akurasi tertinggi sebesar 63,19% dari rasio 7:3.

Penelitian pada analisis sentimen *E-Wallet* pada *Google Play* menggunakan algoritma *Naive Bayes* berbasis *Particle Swarm Optimization* didapatkan nilai akurasi tanpa *feature selection* 82.30% dan setelah ditambahkan penggunaan *feature selection* menjadi 83.60%.

Pada penelitian ini dengan mengkombinasikan antara penggunaan *Feature Selection* dan penggunaan

perbandingan rasio antara data *training* dan data *testing* 9:1; 7:3; 5:5; 3:7; dan 1:9 mendapatkan hasil nilai akurasi tertinggi pada algoritma *Support Vector Machine* dengan *Feature Selection* sebesar 78,77% untuk rasio 5:5 dan algoritma *Naive Bayes* dengan *Feature Selection* sebesar 88.24% pada rasio 9:1.

4. Kesimpulan

Hasil penelitian analisis sentimen pemindahan ibu kota negara pada pengguna *twitter* dengan *feature selection* algoritma *Naive Bayes* dan *Support Vector Machine* dengan *Rapid Miner* didapatkan nilai akurasi terbaik algoritma *Naive Bayes* dengan *feature selection* ada pada rasio 9:1 dengan hasil akurasi 88.24% sedangkan hasil akurasi terbaik algoritma *Support Vector Machine* dengan *Feature Selection* ada pada rasio 5:5 dengan hasil akurasi 78.77%. Nilai akurasi algoritma *Naive Bayes* dengan *feature selection* lebih tinggi 9.47 % dibandingkan akurasi algoritma *Support Vector Machine* dengan *feature selection*. Hasil rata-rata akurasi tertinggi diperoleh algoritma *Naive Bayes* dengan *feature selection* dengan hasil rata-rata akurasi 79.80%, lebih tinggi 1.85% dari algoritma *Support Vector Machine* dengan *feature selection*. Tidak selalu ada korelasi positif antara tingginya nilai akurasi dengan nilai *Area Under Curve* (AUC).

Penggunaan *feature selection* pada algoritma *Support Vector Machine* dan algoritma *Naive Bayes* ditambah dengan penggunaan perbandingan rasio antara data *training* dan data *testing* akan meningkatkan nilai akurasi yang lebih baik. Penelitian sebelumnya pada algoritma *Support Vector Machine* tanpa *feature selection* nilai akurasi sebesar 63,19% pada rasio 7:3 setelah menggunakan *feature selection* nilai akurasi sebesar 78.77% pada rasio 5:5 sedangkan untuk algoritma *Naive Bayes* tanpa *feature selection* nilai akurasi sebesar 62,09% pada rasio 7:3 setelah menggunakan *feature selection* nilai akurasi sebesar 88.24% pada rasio 9:1.

Mayoritas respon masyarakat terkait pemindahan ibu kota negara adalah tidak setuju dengan prosentase perolehan sentimen negatif 75,74% dan perolehan sentimen positif 24,26%.

Untuk penelitian selanjutnya dapat disarankan dengan membandingkan rasio pengambilan data teratas *regular attributes* pada penggunaan *feature selection weight by correlation* atau dapat juga dengan penggunaan *feature selection weight* lainnya selain *feature selection weight by correlation*. Kemudian pengambilan data juga dapat dilakukan dengan lebih banyak jumlahnya untuk mendapatkan nilai akurasi lebih baik.

Daftar Rujukan

- [1] Kusuma, Hendra (26 August 2019). "Resmi! Jokowi Putuskan Ibu Kota RI Pindah ke Kaltim". Diakses pada 25 April 2020 dari <https://finance.detik.com/properti/d-4681152/resmi-jokowi-putuskan-ibu-kota-ri-pindah-ke-kaltim>.

- [2] Pratomo, Harwanto Bimo (2019, Mei 5). 5 Pro dan Kontra Rencana Pemindahan Ibu Kota Presiden Jokowi. Diakses pada 18 Juni 2020, dari <https://www.merdeka.com/uang/5-pro-dan-kontra-rencana-pemindahan-ibu-kota-presiden-jokowi.html>.
- [3] Tyagi, P., & Tripathi, R. C., 2019, A Review towards the *Sentiment Analysis Techniques for the Analysis of Twitter Data*. International Conference on Advanced Computing and Software Engineering 2019, Sultanpur, February 8-9.
- [4] Pamungkas, D. S., Setiyanto, N. A., & Dolphina, E., 2015, Analisis *Sentiment* pada Sosial Media *Twitter* Menggunakan Naïve Bayes Classifier terhadap Kata Kunci “Kurikulum 2013”. *Techno. Com*, No.4, Vol.14, 299-314.
- [5] Kirilenko, A. P., Stepchenkova, S. O., Kim, H., & Li, X., 2018, Automated *sentiment* analysis in tourism: Comparison of approaches. *Journal of Travel Research*, No.8, Vol.57, 1012-1025.
- [6] Liu, Y., Huang, X., An, A., & Yu, X., 2007, ARSA: A *Sentiment*Aware Model for Predicting Sales Performance Using Blogs. In Proceedings of the 30th annual international ACM SIGIR conference on Research and development in information retrieval, New York.
- [7] Taboada, M., Brooke, J., Tofiloski, M., Voll, K. dan Stede, M. (2011). ‘Lexicon- based methods for *sentiment* analysis
- [8] Hidayat, A. N., 2015, Analisis Sentimen Terhadap Wacana Politik Pada Media Masa Online Menggunakan Algoritma *Support Vector Machine* Dan Naive Bayes. *Jurnal Elektronik Sistem Informasi dan Komputer*, No.1, Vol.1, 12-18.
- [9] Ahmad Saleh, 2015, "Klasifikasi Gejala Depresi Pada Manusia dengan Metode *Naive Bayes* Menggunakan Java", Penerbit Andi, Yogyakarta.
- [10] Fiarni, C., Maharani, H., & Pratama, R., 2016, *Sentiment* analysis system for Indonesia online retail shop review using hierarchy *Naive Bayes* technique. In 2016 4th international conference on information and communication Technology, Bandung, May 25-27.
- [11] Buntoro, G.A., 2017, Analisis Sentimen Calon Gubernur DKI Jakarta 2017 Di *Twitter*. *INTEGER: Journal of Information Technology*, No.1, Vol.2.
- [12] Anto Satriyo Nugroho, A. B. (2003). *Support Vector Machine*. Application of *Support Vector Machine* in Bioinformatics.
- [13] William S. Sanders, 2011. "Prediction of Cell Penetrating Peptides by Support Vector Machines", doi: 10.1371/journal.pcbi.1002101.
- [14] Indrayuni, Elly. 2016. Analisa Sentimen Review Hotel Menggunakan Algoritma *Support Vector Machine* Berbasis Particle Swarm Optimization, ISSN: 2338 – 8161. Pontianak : Jurnal Evolusi Volume 4 Nomor 2 – 2016.
- [15] Achyani, Yuni Eka. 2018, Penerapan Metode Particle Swarm Optimization Pada Optimasi Prediksi Pemasaran Langsung, *Jurnal Informatika*, Vol.5 No.1 April 2018, pp. 1~11
- [16] Pamungkas, D. S., Setiyanto, N. A., & Dolphina, E., 2015, Analisis Sentimen pada Sosial Media *Twitter* Menggunakan Naïve Bayes Classifier terhadap Kata Kunci “Kurikulum 2013”. *Techno. Com*, No.4, Vol.14, 299-314.
- [17] Priyono, F., Kanti, S., Dzulfiqar, I., Amirulloh, I., Alvi, A., & Rosiyadi, D., 2016, Analisis Sentimen Media Sosial Opini Ujian Nasional Berbasis Komputer menggunakan Metoda Naive Bayes. *Journal of Electrical And Electronics Engineering*, No.2, Vol.1.
- [18] Mihuandayani, M., Feriyanto, E., Syarham, S., & Kusriani, K., 2018, Opinion Mining pada Komentas *Twitter* E-KTP Menggunakan Naïve Bayes Classifier. *Semnasteknomedia Online*, No.1, Vol.6, 1-2.
- [19] Anggelina, Shella Yuni (2020) Analisis Sentimen Dan Klasifikasi Respon Masyarakat Terhadap Pemindahan Ibu Kota Negara Ke Kalimantan Timur Pada *Twitter*. S1 Sistem Informasi thesis, STMIK Widya Cipta Dharma.
- [20] Amar P. Natasuwarna, 2019, Analisis Sentimen Keputusan Pemindahan Ibukota Negara Menggunakan Klasifikasi Naive Bayes, Seminar Nasional Sistem Informasi dan Tehnik Informasika (2019).
- [21] Santoso, Imam., Gata, Windu., Paryanti, Atik Budi, 2019, Penggunaan *Feature Selection* di Algoritma *Support Vector Machine* untuk Sentimen Analisis Komisi Pemilihan Umum, *Jurnal Resti (Rekayasa Sistem dan Teknologi Informasi)* Vol. 3 No. 3 (2019) 364 – 370.
- [22] Saputra, Suwanda Aditya., Rosiyadi, Didi., Gata, Windu., Husain, Syepri Maulana., 2019, Analisis Sentimen E-Wallet Pada Google Play Menggunakan Algoritma *Naive Bayes* Berbasis Particle Swarm Optimization, *Jurnal Resti (Rekayasa Sistem dan Teknologi Informasi)* Vol. 3 No. 3 (2019) 377 – 382.