

Analisis Sentimen Pemindahan Ibu Kota Pada Twitter Dengan Metode Support Vector Machine

(Center, Bold, Arial 16, maximum 12 words in English, uppercase the first letter)

Tezza Fazar Tri Hidayat^{a1}, Garno^{a2}, dan Azhari Ali Ridha^{b3}
(Center, Arial 10, First name Middle name Last name)

^aTeknik Informatika Fakultas Ilmu Komputer Universitas Singaperbangsa Karawang
Jl. HS. Ronggowaluyo Telukjambe Timur, Karawang, Jawa Barat, Indonesia.

¹1610631170215@student.unsika.ac.id

²garno@staff.unsika.ac.id

(Center, Arial 10, ^aif the authors from the same affiliation)

^bTeknik Informatika Fakultas Ilmu Komputer Universitas Singaperbangsa Karawang
Jl. HS. Ronggowaluyo Telukjambe Timur, Karawang, Jawa Barat, Indonesia.

³azhari.ali@unsika.ac.id

Abstrak

(Center, Bold, Italic, Arial 10)

Relokasi ibu kota Indonesia kini telah diresmikan oleh Presiden Joko Widodo pada 26 Agustus 2019 ke Kalimantan, ini adalah sejarah baru dalam sejarah Indonesia karena belum pernah terjadi sebelumnya, sehingga memunculkan banyak pendapat atau tanggapan dari masyarakat. Analisis sentimen adalah kegiatan yang digunakan untuk menganalisis pendapat atau opini seseorang tentang suatu topik. Twitter adalah media sosial yang digunakan untuk mengekspresikan pendapat pengguna dan menyatukannya pada suatu topik. Support Vector Machine adalah metode text mining yang mencakup metode klasifikasi dan Term Frequency - Inverse Document Frequency adalah metode pembobotan karakter. SVM dan TF-IDF dapat digunakan untuk menganalisis sentimen opini publik tentang topik pemindahan ibukota Indonesia. Tujuan dari penelitian ini adalah untuk mengklasifikasikan opini publik tentang topik memindahkan Ibu Kota Indonesia dari ribuan tweet yang telah dikumpulkan dan disaring. Tweet pada dari 22-29 Maret 2020 telah diproses menjadi 992 tweet dan terdiri dari 221 data dengan label positif dan 771 data negatif. Dan menggunakan metode SVM yang memiliki akurasi 77,72% dan dikombinasikan dengan TFIDF yang meningkatkan akurasinya menjadi 78,33%.

(Justify, Arial 10)

Kata Kunci: Sentimen Analisis, SVM, TF-IDF, Twitter.

(Minimum 5 keywords related to the content and separated by comma, italic)

Abstract

The relocation of the capital city of Indonesia has now been inaugurated by President Joko Widodo on August 26, 2019, to Kalimantan, it is a new history in Indonesian history because it has never happened before, giving rise to many opinions or responses from the public. Sentiment analysis is an activity used to analyze a person's opinions or opinions on a topic. Twitter is a social media used to express user's opinions and united them on a topic. Support Vector Machine is a method of text mining which includes classification methods and Term Frequency - Inverse Document Frequency is a method of weighting characters. SVM and TF-IDF can be used to analyze public opinion sentiments on the topic of moving the Indonesian capital. The purpose of this study is to classify public opinion on the topic of moving the Indonesian capital from thousands of tweets that have been collected and filtered. Tweets that have been processed are 992 tweets from March 22-29, 2020 and, consist of 221 data with positive labels and 771 negative data. Moreover, using the SVM method, which has an accuracy of 77.72% and combined with TFIDF, which increases its accuracy to 78.33%.

Keywords: Sentiment Analysis, Support Vector Machine, Term Frequency – Inverse Document Frequency, Twitter.

1. Pendahuluan

Penggunaan internet di Indonesia tiap tahun selalu meningkat, di mana pada tahun 2018 sebanyak 79,13% di antaranya digunakan untuk mengakses medial sosial seperti Facebook, Instagram, dan Twitter [1]. Twitter merupakan salah satu media sosial terbesar di Indonesia yang banyak digunakan masyarakat Indonesia untuk berbagai keperluan seperti curhat, beropini, atau mengobrol bersama teman. Medial sosial memiliki pengaruh yang besar terhadap penggunaanya yang dapat menentukan kesuksesan atau kehancuran hidup seseorang [2].

Adanya fitur trending topic pada twitter memudahkan untuk mengetahui topik apa yang sedang ramai dibicarakan oleh masyarakat twitter. Salah satunya pemindahan ibu kota, berita pemindahan ibu kota Indonesia dari Jakarta ke pulau Kalimantan yang diresmikan oleh Presiden Joko Widodo pada 26 Agustus 2019 [3]. Berita tersebut membuat heboh masyarakat Indonesia terutama pengguna twitter sehingga menjadi perbincangan dan menjadi trending topic yang di dalamnya banyak yang mengeluarkan pendapat dan opini mereka tentang keputusan tersebut baik opini positif maupun opini negatif. Banyak penelitian yang dilakukan menggunakan twitter untuk mencari data yang didapatkan dari twitter salah satunya adalah *text mining* atau penambangan teks yang digunakan untuk menganalisis opini yang terdapat dalam sebuah topik di twitter. Analisis sentimen atau biasa disebut dengan *opinion mining* merupakan bagian dari penelitian *text mining* yang melakukan proses memahami, mengekstrak dan mengolah data tekstual secara otomatis untuk mendapatkan informasi sentimen yang terkandung dalam sebuah kalimat opini [2].

Banyak metode *text mining* yang dapat digunakan untuk melakukan analisis sentimen di antaranya adalah *Lexicon Based*, *Support Vector Machine*, *K-Nearest Neighbor*, dan *Naïve Bayes Classifier*. Dari banyak metode *text mining* dipilih metode *Support Vector Machine* karena memiliki nilai akurasi yang cukup tinggi dibanding metode yang lain, seperti pada penelitian Rosiyadi (2019) yang membandingkan Kinerja *Algoritma K-Nearest Neighbor*, *Naïve Bayes Classifier* dan *Support Vector Machine* [4], dengan nilai akurasi tertinggi didapatkan dari metode *Support Vector Machine* sebesar 81,58% sedangkan untuk metode *K-Nearest Neighbor*, *Naïve Bayes Classifier* mendapatkan akurasi sebesar 81,32%, dan *Naïve Bayes Classifier* sebesar 78,95%. Lalu pada penelitian selanjutnya yaitu tentang Klasifikasi SMS Spam Menggunakan *Support Vector Machine* oleh Setiyono dan Pardede (2019) [5], mendapatkan hasil akurasi 98,33% dengan menggunakan metode SVM, dan selanjutnya penelitian tentang Perbandingan Akurasi dan Waktu Proses Algoritma K-NN dan SVM dalam Analisis Sentimen Twitter oleh Rangga, Nasution, dan Hayaty (2019) [6], yang mendapatkan nilai akurasi tertinggi oleh SVM sebesar 89,70%. Dari ketiga penelitian di atas dapat disimpulkan bahwa metode SVM memiliki nilai akurasi yang paling tinggi. Selain itu pada penelitian

Selain itu dalam penelitian Honakan, Adiwijaya, dan Al-Faraby (2018) mengungkapkan bahwa pembobotan karakter (*term weight*) dapat meningkatkan akurasi dengan proses pelatihan dan penggunaan fungsi klasifikasi lebih efisien [7]. Sehingga digunakan metode TF-IDF (*Term Frequency- Invers Document Frequency*) untuk meningkatkan akurasi.

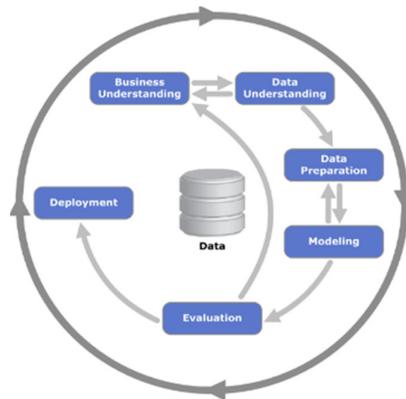
Tujuan dari penelitian ini adalah untuk mengetahui pola opini teks pada sentimen masyarakat terhadap pemindahan ibu kota di twitter dengan menggunakan SVM dan untuk mengetahui hasil peningkatan akurasi menggunakan TF-IDF.

2. Metode Penelitian

2.1. CRISP-DM (Cross-Industry Standard Process for Data Mining)

Metode penelitian yang digunakan untuk menyusun penelitian ini adalah dengan menggunakan CRISP-DM (*Cross-Industry Standard Process for Data Mining*) yang terdiri dari 6 tahapan, antara lain *Business Understanding* (Pemahaman Bisnis), *Data Understanding* (Pemahaman Data), *Data Preparation* (Pengelolaan Data), *Modeling* (Pemodelan), *Evaluation* (Evaluasi), dan *Deployment* (Penyebaran).

Menurut Larose, dalam Setiawan (2016) mengemukakan bahwa terdapat enam fase yang terdapat pada metode CRISP-DM, yaitu sebagai berikut [8]:

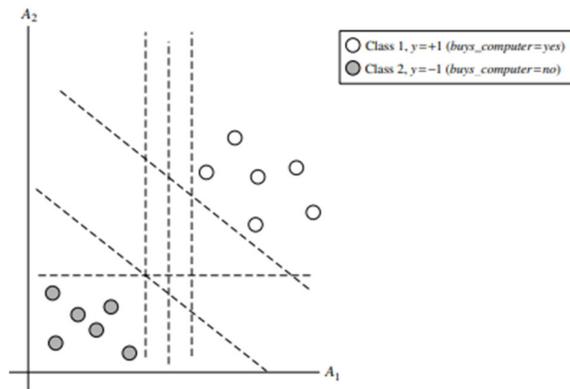


Gambar 1. Merupakan gambar tahapan metode penelitian CRISP-DM.

1. *Business Understanding* (Pemahaman Bisnis) merupakan fase pertama pada metode CRISP-DM. Pada fase ini dibutuhkan pemahaman tentang substansi dari kegiatan data mining yang akan dilakukan, kebutuhan secara detail dalam lingkup bisnis atau unit penelitian secara keseluruhan. Kegiatan yang dilakukan antara lain: menentukan sasaran atau tujuan dan batasan menjadi formula dari permasalahan data mining dan membuat perencanaan strategi serta jadwal penelitian [9].
2. *Data Understanding* (Pemahaman Data) adalah fase mengumpulkan data awal, mempelajari data un-tuk bisa mengenal data yang akan dipakai. Fase ini mencoba untuk mendapatkan pemahaman yang men-dalam tentang data, mengidentifikasi masalah kualitas data, atau untuk mendeteksi adanya bagian yang menarik dari data yang dapat digunakan untuk hipotesa untuk informasi yang tersembunyi [8].
3. *Data Preparation* (Pengolahan Data) adalah fase yang meliputi semua kegiatan untuk membangun dataset akhir (data yang akan diproses pada tahap pemodelan/modeling) dari data mentah. Pada tahap yang ini yaitu mempersiapkan data untuk melakukan langkah-langkah yang disebut dengan text preprocessing [10].
4. *Modeling* (Pemodelan) adalah fase menentukan teknik data mining yang digunakan, menentukan tools data mining, teknik data mining, algoritma data mining, menentukan parameter dengan nilai yang optimal [9].
5. *Evaluation* (Evaluasi) merupakan fase kelima dalam metode CRISP-DM yang di mana pada fase ini dilakukan evaluasi terhadap model yang telah dibentuk pada fase sebelumnya. Fase ini dilakukan secara mendalam dengan tujuan untuk menentukan apakah model dapat mencapai tujuan yang ditetapkan pada fase awal (*Business Understanding*) [12].
6. *Deployment* (Penyebaran). Pada fase ini, semua pengetahuan dan informasi yang telah diperoleh dari fase-fase sebelumnya akan disampaikan dan dipresentasikan dalam bentuk khusus sehingga dapat digunakan oleh pengguna. Tahap deployment dapat berupa pembuatan laporan sederhana atau mengimplementasikan proses data mining yang berulang dalam perusahaan [8].

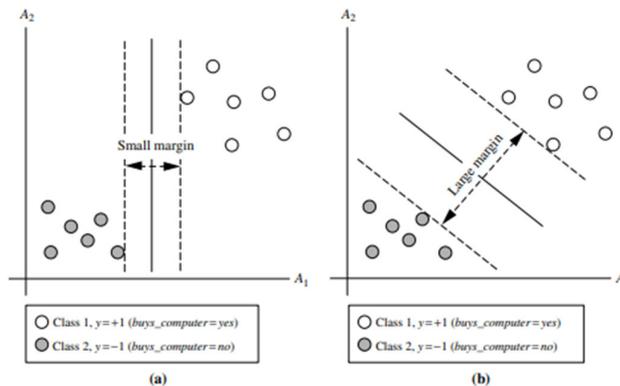
2.2. SVM (*Support Vector Machine*)

Support Vector Machine merupakan salah satu algoritma klasifikasi yang masuk kelas *supervised learning* yang pertama kali diperkenalkan oleh Vapnik pada tahun 1992 sebagai rangkaian harmonis konsep-konsep unggulan dalam bidang *pattern recognition* [11]. *Supervised learning*, merupakan jenis kelas yang di mana dalam implementasinya perlu adanya tahap pelatihan dan pengenalan terhadap suatu objek yang akan dianalisis serta disusul tahap pengujian. Konsep dari SVM adalah dengan mencari *hyperplane* terbaik yang berfungsi sebagai pemisah dua kelas data.



Gambar 2. SVM yang Memisahkan Dua Data dengan *Hyperplane* (Han, 2014:409)

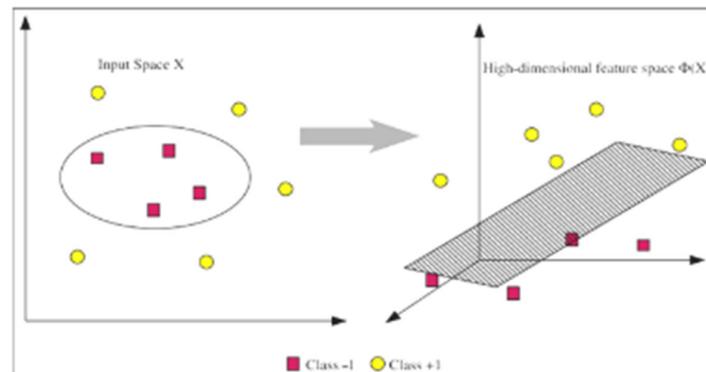
Pada Gambar 2 diatas terdapat 2 buah kelas +1 yang mewakili setuju membeli komputer, dan class -1 tidak setuju membeli komputer, dan memiliki 2 atribut input yaitu A1 dan A2. Dalam gambar diatas bisa dilihat bahwa data 2 dimensi dapat dipisah secara linear, karena bisa menarik garis lurus untuk memisahkan semua data tuple kelas +1 dari semua tuple kelas -1, garis hitam ditengah tersebut merupakan *hyperplane* (Gambar 3) [12]. *Hyperplane* merupakan garis yang dibentuk sebagai pemisah terbaik antara kedua class -1 dan +1 dan dapat dimaksimalkan dengan mengukur *margin* dan mencari titik maksimalnya. *Margin* (garis hitam putus-putus) adalah jarak antara *hyperplane* tersebut dengan tuple terdekat dari masing-masing *class* (Gambar 3) [12]. *Tuple* yang paling dekat ini disebut sebagai *support vector*.



Gambar 3. Dua *Hyperplane* di Satu Data (Han, 2014:410)

Karena banyaknya garis pemisah maka perlu di cari *hyperplane* terbaik, SVM mendekati masalah ini dengan mencari *Maximum Marginal Hyperplane* (MMH). Kedua *hyperplane* dapat dengan benar mengklasifikasikan semua data tuple yang diberikan pada Gambar 3 merupakan gambar jika *hyperplane* digambar terpisah.

Permasalahan klasifikasi pada kenyataannya bersifat non linear. Untuk menyelesaikan problem non linear, SVM dimodifikasi dengan memasukkan fungsi *Kernel*. Konsep dari *Kernel Trick* dalam SVM adalah dengan membuat dimensi baru, jadi data akan dimasukan kedalam dimensi baru dan kemudian diberikan *hyperplane* untuk memisahkan data tersebut. [11].



Gambar 4. Dua Hyperplane di Satu Data (Sosilowati, 2015)

Dengan banyaknya jenis *kernel* pada *Support Vector Machine* tentu tidak semuanya baik dalam menyelesaikan masalah klasifikasi, karena setiap fungsi *kernel* memiliki fungsi yang berbeda tergantung *dataset* yang digunakan. Pada penelitian ini menggunakan *kernel dot* yang dimana *kernel* ini merupakan *kernel bawaan* atau *default* pada *Support Vector Machine* di RapidMiner yang memiliki akurasi tertinggi dibanding *kernel* yang lainnya.

2.3. TF-IDF (Term Frequency-Inverse Document Frequency)

Pembobotan kata (*term weight*) memiliki tujuan untuk memberikan bobot setiap kata (*term*) yang terdapat pada dokumen teks. Salah satu metode populer untuk melakukan pembobotan kata adalah TF-IDF (*Term Frequency-Inverse Document Frequency*) [13].

Berikut merupakan beberapa metode pembobotan menggunakan TF-IDF [13].

1. *Term Frequency (TF)* merupakan frekuensi kemunculan kata pada suatu dokumen teks. TF ($f_{t,d}$) didefinisikan jumlah kemunculan *term* t pada dokumen d . Berikut merupakan persamaan TF :

$$TF_{t,d} = f_{(t,d)} \tag{1}$$

Dengan :

TF = Frekuensi kemunculan kata pada suatu dokumen teks.

f = Jumlah kata pada suatu dokumen.

2. *Invers Document Frequency (IDF)* merupakan frekuensi kemunculan *term* pada keseluruhan dokumen teks. *Term* yang jarang muncul pada keseluruhan dokumen teks memiliki nilai *Invers Document Frequency* lebih besar dibandingkan dengan *term* yang sering muncul. Berikut merupakan persamaan IDF :

$$IDF = \log \frac{N}{DF_t} \tag{2}$$

Dengan :

N = Jumlah dokumen

DF_t = Nilai TF

3. *Term Frequency – Inverse Document Frequency (TF-IDF)* merupakan perhitungan untuk pembobotan. Setelah TF dan DF didapatkan dengan persamaan (1) dan (2) maka selanjutnya mencari TF – IDF. Berikut merupakan persamaan TF – IDF :

$$W_t = TF \times IDF \tag{3}$$

Dengan :

W_t = TF-IDF

TF = Nilai TF

IDF = Nilai IDF

Berikut merupakan contoh pembobotan TF-IDF, perhatian Tabel 2.1 yang berisi contoh 3 buah *dataset* [14].

Tabel 1. Contoh Pembobotan TF-IDF (*Dataset*)

Data 1	Password adalah sekumpulan karakter
Data 2	adalah sekumpulan karakter yang melindungi suatu software
Data 3	Password generator adalah software yang dapat menyusun password

Selanjutnya data pada Tabel 1 dimasukan kedalam Tabel 2 dengan memecahnya menjadi kalimat serta mengabaikan kata yang tidak berarti seperti yang, adalah, dan untuk.

Tabel 2. Contoh Pembobotan TF-IDF (Perhitungan)

Term	TF			df	IDF = $\log n/df_t$			TFIDF $w = tf \times idf$		
	1	2	3		1	2	3	1	2	3
password	1	0	2	2	0,1761	0	0,3522	0,1761	0	0,7044
kumpul	1	1	0	2	0,1761	0,1761	0	0,1761	0,1761	0
karakter	1	1	0	2	0,1761	0,1761	0	0,1761	0,1761	0
lindungi	0	1	0	1	0	0,4771	0	0	0,4771	0
suatu	0	1	0	1	0	0,4771	0	0	0,4771	0
software	0	1	1	2	0	0,1761	0,1761	0	0,1761	0,1761
generator	0	0	1	1	0	0	0,4771	0	0	0,4771
dapat	0	0	1	1	0	0	0,4771	0	0	0,4771
nyusun	0	0	1	1	0	0	0,4771	0	0	0,4771

Pada Tabel 2 merupakan tabel perhitungan TF-IDF, dimana terdapat 3 data (Tabel 1) yang kemudian setiap kata-kata pada Tabel 1 tersebut di pisah, dan di masukan kedalam kolom *term*, kata yang dimasukan adalah kata-kata yang berbeda di setiap kalimat pada 3 data (Tabel 1). Pada kolom TF terdapat 3 kolom yang mewakili 3 kalimat, dimana setiap nilai yang ada mewakili setiap jumlah kata disetiap kalimat. Selanjutnya mengikuti rumus (2) dan selanjutnya (3). Dengan menggunakan metode TF-IDF maka setiap dataset akan dipecah dan diberi nilai setiap katanya, yang biasa disebut *term weight*, hal itulah yang menjadikan nilai akurasi dari SVM meningkat.

3. Hasil dan Pembahasan

3.1. Hasil

Hasil penelitian yang sedang dilakukan adalah bagaimana menganalisis algoritma Support Vector Machine dapat mengklasifikasi opini masyarakat terhadap proses pemindahan ibu kota di Indonesia serta peningkatan akurasi menggunakan pembobotan karakter Term Frequency – Inverse Document Frequency.

1. Business Understanding (Pemahaman Bisnis)

Hasil dari tahapan yang pertama ini meliputi menentukan tujuan bisnis dan menentukan tujuan data mining :

- Menentukan Tujuan Bisnis

Tujuan dilakukannya penelitian mengenai analisis sentimen pemindahan ibu kota untuk mengetahui pola opini masyarakat terhadap pemindahan ibu kota pada twitter. Dimana pada tanggal 26 Agustus 2019 membuat masyarakat banyak beropini mengenai pemindahan ibu kota di Indonesia, baik yang setuju atau pun yang tidak setuju semuanya mengeluarkan pendapatnya di twitter sampai membuat topik pemindahan ibu kota tersebut trending topic atau menjadi berita yang terus ramai di perbincangan. Sampai pada tahun 2020 opini masyarakat terus berlanjut ditambahnya dengan usulan lockdown karena adanya pandemi Covid19 yang menganggap pemerintah lebih mementingkan pemindahan ibu kota di banding keselamatan masyarakat karena tidak ada tindakan *lockdown*.

- Menentukan Tujuan Data Mining

Tujuan data mining pada penelitian ini adalah untuk mengelompokkan opini masyarakat menjadi 2 kelompok yaitu positif dan negatif. Menggunakan metode data mining klasifikasi untuk untuk mengelompokkan opini tersebut.

2. Data Understanding (Pemahaman Data)

Hasil dari data understanding (pemahaman data) meliputi mengumpulkan data awal, mendeskripsikan data, dan memverifikasi kualitas data.

- Mengumpulkan Data Awal

Hasil dari pengumpulan data mulai studi pustaka, observasi, dan mengumpulkan data dari twitter. Studi pustaka meliputi pencarian materi mengenai text mining dari jurnal, internet, buku, skripsi dan lain sebagainya yang relevan. Observasi dilakukan di media sosial twitter dengan mempelajari cara kerja dan penggunaan twitter serta pencarian trending topic dan pencarian kata kunci yang tepat untuk menggali data. Kata kunci yang digunakan untuk menumpulkan dataset di twitter adalah “pindah ibu kota” yang diambil dari tanggal 22-29 Maret 2020 bertepatan dengan ramainya pandemi Covid19 dan isu *lockdown*. Data diambil dari twitter dengan menggunakan API twitter yang diproses oleh RapidMiner.

- Mendeskripsikan Data

Data yang didapatkan cukup banyak yaitu 2500 data lebih tetapi banyak data yang sama atau duplikat sehingga data langsung di filter menjadi 998 data. Sehingga data yang diambil dari twitter pada tanggal 22-29 Maret 2020 sebanyak 998 data.

- Memverifikasi Kualitas Data

Data yang telah di filter menjadi 998 data tersebut diambil dan di verifikasi oleh guru Bahasa Indonesia untuk di berikan label positif dan negatif serta masuk kedalam kata kunci atau kategori yang di cari atau tidak, jika tidak berhubungan dengan kata kunci “pindah ibu kota” maka data akan di buang sehingga data menjadi 992 yang terdiri dari 221 data Positif dan 771 data Negatif.

3. Data Preparation (Pengelolaan Data)

Data preparation atau pengelolaan data merupakan tahapan selanjutnya, pada tahapan ini dilakukan langkah-langkah :

- Membersihkan Data

Membersihkan data dilakukan dengan menggunakan tahapan *pre processing* yang terdiri dari *Cleaning, Case Folding, Tokenizing, Stopword Removal, dan Stemming*. Dimana dalam penggunaan *Stopword Removal* adalah dengan membuat manual daftar kata-kata yang ada dalam dataset dalam bentuk file notepad atau .txt yang kemudian di masukan kedalam Rapidminer yang selanjutnya membuat kalimat di pecah menjadi kata per kata. Sedangkan untuk melakukan *stemming* menggunakan fitur *regular expression* untuk membuat kata menjadi kata dasar. Dan hal tersebut tidak akan dibahas dalam penelitian ini.

Tabel 3. Tabel Perbandingan Sebelum dan Sesudah *Pre-processing*

Text Pre-processing	Sebelum	Sesudah
Cleaning	Hemmmmm usul lock down nasional, anggaran pindah ibu kota sementara ditunda dialihkan utk penanganan wabah covid19 ?? https://t.co/6YH7NNLi6z	Hemmmmm usul lock down nasional, anggaran pindah ibu kota sementara ditunda dialihkan utk penanganan wabah covid19 ??
Case Folding	Hemmmmm usul lock down nasional, anggaran pindah ibu kota sementara ditunda dialihkan utk penanganan wabah covid19 ??	hemmmmm usul lock down nasional, anggaran pindah ibu kota sementara ditunda dialihkan utk penanganan wabah covid19 ??

Tokenizing	hemmmmm usul lock down nasional anggaran pindah ibu kota sementara ditunda dialihkan utk penanganan wabah covid19 ??	hemmmmm usul lock down nasional anggaran pindah ibu kota sementara ditunda dialihkan utk penanganan wabah covid
Stopword Removal	hemmmmm usul lock down nasional anggaran pindah ibu kota sementara ditunda dialihkan utk penanganan wabah covid	“usul”, “lock”, “down”, “nasional”, “anggaran”, “pindah”, “ibu”, “kota”, “sementara”, “ditunda”, “dialihkan”. “penanganan”, “wabah”, “covid”
Stemming	“usul”, “lock”, “down”, “nasional”, “anggaran”, “pindah”, “ibu”, “kota”, “sementara”, “ditunda”, “dialihkan”. “penanganan”, “wabah”, “covid”	usul lock down nasional anggar pindah ibu kota sementara tunda alih tangan wabah covid

- Format Data

Data yang telah dibersihkan selanjutnya diubah di bobotkan menggunakan metode *Term Frequency – Inverse Document Frequency* sebelum melanjutkan ke proses selanjutnya. Pada Gambar 5 terdapat operator *Generate TF-IDF* yang digunakan untuk membobotkan karakter yang telah di *pre processing*. Operator *Nominal To Text* pada Gambar 4 dan 5 digunakan untuk mengubah jenis atribut nominal yang dipilih menjadi teks. Operator ini juga memetakan semua nilai dari atribut ini ke nilai *string* yang sesuai. Data teks diproses untuk mengekstrak kata-kata dan mengirimkan kata.

4. Modeling (Pemodelan)

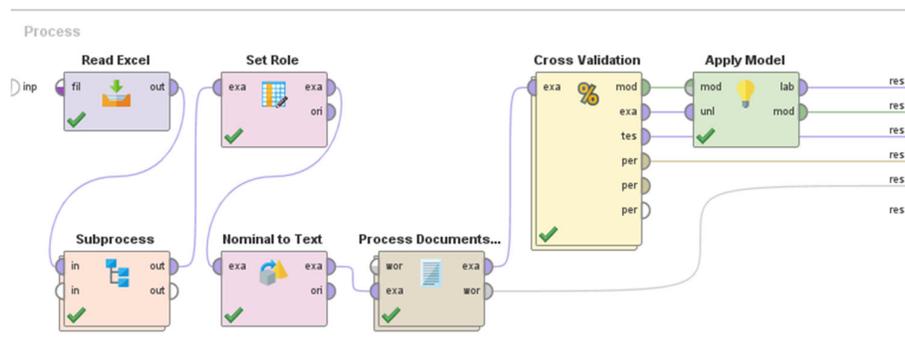
Tahap modeling terdiri dari kegiatan :

- Memilih Teknik Pemodelan

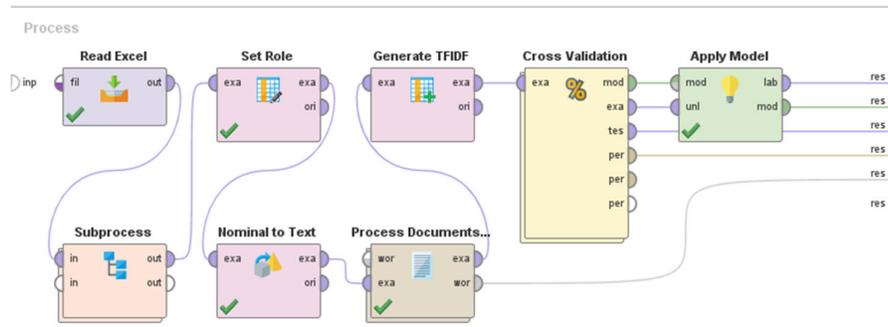
Pemodelan dilakukan dengan menggunakan algoritma klasifikasi *text mining* yaitu *Support Vector Machine* yang digunakan untuk mengklasifikasi opini masyarakat menjadi positif dan negatif pada topik pemindahan ibu kota di *twitter*.

- Membangun Model

Terdapat dua model yang dibangun yaitu model *Support Vector Machine* pada Gambar 5 dan model *Support Vector Machine* dengan pembobotan *Term Frequency - Inverse Document Frequency* pada Gambar 6. SVM bekerja dengan mencari hyperplane atau garis pemisah antara dua data yang dalam penelitian ini adalah positif dan negatif. Digunakannya SVM dan TF-IDF dalam penelitian ini adalah untuk meningkatkan nilai akurasi yang dihasilkan oleh metode SVM dimana dengan menggunakan TF-IDF setiap dataset yang berbentuk kalimat akan di pecah menjadi kata per kata yang kemudian setiap kata yang ada akan di beri bobot (*term weight*) secara otomatis oleh Rapidminer.



Gambar 5. Merupakan gambar model proses menggunakan metode SVM.



Gambar 6. Merupakan gambar model proses menggunakan metode SVM dan TF-IDF.

5. Evaluation (Evaluasi)

Evaluasi merupakan tahap untuk melihat hasil dari model yang telah dibuat menggunakan *Confusion Matrix* serta menggunakan *Cross Validation* (Gambar 5 dan 6). Hasil dari pengujian model yang dilakukan adalah mengklasifikasikan *tweet* opini masyarakat terhadap pemindahan ibu kota Indonesia menjadi positif dan negatif sebagai *dataset* menggunakan algoritma *Support Vector Machine* dan *Support Vector Machine* dengan *Term Frequency – Invers Document Frequency* untuk mengetahui model terbaik pada masing-masing model berikut ini hasil pengujian model pada Rapidminer:

- Mengevaluasi Hasil

Berikut merupakan tabel perbandingan (Tabel 2) dari pengujian dengan menggunakan *Support Vector Machine* dengan metode *Support Vector Machine* menggunakan metode pembobotan karakter *Term Frequency – Inverse Document Frequency*.

Tabel 4. Tabel Perbandingan SVM dengan SVM+TFIDF

Confusion Matrix	SVM	SVM dengan TF-IDF
Accuracy	77,72%	78,33%
Precision	77,78%	78,60%
Recall	99,87%	99,09%
F Measure	87,45%	87,66%

Dapat dilihat pada Tabel 4 yang menunjukkan peningkatan persentase setelah menggunakan pembo-botan karakter *Term Frequency – Inverse Document Frequency*, seperti akurasi yang meningkat menjadi 78,33%, presisi menjadi 78,60%, dan *F-measure* menjadi 87,66%. Fungsi utama pembobotan karakter (*term weight*) atau dalam penelitian ini TF-IDF adalah untuk meningkatkan nilai akurasi, sehingga selain nilai akurasi bisa mengalami penurunan atau ikut naik, pada Tabel 4 nilai Recall turun menjadi 99,09% dan hal tersebut dipengaruhi oleh perubahan jumlah data yang di prediksi benar positif (*True Positif*) atau benar negatif (*True Negative*).

- Proses Peninjauan

Setelah melakukan tahapan-tahapan sebelumnya, mulai dari Business Understanding (Pemahaman Bisnis), Data Understanding (Pemahaman Data), Data Preparation (Pengelolaan Data), Modeling (Pemodelan).

Data yang digunakan merupakan data yang diambil dari twitter dari tanggal 22-29 Maret 2019 dengan jumlah data 2500 lebih, dan setelah dilakukan pembersihan data dengan pre processing dan verifikasi data oleh guru Bahasa menjadi 992 data yang terbagi menjadi 221 data Positif dan 771 data Negatif. Pada pros-es Modeling didapatkan 2 buah model yang pertama Model dengan menggunakan metode SVM (Gambar 5) dan yang kedua model menggunakan metode SVM dan TFIDF (Gambar 6).

6. Deployment (Penyebaran)

Pada penelitian ini tahapan Deployment atau penyebaran yang berupa penyebaran penelitian dalam bentuk laporan akhir. Laporan akhir yang digunakan pada penelitian ini berupa Laporan Penelitian atau Skripsi yang telah di setujui oleh Dosen Pembimbing 1&2 dan Dosen Pengujian 1&2 Skripsi.

3.2. Pembahasan

Pemindahan ibu kota yang diresmikan pada 26 Agustus 2019 menuai banyak respons dari masyarakat baik itu respons positif atau negatif. Proyek tersebut tentu tidak menelan biaya yang sedikit serta membu-tuhkan waktu yang lama untuk proses pembangunannya, dan karena itu akan selalu menjadi perbincangan di masyarakat khususnya di twitter. Trending topic di twitter cukup baik untuk di gunakan sebagai patokan masalah apa yang sedang di bicarakan oleh masyarakat sehingga mudah untuk diketahui seperti peminda-han ibu kota yang pada bulan Maret 2020 kembali trending yang bertepatan dengan adanya pandemi Covid-19. Covid-19 menjadi virus yang baru dan membuat ahli medis pada waktu kebingungan karena minimnya informasi tentang virus itu, lockdown merupakan salah satu cara awal untuk menghentikan penyebaran virus tersebut.

Dengan munculnya Covid-19 membuat topik pemindahan ibu kota kembali ramai di perbincangkan dan mendapatkan banyak respon dari masyarakat tentu yang merespons positif dan negatif. Data untuk penelitian diambil dari twitter pada tanggal 22-29 Maret 2020 yang cukup untuk menjadikannya sebagai dataset penelitian. Data yang didapatkan berjumlah 2500 lebih tetapi banyaknya data karena adanya duplikasi atau retweet pada twitter sehingga untuk menghindari duplikasi data penambahan data langsung di sortir dan mendapatkan 998 dataset. Dataset yang telah didapatkan selanjutnya dilakukan verifikasi oleh ahli Bahasa dalam hal ini guru Bahasa Indonesia yang memverifikasi dataset serta memberikan label positif dan negatif pada dataset sehingga sebanyak 992 data yang dinyatakan berhubungan dengan peminda-han ibu kota di bagi menjadi 2 yaitu 221 data berlabel positif dan 771 data berlabel negatif.

Berdasarkan hasil pengujian yang dilakukan terhadap data opini masyarakat terhadap pemindahan ibu kota Indonesia yang terdiri dari 2 label yaitu positif dan negatif menggunakan tool RapidMiner dengan al-goritma *Support Vector Machine* menunjukkan hasil akurasi sebesar 77,72% dengan hasil *True Positive* 1 data, *True Negative* 770 data, *False Positive* 1 data dan *False Negative* 220 data. Dan hasil peningkatan akurasi dengan menggunakan *Term Frequency – Inverse Document Frequency* yang meningkatkan akurasi menjadi 78,33% dengan hasil *True Positive* 13 data, *True Negative* 764 data, *False Positive* 7 data dan *False Negative* 208 data. Dalam penelitian ini di gunakan metode SVM dengan TF-IDF untuk meningkatkan nilai akurasinya hal tersebut telah di lakukan oleh Honakan, Adiwijaya, & AL-Faraby (2018) yang dimana penggunaan *term weight* atau pembobotan karakter dapat meningkat akurasi pada dataset saat melakukan tahapan *pre-processing* karena yang diukur merupakan nilai dari bobot kata sehingga meningkatkan akurasi pada metode SVM.

4. Kesimpulan

Penerapan algoritma *Support Vector Machine* dengan menggunakan metode *Cross-Industry Standard Process for Data Mining* dengan tool RapidMiner merupakan salah cara untuk menyelesaikan penelitian ini. Pola opini masyarakat menunjukkan respon terhadap pemindahan Ibu Kota dengan menggunakan metode *Support Vector Machine* adalah *True Positive* 1 data, *True Negative* 770 data, *False Positive* 1 data dan *False Negative* 220 data sedangkan dengan menggunakan *Support Vector Machine* dengan *Term Frequency – Inverse Document Frequency* adalah *True Positive* 13 data, *True Negative* 764 data, *False Positive* 7 data dan *False Negative* 208 data. Selanjutnya untuk penerapan *Term Frequency – Inverse Document Frequency* untuk meningkatkan akurasi mendapatkan nilai 0,61% lebih tinggi dibanding tidak menggunakan TF-IDF.

References

- [1] Badan Pusat Statistik Indonesia, *Statistik Telekomunikasi Indonesia*. Jakarta: Badan Pusat Statistik, 2018.
- [2] G. A. Buntoro, "Analisis Sentimen Hatespeech Pada Twitter Dengan Metode Naïve Bayes Classifier Dan Support Vector Machine," *J. Din. Inform.*, vol. 5, no. 2, pp. 1–12, 2016.
- [3] BPMI Setpres, "Presiden Jokowi Berkirim Surat ke DPR," 2019. [Online]. Available: <https://www.presidentri.go.id/siaran-pers/presiden-jokowi-berkirim-surat-ke-dpr/>.
- [4] I. Saputra and D. Rosiyadi, "Perbandingan Kinerja Algoritma K-Nearest Neighbor , Naïve Bayes Classifier dan Support Vector Machine dalam Klasifikasi Tingkah Laku Bully pada Aplikasi Whatsapp," *Fakt. Exacta*, vol. 12, no. 2, pp. 101–111, 2019.
- [5] A. Setiyono and H. F. Pardede, "Klasifikasi Sms Spam Menggunakan Support Vector Machine," *J. PILAR Nusa Mandiri*, vol. 15, no. 2, pp. 275–280, 2019.
- [6] M. R. A. Nasution and M. Hayaty, "Perbandingan Akurasi dan Waktu Proses Algoritma K-NN dan SVM dalam Analisis Sentimen Twitter," *J. Inform.*, vol. 6, no. 2, pp. 212–216, 2019.
- [7] Honakan, Adiwijaya, and S. AL-Faraby, "Analisis Dan Implementasi Support Vector Machine Dengan String Kernel Dalam Melakukan Klasifikasi Berita Berbahasa Indonesia Analysis and Implementation Support Vector Machine With String Kernel for Classification indonesian news," *e-Proceeding Eng.*, vol. 5, no. 1, pp. 1701–1710, 2018.
- [8] R. Setiawan, "Penerapan Data Mining Menggunakan Algoritma K-Means Clustering Untuk Menentukan Strategi Promosi Mahasiswa Baru (Studi Kasus : Politeknik LP3I Jakarta)," *J. Lentera ICT*, vol. 3, no. 1, pp. 76–92, 2016.
- [9] A. Fauzi, A. N. Rais, M. F. Akbar, and W. Gata, "Analisis Sentimen Pada Media Sosial Facebook," in *Seminar Nasional Teknologi Informasi Universitas Ibn Khaldun Bogor*, 2018, pp. 46–54.
- [10] A. P. Fadillah, "Penerapan Metode CRISP-DM untuk Prediksi Kelulusan Studi Mahasiswa Menempuh Mata Kuliah (Studi Kasus Universitas XYZ)," *J. Tek. Inform. dan Sist. Inf.*, vol. 1, no. 3, pp. 260–270, 2015.
- [11] E. Susilowati, M. K. Sabariah, and A. A. Gozali, "Implementasi Metode Support Vector Machine Untuk Melakukan Klasifikasi Kemacetan Lalu Lintas Pada Twitter," *E-Proceeding Eng.*, vol. 2, no. 1, pp. 1478–1484, 2015.
- [12] J. Han, M. Kamber, and J. Pei, *Data mining: Data mining Concepts and Techniques*. 2014.
- [13] Z. Efendi and Mustakim, "Text Mining Classification sebagai Rekomendasi Dosen Pembimbing Tugas Akhir Program Studi Sistem Informasi," in *Seminar Nasional Teknologi Informasi Komunikasi dan Industri*, 2017, pp. 235–242.
- [14] S. Andayani and A. Ryansyah, "Implementasi Algoritma TF-IDF Pada Pengukuran Kesamaan Dokumen," *JuSiTik J. Sist. dan Teknol. Inf. Komun.*, vol. 1, no. 1, p. 53, 2017.