

ANALISIS SENTIMEN KEPUTUSAN PEMINDAHAN IBUKOTA NEGARA MENGGUNAKAN KLASIFIKASI NAIVE BAYES

Amar P. Natusuwarna
STMIK Pontianak
Jurusan Sistem Informasi
e-mail: a.pegirosa@gmail.com

Abstrak

Keputusan pemerintah Indonesia mengenai pemindahan ibukota memperoleh tanggapan secara positif dan negatif oleh masyarakat di media sosial diantaranya melalui Twitter. Data analisis sentimen diambil dari opini masyarakat yang menggunakan bahasa Indonesia dari tweet media sosial Twitter secara crawling. Pencarian kata yang digunakan adalah #IndonesiaIbuKotaBaru, #IbukotaPalangkaraya, dan #YukUsulIbukotaDimana. Tahapan kegiatan penelitian terdiri dari pengumpulan data melalui media sosial Twitter, pre-processing data, seleksi fitur, klasifikasi dengan Naive Bayes, dan evaluasi. Pre-processing data mencakup normalization, case folding, convert emoticon, convert negation, tokenizing, stopword removal, dan stemming. Seleksi fitur adalah dengan memanfaatkan term frequency. Sedangkan untuk evaluasi menggunakan cross validation dan confusion matrix dengan cara mendapatkan perhitungan akurasi. Tujuan penelitian ini adalah melakukan analisis sentimen opini masyarakat sehubungan dengan keputusan pemindahan ibukota Republik Indonesia dengan klasifikasi dua kelas yaitu positif dan negatif. Jumlah data yang digunakan sebanyak 200 data tweet yang terbagi atas 100 opini positif dan 100 opini negatif dengan menggunakan lima rasio perbandingan data latih dan data uji. Penelitian ini menghasilkan akurasi rata-rata sebesar 89,86%.

Kata kunci— Data Mining, Klasifikasi Naive Bayes, Term Frequency, Analisis Sentimen, Ibukota Baru

Abstract

The decision of Indonesia government relates to capital moving receiving positive and negative resposn from the public on social media including Twitter. Sentiment analysis data is captured from public opinion using Indonesian from Twitter social media tweets in a crawling manner. The search words are #IndonesiaIbuKotaBaru, #IbuKotaPalangkaraya, and #YukUsulIbukotaDimana. The Research methods are collecting data through Twitter, pre-processing data, feature selection, classification with Naive Bayes, and evaluation. Pre-processing data consists of normalization, case folding, convert emoticons, convert negation, tokenizing, stopword removal, and stemming. Feature selection involves term frequency. Meanwhile, evaluation covers cross validation and confusion matrix with calculation of accuracy. The purpose of this study is to conduct sentiment analysis of public opinion with the decision of moving the capital city with the classification of two classes: positive and negative. The amount of data utilized is 200 data tweets divided into 100 positive opinions and 100 negative opinions using five ratios of training data and test data. The evaluation result shows an average accuracy of 89.86%.

Keywords— Data Mining, Naive Bayes Classifier, Term Frequency, Sentiment Analysis, New Capital City

1. PENDAHULUAN

Pemerintah Indonesia terhitung tanggal 29 April 2019 telah memutuskan akan melakukan pemindahan ibukota negara dari Jakarta ke luar pulau Jawa. Berita tersebut telah menimbulkan opini negatif dan positif dari masyarakat. Beberapa tagar pada media sosial Twitter bermunculan diantaranya adalah #IndonesiaIbukotaBaru, #IbukotaPalangkaraya, dan #YukUsullbukotaDimana. Analisis sentimen dapat mengklasifikasikan opini-opini tersebut diantaranya menggunakan algoritma data mining seperti Klasifikasi *Naive Bayes* [1]. Kelebihan klasifikasi *Naive Bayes* dibandingkan dengan metode klasifikasi yang lain adalah sederhana, menggunakan teori peluang [1], sehingga metode ini cepat dalam proses pelatihan [2]. Kelebihan lain, akurasi metode ini cukup baik pada data besar seperti analisis sentimen [3]. Analisis sentimen adalah studi komputasi berkaitan dengan opini masyarakat mengomentari suatu kejadian disampaikan dalam bentuk teks [2]. Problem mendasar pada opinion mining adalah mengekstrak opini berdasarkan klasifikasi [3]. Besarnya manfaat bagi pengetahuan dan industri, membuat analisis sentimen berkembang pesat [4].

Sebelum keputusan pemerintah tersebut diambil, berikut beberapa opini positif berdasarkan hasil penelitian. Pemindahan ibukota akan menimbulkan efek pemerataan pembangunan [5], saat ini tingkat ekonomi di pulau Jawa jauh lebih tinggi dibandingkan pulau-pulau lainnya di Indonesia [6]. Ibukota baru dapat dipilih daerah aman terhadap bencana alam [7], berpotensi menyediakan lahan kosong besar hingga dapat bertahan ratusan tahun ke depan [8]. Pembangunan ibukota baru akan menambah lahan pekerjaan bagi masyarakat sekitar [6]. Sebaliknya, diperoleh opini negatif terhadap pemindahan ibukota Jakarta. Pemerintah harus mengeluarkan biaya sekitar seratus trilyun untuk infrastruktur dan mengangkut pegawai pemerintah pusat dan keluarganya [5][6]. Diketahui, APBN pemerintah saat ini minim [7]. Selain itu, masyarakat lokal mengalami kekhawatiran akan menjadi masyarakat terpinggirkan [8]. Persoalan di Jakarta yang menjadi alasan seperti banjir, sampah, dan kemacetan harus diselesaikan bukan dengan cara pemindahan ibukota [6].

Keputusan pemerintah memindahkan ibukota banyak mendapat opini masyarakat diantaranya melalui *Twitter*. *Twitter* adalah media sosial paling populer dalam menyampaikan opini [9]. Penyampaian opini di *Twitter* cenderung menggunakan bahasa informal dan istilah yang dipersingkat [2]. Analisis sentimen merupakan metode yang dapat menyelesaikan persoalan ini dengan melakukan perubahan format yang tidak terstruktur menjadi format yang mempunyai sejumlah kelas yang diinginkan [9]. *Tweet* yang dilakukan oleh masyarakat merupakan sumber data yang valid untuk dilakukan analisis sentimen [10].

Sudah banyak penelitian yang menggunakan analisis sentimen dengan menggunakan Klasifikasi *Naive Bayes*. Beberapa penelitian yang telah menggunakan analisis sentimen berdasarkan opini masyarakat sehubungan dengan keputusan pemerintah dengan menggunakan algoritma Klasifikasi *Naive Bayes* diantaranya adalah analisis sentimen terhadap pelaksanaan kurikulum 2013 [10], analisis sentimen terhadap ujian nasional secara online [11], dan analisis sentimen penerapan e-KTP [12]. 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*. Seleksi fitur menggunakan metode *Term Frequency* [13]. Penelitian ini menganalisa opini yang terdiri dari dua kelas yakni opini positif dan opini negatif [14]. Data diambil dari *Twitter* terdiri dari 100 *tweet* opini positif dan 100 *tweet* opini negatif [15]. Perangkat lunak *RapidMiner* yang cukup handal digunakan pada analisis sentimen [11], dimana dapat mengolah algoritma klasifikasi *Naive Bayes*; seleksi fitur *Term Frequency*; dan evaluasi menggunakan *Cross Validation* dan *Confusion Matrix* [2]. Hasil akhir penelitian ini berupa rata-rata akurasi dari klasifikasi *Naive Bayes* dan seleksi fitur *Term Frequency* dengan lima rasio perbandingan antara data latih dan data uji yakni rasio 90:10, 80:20, 70:30, 60:40, dan 50:50 [16].

2. METODE PENELITIAN

Penelitian ini menggunakan metodologi kuantitatif dengan cara mengumpulkan data opini masyarakat berbahasa Indonesia diambil dari media sosial *Twitter*. Untuk memperoleh analisis sentimen, ada lima langkah penelitian yang dilakukan terdiri dari: *crawling*, *pre-processing*, *feature selection*, *classification* dan *evaluation* [12]. Langkah pertama analisis sentimen ini adalah dengan melakukan *crawling* [12], menggunakan operator search *Twitter* pada *RapidMiner* versi 9.201. Langkah kedua, melakukan data *pre-processing* yang terdiri dari *normalization*, *case folding*, *convert emoticon*, *convert negation*, *tokenizing*, *stopword removal*, dan *stemming* [12]. *Normalization* mengubah bahasa informal, istilah-istilah yang disingkat menjadi bahasa dan istilah sesuai KBBI, dan menghilangkan karakter seperti tanda @, RT, URL, *hashtag* di mana komponen tersebut tidak berpengaruh pada sentimen [4]. *Case folding* merubah huruf besar menjadi huruf kecil [11]. *Convert emoticon* merubah ikon menjadi kata sesuai arti simbolnya [12], tetapi *emoticon* dapat diabaikan apabila mengganggu proses analisis sentimen [4]. *Convert negation* adalah kata negatif seperti *tidak*, *bukan*, *jangan*, kemudian kata tersebut dirubah dalam bentuk positif seperti *tidak mudah* dirubah menjadi *sulit* [13]. *Tokenizing* memecah sekumpulan yang terdiri dari kata-kata menjadi satuan kata, termasuk menghilangkan karakter titik, koma, garis miring, spasi dan angkat [1][3]. *Stopword removal* adalah meniadakan kata-kata yang dianggap tidak mempunyai efek dalam proses klasifikasi [10]. Sedangkan *stemming* mengembalikan setiap kata menjadi hanya kata dasarnya saja [15]. Langkah ketiga adalah *feature selection* menggunakan *Term Frequency*[1]. *Term Frequency* adalah jumlah kemunculan sebuah kata atau istilah dalam suatu dokumen [13]. Jumlah kemunculan suatu *term* atau istilah semakin besar, maka bobot memberikan nilai kesesuaian semakin besar [12]. *Term Frequency* dapat ditunjukkan dengan persamaan (1). Dimana tf adalah *Term Frequency*, $W_{tf,d}$ merupakan bobot tf yang dihitung secara logaritma [13].

$$W_{tf,d} = \begin{cases} 1 + \log_{10} tf_{t,d}, & \text{if } tf_{t,d} > 0 \\ 0, & \text{otherwise} \end{cases} \quad (1)$$

Langkah keempat adalah dengan melakukan klasifikasi *Naïve Bayes* yang mengacu kepada peluang. Klasifikasi *Naïve Bayes* mengansumsikan bahwa sebuah variabel tidak ada hubungan dengan atribut variabel lain. Berikut di bawah ini persamaan (2) Klasifikasi *Naïve Bayes*, X menyatakan kelas, $P(X_i|Y_i)$ menyatakan peluang kata ke i pada kelas [9][10].

$$P(X|y_j) = \prod_{i=1}^m P(x_i|y_i) \quad (2)$$

Langkah terakhir melakukan evaluasi menggunakan *Cross Validation* dan *Confusion Matrix*. Evaluasi performa dapat dilihat dari hasil akurasi [1]. Tabel 1 menunjukkan *Confusion Matrix* yang terdiri dari aktual positif, aktual negatif, prediksi positif, prediksi negatif.

Tabel 1. Confusion Matrix [1][4]

		Prediksi	
		Negatif	Positif
Aktual	Negatif	a	b
	Positif	c	d

Seperti yang terlihat pada Tabel 1, simbol a menunjukkan jumlah prediksi yang benar bahwa data adalah negatif. Simbol b memperlihatkan jumlah prediksi yang keliru bahwa data adalah positif. Simbol c memperlihatkan jumlah prediksi keliru bahwa data adalah negatif. Simbol d menunjukkan jumlah prediksi yang benar bahwa data adalah positif [4]. Berikut adalah persamaan akurasi (AC) yang merupakan tingkat kedekatan data uji dengan data latihan [1][14]. Performa akurasi dapat dihitung dengan persamaan (3) di bawah ini.

$$AC = \frac{a + d}{a + b + c + d} \quad (3)$$

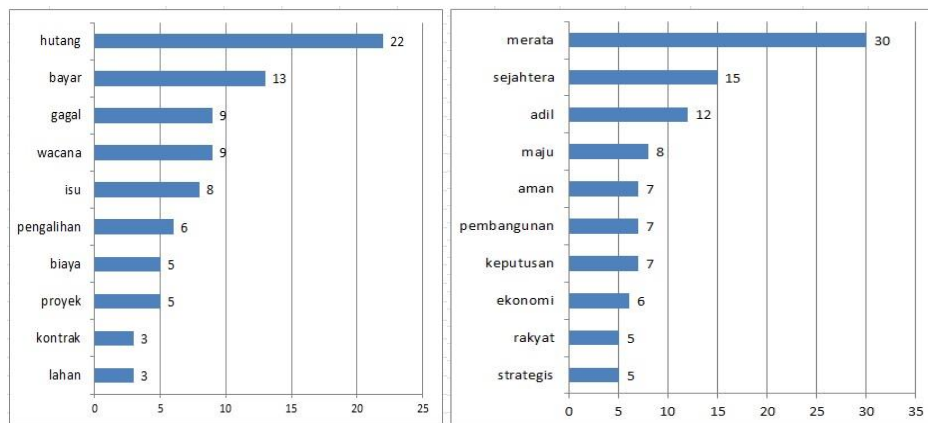
3. HASIL DAN PEMBAHASAN

Data *tweet* dalam bahasa Indonesia diambil dari 29 April sampai dengan 1 Mei 2019 dengan jumlah total sebanyak 12.725 *tweet*. Data diperoleh berdasarkan *tweet* yang dilakukan masyarakat pada *Twitter* dengan tagar #IbukotaPalangkaraya, #IndonesiaIbukotaBaru, dan #YukUsulIbukota. Dari sejumlah *tweet* tersebut, total opini yang diambil pada pengujian ini adalah 200 *tweet* yang dilakukan secara acak hingga diperoleh 100 *tweet* sentimen positif dan 100 *tweet* sentimen negatif. Data *Twitter* tersebut kemudian dilakukan pengujian analisis sentimen dengan klasifikasi *Naïve Bayes* yang menghasilkan nilai akurasi. Tabel 2 menunjukkan contoh hasil *tweet* yang berhasil dikeluarkan oleh *RapidMiner Studio* setelah mengambil data dari *Twitter Search API*. Opini dari *tweet* tersebut ditambahkan secara manual pemberian atribut sentimen dengan nilai positif atau negatif.

Tabel 2. Contoh Hasil Perolehan Data *Tweet* dan Atribut Sentimen

No	Tweet	Sentimen
1	Menurut sy pindah ibukota baik utk pemerataan pembangunan walaupun itu tidaklah mudah ☺ #IndonesiaIbukotaBaru http://t.co//JdOfINpDO	Positif
2	@jokossan: mindahin ibukota ... cari dana tambahan quota haji aja bingung	Negatif

Gambar 1 di bawah ini adalah jumlah kata terbanyak hasil *pre-processing* yang dilakukan secara manual. Dimana gambar sebelah kiri menunjukkan jumlah 10 kata terbanyak untuk opini positif dan gambar sebelah kanan menunjukkan jumlah 10 kata terbanyak untuk opini negatif.

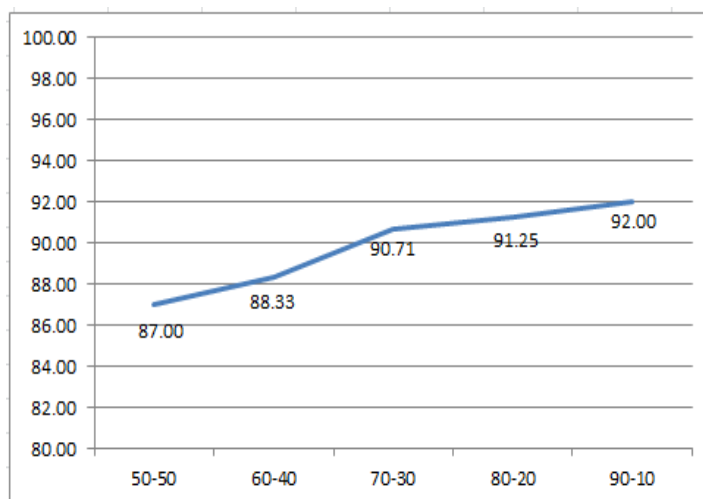


Gambar 1. Grafik Jumlah 10 Kata Terbanyak Opini Positif (kanan) dan Opini Negatif(kiri)

Pada Gambar 1 untuk opini positif, kata yang paling sering muncul adalah *merata* dengan kemunculan 30 kali, diikuti oleh *sejahtera*, *adil*, *maju*, *aman*, *pembangunan*, *keputusan*, *ekonomi*, dan *strategis*. Sementara itu untuk opini negatif, kata yang paling sering muncul adalah *hutang* dengan kemunculan 22 kata, diikuti oleh *bayar*, *gagal*, *wacana*, *isu*, *pengalihan*, *biaya*, *proyek*, *kontrak* dan *lahan*. Opini positif menunjukkan penerimaan masyarakat pendukung pindah ibukota bahwa pemindahan tersebut dapat meningkatkan pemerataan, kesejahteraan, dan keadilan baik untuk wilayah maupun masyarakat sekitarnya. Sehingga, kemajuan itu tidak hanya terpaku di pulau Jawa saja [6]. Sementara itu, opini negatif menunjukkan penolakan sebagian masyarakat mengenai pemindahan ibukota terutama berkaitan hutang baru, minimnya anggaran APBN untuk

pembayaran, dan ada kemungkinan gagalnya proses pemindahan sehubungan dengan permasalahan ekonomi tersebut [7].

Evaluasi menggunakan algoritma klasifikasi *Naive Bayes* dilakukan dengan membandingkan kelima rasio antara data latih dan data uji yakni 90:10, 80:20, 70:30, 60:40, dan 50:50. Hasil pengujian performa akurasi dari pengolahan menggunakan klasifikasi *Naive Bayes* dan seleksi fitur *Term Frequency* seperti yang dapat dilihat pada Gambar 2.



Gambar 2. Akurasi dengan Lima Rasio Data Latih dan Data Uji

Gambar 2 menunjukkan secara umum grafik berupa akurasi yang mengalami kenaikan nilai berbanding lurus dengan kenaikan data latih. Grafik akurasi secara konsisten mengalami kenaikan secara gradual bermula dari 87,00% pada rasio 50:50 menjadi 92,00% pada rasio 90:10, diperoleh rata-rata akurasi menjadi 89,86%. Oleh sebab itu, rasio 90:10 [11] dan rasio 70:30 [13] sering digunakan pada penelitian untuk mendapatkan nilai akurasi.

4. KESIMPULAN

Penggunaan metode klasifikasi *Naive Bayes* cukup baik pada analisis sentimen mengenai keputusan pemerintah Indonesia akan pindah ibukota dari Jakarta ke kota lain di luar pulau Jawa. Analisis sentimen pada penelitian ini menggunakan data dari *Twitter* yang berisi opini negatif dan opini positif dari masyarakat. Opini yang disampaikan oleh masyarakat baik opini positif ataupun opini negatif melalui *Twitter* setelah pengumuman pemindahan ibukota memiliki kesamaan dengan pendapat peneliti yang melakukan pengkajian sebelum pengumuman pemindahan ibukota tersebut. Untuk merubah data tidak terstruktur dari *Twitter* menjadi terstruktur dilakukan *pre-processing* supaya dapat diolah menggunakan algoritma *Naive Bayes*. Ketika dilakukan proses manual pada *pre-processing* 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%.

5. SARAN

Penelitian ini dapat dilanjutkan dengan melakukan analisis sentimen opini masyarakat berupa *tweet* diambil dari *Twitter* ketika pemerintah Indonesia telah memutuskan lokasi yang akan menjadi ibukota baru. Metode algoritma klasifikasi *Naive Bayes* dan seleksi fitur *Term*

Frequency yang terbukti menghasilkan akurasi yang baik dapat digunakan kembali. Selain metode algoritma klasifikasi *Naive Bayes*, algoritma klasifikasi lain seperti *Support Vector Machine* dapat juga dimanfaatkan. Sementara itu untuk seleksi fitur selain *Term Frequency*, seleksi fitur lainnya yang dapat digunakan seperti *Term Frequency Inverse Document Frequency (TF-IDF)*.

DAFTAR PUSTAKA

- [1] 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.
- [2] 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.
- [3] 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.
- [4] Buntoro, G.A., 2017, Analisis Sentimen Calon Gubernur DKI Jakarta 2017 Di Twitter. *INTEGER: Journal of Information Technology*, No.1, Vol.2.
- [5] Taufiq, M., 2017, Pemindahan Ibu Kota dan Potensi Konektivitas Pemerataan Ekonomi. *Prosiding Semas Pemindahan Ibu Kota Negara 2017*, Palangkaraya, Dec 13.
- [6] Hutasoit, W. L., 2019, Analisa Pemindahan Ibukota Negara. *Dedikasi*, No.2, Vol.39, 108-128.
- [7] Septiana, D., Sumarlam, S., 2018, Palangka Raya the Capital City of Indonesia: Critical Discourse Analysis on News about Moving the Capital City from Jakarta. In *International Seminar on Recent Language, Literature, and Local Cultural Studies (BASA 2018)*, Solo Baru, October 19-20.
- [8] Yahya, M., 2018, Pemindahan Ibu Kota Negara Maju dan Sejahtera. *Jurnal Studi Agama dan Masyarakat*, No.1, Vol.14, 21-30.
- [9] 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.
- [10] Pamungkas, D. S., Setiyanto, N. A., & Dolphina, E., 2015, Analisis Sentiment pada Sosial Media Twitter Menggunakan Naive Bayes Classifier terhadap Kata Kunci "Kurikulum 2013". *Techno. Com*, No.4, Vol.14, 299-314.
- [11] 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.
- [12] Mihuandayani, M., Feriyanto, E., Syarham, S., & Kusriani, K., 2018, Opinion Mining pada Komentas Twitter E-KTP Menggunakan Naive Bayes Classifier. *Semnasteknomedia Online*, No.1, Vol.6, 1-2.
- [13] Rofiqoh, U., Perdana, R. S., & Fauzi, M. A., 2017, Analisis Sentimen Tingkat Kepuasan Pengguna Penyedia Layanan Telekomunikasi Seluler Indonesia Pada Twitter Dengan Metode Support Vector Machine dan Lexicon Based Features. *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, No.12, Vol.1, 1725-1732.
- [14] Antinasari, P., Perdana, R. S., & Fauzi, M. A., 2017, Analisis Sentimen Tentang Opini Film Pada Dokumen Twitter Berbahasa Indonesia Menggunakan Naive Bayes Dengan Perbaikan Kata Tidak Baku. *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, No.12, Vol.1, 1733-1741.

- [15] Fanissa, S., Fauzi, M. A., & Adinugroho, S., 2018, Analisis Sentimen Pariwisata di Kota Malang Menggunakan Metode Naive Bayes dan Seleksi Fitur Query Expansion Ranking. *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, No.8, Vol.2, 2766-2770.
- [16] Perdana, R.S., 2018, Penerapan Sentimen Analisis Acara Televisi Pada Twitter Menggunakan Support Vector Machine dan Algoritma Genetika sebagai Metode Seleksi Fitur. *Jurnal Pengembangan Teknologi Informasi dan Komputer*, No.3, Vol.2, 998-1007.