

ANALISIS SENTIMEN REGISTRASI ULANG KARTU SIM PADA TWITTER
MENGUNAKAN ALGORITMA SVM DAN K-NN

Mira Kusmira

Sistem Informasi¹,
STMIK Nusa Mandiri Jakarta ¹,
<http://www.nusamandiri.ac.id>
mira.mik@nusamandiri.ac.id ¹,

Abstract— Sentiment analysis is the process of analyzing, understanding, and classifying opinions, evaluations, assessments, attitude, and emotions to an entity such as events re-register Sim Card discussed in society or internet, especially Twitter. everyone is free to opine on re-register Sim Card so many opinions. This study uses English text contained on the twitter of opinions. So classifying the category positive, negative done automatically with Analyze Sentiment using Rapid Miner. The methods used in this study for process Sentiment Analysis using the method k-Nearest Neighbor (k-NN) and Support Vector Machine (SVM) in better regard than methods k-Nearest Neighbor in this research because it produces Accuracy 78.97% and AUC 0.851%.

Keywords: Sentiment Analysis, Re-register Sim Card, Twitter, SVM, k-NN

Abstrak— Analisis sentimen adalah proses menganalisa, memahami dan mengklasifikasikan pendapat, evaluasi, penilaian, sikap dan emosi terhadap suatu entitas seperti pada peristiwa Registrasi Ulang Kartu SIM yang ramai menjadi perbincangan di dunia nyata maupun media sosial, salah satunya adalah *twitter*. Semua orang dengan bebas memberikan pendapatnya atau beropini tentang registrasi ulang kartu SIM sehingga memunculkan banyak opini. Penelitian ini menggunakan teks bahasa Indonesia yang terdapat pada *twitter* berupa opini sehingga pengklasifikasian kategori *Positive*, *Negative* dilakukan secara manual. Metode yang digunakan dalam penelitian ini untuk proses klasifikasi analisis *sentiment* menggunakan metode *k-Nearest Neighbor (k-NN)*, dan *Support Vector Machine (SVM)* hasil dari penelitian ini adalah analisis sentimen terhadap registrasi ulang kartu SIM, Metode SVM dianggap lebih baik dari pada metode k-NN dalam penelitian ini karena menghasilkan *Accuracy* 78.97% dan *AUC* 0.851

Kata Kunci: Analisis Sentiment, Registrasi Ulang Kartu SIM, *Twitter*, *SVM*, *k-NN*

PENDAHULUAN

Sim Card adalah sebuah kartu pintar seukuran prangko yang ditaruh di telepon genggam yang menyimpan kunci pengenalan jasa telekomunikasi (Gusriani, 2017). Baru-baru ini Kementerian Kominfo memberlakukan peraturan registrasi ulang kartu SIM prabayar dengan validasi data dukcapil. Demi peningkatan perlindungan hak pelanggan jasa telekomunikasi, kementerian komunikasi dan informasi akan memberlakukan registrasi nomor pelanggan yang divalidasi dengan nomor induk kependudukan (NIK) dan Nomor Kartu Keluarga (KK).

Registrasi ulang kartu SIM diberlakukan dari tanggal 31 oktober 2017 sampai 28 februari 2018. (KOMINFO). Sejak diberlakukan tanggal registrasi dan ditetapkan dengan segala persyaratannya, masalah registrasi ini mulai banyak diperbincangkan, baik di dunia nyata maupun

dunia maya. Semua orang bebas berpendapat atau beropini tentang registrasi ulang kartu SIM sehingga memunculkan banyak opini, tidak hanya opini yang bersifat positif atau netral tapi juga yang negatif. Mereka tuangkan opini tersebut di media sosial salah satunya adalah media sosial *twitter*. Media sosial khususnya *twitter* sekarang ini menjadi perangkat komunikasi yang sangat populer dikalangan pengguna internet. Pada konferensi resmi pengembangan *twitter* Chrip 2010, perusahaan tersebut menyebutkan bahwa pada bulan april 2010, *Twitter* memiliki 106 juta akun dan sebanyak 180 juta pengunjung unik setiap bulannya. Jumlah pengguna *twitter* disebutkan terus meningkat 300.000 user setiap harinya (Buntoro, 2016)

Analisis *sentiment* atau *opinion mining* merupakan proses memahami, mengekstrak dan mengolah data tekstual secara otomatis untuk mendapatkan informasi *sentiment* yang

terkandung dalam suatu kalimat opini. (Mahardhika & Zuliarso, 2018).

Dengan demikian sangat penting untuk melakukan analisis sentimen terhadap opini *public* mengenai registrasi ulang kartu SIM, untuk mengetahui sentimen terhadap registrasi ulang kartu SIM.

k-Nearest Neighbor (k-NN) adalah mengklasifikasikan obyek berdasarkan atribut dan *training sample*, diberikan titik *query*, akan ditemukan sejumlah k obyek atau (titik *training*) yang paling dekat dengan titik *Query*. Klasifikasi menggunakan voting terbanyak diantara klasifikasi ketetanggaaan sebagai nilai prediksi dari *query* instance yang baru.

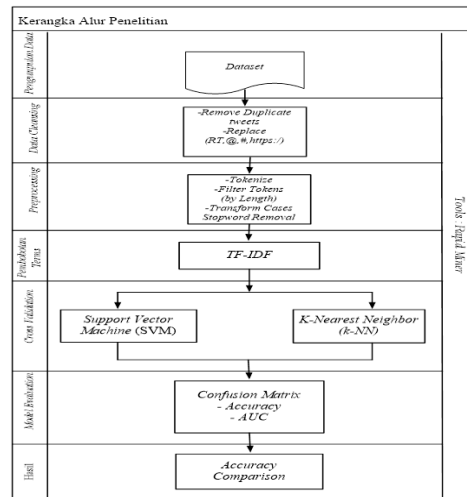
Support Vector Machine (SVM) adalah seperangkat metode pembelajaran terbimbing yang menganalisis data dan mengenali pola, digunakan untuk klasifikasi dan analisis regresi. SVM standar mengambil himpunan data input, dan memprediksi, untuk setiap masukan yang diberikan, kemungkinan masukan adalah anggota dari salah satu kelas dari dua kelas yang ada, yang membuat sebuah SVM sebagai penggolong nonprobabilistik linier biner, karena SVM adalah sebuah pengklasifikasian, kemudian diberi suatu himpunan pelatihan, masing-masing ditandai sebagai milik salah satu dari dua kategori, suatu algoritma pelatihan SVM membangun sebuah model yang memprediksi apakah data yang baru jatuh ke dalam suatu kategori atau yang lain.

Penelitian sebelumnya telah banyak dilakukan dengan menerapkan metode *text mining* oleh peneliti dalam menganalisa sentimen. Penerapan algoritma *Support Vector Machine* (SVM) dalam menganalisa Sentimen dan Klasifikasi Kategori Terhadap Tokoh Publik Pada *Twitter* (Hidayatullah & Azhari, 2014), dan menganalisa sentimen terhadap pasar otomotif mobil dengan menggunakan algoritma *Naïve Bayes*. Dimana hasilnya cukup baik (Rustiana & Rahayu, 2017)

Pada kesempatan ini penelitian yang akan dilakukan adalah membandingkan dua metode klasifikasi, diantaranya yaitu *Support Vector Machine* (SVM), dan *k-Nearest Neighbor* (k-NN). Dua algoritma tersebut digunakan dalam menganalisa sentimen registrasi ulang kartu SIM. untuk mengetahui sentimen opini publik terhadap registrasi ulang kartu SIM.

BAHAN DAN METODE

Berikut ini adalah metode yang digunakan untuk proses analisis sentimen yang digunakan dalam penelitian ini.



Sumber:

Gambar 1. Kerangka Alur Penelitian

Data yang digunakan pada penelitian ini merupakan data sentiment yang berasal dari twett yang berisi opini dengan kata kunci "registrasi kartu SIM". Data yang diambil hanya tweet dalam bahasa Indonesia, sebanyak 290 data. Data tersebut terbagl menjadi opini positif dan opini negatif.

Pengolahan Data Awal

Tahapan yang dilakukan maka data perlu kita berikan analisa sentimen secara manual dengan membaca maksud kalimat yang ada dalam sentimen tersebut, sehingga dapat diklasifikasikan kalimat tersebut merupakan sentimen positif atau negatif, dilanjutkan dengan tahapan data *cleansing*, Tahapan tersebut mengubah beberapa karakter pada data *tweet* kedalam kamus bahasa indonesia. Contohnya "RT" diganti dengan "retweet", "@" diganti dengan "at". Selain itu, ada juga atribut-atribut data yang tidak relevan, data tersebut lebih baik dibuang karena keberadaannya bisa mengurangi mutu atau akurasi.

Metode Yang Diusulkan

Pada penelitian ini peneliti mengusulkan dalam analisis sentimen guna mengetahui sentimen pengguna media sosial *twitter*, dengan menggunakan metode klasifikasi dinataranya algoritma *Support Vector Machine* (SVM), dan algoritma *k-Nearest Neighbor*(k-NN).

Experimen dan Hasil Penelitian

Dalam melakukan pengujian model *software* yang digunakan untuk mengambil data dan alat

bantu untuk menghitung tingkat akurasi adalah *RapidMiner* versi 7.6.

Evaluasi dan Validasi Pengujian

Dalam penelitian ini penulis menggunakan validasi standar yaitu *10 folds cross-validation* dimana proses ini membagi data secara acak ke dalam 10 bagian. Proses pengujian dimulai dengan pembentukan model dengan data pada bagian pertama. Model yang terbentuk akan diujikan pada 9 bagian data sisanya. Proses yang dilakukan setelah melakukan pengujian yaitu mengukur *performance* dari algoritma klasifikasi *text mining* yang dipakai. Dalam penelitian ini *performance* diukur menggunakan *Accuracy* dan *AUC (Area Under Curve)*.

HASIL DAN PEMBAHASAN

Hasil Penelitian

Data di proses untuk menyamakan bentuk kata, menghilangkan karakter –karakter selain huruf, dan mengurangi volume kosakata proses ini terdiri dari lima tahapan. Ke lima tahapan tersebut adalah *cleansing, tokenize, filter (by length), stopword removal, Transform Cases* dan metode yang digunakan dalam menganalisis data tersebut menggunakan algoritma *Support Vector Machine (SVM)* dan *k-Nearest Neighbor (k-NN)*.

Hasil Experimen Pengujian Igoritma

1. Metode Support Vector Machine

Metode *Support Vector Machine* dalam penelitian ini ditentukan dengan cara melakukan uji coba memasukan nilai C, dan Epsilon. Berikut ini adalah hasil dari percobaan yang telah dilakukan untuk penentuan nilai *training Cycles SVM*

Tabel 1 Experimen Penentuan Nilai Training Cycles SVM

C	Epsilon	SVM	
		Accuracy	AUC
0	0	75.86%	0.87
0.1	0.1	74.83%	0.834
0.2	0.2	74.83%	0.828
0.3	0.3	74.14%	0.813
0.4	0.4	75.86%	0.823
0.5	0.5	75.86%	0.84
0.6	0.6	73.79%	0.823
0.7	0.7	76.21%	0.829
0.8	0.8	74.83%	0.814

0.9	0.9	73.45%	0.829
0	0.1	78.97%	0.851
1	1	48.28%	0.5
1	0	76.55%	0.84

Sumber: (Kusmira, 2019)

Hasil pengujian menunjukkan bahwa penerapan metode *Support Vector Machine (SVM)*, pada tabel tersebut yang paling tinggi dengan penentuan hasil dari **C=0.0** dan **Epsilon 0.1** menghasilkan **Accuracy = 78.97** dan **AUC= 0.851**

2. k-Nearest Neighbor (k-NN)

Metode *k-Nearest Neighbor* dalam penelitian ini ditentukan dengan cara melakukan uji coba memasukan nilai k yang digunakan menyatakan jumlah tetangga terdekat yang dilibatkan dalam penentuan prediksi label kelas pada uji data.

Tabel 2 Experimen Penentuan Nilai Training k-NN

K	k-NN	
	Accuracy	AUC
1	69.66%	0.5
2	74.14%	0.743
3	76.21%	0.803
4	75.45%	0.805
5	71.03%	0.5
6	75.17%	0.829
7	76.52%	0.85
8	73.79%	0.836
9	75.52%	0.85
10	76.55%	0.852

Sumber: (Kusmira, 2019)

Hasil pengujian menunjukkan bahwa penerapan metode *k-Nearest Neighbor* Pada Tabel tersebut dengan penentuan nilai **k= 10** menghasilkan **Accuracy = 76.55%** dan **AUC = 0.852**

Analisa Evaluasi Hasil dan Validasi Model

Validasi digunakan untuk memperoleh prediksi menggunakan model yang ada dan kemudian membandingkan hasil tersebut dengan hasil yang sudah diketahui, ini mewakili langkah paling penting dalam proses membangun sebuah model *Cross Validation* adalah teknik validasi dengan membagi data secara acak ke dalam k bagian dan masing-masing bagian akan dilakukan proses klasifikasi dalam *Cross Validation*, jumlah tetap lipatan atau partisi dari data ditentukan sendiri. Cara standar untuk memprediksi *error rate* dari teknik pembelajaran sebuah *sample data*

adalah dengan menggunakan *ten fold cross validation*.

Hasil dari pengujian model yang dilakukan adalah mengklasifikasikan opini publik positif, opini negatif menggunakan *Support Vector Machine (SVM)*, dan *k-Nearest Neighbor (k-NN)* untuk mendapatkan nilai akurasi terbaik pada *Framework Rapid Miner*.

Pada pengujian ini, data yang digunakan adalah data bersih yang telah melalui *preprocessing*. Proses validasi terdiri dari data training dan data testing. berikut adalah hasil perhitungan pada tingkat *accuracy*, *class precision*, dan *class recall* serta hasil dari kurva ROC

A. Confusion Matrix Support Vector Machine (SVM)

Tabel 3 Confusion Matrix Support Vector Machine (SVM)

accuracy: 78.97% +/- 6.44%
(mikro:78.97%)

Confusion Matrix:

	True Negatif	True Positif	class precision
pred. Negatif	103	19	84.43%
pred. Positif	42	126	75.00%
class recall	71.03%	86.90%	

Sumber: (Kusmira, 2019)

$$Acc (Accuracy) = \frac{103+126}{103+19+42+126} = \frac{229}{290} \times 100\% = 78.97\%$$

Berdasarkan uji coba menggunakan SVM nilai *accuracy* yang diperoleh yaitu 78.97% dari 290 data opini mengenai registrasi ulang kartu SIM yang telah dibersihkan melalui proses data *cleansing* dan data *preprocessing* menggunakan *RapidMiner*. Hasilnya menunjukkan bahwa klasifikasi opini negatif yang sesuai prediksi yaitu 103, dan opini negatif yang termasuk kedalam prediksi positif yaitu 42. Sedangkan opini positif yang termasuk kedalam prediksi positif 126, Selanjutnya positif yang termasuk kedalam prediksi negatif 19.

B. Kurva ROC Support Vector Machine (SVM)

Kurva ROC *Support Vector Machine* yang ditunjukkan pada gambar 4.1 dapat mengekspresikan *Confusion matrix* dari tabel 4.7 kurva ROC tersebut diartikan dengan *false* Positif untuk *garis horizontal* dan *True* Positif untuk *garis vertical* dengan nilai *AUC=0.851*



Sumber: (Kusmira, 2019)

Gambar 2 Kurva ROC Support Vector Machine (SVM)

C. Confusion Matrix k-Nearest Neighbor

Confusion Matrix memberikan keputusan yang diperoleh dalam *training* dan *testing* dan memberikan penilaian *performance* klasifikasi berdasarkan kelas yang telah di uji. Berikut *accuracy* yang diambil pada *experiment* sebelumnya adalah *k= 10* karena *accuracynya* paling tinggi.

Tabel 4 Confusion Matrix k-Nearest Neighbor (k-NN)

accuracy: 76.55% +/- 7.36% (mikro:76.55%)

Confusion Matrix:

	True Negatif	True Positif	class precision
pred. Negatif	116	39	74.84%
pred. Positif	29	106	78.52%
class recall	80.00%	73.10%	

Sumber: (Kusmira, 2019)

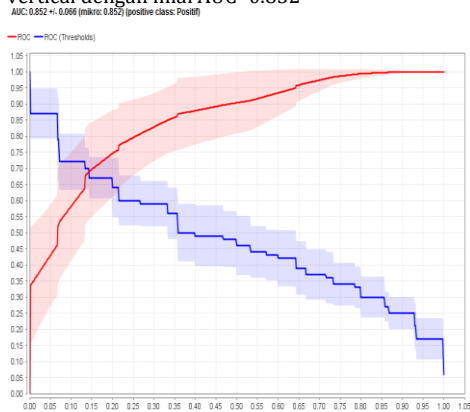
$$Acc (Accuracy) = \frac{116+106}{116+39+29+106} = \frac{222}{290} \times 100\% = 76.55\%$$

Akurasi yang diperoleh yaitu 76.55% dari 290 data opini mengenai registrasi ulang kartu SIM yang telah dibersihkan melalui proses data *cleansing* dan data *preprocessing* menggunakan *RapidMiner*.

Hasilnya menunjukkan bahwa klasifikasi opini negatif yang sesuai prediksi yaitu 116, opini *negative* yang termasuk kedalam prediksi *positive* yaitu 29. Selanjutnya untuk klasifikasi opini positif yang sesuai prediksi yaitu 106 dan opini positif yang termasuk kedalam prediksi *negative* yaitu 39.

D. Kurva ROC k-Nearest Neighbor (k-NN)

Kurva ROC *k- Nearest Neighbor (k-NN)* yang ditunjukkan pada gambar 4.7 dapat mengekspresikan *Confusion matrix* dari tabel 4.8 kurva ROC tersebut diartikan dengan false Positif untuk garis horizontal dan True Positif untuk garis vertical dengan nilai AUC=0.852



Sumber: (Kusmira, 2019)
Gambar 3 Kurva ROC k-Neares Neighbor (k-NN)

Dari hasil uji coba yang telah dilakukan diatas yaitu antara SVM dan k-NN maka dapat disimpulkan Algoritma k-NN Lebih baik dibandingkan dengan algoritma SVM meskipun perbedaannya tidak terlalu jauh. Algoritma SVM mendapatkan nilai **Accuracy 78.97%** dan **AUC 0.851** hasil *Accuracy* dan *AUC* ini lebih tinggi dibanding uji coba dengan menggunakan Algoritma *k-NN* yang mendapatkan **Accuracy 76.55%** dan **AUC 0.852**.

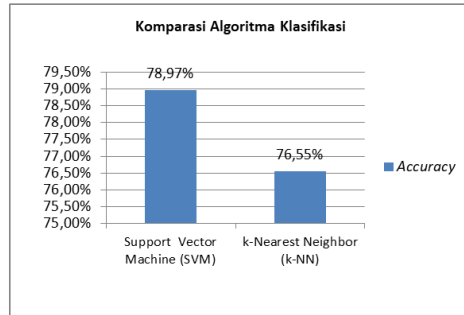
4.3 Hasil Perbandingan Komparasi Algoritma SVM dan k-NN

Adapun perbandingan hasil komparasi *accuracy* dan *AUC* algoritma yang sudah dilakukan pengujiannya sebagai berikut:

Tabel 5 Komparasi Accuracy dan AUC Algoritma Klasifikasi

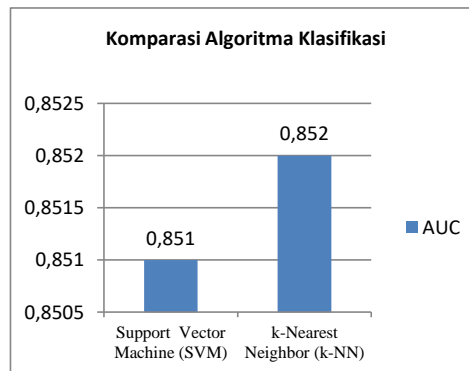
Algoritma	Accuracy	AUC
Support Vector Machine (SVM)	78.97%	0.851
k-Nearest Neighbor	76.55%	0.852

Sumber: (Kusmira, 2019)



Sumber: Hasil Penelitian (Kusmira 2019)
Gambar 5 Diagram Komparasi Nilai Accuracy Algoritma Klasifikasi

Comment [F1]: Harus ada sumber. Sumber harus ada di daftar Referensi. Perhatikan penulisan sumber pada paper yang sudah terbit



Sumber: (Kusmira, 2019)
Gambar 4 Diagram Komparasi Nilai AUC Algoritma Klasifikasi

Berdasarkan pengujian yang telah dilakukan terhadap opini *public* mengenai “registrasi ulang kartu SIM” yang telah dikumpulkan melalui data dari *twitter* secara acak sebanyak 290 data (145 positif dan 145 negatif) dalam bahasa Indonesia. Hasil berupa hipotesa awal tidak sesuai dengan hasil akhir karena baik algoritma SVM maupun k-NN menghasilkan nilai *accuracy* yang hampir sama terhadap *dataset* yang di uji. Dalam penelitian ini, *accuracy* SVM 78.97% nilai *accuracy* ini lebih tinggi dibandingkan dengan *accuracy* yang dihasilkan k-NN yaitu sebesar 76.55 %. Sedangkan untuk nilai AUC yang dihasilkan SVM dengan nilai C=0.0 dan Epsilon=1.0 dengan nilai AUC 0.851%, sedangkan untuk nilai AUC yang dihasilkan k-NN yaitu dengan k=10 menghasilkan AUC= 0.852%

Hal ini membuktikan dengan menggunakan *text mining* bahasa indonesia bahwa dengan menggunakan algoritma SVM maupun k-NN tetap *accuracy* dan AUC yang dihasilkan tidak jauh berbeda.

Pengujian SVM lebih lama dibandingkan dengan penggunaan algoritma k-NN. Karena SVM

bekerja menggunakan parameter nilai C dan Epsilon yang mencari *hyperplane hyperplane* yang terbaik meskipun pada penelitian ini hasil nilai SVM lebih rendah dibandingkan dengan nilai yang dihasilkan k-NN.

Model klasifikasi *text* pada opini *public* mengenai "registrasi ulang kartu SIM" dilakukan secara manual dengan memberikan penilaian apakah kalimat tersebut termasuk kategori sentimen positif atau sebaliknya sentimen negatif, kemudian *text* tersebut disimpan dalam suatu *dataset* untuk pengolahan selanjutnya yaitu data *cleansing* dan *preprocessing*, dimana proses ini mengubah atau menghapus karakter-karakter yang tidak sesuai dengan analisa *sentiment* atau tidak terdaftar dalam kamus bahasa Indonesia. Setelah proses data *cleansing* lalu pemisahan kata-kata, masing-masing kata memiliki bobot sehingga dapat dilihat kata mana saja yang berhubungan dengan *sentiment* yang sering muncul dan memiliki bobot tertinggi. Selanjutnya dilakukan *cross validation* dengan menggunakan 2 algoritma yaitu *Support Vector Machine* dan *k-Nearest Neighbor* (k-NN), Dengan ini dapat disimpulkan bahwa algoritma *Support Vector Machine* (SVM) memiliki *accuracy* yang lebih baik dibandingkan dengan algoritma *k-Nearest Neighbor* (k-NN).

KESIMPULAN

Berdasarkan hasil penelitian dapat ditarik kesimpulan Penggunaan algoritma *Support Vector Machine* (SVM) akurasiya lebih tinggi dibandingkan algoritma *k-Nearest Neighbor* (k-NN) untuk melakukan klasifikasi sentimen analisis teks berbahasa Indonesia dalam analisis sentimen registrasi kartu SIM dan *Support Vector Machine* (SVM) menghasilkan *Accuracy* = 78.97% dan AUC = 0.851.

REFERENSI

- Buntoro, G. A. (2016). ANALISIS SENTIMEN HATESPEECH PADA TWITTER DENGAN METODE NAÏVE BAYES CLASSIFIER DAN SUPPORT VECTOR MACHINE. *Jurnal Dinamika Informatika*, 5(2). Retrieved from <http://ojs.upy.ac.id/ojs/index.php/dinf/article/view/975>
- Gusrion, D. D. (2017). PERANGKAT JARINGAN MOBILE PADA KENDARAAN LAYANAN KELILING. *Komputer Teknologi Informasi*, 4(2). Retrieved from <http://lppm.upiypk.ac.id/komtekinfo/index.php/KOMTEKINFO/article/view/113>
- Hidayatullah, A. F., & Azhari, A. S. (2014). ANALISIS SENTIMEN DAN KLASIFIKASI KATEGORI TERHADAP TOKOH PUBLIK PADA TWITTER. In *Seminar Nasional Informatika (SEMNASIF)* (pp. 115-122). Yogyakarta: UPN "Veteran" Yogyakarta. Retrieved from <http://jurnal.upnyk.ac.id/index.php/semnasif/article/view/1016>
- Kusmira, M. (2019). *Laporan Akhir Penelitian PDY: Analisis Sentimen Registrasi Ulang Kartu SIM Pada Twitter Menggunakan Algoritma SVM Dan K-NN*. Jakarta.
- Mahardhika, Y. S., & Zuliarso, E. (2018). ANALISIS SENTIMEN TERHADAP PEMERINTAHAN JOKO WIDODO PADA MEDIA SOSIAL TWITTER MENGGUNAKAN ALGORITMA NAIVES BAYES CLASSIFIER. In *SINTAK* (pp. 409-413). Semarang: Fakultas Teknologi Informasi Universitas Stikubank. Retrieved from <https://unisbank.ac.id/ojs/index.php/sintak/article/view/6651>
- Rustiana, D., & Rahayu, N. (2017). ANALISIS SENTIMEN PASAR OTOMOTIF MOBIL: TWEET TWITTER MENGGUNAKAN NAÏVE BAYES. *Simetris: Jurnal Teknik Mesin, Elektro Dan Ilmu Komputer*, 8(1), 113-120. <https://doi.org/10.24176/simet.v8i1.841>