

IMPLEMENTASI SUPPORT VECTOR MACHINE UNTUK ANALISIS SENTIMEN PENGGUNA TWITTER TERHADAP PELAYANAN TELKOM DAN BIZNET

Fadholi Fat Haranto¹; Bety Wulan Sari² *

¹Informatika, ²Sistem Informasi
Universitas AMIKOM Yogyakarta
www.amikom.ac.id

¹fadholi.haranto@students.amikom.ac.id, ²bety@amikom.ac.id

*Correspondent Author

Abstract— *Social media is a media that can be used for expression by its users. Twitter is quite popular and is often used in Indonesia, Twitter users can express and aspire without restrictions. Tweets in the form of expressions and aspirations written by Twitter users can be used to review a product or service. In this study, researchers used text mining techniques by applying the Support Vector Machine algorithm which is used for twitter user sentiment analysis of Telkom and Biznet services. Data on Telkom and Biznet services will be calculated in this study with a number of datasets of 500 tweets originating from twitter crawling data, there are 250 tweets that are used as datasets for each object. Some of the data will be used for training data and testing data in the modeling process using the Support Vector Machine algorithm. The method used for testing the model is Confusion Matrix while K-Fold Cross Validation is intended to divide the training data and testing data according to the folds used. The test results obtained using the K-Fold Cross Validation and Confusion Matrix methods on the model are made using the Support Vector Machine algorithm which results in 79.6% accuracy, 76.5% precision, 72.8% recall, and F1-score accuracy 74,6% for Telkom, and accuracy of 83.2%, precision 78.8%, recall 71.6%, and F1-score 75% for Biznet.*

Keywords: *Sentiment Analysis, Support Vector Machine, Classification, Biznet, Telkom.*

Intisari— Sosial media merupakan suatu media yang dapat digunakan untuk berekspresi oleh penggunanya. Twitter cukup populer dan sering digunakan di Indonesia, pengguna twitter dapat berekspresi dan beraspirasi tanpa adanya batasan. *Tweet* yang berupa ekspresi dan aspirasi yang ditulis oleh pengguna twitter dapat digunakan untuk ulasan sebuah produk atau layanan. Pada penelitian ini, peneliti menggunakan teknik *text mining* dengan menerapkan algoritma *Support Vector Machine* yang dipergunakan untuk analisis sentimen pengguna twitter terhadap pelayanan

Telkom dan Biznet. Data pada pelayanan Telkom dan Biznet akan dilakukan perhitungan pada penelitian ini dengan jumlah dataset sebanyak 500 *tweet* yang berasal dari *crawling* data twitter, terdapat 250 *tweet* yang dijadikan dataset pada masing-masing objek. Sejumlah data tersebut akan dipergunakan untuk data *training* serta data *testing* dalam proses pembuatan model menggunakan algoritma *Support Vector Machine*. Metode yang digunakan untuk pengujian model adalah *Confusion Matrix* sedangkan *K-Fold Cross Validation* ditujukan untuk untuk membagi data *training* dan data *testing* sesuai lipatan yang digunakan. Hasil pengujian yang diperoleh menggunakan metode *K-Fold Cross Validation* dan *Confusion Matrix* pada model yang dibuat menggunakan algoritma *Support Vector Machine* yang memberikan hasil nilai *accuracy* 79,6%, *precision* 76,5%, *recall* 72,8% , dan *F1-score* 74,6% untuk Telkom, serta *accuracy* 83,2%, *precision* 78,8%, *recall* 71,6%, dan *F1-score* 75% untuk Biznet.

Kata Kunci: *Analisis Sentimen, Support Vector Machine, Klasifikasi, Biznet, Telkom.*

PENDAHULUAN

Jumlah pengguna internet di Indonesia bertumbuh dengan pesat sebesar 10,12% dari tahun 2017 saat angka penetrasi sebanyak 54,46%. Hasil tersebut diperoleh dari Polling Indonesia dengan Asosiasi Penyelenggara Jasa Internet Indonesia. (Pratomo, 2019)

Saat ini media sosial seperti Facebook, Twitter, LinkedIn, YouTube menjadi bagian dari kehidupan manusia termasuk dalam hal berinteraksi dengan lainnya. Selain digunakan untuk interaksi sosial, media ini juga sebagai identitas diri ke masyarakat umum. (Susilo & Rochimah, 2013)

Dalam media sosial, setiap status yang diunggah oleh pengguna, tidak semuanya bermakna akademis. Sehingga dibutuhkan klasifikasi dokumen untuk dapat mengetahui perbedaan status bertopik akademis dari yang non-akademis. Status-status ini juga mempunyai sentimen dari penulisnya. Dengan melakukan analisis sentimen, dapat diketahui informasi tentang emosi pengguna (Susilo & Rochimah, 2013)

Opinion mining atau analisis sentiment merupakan bagian dari *text mining* yang melakukan studi mengenai opini orang-orang, sentimen, evaluasi, tingkah laku dan emosi terhadap suatu entitas seperti produk, layanan, organisasi, individu, permasalahan, topik, acara dan atribut-atributnya (Liu, 2012)

Hal ini membuat produk-produk besar maupun kecil, luar negeri maupun dalam negeri kemudian berlomba-lomba melakukan pemasaran via Twitter ke pasar Indonesia. Mereka membuat akun twitter dan kemudian menunggu *reply* atau *mention* dari pengguna lain untuk mengetahui bagaimana respon atas produk mereka (Monarizqa, Nugroho, & Hantono, 2014)

Hal yang paling penting untuk melakukan analisis komentar yang terdapat di Twitter adalah dengan melakukan analisis sentimen. Setelah dilakukan analisis sentimen, lalu diterjemahkan sehingga menjadi sesuatu yang lebih berarti yaitu salah satunya dalam bentuk rating. Rating ini menjadi penting bagi ranah bisnis karena bisa dijadikan salah satu parameter kesuksesan. (Monarizqa et al., 2014).

Banyak algoritma yang digunakan untuk melakukan analisis sentimen, disini peneliti akan menerapkan algoritma *Support Vector Machine* (SVM).

SVM merupakan salah satu metode klasifikasi dengan menggunakan metode *machine learning (supervised learning)* yang memprediksi kelas berdasarkan pola dari hasil proses *training* yang diciptakan oleh Vladimir Vapnik. Klasifikasi dilakukan dengan garis pembatas (*hyperlane*) yang memisahkan antara kelas opini positif dan opini negatif. Secara intuitif, suatu garis pembatas yang baik adalah yang memiliki jarak terbesar ke titik data pelatihan terdekat dari setiap kelas, karena pada umumnya semakin besar *margin*, semakin rendah *error* generalisasi dari pemilahan. Margin adalah jarak dari suatu titik vektor di suatu kelas terhadap *hyperplane* (Santoso, Virginia, & Lukito, 2017)

Penelitian lainnya dilakukan oleh (Preety & Dahiya, 2015) yang berjudul *Sentimen Analysis Using SVM and Naive Bayes Algorithm*. Penelitian ini membandingkan metode *Support Vector Machine* dengan metode *Naive Bayes* dimana

metode *Support Vector Machine* memiliki hasil akhir klasifikasi lebih tinggi daripada metode *Naive Bayes* yaitu 83.59% banding 79.66%.

Oleh karena itu peneliti bermaksud membuat sebuah aplikasi analisis sentimen pengguna twitter terhadap pelayanan Telkom dan Biznet menggunakan algoritma *Support Vector Machine* yang digunakan oleh Telkom dan Biznet, dengan tujuan mengetahui presentase sentimen negatif dan positif dari Telkom dan Biznet sehingga kedua perusahaan tersebut bisa meningkatkan pelayanan yang lebih baik.

BAHAN DAN METODE

Penelitian ini melakukan beberapa tahapan dengan metode sebagai berikut :

1. Pengumpulan data

Data *tweet* diambil dari twitter, pengambilan data dilakukan dengan cara *crawling* data yang ada di twitter yang berhubungan dengan Telkom dan Biznet. Data yang diambil adalah 500 data, pada masing-masing objek memiliki 250 *tweet* yang digunakan sebagai data *training*. 250 data Telkom sebanyak 103 *tweet* positif serta 147 *tweet* negatif, sedangkan Biznet terdapat 88 *tweet* positif serta 162 *tweet* negatif.

2. Preprocessing

Preprocessing menjadi proses awal sebelum melakukan klasifikasi. Proses ini digunakan untuk membersihkan data dari *noise* dan siap digunakan pada proses selanjutnya, berikut merupakan tahap-tahap *preprocessing*:

a. Cleaning

Pada proses *cleaning* kalimat dibersihkan dari *hashtag*, *mention*, dan juga tanda baca.

b. Case Folding

Merupakan proses penggantian huruf dari huruf yang bercampur (*lowercase* dan *uppercase*) menjadi semua huruf kecil.

c. Tokenization

Tokenization ialah proses perubahan kalimat menjadi kata.

d. Stopword

Merupakan penghilangan kata yang sering berbobot namun sering muncul.

3. Transformation

Term Frequency-Inverse Document Frequency (TF-IDF) digunakan pada proses *transformation* dengan cara merubah kalimat atau kata menjadi angka agar bisa diklasifikasi dan menggunakan *Cosine Similarity* untuk melihat kemiripan antar kata atau dokumen.

4. Klasifikasi

Untuk klasifikasi *text* menggunakan algoritma *Support Vector Machine* yang ditujukan untuk membuat model klasifikasi.

5. Evaluasi dan Hasil Validasi

Metode untuk mengevaluasi algoritma yang digunakan yaitu dengan *k-fold (10-fold) cross validation* serta *confusion matrix* yang menghasilkan *accuracy, precision, recall, f1-score*.

HASIL DAN PEMBAHASAN

Penerapan algoritma *Support Vector Machine* terhadap hasil dari analisis sentiment menghasilkan penjelasan di bawah ini.

Untuk mengevaluasi program peneliti menggunakan metode *10-Fold Cross Validation*, dan *Confussion matrix*. *Confussion Matrix* digunakan untuk mencari nilai *accuracy, precision, recall*, dan *F1-score*, sedangkan *10-Fold Cross Validation* digunakan untuk membagi dataset ke dalam data *training* dan data *testing*. Evaluasi ini ditujukan untuk menentukan presentase keakuratan dari suatu model yang dibuat. Dalam pengujian ini akan dilakukan prediksi terhadap data testing menggunakan data *training* hasil dari *Cross Validation*. Berikut merupakan tabel yang menunjukkan performa dari tiap-tiap *fold* yang diuji.

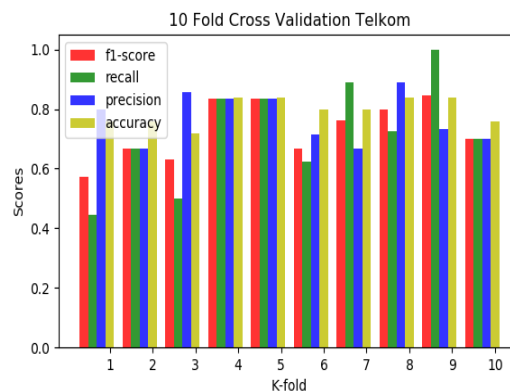
Tabel 1. Hasil 10 Fold Cross Validation Telkom

Fold	Parameter	Hasil(%)
Fold 1	Accuracy	76
	Precision	80
	Recall	44
	F1-score	57
	Jumlah Data	25
Fold 2	Accuracy	76
	Precision	67
	Recall	67
	F1-score	67
	Jumlah Data	25
Fold 3	Accuracy	72
	Precision	86
	Recall	50
	F1-score	63
	Jumlah Data	25
Fold 4	Accuracy	84
	Precision	83
	Recall	83
	F1-score	83
	Jumlah Data	25
Fold 5	Accuracy	84
	Precision	83
	Recall	83
	F1-score	83
	Jumlah Data	25
Fold 6	Accuracy	80
	Precision	71
	Recall	63
	F1-score	67
	Jumlah Data	25
Fold 7	Accuracy	80
	Precision	67
	Recall	89

Fold	Parameter	Hasil(%)
Fold 8	F1-score	76
	Jumlah Data	25
	Accuracy	84
	Precision	89
	Recall	73
Fold 9	F1-score	80
	Jumlah Data	25
	Accuracy	84
	Precision	73
	Recall	100
Fold 10	F1-score	85
	Jumlah Data	25
	Accuracy	76
	Precision	70
	Recall	70
Rata-Rata	F1-score	70
	Jumlah Data	25
	Accuracy	83,2
	Precision	78,8
	Recall	71,6
Jumlah Data		250

Sumber : (Haranto & Sari, 2019)

Dari Tabel 1, akurasi tertinggi berada di fold ke-4, fold ke-5, fold ke-8 dan fold ke-9 dengan nilai 84%. Pada fold ke-4 dan fold ke-5 memiliki nilai *precision* 83%, *recall* 83%, dan *F1-score* 83%, fold ke-8 memiliki nilai *precision* 89%, *recall* 73%, dan *F1-score* 80%, fold ke-9 memiliki nilai *precision* 73%, *recall* 100%, dan *F1-score* 85%. Adapun hasil akurasi terendah yang dihasilkan oleh model yaitu 72% yang berada pada fold ke-3. Pada fold ke-3 memiliki nilai *precision* 86%, *recall* 50%, dan *F1-score* 63%. Hasil performa dari *confusion matrix* pada 10-fold Cross Validation memiliki rata-rata nilai *accuracy* 79,6%, *precision* 76,5%, *recall* 72,8%, dan 74,6%. Data di masukkan ke dalam grafik bar chart, sehingga dapat divisualisasikan seperti pada Gambar 1.



Rata-rata Accuracy : 79.6%, Precision : 76.5%, Recall : 72.8%, F1-Score : 74.6%
 jika accuracy, precision, recall, f1-score memiliki nilai yang tinggi maka model dapat mengklasifikasi dengan baik

Sumber : (Haranto & Sari, 2019)

Gambar 1. Hasil 10 Fold Cross Validation Telkom

Tabel 2. Hasil 10 Fold Cross Validation Biznet

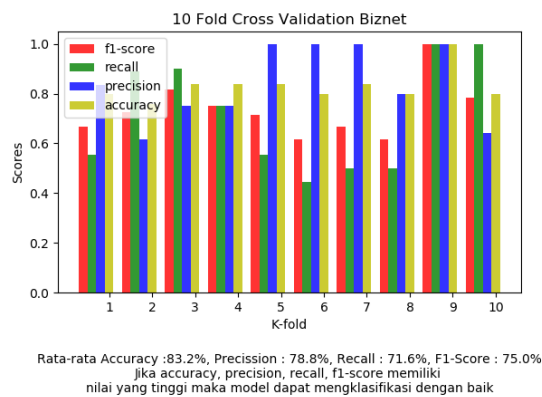
Fold	Parameter	Hasil(%)
Fold 1	Accuracy	80
	Precision	83
	Recall	56
	F1-score	67
	Jumlah Data	25
Fold 2	Accuracy	76
	Precision	62
	Recall	89
	F1-score	73
	Jumlah Data	25
Fold 3	Accuracy	84
	Precision	75
	Recall	90
	F1-score	82
	Jumlah Data	25
Fold 4	Accuracy	84
	Precision	75
	Recall	75
	F1-score	75
	Jumlah Data	25
Fold 5	Accuracy	84
	Precision	100
	Recall	56
	F1-score	71
	Jumlah Data	25
Fold 6	Accuracy	80
	Precision	100
	Recall	44
	F1-score	62
	Jumlah Data	25
Fold 7	Accuracy	84
	Precision	100
	Recall	50
	F1-score	67
	Jumlah Data	25
Fold 8	Accuracy	80
	Precision	80
	Recall	50
	F1-score	62
	Jumlah Data	25
Fold 9	Accuracy	100
	Precision	100
	Recall	100
	F1-score	100
	Jumlah Data	25
Fold 10	Accuracy	80
	Precision	64
	Recall	100
	F1-score	78
	Jumlah Data	25
Rata-Rata	Accuracy	83,2
	Precision	78,8

Fold	Parameter	Hasil(%)
	Recall	71,6
	F1-score	75
	Jumlah Data	250

Sumber : (Haranto & Sari, 2019)

Dari

Tabel 2, akurasi tertinggi berada pada fold ke-9 dengan nilai 96%. Pada fold ke 9, memiliki nilai *precision* 100%, *recall* 100%, dan *F1-score* 100%. Adapun hasil akurasi terendah yang dihasilkan oleh model yaitu 76% yang berada pada fold ke-2. Pada fold ke-2 memiliki nilai *precision* 62%, *recall* 82%, dan *F1-score* 73%. *Confusion matrix* memberikan hasil rata-rata performa pada 10-fold Cross Validation memiliki nilai *accuracy* 83,2%, *precision* 78,8%, *recall* 71,6%, dan *F1-score* 75%. Data di masukkan ke dalam grafik bar chart, sehingga dapat divisualisasikan seperti pada Gambar 2



Sumber : (Haranto & Sari, 2019)

Gambar 2. Hasil 10 Fold Cross Validation Biznet

Tabel 3. Confusion Matrix Telkom

	Prediksi Negatif	Prediksi Positif	Jumlah
Sebenarnya Negatif	124	23	147
Sebenarnya Positif	28	75	103
Jumlah	152	98	250

Sumber : (Haranto & Sari, 2019)

Pada

Tabel 3 terdapat 152 data yang diprediksi negatif, dan 98 data yang diprediksi positif. Namun dari 152 data yang diprediksi negatif terdapat kesalahan prediksi sebanyak 28 data, yang seharusnya berada pada kelas positif. Untuk 98

data yang diprediksi pada kelas positif terdapat 23 data salah prediksi dan seharusnya berada di kelas negatif.

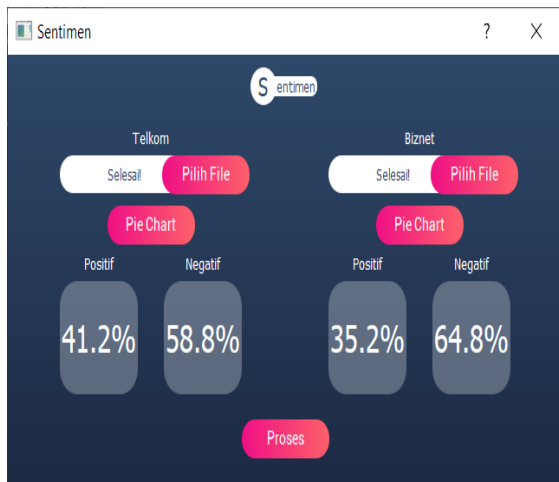
Tabel 4. Confusion Matrix Biznet

	Prediksi Negatif	Prediksi Positif	Jumlah
Sebenarnya Negatif	145	17	162
Sebenarnya Positif	25	63	88
Jumlah	170	80	250

Sumber : (Haranto & Sari, 2019)

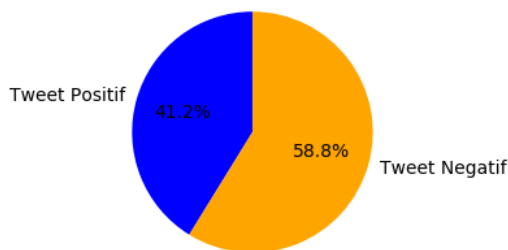
Pada Tabel 4 terdapat 170 data yang diprediksi negatif, dan 80 data yang diprediksi positif. Namun dari 170 data yang diprediksi negatif terdapat kesalahan prediksi sebanyak 25 data, yang seharusnya berada pada kelas positif. Untuk 80 data yang diprediksi pada kelas positif terdapat 17 data salah prediksi dan seharusnya berada di kelas negatif.

Berikut merupakan hasil sentimen Telkom dan Biznet menggunakan data baru yang diambil dari twitter.



Sumber : (Haranto & Sari, 2019)
Gambar 3. Hasil Sentimen

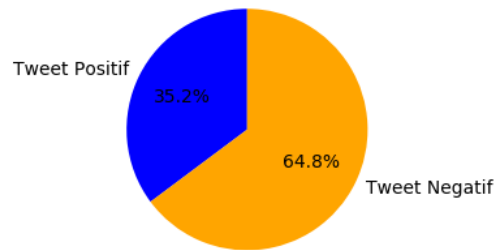
Presentase Prediksi Telkom



Dari hasil diagram diatas dapat disimpulkan bahwa sentimen masyarakat kepada Telkom adalah negatif
Sumber : (Haranto & Sari, 2019)

Gambar 4. Pie chart prediksi Telkom

Presentase Prediksi Biznet



Dari hasil diagram diatas dapat disimpulkan bahwa sentimen masyarakat kepada Biznet adalah negatif
Sumber : (Haranto & Sari, 2019)

Gambar 5. Pie chart prediksi Telkom

Dari diagram *pie* diatas bisa disimpulkan bahwa bagian yang berwarna biru merupakan presentase positif dan bagian warna oranye adalah persentase negatif. Pada Gambar 4 diagram *pie* Telkom mendapat nilai positif sebesar 41,2% dan negatif sebesar 58,8% yang menandakan sentimen masyarakat kepada Telkom adalah negatif. Sedangkan pada Gambar 5 diagram *pie* Biznet mendapat nilai positif sebesar 35,2% dan negatif sebesar 64,6% yang menandakan sentimen masyarakat kepada Biznet adalah negatif. Untuk melihat mana yang lebih bagus pelayanannya bisa dilihat presentase pada diagram *pie* kedua perusahaan tersebut. Karena Telkom dan Biznet sama-sama mendapatkan sentimen negatif untuk melihat mana yang lebih bagus kita bisa melihat pada presentase positif, dari hasil di atas Telkom lebih baik dibandingkan Biznet karena memiliki presentase positif lebih banyak dari Telkom yaitu 58,8%, sedangkan Biznet hanya 35,2%. Untuk presentase positif dan negatif dapat berubah sewaktu-waktu sesuai dengan opini yang diberikan masyarakat kepada Biznet maupun Telkom.

KESIMPULAN

Penelitian sentimen pengguna twitter terhadap Biznet dan Telkom menggunakan Algoritma *Support Vector Machine* mendapatkan hasil persentase nilai positif sebesar 41,2% dan negatif sebesar 58,8% untuk Telkom dan persentase nilai positif sebesar 35,2% dan negatif sebesar 64,8% untuk Biznet. Dari presentase yang didapatkan dapat disimpulkan bahwa Telkom lebih baik pelayanannya dibandingkan Biznet dikarenakan presentase positif Telkom lebih banyak dari Biznet. Untuk presentase positif dan negatif dapat berubah sewaktu-waktu sesuai dengan opini yang diberikan masyarakat kepada Biznet maupun Telkom.

Untuk mengevaluasi model *Support Vector Machine* yang digunakan, diterapkan metode *K-10 Fold Cross Validation* dan *Confusion Matrix* yang

menghasilkan rata-rata nilai *accuracy* 79,6%, *precision* 76,5%, *recall* 72,8%, dan *f1-score* sebesar 74,6% hasil dari Telkom, dan nilai *accuracy* 83,2%, *precision* 78,8%, *recall* 71,6%, dan *f1-score* 75% hasil dari Biznet. Dengan demikian Algoritma *Support Vector Machine* cocok dan dapat dipergunakan untuk analisis sentimen terhadap data *tweet* Telkom dan Biznet.

PENGLASIFIKASIAN TOPIK DAN ANALISIS SENTIMEN DALAM MEDIA SOSIAL. In *SNASTI 2013* (pp. SC1-SC10). Surabaya: STIKOM Surabaya. Retrieved from <https://core.ac.uk/download/pdf/35357695.pdf>

REFERENSI

- Haranto, F. F., & Sari, B. W. (2019). *Laporan Akhir Penelitian: Implementasi Support Vector Machine Untuk Analisis Sentimen Pengguna Twitter Terhadap Pelayanan Telkom Dan Biznet*. Yogyakarta.
- Liu, B. (2012). Sentiment Analysis and Opinion Mining. *Synthesis Lectures on Human Language Technologies*, 5(1), 1-167. <https://doi.org/10.2200/S00416ED1V01Y201204HLT016>
- Monarizqa, N., Nugroho, L. E., & Hantono, B. S. (2014). PENERAPAN ANALISIS SENTIMEN PADA TWITTER BERBAHASA INDONESIA SEBAGAI PEMBERI RATING. *Jurnal Penelitian Teknik Elektro Dan Teknologi Informasi*, 1(3). Retrieved from <http://ejpteti.jteti.ugm.ac.id/index.php/JPTE TI/article/view/31>
- Pratomo, Y. (2019, May 16). APJII: Jumlah Pengguna Internet di Indonesia Tembus 171 Juta Jiwa. *Tekno.Kompas.Com*, p. Internet. Retrieved from <https://tekno.kompas.com/read/2019/05/16/03260037/apjii-jumlah-pengguna-internet-di-indonesia-tembus-171-juta-jiwa>
- Preety, P., & Dahiya, S. (2015). SENTIMENT ANALYSIS USING SVM AND NAÏVE BAYES ALGORITHM. *International Journal of Computer Science and Mobile Computing*, 4(9), 212-219. Retrieved from <https://ijcsmc.com/docs/papers/September 2015/V4I9201545.pdf>
- Santoso, V. I., Virginia, G., & Lukito, Y. (2017). PENERAPAN SENTIMENT ANALYSIS PADA HASIL EVALUASI DOSEN DENGAN METODE SUPPORT VECTOR MACHINE. *Jurnal Transformatika*, 14(2), 72. <https://doi.org/10.26623/transformatika.v14i2.439>
- Susilo, T. H., & Rochimah, S. (2013).