

ANALISIS SENTIMEN TERKAIT VAKSIN COVID-19 PADA DATA TWITTER MENGUNAKAN *SUPPORT VECTOR MACHINE*

Muhammad Dani Mulyawan¹, Isnandar Slamet², Respatiwan³

¹ Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Sebelas Maret
email: mmdeenm@student.uns.ac.id

² Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Sebelas Maret
email: penulis: isnandarlamet@staff.uns.ac.id

³ Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Sebelas Maret
email: respatiwan@staff.uns.ac.id

Abstract

Pemerintah Indonesia telah mengumumkan bahwa Indonesia telah terkena dampak dari Covid-19 sejak 2 Maret 2020. Upaya telah dilakukan oleh berbagai negara termasuk Indonesia, namun dengan beredarnya persepsi, rumor, dan banyak informasi yang masuk ke masyarakat mengenai vaksin, menyebabkan timbulnya pro dan kontra terkait upaya vaksinasi yang dilakukan pemerintah. Oleh karena itu, analisis sentimen mengenai kebijakan yang dilakukan pemerintah terkait vaksin dirasa perlu. Tujuan penelitian ini adalah mengetahui kecenderungan sentimen terkait vaksin Covid-19 di Indonesia berdasarkan hasil dari klasifikasi. Metode klasifikasi yang digunakan pada penelitian ini adalah Support Vector Machine (SVM). Metode klasifikasi SVM dipilih karena memiliki kemampuan generalisasi dalam mengklasifikasikan suatu pattern, tidak termasuk data yang dipakai dalam fase pembelajaran metode tersebut. Klasifikasi menggunakan SVM kernel linear dengan pembobotan TF-IDF dan pembagian data menggunakan K-fold cross validation dengan nilai k=10. Klasifikasi dilakukan menjadi dua kelas, yaitu positif dan negatif. Setelah dilakukan pra-proses dan klasifikasi, didapatkan nilai f1, akurasi, presisi, dan recall sebagai nilai acuan dalam mengevaluasi klasifikasi. Hasil penelitian menunjukkan bahwa SVM cukup baik dalam mengklasifikasikan data pada penelitian ini dengan nilai f1= 89.219%, akurasi= 82.738%, presisi= 83.333%, dan recall= 96%. Nilai tersebut dapat dikategorikan baik, sehingga SVM dapat dikatakan baik untuk mengklasifikasikan data sentimen terkait vaksin Covid-19 di Indonesia. Penelitian lanjutan dapat dilakukan dengan data tweet yang lebih kaya, pra-proses yang lebih mendalam, serta membandingkannya dengan teknik klasifikasi lainnya guna mendapat nilai evaluasi klasifikasi yang lebih baik.

Keywords: *analisis sentimen, klasifikasi, vaksin covid-19, Twitter, SVM*

1. PENDAHULUAN

Awal tahun 2020, dunia dikejutkan dengan munculnya suatu wabah yang kemudian menyebar dengan cepat ke lebih dari 190 negara dan teritori. Wabah ini diberi nama *coronavirus disease 2019* (Covid-19) yang disebabkan oleh *Severe Acute Respiratory Syndrome Coronavirus-2* (SARS-CoV-2). Sejak 12 Maret 2020, *World Health Organization* (WHO) telah menetapkan wabah virus ini menjadi pandemi. Peran pemerintah dalam mengambil kebijakan adalah kunci utama dalam penanganan pandemi ini. Oleh karena itu, perlu segera dilakukan intervensi tidak hanya dari sisi penerapan protokol kesehatan namun juga diperlukan intervensi lain yang efektif untuk

memutuskan mata rantai penularan penyakit, yaitu melalui upaya vaksinasi. Upaya telah dilakukan oleh berbagai negara termasuk Indonesia, namun dengan beredarnya persepsi, rumor, dan banyak informasi yang masuk ke masyarakat mengenai vaksin, menyebabkan timbulnya pro dan kontra terkait upaya vaksinasi yang dilakukan pemerintah. Oleh karena itu, analisis sentimen mengenai kebijakan yang dilakukan pemerintah terkait vaksin dirasa perlu untuk mengetahui kecenderungan sentimen terkait vaksin Covid-19 di Indonesia.

Data yang dirilis oleh Badan Pusat Statistik pada tahun 2019 sekitar 82.66% masyarakat perkotaan maupun pedesaan yang usianya lima tahun ke atas mengakses internet

untuk keperluan media sosial (BPS, 2019). Media sosial merupakan alat yang digunakan untuk berbagi informasi dengan cepat dan mudah. Twitter adalah media sosial yang memiliki karakteristik dan format yang unik dengan simbol ataupun aturan khusus. Pengguna Twitter hanya dapat mengirim dan membaca pesan blog seperti pada umumnya dengan batas maksimal sejumlah 140 karakter, pesan tersebut dikenal dengan *tweet* (Zhang *et al.*, 2011). *Tweet* yang di-posting pengguna beraneka ragam, dapat berupa pendapat, fakta, saran, atau kritik terhadap sesuatu. Pink (2018) mengungkapkan bahwa

Twitter merupakan sumber mengukur pendapat dan kondisi emosional yang terbaik. Hampir satu miliar manusia memiliki akun, dan mereka memposting sekitar 6000 *tweets* per detik. Jumlah pesan mini yang sangat besar ini telah menghasilkan lautan data yang bisa digunakan oleh ilmuwan untuk memahami perilaku manusia. Metode klasifikasi yang digunakan pada penelitian ini adalah *Support Vector Machine* (SVM). Metode klasifikasi SVM dipilih karena memiliki kemampuan generalisasi dalam mengklasifikasikan suatu pattern, tidak termasuk data yang dipakai dalam fase pembelajaran metode tersebut (Nugroho *et al.*, 2003). SVM merupakan metode klasifikasi yang mempunyai keunggulan dari metode lain, yaitu selain menggunakan menggunakan jarak sebagai penentunya, metode ini juga menggunakan vektor sebagai syaratnya, sehingga akurasi yang dihasilkan akan lebih baik dibandingkan dengan metode lain.

2. KAJIAN LITERATUR DAN PEGEMBANGAN HIPOTESIS

Analisis Sentimen.

Analisis sentimen adalah salah satu cabang penelitian dari *Text Mining* yang berguna untuk mengklasifikasi dokumen teks berupa opini berdasarkan sentimen. Analisis sentimen dapat digunakan untuk memberikan nilai positif, negatif, atau netral pendapat seseorang tentang topik yang berbeda yang ada dalam *tweet* (Sunni dan Widyanoro, 2012). Analisis sentimen dapat disebut juga *opinion mining* yang bertujuan untuk menganalisis, memahami, mengolah dan mengestrak data tekstual yang berupa opini

terhadap entitas tentang topik tertentu agar mendapatkan suatu informasi.

Term Frequency - Inverse Document Frequency (TF-IDF)

Metode TF-IDF merupakan suatu cara untuk memberikan bobot hubungan suatu kata (*term*) terhadap dokumen. Metode ini menggabungkan dua konsep untuk perhitungan bobot, yaitu frekuensi kemunculan sebuah kata di dalam sebuah dokumen tertentu dan inverse frekuensi dokumen yang mengandung kata tersebut. Frekuensi kemunculan kata di dalam dokumen yang diberikan menunjukkan seberapa penting kata itu di dalam dokumen tersebut. Frekuensi dokumen yang mengandung kata tersebut menunjukkan seberapa umum kata tersebut. Sehingga bobot hubungan antara sebuah kata dan sebuah dokumen akan tinggi apabila frekuensi kata tersebut tinggi di dalam dokumen dan frekuensi keseluruhan dokumen yang mengandung kata tersebut yang rendah pada kumpulan dokumen (Nurjannah *et.al.*, 2013). Nilai TF-IDF didapatkan dengan mengalikan dua persamaan yang diformulasikan sebagai berikut:

$$tf = 0.5 + 0.5 \times \frac{tf}{\max(tf)}$$

$$idf_t = \log\left(\frac{D}{df_t}\right)$$

$$W_{d,t} = tf_{d,t} \times IDF_{d,t}$$

Keterangan:

tf: banyaknya term yang dicari dalam sebuah dokumen

D: total dokumen

df_t: jumlah dokumen yang mengandung term *t*

IDF: *Inversed Document Frequency*

d: dokumen ke- *d*

t: kata ke-*t* dari kata kunci

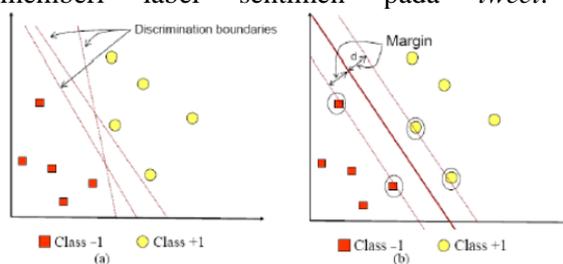
W_{d,t}: bobot dokumen ke-*d* terhadap kata ke-*t*

Support Vector Machine (SVM)

SVM merupakan salah satu metode klasifikasi dengan menggunakan *machine learning (supervised learning)* yang memprediksi kelas berdasarkan model atau pola dari hasil proses *training*. Klasifikasi dilakukan dengan mencari *hyperplane* atau garis pembatas (*decision boundary*) yang memisahkan antara suatu kelas dengan kelas

lain, yang dalam kasus ini garis tersebut berperan memisahkan *tweet* bersentimen positif (berlabel +1) dengan *tweet* bersentimen negatif (berlabel 0). SVM melakukan pencarian nilai *hyperplane* dengan menggunakan *support vector* dan nilai margin.

Ilustrasi metode SVM dapat dilihat pada Gambar 2.1, yang menunjukkan penyebaran data yang digambarkan dengan warna merah (kotak) dan warna kuning (lingkaran). Data berwarna merah merupakan anggota dari kelas -1 dan data berwarna kuning adalah anggota dari kelas +1. Masalah utama dari klasifikasi adalah mencari *hyperplane* pemisah antara kedua kelas. Penelitian ini mempunyai data masukan yang memiliki representasi vektor didapatkan dari proses pembobotan (Nugroho *et al.*, 2003). *Training* data dilakukan pada klasifikasi SVM, maka akan menghasilkan sebuah nilai atau pola yang akan digunakan pada proses pengujian untuk proses SVM, yang bertujuan memberi label sentimen pada *tweet*.



Gambar 2.1. Ilustrasi SVM

K-Fold Cross Validation.

Teknik pembagian data yang kerap digunakan untuk mengestimasi performa atau kualitas suatu model adalah *K-fold cross validation*, yang bertujuan untuk mendapatkan model data latih yang baik. *Dataset* akan dibagi sebanyak k *subset* data yang jumlahnya sama besar dan melakukan uji coba dengan satu *subset* data menjadi data uji dan *subset* data lainnya menjadi data latih (Zaki dan Meira, 2014). *K-Fold Cross Validation* akan membagi data ke dalam k buah partisi atau disebut dengan fold $K_1; K_2; \dots; K_k$ dengan ukuran kelas yang sama. Pelatihan dan pengujian dilakukan sebanyak k kali, dalam perulangan ke- i partisi K_1 akan menjadi data uji, $K_2; K_3; \dots; K_k$ akan menjadi data latih dan seterusnya (Han *et al.*, 2011).

Confusion Matrix.

Confusion matrix atau juga disebut matriks klasifikasi adalah suatu alat visual yang digunakan dalam *supervised learning* (Manning *et al.*, 2008). *Confusion matrix* berisi jumlah kasus yang diklasifikasikan benar dan jumlah kasus yang diklasifikasikan salah. Evaluasi atau pengujian dilakukan pada data uji terhadap data yang sudah dilatih yang divisualisasikan ke dalam *confusion matrix*. Nilai akurasi dari evaluasi mengukur efektivitas dari sistem klasifikasi. Semakin tinggi nilai akurasi maka semakin baik sistem klasifikasi, karena pada analisis ini menggunakan dua kelas, yaitu positif dan negatif maka tabel *confusion matrix*-nya berbentuk 2×2 yang terdiri dari *predict class* dan *actual class*.

Tabel 2.1. *Confusion Matrix 2x2*

		<i>Predict Class</i>	
		Kelas A	Kelas B
<i>Actual Class</i>	Kelas A	AA	AB
	Kelas B	BA	BB

Kelas A merupakan representasi dari sentimen dengan kelas positif dan kelas B untuk sentimen yang dengan kelas negatif. Akurasi merupakan nilai yang menggambarkan seberapa akurat model dapat mengklasifikasikan dengan benar. Akurasi merupakan rasio prediksi benar (positif dan negatif) dengan keseluruhan data, dengan kata lain, akurasi merupakan tingkat kedekatan nilai prediksi dengan nilai aktual (sebenarnya). Berikut merupakan persamaan untuk menghitung nilai akurasi:

$$\text{Akurasi} = \frac{AA + BB}{AA + AB + BA + BB}$$

Evaluasi klasifikasi juga dapat dilakukan dengan melihat nilai presisi dan *recall*. Presisi merupakan probabilitas bahwa sebuah item yang terpilih adalah relevan dan menggambarkan tingkat keakuratan antara data yang diminta dengan hasil prediksi yang diberikan oleh model. Presisi merupakan rasio prediksi benar positif dibandingkan dengan keseluruhan hasil yang diprediksi positif. Dari semua kelas positif yang telah di prediksi dengan benar, berapa banyak data yang benar-benar positif. Nilai presisi tiap kelas dapat diperoleh sebagai berikut:

$$\text{Presisi } i = \frac{A_i}{A_i + B_i}$$

Jumlah presisi dapat diperoleh dengan jumlah presisi tiap kelas, lalu dibagi

jumlah kelasnya. Sedangkan *recall* adalah rasio dari item yang relevan yang dipilih terhadap total jumlah item yang relevan. *Recall* menggambarkan keberhasilan model dalam menemukan kembali sebuah informasi. Maka, *recall* merupakan rasio prediksi benar positif dibandingkan dengan keseluruhan data yang benar positif. Nilai *recall* dapat diperoleh dengan persamaan sebagai berikut:

$$Recall\ i = \frac{iA}{iA + iB}$$

Sama halnya seperti presisi, untuk menghitung nilai keseluruhan dari *recall* adalah total *recall* tiap kelas dibagi dengan jumlah kelas (Iskandar dan Suprpto, 2015). Selanjutnya ada nilai *f1* yang merupakan perbandingan rata-rata presisi dan *recall* yang dibobotkan, dengan persamaan sebagai berikut:

$$f1 = \frac{2 \times (recall \times presisi)}{(recall + presisi)}$$

3. METODE PENELITIAN

Penelitian ini merupakan studi kasus pada data Twitter mengenai vaksin Covid-19 di Indonesia untuk menghasilkan klasifikasi dari sentimen masyarakat. Seluruh proses analisis yang dilakukan menggunakan Bahasa pemrograman Python. Berikut merupakan langkah-langkah yang akan dilakukan pada penelitian ini.

Persiapan Data

Data yang digunakan merupakan data sekunder melalui *Application Programming Interface* (API) yang disediakan Twitter kemudian data tersebut dilakukan proses memberikan label (positif, negatif, dan netral) secara manual (bersifat subjektif). Data yang terkumpul dan siap diolah berbentuk *tweets* berjumlah 1675 data *tweets* yang merupakan cuitan pengguna Twitter di Indonesia terkait vaksin Covid-19 dengan rentang waktu pengambilan data sejak 10 Januari 2021- 10 April 2021.

Studi Pustaka

Pada tahapan ini merupakan proses pengumpulan data dan pengumpulan informasi metode yang akan digunakan dengan cara mempelajari berbagai referensi (buku, laporan, hasil penelitian, jurnal, ataupun artikel) yang berhubungan dengan masalah yang diteliti.

Praproses

Sebelum data dilakukan klasifikasi perlu dilakukan praproses untuk mengubah bentuk dokumen menjadi data yang terstruktur sesuai kebutuhannya agar dapat diolah lebih lanjut dalam proses *text mining*. Tahapan praproses teks dalam klasifikasi bertujuan untuk meningkatkan akurasi klasifikasi data. Praproses dalam *text mining* cukup rumit karena dalam Bahasa Indonesia terdapat berbagai aturan penulisan kalimat maupun pembentukan kata berimbuhan. Urutan tahapan dalam praproses meliputi: pembersihan data yaitu dengan cara menghilangkan *tweet* yang duplikat. *Tweet* yang duplikat dalam Twitter biasanya berupa *retweet* dengan simbol “RT”, maka *tweet* tersebut akan dihapus. Selain itu, *tweet* dengan bahasa asing akan dihapus, karena data yang digunakan dalam penelitian ini adalah data Twitter Bahasa Indonesia, menghilangkan link URL dan *username*, melakukan *case folding*, yaitu mengubah semua teks dengan huruf kecil (non-kapital) serta menghilangkan tanda baca, *whitespace* (spasi, *tab*, *newline*) digunakan sebagai pemisah antara kata yang akan dipotong, menghapus kata pada *tweet* yang terdapat dalam daftar *stopwords* (*term* atau kata-kata yang tidak berhubungan dengan dokumen meskipun kata tersebut sering muncul di dalam dokumen, tetapi jika dihapus tidak mengubah makna dari *tweet*), melakukan *stemming* untuk menghilangkan kata imbuhan dan mendapatkan kata dasar, mengubah data *tweet* kedalam bentuk frekuensi kemunculan kata (Kurniawan, 2017). Data yang telah melalui praproses, selanjutnya akan diberi label secara manual sesuai dengan kelasnya. Penentuan sentimen secara manual ini bersifat subjektif, karena ditentukan menurut pendapat dari peneliti berdasarkan kelas sentimen yang telah ditentukan sebelumnya yaitu positif, negatif, dan netral. Proses ini dilakukan agar dapat mengetahui pengelompokan sentimen pada data yang diperoleh sebelumnya.

K-Fold Cross Validation

Teknik pembagian data dilakukan dengan *K-fold cross validation* dengan menggunakan nilai $k=10$.

Klasifikasi Data

Teknik klasifikasi yang digunakan adalah SVM. Setelah didapatkan hasil

klasifikasi maka akan dilakukan evaluasi hasil klasifikasi, hingga akhirnya menginterpretasikannya dan menarik kesimpulan.

Evaluasi Klasifikasi

Langkah selanjutnya adalah melakukan evaluasi. Evaluasi bertujuan untuk mengetahui tingkat akurasi hasil klasifikasi data menggunakan metode SVM. Evaluasi dapat diukur dengan menghitung nilai akurasi dari hasil klasifikasi. Nilai akurasi dari klasifikasi dapat diperoleh dengan menggunakan *confusion matrix*.

4. HASIL DAN PEMBAHASAN

Sebelum melakukan analisis, langkah pertama adalah mempersiapkan data yang siap untuk diolah.

Persiapan Data

Data dikumpulkan menggunakan bantuan *package* Python bernama *tweepy* melalui API Twitter Bahasa Indonesia dengan kata kunci “vaksin” tanpa menyertakan data *retweet*. Data yang terkumpul disimpan dalam bentuk *comma separated values* (csv) untuk nantinya dipilah dan diberi label secara manual, hingga terkumpul data berjumlah 1675 *tweets* yang siap diolah ke tahap selanjutnya.

Hasil dari pelabelan data adalah sebagai berikut:

Tabel 4.1. Persebaran Data tiap Kelas

Label	Jumlah Data
-1 (negatif)	431
0 (netral)	661
1 (positif)	584

Data yang telah dibentuk ke dalam tiga kelas tersebut, diubah menjadi dua kelas dengan menggabungkan *tweet* berlabel netral dengan data berlabel positif, dan menjadikan bentuk data baru yang akan diolah adalah 1245 data berlabel 1 (positif) dan data berlabel 0 (negatif).

Praproses Data

Data yang telah diolah sebelumnya masih terdapat komponen-komponen yang tidak diperlukan atau akan mengurangi keakuratan dari hasil klasifikasi. Oleh karena itu diperlukan pembersihan data yang berupa:

1. melakukan *case folding*, yaitu mengubah semua teks dengan huruf kecil (non-kapital)
 2. menghilangkan tanda baca
 3. menghapus ASCII (*American Standard Code for Information Interchange*) dan Unicode yang merupakan bentuk dari kode karakter menghilangkan tanda baca, *whitespace* (spasi, *tab*, *newline*)
 4. menghapus kata pada *tweet* yang terdapat dalam daftar *stopwords* (*term* atau kata-kata yang tidak berhubungan dengan dokumen meskipun kata tersebut sering muncul di dalam dokumen, tetapi jika dihapus tidak mengubah makna dari *tweet*)
 5. serta melakukan *stemming* untuk menghilangkan kata imbuhan dan mendapatkan kata dasar
- Setelah melakukan pembersihan terhadap data, maka akan dilakukan pembagian data uji dan data latih dengan ukuran data uji sebesar 0.1.

Klasifikasi Data

Sebelum melakukan klasifikasi, data teks perlu dirubah menjadi vektor dan dilakukan pembobotan kata dengan TF-IDF. Klasifikasi dilakukan menggunakan metode SVM dengan kernel linear dan menggunakan bantuan *package* dari Python, yaitu *sklearn*. Pada proses klasifikasi data diuji dengan menggunakan metode *10 fold cross validation*.

Evaluasi Klasifikasi

Hasil analisis klasifikasi menggunakan teknik SVM dengan penggunaan data sebanyak 1675 selanjutnya dievaluasi. Evaluasi klasifikasi dilakukan menggunakan nilai *f1*, akurasi, presisi, dan *recall*, berikut tabel dari nilai-nilai komponen evaluasi klasifikasi:

Tabel 4.2. Nilai Hasil Evaluasi Klasifikasi

f1	akurasi	presisi	recall
89.219%	82.738%	83.333%	96%

Hasil dari evaluasi klasifikasi menunjukkan nilai yang bagus dalam proses klasifikasi. Evaluasi klasifikasi juga dapat divisualisasikan dalam bentuk *confusion matrix* sebagai berikut:

Tabel 4.3. *Confusion Matrix*

		Predict Class	
		Kelas A	Kelas B
Actual Class	Kelas A	120	24
	Kelas B	5	5

Kelas A merupakan representasi dari sentimen dengan kelas positif dan kelas B untuk sentimen yang dengan kelas negatif. Selanjutnya akan dicoba memasukkan data sentimen baru dan melihat apakah data tersebut diklasifikasikan dengan tepat. Kalimat pertama yang akan diuji adalah “mari sukseskan program vaksin” dan kalimat kedua adalah “masih takut akan efek samping dari vaksin”. Hasil keluaran menunjukkan data masukan berupa kalimat pertama adalah nilai bobot 1 dan kalimat kedua adalah 0 yang menunjukkan kalimat pertama merupakan kalimat dengan sentimen positif dan kalimat kedua masuk dalam kalimat dengan sentimen negatif.

5. KESIMPULAN

Analisis sentiment terkait vaksin Covid-19 pada data Twitter dilakukan dengan teknik klasifikasi SVM kernel linear dengan pembobotan TF-IDF dan pembagian data menggunakan *K-fold cross validation* dengan nilai $k=10$. Klasifikasi dilakukan menjadi dua kelas, yaitu positif dan negatif. Setelah dilakukan praproses dan klasifikasi, didapatkan nilai $f1$, akurasi, presisi, dan *recall* sebagai nilai acuan dalam mengevaluasi klasifikasi dengan nilai $f1= 89.219\%$, akurasi= 82.738% , presisi= 83.333% , dan *recall*= 96% . Nilai tersebut dapat dikategorikan baik, sehingga SVM dapat dikatakan baik untuk mengklasifikasikan data sentimen terkait vaksin Covid-19 di Indonesia.

6. REFERENSI

Badan Pusat Statistik, *Statistik Telekomunikasi Indonesia 2018*, Badan Pusat Statistik Indonesia, Jakarta, 2019.

- Han J., Kamber M., and Pei J., *Data Mining Concepts and Techniques Third Edition*, Elsevier Inc, Waltham, 2011.
- Iskandar D. dan Suprpto Y.K., Perbandingan Akurasi Klasifikasi Tingkat Kemiskinan antara Algoritma C 4.5 dan *Naïve Bayes*, *Jurnal Ilmiah NERO 2* (2015), no. 01, 37-43.
- Kurniawan T., *Implementasi Text Mining Pada Analisis Sentimen Pengguna Twitter terhadap Media Mainstream Menggunakan Naïve Bayes Classifier dan Support Vector Machine*. Institut Teknologi Sepuluh Nopember, Surabaya, 2017.
- Manning C. D., Raghavan P., dan Schütze H., *An Introduction to Information Retrieval*, London, Cambridge University Press Cambridge, 2008.
- Nugroho, Witarto, dan Handoko, Application of Support Vector Machine in Bioinformatics, *Proceeding of Indonesian Scientific Meeting in Central* (2003), 19-27.
- Nurjannah M., Hamdani, Astuti I. F., Penerapan Algoritma *Term Frequency - Inverse Document Frequency* (TF-IDF) untuk Text Mining, *Jurnal Informatika Mulawarman 8* (2013), no. 3, 110-113.
- Pink D. H., *When*, Elex Media Komputindo, Jakarta, 2018
- Sunni I. dan Widyanoro D. H., Analisis Sentimen dan Ekstraksi Topik Penentu Sentimen pada Opini Terhadap Tokoh Publik, *Jurnal Sarjana Institut Teknologi Bandung Bidang Teknik Elektro dan Informatika I* (2012), no 2, 200-206.
- Zaki. J. M. and Meira JR, *Data Mining and Analysis Fundamental Concepts and Algorithms*. Cambridge University Press, Cambridge, 2014.
- Zhang, L., Ghosh R., Dekhil M., Hsu M., and Liu B., *Combining Lexicon-based and Learning based Methods for Twitter Sentiment Analysis* dalam: HP

Laboratories, <http://www.hpl.hp.com/techreports/2011/HPL-2011-89.pdf>, diakses pada tanggal 3 Juni 2020.