

**IMPLEMENTASI *FEATURE SELECTION* DAN PEMBOBOTAN TFIDF
PADA KLASIFIKASI *TWEETS* INFORMASI KEHILANGAN DENGAN
NAÏVE BAYES**

SKRIPSI

diajukan untuk memenuhi sebagian syarat untuk memperoleh gelar
Sarjana Komputer Program Studi Ilmu Komputer



oleh

BURHANUDIN

NIM 1504286

**PROGRAM STUDI ILMU KOMPUTER
DEPARTEMEN PENDIDIKAN ILMU KOMPUTER
FAKULTAS PENDIDIKAN MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS PENDIDIKAN INDONESIA**

2022

**IMPLEMENTASI *FEATURE SELECTION* DAN PEMBOBOTAN TF-IDF
PADA KLASIFIKASI *TWEETS* INFORMASI KEHILANGAN DENGAN
NAÏVE BAYES**

Oleh

Burhanudin

NIM 1504286

Sebuah Skripsi yang Diajukan untuk Memenuhi Salah Satu Syarat Memperoleh
Gelar Sarjana Komputer di Fakultas Pendidikan Matematika dan Ilmu
Pengetahuan Alam

© Burhanudin 2022

Universitas Pendidikan Indonesia

Agustus 2022

Hak Cipta Dilindungi Undang-Undang

Skripsi ini tidak boleh diperbanyak seluruhnya atau sebagian, dengan dicetak
ulang, difotokopi, atau cara lainnya tanpa izin dari penulis

BURHANUDIN

1504286

**IMPLEMENTASI *FEATURE SELECTION* DAN PEMBOBOTAN TF-IDF
PADA KLASIFIKASI *TWEETS* INFORMASI KEHILANGAN DENGAN
NAÏVE BAYES**

DISETUJUI DAN DISAHKAN OLEH PEMBIMBING

Pembimbing I,



Yaya Wihardi, M.Kom.

NIP. 198903252015041001

Pembimbing II,



Rosa Ariani Sukanto, M.T.

NIP. 198109182009122003

Mengetahui,

Ketua Program Studi Ilmu Komputer



Dr. Rani Megasari, M.T.

NIP. 198705242014042002

PERNYATAAN

Dengan ini saya menyatakan bahwa skripsi dengan judul “Implementasi *Feature Selection* dan Pembobotan TFIDF Pada Klasifikasi *Tweets* Informasi Kehilangan Dengan Naïve Bayes” ini beserta seluruh isinya adalah benar-benar karya saya sendiri. Saya tidak melakukan penjiplakan atau pengutipan dengan cara-cara yang tidak sesuai dengan etika ilmu yang berlaku dalam masyarakat keilmuan. Atas pernyataan ini, saya siap menanggung risiko/sanksi apabila di kemudian hari ditemukan adanya pelanggaran etika keilmuan atau ada klaim dari pihak lain terhadap keaslian karya saya ini.

Bandung, Agustus 2022
Yang membuat pernyataan,



Burhanudin

KATA PENGANTAR

Puji dan syukur penulis panjatkan kehadirat Allah SWT. karena berkat kehendak dan karunia-Nya penulis dapat menyelesaikan skripsi yang berjudul “Implementasi *Feature Selection* dan Pembobotan TFIDF Pada Klasifikasi *Tweets* Informasi Kehilangan Dengan Naïve Bayes” ini dapat terselesaikan.

Penyusunan skripsi ini ditujukan untuk memenuhi dan melengkapi salah satu syarat untuk mendapatkan gelar sarjana komputer pada Program Studi Ilmu Komputer, Fakultas Pendidikan Matematika dan Ilmu Pengetahuan Alam Universitas Pendidikan Indonesia.

Penulis menyadari bahwa dalam penyusunan skripsi ini masih terdapat banyak kekurangan dan keterbatasan yang perlu disempurnakan. Oleh karena itu, penulis sangat mengharapkan saran maupun kritik yang membangun agar tidak terjadi kesalahan yang sama di kemudian hari dan dapat meningkatkan kualitas ke tahap lebih baik.

Bandung, Agustus 2022

Penulis

UCAPAN TERIMA KASIH

Penulis tentunya tidak dapat menyelesaikan penelitian ini tanpa bantuan dan dukungan dari pihak-pihak yang telah membantu baik langsung ataupun tidak. Maka dari itu pada kesempatan ini, penulis dengan senang hati mengucapkan terima kasih kepada:

1. Kedua orang tua penulis, Bapak M. Taufikurachman dan Ibu Siti Karomah yang senantiasa memberikan dukungan baik doa, semangat, dan materi demi kelancaran penyusunan skripsi ini.
2. Kakak-kakak penulis yaitu M. Falakhudin Sodiq, M. Syamsul Hidayat, Abdul Aziz, dan Ahmad Fauzan yang selalu memberikan dukungan semangat, doa, dan materi kepada penulis dalam penyusunan skripsi ini.
3. Dosen Pembimbing penulis yaitu Bapak Yaya Wihardi, M.Kom. dan Ibu Rosa Ariani Sukamto, M.T. yang telah memberikan bimbingan, pengarahan serta pengetahuan yang berharga dalam penyusunan skripsi ini.
4. Dosen Pembimbing Akademik yaitu Bapak Lala Septem Riza Ph.D., yang telah memberi bimbingan, arahan, dan motivasi kepada penulis selama menjadi mahasiswa Ilmu Komputer.
5. Seluruh dosen yang telah membimbing dan memberikan ilmu yang bermanfaat pada penulis semasa kuliah dan staf administrasi Departemen Pendidikan Ilmu Komputer yang telah memberikan informasi akademik selama masa perkuliahan.
6. Teman-teman penulis, Rendy Fathagrap, M. Naufal Fazanadi, Rizki Nugraha, dan Fadhil Farras H. N. yang telah membantu dan memberikan dukungan dalam proses penelitian ini.
7. Teman-teman seperjuangan dalam pengerjaan skripsi, Agung Sugiarto, M Ammar F.R., dan Dimas Saptahadi yang telah saling memberikan semangat dalam menyelesaikan studi.
8. Herlina Andriani yang telah menjadi tempat berkeluh kesah, berbagai saran serta nasihat, penyemangat, pendukung dan teman diskusi penulis dari awal perkuliahan hingga penulis menyelesaikan skripsi ini.
9. Semua rekan-rekan ilmu komputer angkatan 2015 yang telah bersama-sama melewati perjalanan yang dilalui selama perkuliahan.

10. Departemen Komunikasi dan Informasi dan Badan Eksekutif Mahasiswa Keluarga Mahasiswa Komputer yang telah memberikan banyak pengalaman selama perkuliahan.

Tidak ada kata yang dapat menggambarkan rasa terima kasih penulis atas semua dukungan yang telah diberikan, semoga senantiasa diberikan kebaikan, kebahagiaan, dan kesehatan oleh Allah SWT.

IMPLEMENTASI *FEATURE SELECTION* DAN PEMBOBOTAN TF-IDF PADA KLASIFIKASI *TWEETS* INFORMASI KEHILANGAN DENGAN NAÏVE BAYES

Oleh

Burhanudin – beurhn@student.upi.edu

1504286

ABSTRAK

Media sosial Twitter merupakan tempat berbagi informasi ke sesama pengguna dengan bentuk informasi teks. Salah satu informasi yang banyak dibagikan adalah informasi kehilangan. Informasi kehilangan dapat diklasifikasikan berdasarkan objeknya, yaitu kehilangan barang dan dokumen, kehilangan kendaraan, orang hilang, dan bukan informasi kehilangan. Tahapan yang dilakukan untuk mengklasifikasikan informasi kehilangan pada penelitian ini adalah melakukan *preprocessing data*, menghitung bobot TF, menghitung bobot fitur menggunakan metode *feature selection* Information Gain, Chi-Square, dan Document Frequency, menyeleksi fitur berdasarkan *threshold*, menghitung bobot TFIDF, dan melakukan klasifikasi menggunakan metode Naive Bayes. Penelitian ini akan menguji perbandingan akurasi berdasarkan kombinasi pembobotan dan metode *feature selection* yang digunakan. Hasil pengujian menunjukkan bahwa kombinasi metode Chi-Square – TFIDF dan Chi-Square - TF dengan penggunaan fitur sebesar 20% dari total fitur memiliki akurasi paling tinggi sebesar 85,09%. Hasil ini menunjukkan bahwa penggunaan *feature selection* dapat meningkatkan akurasi pada klasifikasi data teks. Sedangkan penggunaan bobot TFIDF secara keseluruhan memiliki hasil yang kurang baik jika dibandingkan dengan penggunaan bobot TF.

Kata Kunci : Informasi Kehilangan, Feature Selection, Information Gain, Chi-Square, Documentg Frequency, TFIDF, Naïve Bayes

**IMPLEMENTASI *FEATURE SELECTION* DAN PEMBOBOTAN TF-IDF
PADA KLASIFIKASI *TWEETS* INFORMASI KEHILANGAN DENGAN
NAÏVE BAYES**

Arranged by

Burhanudin – beurhn@student.upi.edu

1504286

ABSTRACT

Twitter is a social media to share information with other users in the form of text. One of the information that is widely shared is lost information. Lost information can be classified based on its object, namely lost goods and documents, lost vehicles, missing people, and not lost information. The steps taken to classify loss information in this study are preprocessing data, calculating TF weights, calculating feature weights using the Information Gain, Chi-Square, and Document Frequency feature selection methods, selecting features based on thresholds, calculating TFIDF weights, and classifying using Naive Bayes method. This study will examine the comparison of accuracy based on the combination of weighting and feature selection method used. The test results show that the combination of the Chi-Square – TFIDF and Chi-Square - TF methods with the use of features of 20% of the total features has the highest accuracy of 85.09%. These results indicate that the use of feature selection can improve the accuracy of text data classification. While the use of TFIDF weights as a whole has poor results when compared to the use of TF weights.

Keyword : Lost Information, Feature Selection, Information Gain, Chi-Square, Documentg Frequency, TFIDF, Naïve Bayes

DAFTAR ISI

PERNYATAAN.....	iii
KATA PENGANTAR	iv
UCAPAN TERIMA KASIH.....	v
ABSTRAK.....	vii
<i>ABSTRACT</i>	viii
DAFTAR ISI.....	ix
DAFTAR TABEL.....	xiii
DAFTAR GAMBAR	xv
BAB I PENDAHULUAN.....	1
1.1 Latar Belakang Penelitian.....	1
1.2 Rumusan Masalah Penelitian.....	5
1.3 Tujuan Penelitian	5
1.4 Manfaat Penelitian	6
1.5 Batasan Masalah	6
1.6 Struktur Organisasi Penelitian	6
BAB II KAJIAN PUSTAKA.....	9
2.1 Peta Literatur.....	9
2.2 Twitter.....	10
2.3 Preprocessing	12
2.4 Feature Selection.....	16
2.4.1 Information Gain	17

2.4.2 Chi-Square	20
2.4.3 Document Frequency	23
2.5 Pembobotan TFIDF	24
2.6 Naïve Bayes	28
2.7 Penelitian Terkait	32
BAB III METODOLOGI PENELITIAN	37
3.1 Desain Penelitian	37
3.1.1 Perumusan Masalah	37
3.1.2 Studi Literatur	38
3.1.3 Pengumpulan Data dan <i>Labeling</i>	38
3.1.4 Praproses Data	38
3.1.5 Pembangunan Model Seleksi Fitur	39
3.1.6 Klasifikasi	39
3.1.7 Pengembangan Perangkat Lunak Klasifikasi Informasi Kehilangan	39
3.1.8 Pembuatan Skenario Eksperimen dan Pengujian Eksperimen	40
3.1.9 Hasil Eksperimen	41
3.1.10 Analisis dan Kesimpulan	41
3.2 Metode Penelitian	41
3.2.1 Metode Pengumpulan Data	41
3.2.2 Metode Pengembangan Perangkat Lunak	41
3.3 Perangkat dan Data Penelitian	43
3.3.1 Perangkat Penelitian	43
3.3.2 Data Penelitian	44

BAB IV HASIL DAN PEMBAHASAN	45
4.1 Tahap Analisis	45
4.1.1 Analisis Umum	45
4.1.2 Analisis Data.....	45
4.1.3 Analisis Kebutuhan Perangkat.....	46
4.2 Tahap Perancangan	46
4.2.1 Alur Pengumpulan Data	47
4.2.2 Alur <i>Training</i> Data	48
4.2.3 Alur <i>Testing</i> Data.....	50
4.3 Tahap Implementasi.....	51
4.3.1 Persiapan Data	51
4.3.2 Pemerataan Data	52
4.3.3 <i>Preprocessing</i> Data.....	53
4.3.4 Pembobotan TF.....	54
4.3.5 Penghitungan <i>Feature Selection</i>	55
4.3.6 Pemilihan Fitur yang Akan Diproses.....	66
4.3.7 Pembobotan TFIDF	70
4.3.8 Klasifikasi	71
4.3.9 Evaluasi <i>Cross Validation</i>	71
4.4 Tahap Testing.....	73
4.5 Eksperimen	77
4.5.1 Skenario Eksperimen.....	77
4.5.2 Hasil Eksperimen.....	77

4.6 Analisis Hasil Eksperimen	80
4.6.1 Analisis Hasil Eksperimen Tanpa <i>Feature Selection</i>	80
4.6.2 Analisis Hasil Eksperimen Information Gain.....	80
4.6.3 Analisis Hasil Eksperimen Chi-Square	83
4.6.4 Analisis Hasil Eksperimen Document Frequency	86
BAB V KESIMPULAN DAN SARAN.....	89
5.1 Kesimpulan	89
5.2 Saran	90
DAFTAR PUSTAKA	91
LAMPIRAN.....	96

DAFTAR TABEL

Tabel 1.1 Contoh <i>tweets</i> kehilangan	3
Tabel 2.1 Contoh <i>tokenization</i> kalimat	13
Tabel 2.2 Contoh stop word elimination.....	14
Tabel 2.3 Contoh <i>stemming</i> menggunakan <i>library</i> Sastrawi	16
Tabel 2.4 Contoh Dokumen untuk Seleksi Fitur.....	18
Tabel 2.5 Jumlah Kemunculan Kata pada Dokumen Seleksi Fitur	19
Tabel 2.6 Distribusi Fitur Terhadap Setiap Kelas untuk Kata ‘jalan’	19
Tabel 2.7 <i>Observed Value</i> untuk kata ‘jalan’	21
Tabel 2.8 <i>Expected Value</i> untuk kata ‘jalan’	22
Tabel 2.9 Representasi Fitur dan Kelas.....	22
Tabel 2.10 Nilai Document Frequency	23
Tabel 2.11 Contoh Dokumen Kehilangan untuk Pembobotan TF-IDF	25
Tabel 2.12 <i>Term Frequency</i> Dokumen Kehilangan.....	25
Tabel 2.13 Nilai DF dan IDF dari Dokumen Kehilangan	26
Tabel 2.14 Nilai bobot TF-IDF dari Dokumen Kehilangan.....	27
Tabel 2.15 Contoh <i>Dataset</i> Klasifikasi Naïve Bayes.....	28
Tabel 2.16 Penghitungan Probabilitas Kata Pada <i>Datatest</i>	29
Tabel 2.17 Penghitungan Probabilitas Kata dengan <i>Laplacian Correction</i>	31
Tabel 3.1 Skenario Eksperimen Penelitian	40
Tabel 4.1 Pembagian Kategori Dataset.....	47
Tabel 4.2 Skenario eksperimen	77

Tabel 4.3 Hasil eksperimen Information Gain dengan TF.....	78
Tabel 4.4 Hasil eksperimen Information Gain dengan TFIDF	78
Tabel 4.5 Hasil eksperimen Chi-Square dengan TF	79
Tabel 4.6 Hasil eksperimen Chi-Square dengan TFIDF	79
Tabel 4.7 Hasil eksperimen Document Frequency dengan TF	79
Tabel 4.8 Hasil eksperimen Document Frequency dengan TFIDF.....	80

DAFTAR GAMBAR

Gambar 2.1 Peta literatur penelitian.....	9
Gambar 2.2 Perbandingan Performa Klasifikasi (Chamansingh & Hosein, 2016)33	
Gambar 3.1 Desain Penelitian.....	37
Gambar 3.2 Model <i>Waterfall</i> (Sommerville, 2010).....	42
Gambar 4.1 Alur pengumpulan data	47
Gambar 4.2 Alur <i>Training Data</i>	48
Gambar 4.3 Alur <i>Testing</i>	50
Gambar 4.4 Kode program untuk memasukkan dan menampilkan dataset.....	51
Gambar 4.5 Contoh 5 teratas dataset informasi kehilangan.....	52
Gambar 4.6 Kode program untuk pemerataan data	52
Gambar 4.7 Kode program <i>preprocessing</i> data	53
Gambar 4.8 Kode program pembobotan TF	54
Gambar 4.9 Contoh hasil pembobotan TF	54
Gambar 4.10 Kode program pembuatan <i>dataframe</i> kategori kemunculan fitur ...	55
Gambar 4.11 Contoh hasil penghitungan kategori kemunculan fitur	55
Gambar 4.12 Kode program untuk menggabungkan indeks 1 sampai terakhir....	56
Gambar 4.13 Hasil penghitungan kategori kemunculan fitur setelah digabung ...	57
Gambar 4.14 Kode program untuk membuat <i>observed value</i>	57
Gambar 4.15 Contoh hasil <i>observed value</i> dari lima fitur pertama	58
Gambar 4.16 Kode program untuk menghitung <i>expected value</i>	58
Gambar 4.17 Contoh hasil <i>expected value</i> dari lima fitur pertama.....	59

Gambar 4.18 Kode program untuk menghitung nilai Chi-Square	59
Gambar 4.19 Contoh nilai Chi-Square dari lima fitur pertama.....	60
Gambar 4.20 Kode program untuk menghitung nilai <i>entropy</i>	61
Gambar 4.21 Contoh nilai <i>entropy</i> untuk lima fitur pertama.....	62
Gambar 4.22 Kode program untuk menghitung nilai Information Gain	62
Gambar 4.23 Contoh nilai Information Gain untuk lima fitur awal	63
Gambar 4.24 Kode program untuk menghitung nilai Document Frequency.....	63
Gambar 4.25 Contoh hasil nilai Document Frequency untuk lima fitur awal	64
Gambar 4.26 Kode program untuk menghitung nilai IDF	64
Gambar 4.27 Contoh hasil nilai IDF untuk lima fitur awal	65
Gambar 4.28 Kode program untuk menyatukan seluruh nilai <i>feature selection</i> ...	65
Gambar 4.29 Contoh <i>dataframe</i> lima terbesar nilai Information Gain.....	66
Gambar 4.30 Kode program pemisahan data yang dipilih dan tidak dipilih	67
Gambar 4.31 Kode program untuk menghapus fitur yang tidak digunakan	68
Gambar 4.32 Contoh dokumen hasil pemilihan fitur dengan Information Gain ..	69
Gambar 4.33 Kode program untuk membuat TF dari hasil seleksi fitur	69
Gambar 4.34 Kode Program untuk menghitung TFIDF	70
Gambar 4.35 Kode program untuk membuat model klasifikasi	71
Gambar 4.36 Kode program untuk evaluasi model klasifikasi	72
Gambar 4.37 Contoh hasil evaluasi model klasifikasi	72
Gambar 4.38 Kode program untuk membuat variabel pengujian	73
Gambar 4.39 Kode program untuk <i>preprocessing</i> data masukan baru.....	74
Gambar 4.40 Kode program untuk mengeliminasi fitur yang tidak digunakan.....	74

Gambar 4.41 Kode program untuk membuat TF dan TFIDF data masukan baru	75
Gambar 4.42 Kode program untuk prediksi kelas data masukan baru	76
Gambar 4.43 Contoh hasil prediksi kelas data masukan baru	76
Gambar 4.44 Hasil evaluasi Information Gain dengan TF	81
Gambar 4.45 Hasil evaluasi Information Gain dengan TFIDF	82
Gambar 4.46 Hasil evaluasi Chi-Square dengan TF	84
Gambar 4.47 Hasil evaluasi Chi-Square dengan TFIDF	85
Gambar 4.48 Hasil evaluasi Document Frequency dengan TF	86
Gambar 4.49 Hasil evaluasi Document Frequency dengan TFIDF	87

DAFTAR PUSTAKA

- Agusta, L. (2009). Perbandingan Algoritma Stemming Porter Dengan Algoritma Nazief & Adriani Untuk Stemming Dokumen Teks Bahasa Indonesia. Konferensi Nasional Sistem Dan Informatika, (Hal. 196-201). Bali.
- Aini, S. H., Sari, Y. A., & Arwan, A. (2018). Seleksi Fitur Information Gain untuk Klasifikasi Penyakit Jantung Menggunakan Kombinasi Metode K-Nearest Neighbor Dan Naïve Bayes. *Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputer*, 2(9), 2546-2554.
- Azhagusundari, B., & Thanamani, A. S. (2013). Feature Selection Based On Information Gain. *International Journal Of Innovative Technology And Exploring Engineering (IJITEE)*, 2(2), 18-21.
- Chamansingh, N., & Hosein, P. (2016). Efficient Sentiment Classification Of Twitter Feeds. *International Conference On Knowledge Engineering And Applications* (Hal. 78-82). Singapore: IEEE.
- Chandani, V., Wahono, R. S., & Purwanto. (2015, Februari). Komparasi Algoritma Klasifikasi Machine Learning Dan Feature Selection Pada Analisis Sentimen Review Film. *Journal Of Intelligent Systems*, 1(1), 56-60.
- Chandrashekar, G., & Sahin, F. (2014). A Survey On Feature Selection Methods. *Computers And Electrical Engineering*, 40(1), 16-28.
- Cherian, V., & Bindu, M. (2017). Heart Disease Prediction Using Naïve Bayes Algorithm And Laplace Smoothing Technique. *International Journal Of Computer Science Trends And Technology (IJCST)*, 5(2), 68-73.
- Furqon, R. M., & Setiawan, E. B. (2020). Deteksi Berita Rumor Pada Sosial Media Twitter Menggunakan Metode Naive Bayes Multinomial Dengan Pembobotan TF-IDF. *E-Proceeding Of Engineering*, 7(2), 7916-7925.

- Gligoric, K., Anderson, A., & West, R. (2018). How Constraints Affect Content: The Case Of Twitter's Switch From 140 To 280 Characters*. *International Conference On Web And Social Media* (Hal. 596-599). California: AAAI Press.
- Guo, A., & Yang, T. (2016). Research And Improvement Of Feature Words Weight Based On TFIDF Algorithm. *IEEE Information Technology, Networking, Electronic And Automation Control Conference* (Hal. 415-419). Chongqing, China: IEEE.
- Jin, C., Ma, T., Hou, R., Tang, M., Tian, Y., Al-Dhelaan, A., & Al-Rodhaan, M. (2015). Chi-Square Statistics Feature Selection Based On Term Frequency And Distribution For Text Categorization. *IETE Journal Of Research*, 61(4), 351-362.
- Jing, L.-P., Huang, H.-K., & Shi, H.-B. (2002). Improved Feature Selection Approach TFIDF In Text Mining. *Proceedings Of The First International Conference On Machine Learning And Cybernetic*, (Hal. 944-946). Beijing.
- Kalokasari, D. H., Shofi, D. I., & Setyaningrum, A. H. (2017). Implementasi Algoritma Multinomial Naive Bayes Classifier Pada Sistem Klasifikasi Surat Keluar (Studi Kasus : Diskominfo Kabupaten Tangerang). *Jurnal Teknik Informatika*, 10(2).
- Kumar, V., & Minz, S. (2014, Juni). Feature Selection: A Literature Review. *Smart Computing Review*, 4(3), 211-229.
- Lei, S. (2012). A Feature Selection Method Based On Information Gain And Genetic Algorithm . *International Conference On Computer Science And Electronics Engineering* (Hal. 355-358). China: IEEE.
- Li, R., Lei, K. H., Khadiwala, R., & Chang, K. C.-C. (2012). Tedas: A Twitter-Based Event Detection And Analysis System. *IEEE 28th International Conference On Data Engineering* (Hal. 1273-1276). Washington DC: IEEE.

- Mccallum, A., & Nigam, K. (1998). A Comparison Of Event Models For Naive Bayes Text Classification. *AAAI-98 Workshop On Learning For Text Categorization*, 752(1), 41-48.
- Negara, A. B., Muhandi, H., & Putri, I. M. (2020). Analisis Sentimen Maskapai Penerbangan Menggunakan Metode Naive Bayes Dan Seleksi Fitur Information Gain. *Jurnal Teknologi Informasi Dan Ilmu Komputer (JTIK)*, 7(3), 599-606.
- Pandis, N. (2016). The Chi-Square Test. *American Journal Of Orthodontics And Dentofacial Orthopedics*, 150(5), 898-899.
- Puspitadewi, I., Erwina, W., & Kurniasih, N. (2016, Juni). Pemanfaatan “Twitter Tmcpoldametro” Dalam Memenuhi Kebutuhan Informasi Para Pengguna Jalan Raya. *Jurnal Kajian Informasi & Perpustakaan*, 4(1), 21-28.
- Putri, W. S., Nurwati, N., & Budiarti, M. (2016). Pengaruh Media Sosial Terhadap Perilaku Remaja. *Prosiding Penelitian & Pengabdian Kepada Masyarakat*, 3, Hal. 47-51.
- Rahman, A., Wiranto, & Doewes, A. (2017, June). Online News Classification Using Multinomial Naive Bayes. *Itsmart: Jurnal Ilmiah Teknologi Dan Informasi*, 6(1), 32-38.
- Rana, R., & Singhal, R. (2015). Chi-Square Test And Its Application In Hypothesis Testing. *Journal Of The Practice Of Cardiovascular Sciences*, 1(1), 69-71.
- Rasywir, E., & Purwarianti, A. (2016). Eksperimen Pada Sistem Klasifikasi Berita Hoax Berbahasa Indonesia Berbasis Pembelajaran Mesin. *Jurnal Cybermatika*, 3(2), 1-8.
- Sarasnomo, M., Muljono, & Soeleman, M. A. (2022). Comparison Of Information Gain And Chi-Square Selection Features For Performance Improvement Of Naive Bayes Algorithm On Determining Students With No Pip Recipients At SMKN 1 Brebes. *Jurnal Ilmiah Indonesia*, 7(4), 3553-3574.

- Schneider, K.-M. (2005). Techniques For Improving The Performance Of Naive Bayes For Text Classification. *International Conference On Intelligent Text Processing And Computational Linguistics*, (Hal. 682-693).
- Setiyaningrum, Y. D., Herdajanti, A. F., Supriyanto, C., & Muljono. (2019). Classification Of Twitter Contents Using Chi-Square And K-Nearest Neighbour Algorithm. *International Seminar On Application For Technology Of Information And Communication (ISEMANTIC)* (Hal. 78-81). Semarang: IEEE.
- Shaltout, N. A., El-Hefnawi, M., Rafea, A., & Moustafa, A. A. (2014). Information Gain As A Feature Selection Method For The Efficient Classification Of Influenza Based On Viral Host. *Proceedings Of The World Congress On Engineering*, 1, Hal. 625-631.
- Sihwi, S. W., Jati, I. P., & Anggrainingsih, R. (2018). Twitter Sentiment Analysis Of Movie Reviews Using Information Gain And Naïve Bayes Classifier. *International Seminar On Application For Technology Of Information And Communication (ISEMANTIC)* (Hal. 190-195). Semarang: IEEE.
- Sommerville, I. (2010). *Software Engineering*. New York: Addison-Wesley.
- Srividhya, V., & Anitha, R. (2010). Evaluating Preprocessing Techniques In Text Categorization. *International Journal Of Computer Science And Application Issue*, 47(11), 49-51.
- Sun, Y., Wong, A. K., & Kamel, M. S. (2009). Classification Of Imbalanced Data: A Review. *International Journal Of Pattern Recognition And Artificial Intelligence*, 23(4), 687-719.
- Thelwall, M., Buckley, K., & Paltoglou, G. (2011, Februari). Sentiment In Twitter Events. *Journal Of The American Society For Information Science And Technology*, 62(2), 406-418.

- Twitter. (2019). Q1 2019 Selected Company Metrics And Financials. Twitter. Dipetik January 7, 2020, Dari https://s22.q4cdn.com/826641620/files/doc_financials/2019/q1/q1-2019-selected-company-metrics-and-financials.pdf
- Uysal, A. K., & Gunal, S. (2014, Januari). The Impact Of Preprocessing On Text Classification. *Information Processing & Management*, 50(1), 104-112.
- Vijayarani, D. S., Ilamathi, M. J., & Nithya, M. (2015). Preprocessing Techniques For Text Mining - An Overview. *International Journal Of Computer Science & Communication Networks*, 5(1), 7-16.
- Vijayarani, S., & Janani, R. (2016, Januari). Text Mining: Open Source Tokenization Tools – An Analysis. *Advanced Computational Intelligence: An International Journal (Aci)*, 3(1), 37-47.
- Wahyudi, D., Susyanto, T., & Nugroho, D. (2017). Implementasi Dan Analisis Algoritma Stemming Nazief & Adriani Dan Porter Pada Dokumen Berbahasa Indonesia. *Jurnal Ilmiah Sinus*, 15(2), 49-56.
- We Are Social. (2021). Digital 2021 : Indonesia. Hootsuite. Dipetik Maret 4, 2020, Dari <https://datareportal.com/reports/digital-2020-indonesia>
- Yang, Y., & Pedersen, J. O. (1997). A Comparative Study On Feature Selection In Text Categorization. *LCML*, 97(35), 412-420.
- Yu, B., & Xu, Z.-B. (2008). A Comparative Study For Content-Based Dynamic Spam Classification Using Four Machine Learning Algorithms. *Knowledge-Based Systems*, 21(4), 355-362.