

**IMPLEMENTASI PEMBOBOTAN *BOW* DAN *TF-IDF*
PADA ALGORITMA *RANDOM FOREST* UNTUK ANALISIS SENTIMEN
(Studi Kasus : Program Merdeka Belajar Kampus Merdeka)**

SKRIPSI

Diajukan untuk memenuhi sebagai syarat untuk memperoleh gelar
Sarjana Matematika Program Studi Matematika



Oleh

Azri Ansori Rahman

1906318

**PROGRAM STUDI MATEMATIKA
FAKULTAS PENDIDIKAN MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS PENDIDIKAN INDONESIA
2023**

LEMBAR HAK CIPTA

IMPLEMENTASI PEMBOBOTAN *BOW* DAN *TF-IDF* PADA ALGORITMA *RANDOM FOREST* UNTUK ANALISIS SENTIMEN

(Studi Kasus : Program Merdeka Belajar Kampus Merdeka)

Oleh :

Azri Ansori Rahman

NIM 1906318

Sebuah skripsi yang diajukan untuk memenuhi salah satu syarat memperoleh gelar Sarjana Matematika pada Fakultas Pendidikan Matematika dan Ilmu Pengetahuan Alam

© Azri Ansori Rahman 2023

Universitas Pendidikan Indonesia

Agustus 2023

Hak cipta dilindungi undang-undang

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

LEMBAR PENGESAHAN

AZRI ANSORI RAHMAN

**IMPLEMENTASI PEMBOBOTAN *BOW* DAN *TF-IDF*
PADA ALGORITMA *RANDOM FOREST* UNTUK ANALISIS SENTIMEN
(Studi Kasus : Program Merdeka Belajar Kampus Merdeka)**

Disetujui dan disahkan oleh pembimbing :

Pembimbing I

 7/8/2023

Dr. Dadan Dasari, M.Si.

NIP. 196407171991021001

Pembimbing II



Dr. Lukman, S.Si., M.Si.

NIP. 196801281994021001

Mengetahui,

Ketua Program Studi Matematika



Dr. Kartika Yulianti, S.Pd., M.Si.

NIP. 198207282005012001

SURAT PERNYATAAN

Dengan ini saya menyatakan bahwa skripsi dengan judul “Implementasi Pembobotan *BoW* dan *TF-IDF* Pada Algoritma *Random Forest* Untuk Analisis Sentimen (Studi Kasus : Program Merdeka Belajar Kampus Merdeka)” 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 yang dijatuhkan kepada saya apabila kemudian ditemukan adanya pelanggaran terhadap etika keilmuan dalam karya saya ini, atau ada klaim dari pihak lain terhadap keaslian karya saya ini.

Bandung, Agustus 2023

Yang membuat pernyataan,



Azri Ansori Rahman

1906318

KATA PENGANTAR

Puji syukur penulis ucapkan kepada Allah SWT karena berkat rahmat dan karunia-Nya penulis dapat menyelesaikan penyusunan skripsi yang berjudul “Implementasi Pembobotan Fitur *BoW* dan *TF-IDF* Pada Algoritma *Random Forest* Untuk Analisis Sentimen (Studi Kasus : Program Merdeka Belajar Kampus Merdeka)” sebagai salah satu syarat memperoleh gelar sarjana Matematika di Universitas Pendidikan Indonesia.

Dalam perjalanan penyusunan skripsi ini, penulis ingin mengucapkan terima kasih yang setinggi-tingginya kepada berbagai pihak yang telah memberikan dukungan, bimbingan, dan dorongan dalam menjalani setiap langkah dari proses ini.

Skripsi ini disusun dengan segala keterbatasan dan kekurangan yang mungkin ada. Oleh karena itu, saran, kritik, dan masukan yang membangun sangat penulis harapkan demi perbaikan di masa yang akan datang

Bandung, Agustus 2023



Azri Ansori Rahman

1906318

UCAPAN TERIMAKASIH

Penulis menyadari bahwa penyelesaian skripsi ini tidak akan tercapai tanpa bantuan, bimbingan, dukungan dan doa yang diberikan oleh berbagai pihak. Oleh karena itu, pada kesempatan ini, penulis dengan tulus ingin mengungkapkan rasa terima kasih yang mendalam kepada :

1. Bapak Dr. Dadan Dasari, M.Si selaku dosen pembimbing I yang telah memberikan waktunya untuk memberikan bimbingan, kritik dan saran selama penulis mengerjakan skripsi ini mulai dari awal hingga selesai
2. Bapak Dr. Lukman, S.Si., M.Si selaku dosen pembimbing II yang telah memberikan waktunya untuk memberikan bimbingan, kritik dan saran selama penulis mengerjakan skripsi ini mulai dari awal hingga selesai
3. Ibu Fitriani Agustina, S.Si., M.Si., selaku Dosen Pembimbing akademik yang telah memberikan waktunya untuk membantu pendampingan akademik perkuliahan.
4. Rekan – rekan mahasiswa Matematika UPI 2019 yang telah berjuang bersama dan saling memberi dukungan selama kuliah
5. Muflih Abror Zen, sahabat penulis yang memberikan nasihat dan motivasi yang memberikan semangat positif semasa kuliah
6. Asep Saefuddin Ash Shidiq yang telah membantu dalam penyusunan skripsi dan senantiasa memberikan semangat
7. Kedua orang tua yang selalu memberikan dukungan dan doa

Semoga Allah SWT memberikan balasan atas kebaikan yang telah diberikan oleh semua pihak. Semoga skripsi ini memberikan manfaat dalam perkembangan ilmu pengetahuan.

ABSTRAK

Program Merdeka Belajar Kampus Merdeka adalah sebuah program yang dicanangkan oleh Kemendikbudristek untuk mengembangkan kompetensi di berbagai bidang ilmu guna persiapan menghadapi tantangan dunia kerja. Namun pelaksanaan program Merdeka Belajar Kampus Merdeka ini menuai pro dan kontra. *Twitter* menjadi salah satu media sosial yang mewadahi pendapat masyarakat mengenai program Merdeka Belajar Kampus Merdeka. Penelitian ini bertujuan untuk memperoleh hasil klasifikasi sentiment terkait kebijakan Merdeka Belajar Kampus Merdeka di *Twitter*. Data penelitian ini adalah *tweet* dengan kata kunci “MBKM”, “MSIB”, dan “Kampus Merdeka”. Algoritma klasifikasi yang digunakan adalah *Random Forest* dengan pembobotan fitur *Bag of Words* dan *TF-IDF*. Diperoleh hasil klasifikasi pada data dengan pembobotan fitur *Bag of Words* terdiri dari 31% sentimen positif dan 69% sentimen negatif dengan akurasi model 87,01%, presisi 92%, *recall* 81% dan *F-Measure* 84%. Sedangkan pada data dengan pembobotan fitur *TF-IDF* diperoleh klasifikasi dengan 32% sentimen positif dan 68% sentimen negatif dengan akurasi model 92,21%, presisi 93%, *recall* 90% dan *F-Measure* 91%. Berdasarkan hasil evaluasi menggunakan kurva *AUC-ROC*, kombinasi algoritma *Random Forest* dengan pembobotan fitur *Bag of Words* merupakan klasifikasi yang baik. Sedangkan kombinasi algoritma *Random Forest* dengan pembobotan fitur *TF-IDF* berdasarkan kurva *AUC-ROC* merupakan klasifikasi yang sangat baik

Kata Kunci : Algoritma *Random Forest*, *Bag of Words*, *TF-IDF*, Merdeka Belajar Kampus Merdeka

ABSTRACT

Merdeka Belajar - Kampus Merdeka Program is a program launched by the Ministry of Education and Culture and Research to develop competencies in various fields of science in preparation for the challenges of work. However, the implementation of the Merdeka Belajar - Kampus Merdeka program has drawn pros and cons. Twitter is one of the social media that accommodates public opinions about the Merdeka Belajar - Kampus Merdeka program. This study aims to obtain sentiment classification results related to the Merdeka Belajar - Kampus Merdeka policy on Twitter. This research data are tweets with the keywords "MBKM", "MSIB", and "Merdeka Kampus". The classification algorithm used is Random Forest with Bag of Words and TF-IDF. The classification results obtained on data with Bag of Words consist of 31% positive sentiment and 69% negative sentiment with 87.01% model accuracy, 92% precision, 81% recall and 84% F-Measure. Meanwhile, data with TF-IDF obtained classification with 32% positive sentiment and 68% negative sentiment with model accuracy 92.21%, precision 93%, recall 90% and F-Measure 91%. Based on the evaluation results using the AUC-ROC curve, the combination of the Random Forest algorithm with Bag of Words is a good classification. While the combination of the Random Forest algorithm with TF-IDF based on the AUC-ROC curve is an excellent classification.

Keywords : *Random Forest algorithm, Bag of Words, TF-IDF, Merdeka Belajar Kampus Merdeka*

DAFTAR ISI

LEMBAR HAK CIPTA.....	i
LEMBAR PENGESAHAN	ii
SURAT PERNYATAAN	iii
KATA PENGANTAR	iv
UCAPAN TERIMAKASIH	v
ABSTRAK.....	vii
<i>ABSTRACT</i>	vii
DAFTAR ISI.....	viii
DAFTAR GAMBAR	xii
DAFTAR TABEL.....	xiii
BAB I PENDAHULUAN.....	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah Penelitian	3
1.3 Tujuan Penelitian.....	3
1.4 Manfaat Penelitian.....	4
1.5 Batasan Masalah.....	4
1.6 Sistematika Penulisan.....	4
BAB II KAJIAN PUSTAKA.....	6
2.1 <i>Data Mining</i>	6
2.2 <i>Text Mining</i>	6
2.3 Praproses Data.....	7
2.4 Seleksi Fitur.....	10
2.4.1 <i>Chi-Square (χ^2)</i>	10
2.5 Klasifikasi.....	14
2.6 Ekstraksi Fitur	14
2.6.1 <i>Bag of Words</i>	15
2.6.2 <i>TF-IDF</i>	16
2.7 <i>K-Fold Cross Validation</i>	19
2.8 <i>Algoritma Random Forest</i>	20
2.9 <i>Confusion Matrix</i>	22

2.10	<i>AUC-ROC (Area Under the Curve-Receiver Operating Characteristic Curve)</i>	23
BAB III METODOLOGI PENELITIAN		25
3.1	Sumber Data.....	25
3.2	Tahap Penelitian.....	26
3.2.1	Tahapan Awal	26
3.2.2	Praproses Data	28
3.2.3	<i>Data Labelling</i>	29
3.2.4	Seleksi Fitur	30
3.2.5	Pembobotan Fitur.....	30
3.2.6	Klasifikasi Teks	31
3.2.7	Analisis Hasil.....	32
BAB IV HASIL DAN PEMBAHASAN		34
4.1	Deskripsi Data	34
4.2	Praproses Data.....	35
4.2.1	Membuka Dataset	35
4.2.2	<i>Data Cleansing</i>	36
4.2.3	<i>Case Folding</i>	38
4.2.4	<i>Slang Word Standarization</i>	38
4.2.5	<i>Stemming</i>	39
4.2.6	<i>Stopword Removal</i>	40
4.2.7	Menghapus Kata yang Kurang dari 2 Huruf.....	41
4.2.8	<i>Tokenizing</i>	42
4.3	<i>Labelling</i>	42
4.4	Seleksi Fitur <i>Chi-Square</i>	47
4.5	Pembobotan Fitur	51
4.5.1	<i>Bag of Words (BoW)</i>	51
4.5.2	<i>TF-IDF</i>	53
4.6	<i>Splitting Data Training dan Data Testing</i>	54
4.7	Model Klasifikasi <i>Random Forest</i>	56
4.8	Evaluasi Model.....	60
4.9	<i>K-Fold Cross Validation</i>	65
4.10	<i>AUC ROC</i>	69

BAB V KESIMPULAN DAN SARAN	73
5.1 Kesimpulan.....	73
5.2 Saran.....	74
DAFTAR PUSTAKA	75
LAMPIRAN.....	79

DAFTAR GAMBAR

Gambar 2. 1 Skema Kinerja Algoritma <i>Random Forest</i>	20
Gambar 3. 1 Alur Penelitian	25
Gambar 4. 1 <i>Code</i> Proses Pengambilan Data	34
Gambar 4. 2 Membuka Dataset CSV	36
Gambar 4. 3 <i>Code Data Cleansing</i>	36
Gambar 4. 4 Hasil Proses <i>Data Cleansing</i>	37
Gambar 4. 5 Proses Penghapusan Angka	37
Gambar 4. 6 Proses <i>Case Folding</i>	38
Gambar 4. 7 Proses <i>Slang Word Standarization</i>	39
Gambar 4. 8 Proses <i>Stemming</i>	40
Gambar 4. 9 <i>Code Stopword Removal</i>	40
Gambar 4. 10 Proses <i>Stopword Removal</i>	41
Gambar 4. 11 Proses Penghapusan Kata yang Kurang dari 2 Huruf	41
Gambar 4. 12 Proses <i>Tokenizing</i>	42
Gambar 4. 13 <i>Code Labelling</i>	43
Gambar 4. 14 Proses <i>Labelling</i>	43
Gambar 4. 15 Hasil <i>Labelling</i>	44
Gambar 4. 16 <i>Drop Data</i> Ulasan Netral.....	44
Gambar 4. 17 <i>Code</i> Pembuatan <i>Pie Chart</i> Ulasan Positif Negatif	45
Gambar 4. 18 <i>Pie Chart</i> Ulasan Positif Negatif	45
Gambar 4. 19 Hasil Visualisasi Word Cloud Positif (A) dan Negatif (B).....	46
Gambar 4. 20 Proses Penghitungan Nilai Kontingensi.....	48
Gambar 4. 21 Proses Penghitungan <i>Chi-Square</i>	49
Gambar 4. 22 Proses Pemilihan Kata	50
Gambar 4. 23 Proses Seleksi Fitur <i>Chi-Square</i>	51
Gambar 4. 24 <i>Code</i> Proses <i>BoW</i>	52
Gambar 4. 25 Tampilan Bobot Kata Setelah Pembobotan <i>BoW</i>	52
Gambar 4. 26 <i>Code</i> Proses <i>TF-IDF</i>	53
Gambar 4. 27 Tampilan Bobot Kata Setelah Pembobotan <i>TF-IDF</i>	54
Gambar 4. 28 <i>Spiting Data Training</i> dan <i>Data Testing</i> Untuk Data dengan Pembobotan <i>BoW</i> . 55	
Gambar 4. 29 <i>Splitting Data Training</i> dan <i>Data Testing</i> Untuk Data dengan Pembobotan <i>TF-IDF</i>	55
Gambar 4. 30 Model Random Forest Untuk Data dengan Pembobotan <i>BoW</i>	56
Gambar 4. 31 Model Random Forest Untuk Data dengan Pembobotan <i>TF-IDF</i>	56
Gambar 4. 32 Hasil Klasifikasi	57
Gambar 4. 33 Visualisasi Pohon Klasifikasi untuk Data <i>BoW</i>	58
Gambar 4. 34 Visualisasi Pohon Klasifikasi untuk Data <i>TF-IDF</i>	59
Gambar 4. 35 Akurasi Klasifikasi.....	60
Gambar 4. 36 <i>Code Confusion Matrix</i> Pada Data dengan Pembobotan Fitur <i>BoW</i>	61
Gambar 4. 37 <i>Confusion Matrix</i> Data dengan Pembobotan Fitur <i>BoW</i>	61

Gambar 4. 38 <i>Code Confusion Matrix</i> Pada Data dengan Pembobotan Fitur <i>TF-IDF</i>	62
Gambar 4. 39 <i>Confusion Matrix Data</i> dengan Pembobotan Fitur <i>TF-IDF</i>	63
Gambar 4. 40 <i>Classification Report</i>	63
Gambar 4. 41 Evaluasi dengan <i>Cross Validation</i>	66
Gambar 4. 42 Metrik Evaluasi <i>Cross Validation</i>	67
Gambar 4. 43 <i>Code Visualisasi AUC-ROC</i> untuk Data dengan Pembobotan <i>BoW</i>	69
Gambar 4. 44 Kurva <i>AUC-ROC</i> untuk Data dengan Pembobotan <i>BoW</i>	70
Gambar 4. 45 <i>Code Visualisasi AUC-ROC</i> untuk Data dengan Pembobotan <i>TF-IDF</i>	71
Gambar 4. 46 Kurva <i>AUC-ROC</i> untuk Data dengan Pembobotan <i>TF-IDF</i>	71

DAFTAR TABEL

Tabel 2. 1 Contoh <i>data cleaning</i>	7
Tabel 2. 2 Contoh <i>case folding</i>	8
Tabel 2. 3 Contoh <i>stopword removal</i>	8
Tabel 2. 4 Contoh <i>stemming</i>	9
Tabel 2. 5 Contoh <i>Slang Word Standarization</i>	9
Tabel 2. 6 Contoh <i>tokenizing</i>	10
Tabel 2. 7 Kontingensi Perhitungan Chi-Square (χ^2)	11
Tabel 2. 8 Contoh Dokumen Untuk Seleksi Fitur.....	12
Tabel 2. 9 Kontingensi Untuk kata ‘budaya’	12
Tabel 2. 10 Distribusi <i>Chi-Square</i> (χ^2).....	13
Tabel 2. 11 Vektor <i>Bag of Words</i>	15
Tabel 2. 12 Perhitungan <i>TF</i>	17
Tabel 2. 13 Perhitungan <i>IDF</i>	17
Tabel 2. 14 Perhitungan <i>TF-IDF</i>	18
Tabel 2. 15 <i>K-fold Cross Validation</i>	19
Tabel 2. 16 <i>Confusion Matrix</i>	22
Tabel 3. 1 Contoh Data <i>Tweet</i>	26
Tabel 3. 2 Skenario Pengujian Model dengan <i>Splitting Data 90:10</i>	32
Tabel 3. 3 Skenario Pengujian Model dengan <i>10-Fold Cross Validation</i>	33
Tabel 4. 1 Struktur Data Sebelum Praproses	35
Tabel 4. 2 Perbandingan Performa Klasifikasi dengan <i>Split Data 90:10</i>	65
Tabel 4. 3 Evaluasi Model dengan <i>10-Fold Cross Validation</i>	67

DAFTAR PUSTAKA

- Afdhal, I., Kurniawan, R., Iskandar, I., Salambue, R., Budianita, E., & Syafria, F. (2022). Penerapan Algoritma Random Forest Untuk Analisis Sentimen Komentar Di YouTube Tentang Islamofobia. *Penerapan Algoritma Random Forest Untuk Analisis Sentimen Komentar Di YouTube Tentang Islamofobia*, 5(1), 122-130.
- Ahuja, R., Chug, A., Kohli, S., Gupta, S., & Ahuja, P. (2019). The impact of features extraction on the sentiment analysis. *Procedia Computer Science*, 152, 341-348.
- Al Amrani, Y., Lazaar, M., & El Kadiri, K. E. (2018). Random forest and support vector machine based hybrid approach to sentiment analysis. *Procedia Computer Science*, 127, 511-520.
- Altamevia, F., Wijaya, H. O. L., & Elmayati, E. (2023). Analisis Pola Penjualan Obat di Apotek Srikandi Menggunakan Algoritma Supervised Learning. *Kesatria: Jurnal Penerapan Sistem Informasi (Komputer dan Manajemen)*, 4(1), 170-176.
- Asghar, M. Z., Khan, A., Ahmad, S., & Kundi, F. M. (2014). A review of feature extraction in sentiment analysis. *Journal of Basic and Applied Scientific Research*, 4(3), 181-186.
- Breiman, L. (2001). Random forests. *Machine learning*, 45, 5-32.
- Darwis, D., Pratiwi, E. S., & Pasaribu, A. F. O. (2020). Penerapan Algoritma Svm Untuk Analisis Sentimen Pada Data Twitter Komisi Pemberantasan Korupsi Republik Indonesia. *Jurnal Ilmiah Edutic: Pendidikan dan Informatika*, 7(1), 1-11.
- Deepu, S., Pethuru, R., & Rajaraajeswari, S. (2016). A Framework for Text Analytics using the Bag of Words (BoW) Model for Prediction. *International Journal of Advanced Networking & Applications (IJANA)*, 2(1), 320-323.
- Dellia, P., & Tjahyanto, A. (2017). Tax Complaints Classification on Twitter Using Text Mining. *IPTEK Journal of Science*, 2(1), 11-15.
- Dias Canedo, E., & Cordeiro Mendes, B. (2020). Software requirements classification using machine learning algorithms. *Entropy*, 22(9), 1057.
- Efrizoni, L., Defit, S., Tajuddin, M., & Anggrawan, A. (2022). Komparasi Ekstraksi Fitur dalam Klasifikasi Teks Multilabel Menggunakan Algoritma Machine Learning. *MATRIK: Jurnal Manajemen, Teknik Informatika dan Rekayasa Komputer*, 21(3), 653-666.
- Febriyani, E., & Februariyanti, H. (2023). Analisis Sentimen Terhadap Program Kampus Merdeka Menggunakan Algoritma Naive Bayes Classifier Di Twitter. *Jurnal Tekno Kompak*, 17(1), 25-38.
- Gautam, A. K., & Bansal, A. B. H. I. S. H. E. K. (2022). Performance analysis of supervised machine learning techniques for cyberstalking detection in social media. *Journal of Theoretical and Applied Information Technology*, 100(2), 449-461.

- Gifari, O. I., Adha, M., Hendrawan, I. R., & Durrand, F. F. S. (2022). Analisis Sentimen Review Film Menggunakan TF-IDF dan Support Vector Machine. *Journal of Information Technology*, 2(1), 36-40.
- Gorunescu, F. (2010). *Data Mining Concepts, Models and Techniques*. Springer International Publishing. <https://doi.org/10.1007/978-3-642-19721-5>
- Habib, M. W., & Sultani, Z. N. (2021). Twitter sentiment analysis using different machine learning and feature extraction techniques. *Al-Nahrain Journal of Science*, 24(3), 50-54.
- Harahap, R. N., & Muslim, K. (2020). Peningkatan akurasi pada prediksi kepribadian mbti pengguna twitter menggunakan augmentasi data. *Jurnal Teknologi Informasi dan Ilmu Komputer*, 7(4), 815.
- Hung, L. P., Alfred, R., & Ahmad Hijazi, M. H. (2015). A review on feature selection methods for sentiment analysis. *Advanced Science Letters*, 21(10), 2952-2956.
- Ipmawati, J. (2016). Komparasi teknik klasifikasi teks mining pada analisis sentimen. *Indonesian Journal of Networking and Security (IJNS)*, 6(1).
- Kwon, Y. M., Jun, S. H., Gal, W. M., & Lim, M. J. (2018). The performance comparison of the classifiers according to binary bow, count bow and Tf-Idf feature vectors for malware detection. *Int. J. Eng. Technol*, 7(3), 15-22.
- Maria Navin, J. R., & Pankaja, R. (2016). Performance analysis of text classification algorithms using confusion matrix. *Int. J. Eng. Tech. Res. IJETR*, 6, 75-78.
- Maulana, M. I., & Soebroto, A. A. (2019). Klasifikasi Tingkat Stres Berdasarkan Tweet pada Akun Twitter menggunakan Metode Improved k-Nearest Neighbor dan Seleksi Fitur Chi-square. *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer e-ISSN*, 2548, 964X.
- Naufal, M. F., & Kusuma, S. F. (2022). Analisis Sentimen pada Media Sosial Twitter Terhadap Kebijakan Pemberlakuan Pembatasan Kegiatan Masyarakat Berbasis Deep Learning. *Jurnal Edukasi dan Penelitian Informatika*, 8(1), 44-49.
- Nugroho, Y. S., & Emiliyawati, N. (2017). Sistem Klasifikasi Variabel Tingkat Penerimaan Konsumen Terhadap Mobil Menggunakan Metode Random Forest. *Jurnal Teknik Elektro*, 9(1), 24-29.
- Parlar, T., & Özel, S. A. (2016, August). A new feature selection method for sentiment analysis of Turkish reviews. In *2016 International Symposium on INnovations in Intelligent SysTems and Applications (INISTA)* (pp. 1-6). IEEE.
- Parlar, T., Özel, S. A., & Song, F. (2018). QER: a new feature selection method for sentiment analysis. *Human-centric Computing and Information Sciences*, 8, 1-19.
- Pohan, R. F. R., Ratnawati, D. E., & Arwani, I. (2022). Implementasi Algoritma Support Vector Machine dan Model Bag-of-Words dalam Analisis Sentimen mengenai PILKADA 2020

- pada Pengguna Twitter. *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer e-ISSN, 2548, 964X*.
- Pratomo, S. A., Al Faraby, S., & Purbolaksono, M. D. (2021). Analisis Sentimen Pengaruh Kombinasi Ekstraksi Fitur Tf-idf Dan Lexicon Pada Ulasan Film Menggunakan Metode Knn. *eProceedings of Engineering, 8(5)*.
- Pratomo, S. A., Al Faraby, S., & Purbolaksono, M. D. (2021). Analisis Sentimen Pengaruh Kombinasi Ekstraksi Fitur Tf-idf Dan Lexicon Pada Ulasan Film Menggunakan Metode Knn. *eProceedings of Engineering, 8(5)*.
- Rahayu, I. P., Fauzi, A., & Indra, J. (2022). Analisis Sentimen Terhadap Program Kampus Merdeka Menggunakan Naive Bayes Dan Support Vector Machine. *Jurnal Sistem Komputer dan Informatika (JSON), 4(2), 296-301*.
- Ramli, R. G., & Sibaroni, Y. (2022). Klasifikasi Topik Twitter menggunakan Metode Random Forest dan Fitur Ekspansi Word2Vec. *eProceedings of Engineering, 9(1)*.
- Rozaq, A., Yunitasari, Y., Sussolaikah, K., Sari, E. R. N., & Syahputra, R. I. (2022). Analisis Sentimen Terhadap Implementasi Program Merdeka Belajar Kampus Merdeka Menggunakan Naïve Bayes, K-Nearest Neighbors Dan Decision Tree. *Jurnal Media Informatika Budidarma, 6(2), 746-750*.
- Sabri, T., El Beggar, O., & Kissi, M. (2022). Comparative study of Arabic text classification using feature vectorization methods. *Procedia Computer Science, 198, 269-275*.
- Setiowati, Y., & Helen, A. (2018). Klasifikasi Analisis Sentimen Mengenai Hotel Di Yogyakarta. *SCAN-Jurnal Teknologi Informasi Dan Komunikasi, 13 (1)*.
- Setiowati, Y., & Helen, A. (2018). Klasifikasi Analisis Sentimen Mengenai Hotel di Yogyakarta. *Scan: Jurnal Teknologi Informasi dan Komunikasi, 13(1), 1-10*.
- Soumya George, K., & Joseph, S. (2014). Text classification by augmenting bag of words (BOW) representation with co-occurrence feature. *IOSR Journal of Computer Engineering, 16(1), 34-38*.
- Sugiarto, D., Utami, E., & Yaqin, A. (2022). Perbandingan Kinerja Model TF-IDF dan BOW untuk Klasifikasi Opini Publik Tentang Kebijakan BLT Minyak Goreng. *JURNAL TEKNIK INDUSTRI, 12(3), 272-277*.
- Suhasini, V., & Vimala, N. (2021). A Hybrid TF-IDF and N-Grams Based Feature Extraction Approach for Accurate Detection of Fake News on Twitter Data. *Turkish Journal of Computer and Mathematics Education, 12(6), 5710-5723*.
- Trstenjak, B., Mikac, S., & Donko, D. (2014). KNN with TF-IDF based framework for text categorization. *Procedia Engineering, 69, 1356-1364*.
- Undamayanti, E., Hermanto, T. I., & Kaniawulan, I. (2022). Analisis Sentimen Menggunakan Metode Naive Bayes Berbasis Particle Swarm Optimization Terhadap Pelaksanaan

- Program Merdeka Belajar Kampus Merdeka. *J-SAKTI (Jurnal Sains Komputer dan Informatika)*, 6(2), 916-930.
- Wandani, A., Fauziah, F., & Andrianingsih, A. (2021). Sentimen Analisis Pengguna Twitter pada Event Flash Sale Menggunakan Algoritma K-NN, Random Forest, dan Naive Bayes. *J-SAKTI (Jurnal Sains Komputer dan Informatika)*, 5(2), 651-665.
- Yan, D., Li, K., Gu, S., & Yang, L. (2020). Network-based bag-of-words model for text classification. *IEEE Access*, 8, 82641-82652.
- Yuan, H., Tang, Y., Sun, W., & Liu, L. (2020). A detection method for android application security based on TF-IDF and machine learning. *Plos one*, 15(9), e0238694.
- Zhafira, D. F., Rahayudi, B., & Indriati, I. (2021). Analisis Sentimen Kebijakan Kampus Merdeka Menggunakan Naive Bayes dan Pembobotan TF-IDF Berdasarkan Komentar pada Youtube. *Jurnal Sistem Informasi, Teknologi Informasi, dan Edukasi Sistem Informasi*, 2(1).