

**PERBANDINGAN METODE *WORD2VEC* DAN TF-IDF DENGAN SVM  
UNTUK KLASIFIKASI TEKS PADA MEDIA SOSIAL *TWITTER* (STUDI  
KASUS PEMILIHAN LEGISLATIF 2024)**

**SKRIPSI**

Diajukan untuk memenuhi sebagian syarat untuk memperoleh gelar Sarjana  
Komputer Program Studi Rekayasa Perangkat Lunak



Oleh:

Mujtahidul Haq Mahyunda

NIM 1909410

**PROGRAM STUDI REKAYASA PERANGKAT LUNAK**

**KAMPUS UPI DI CIBIRU**

**UNIVERSITAS PENDIDIKAN INDONESIA**

**2023**

**PERBANDINGAN METODE *WORD2VEC* DAN TF-IDF DENGAN SVM  
UNTUK KLASIFIKASI TEKS PADA MEDIA SOSIAL *TWITTER* (STUDI  
KASUS PEMILIHAN LEGISLATIF 2024)**

Oleh

Mujtahidul Haq Mahyunda

Diajukan untuk memenuhi sebagian syarat untuk memperoleh gelar Sarjana  
Komputer Program Studi Rekayasa Perangkat Lunak

© Mujtahidul Haq Mahyunda

Universitas Pendidikan Indonesia

Agustus 2023

© Hak Cipta Dilindungi Undang-Undang

Skripsi ini tidak boleh diperbanyak seluruhnya atau sebagian, dengan cara dicetak  
ulang, difotokopi atau dengan cara lain tanpa seizin dari peneliti.

## HALAMAN PENGESAHAN

MUJTAHIDUL HAQ MAHYUNDA

PERBANDINGAN METODE *WORD2VEC* DAN TF-IDF DENGAN SVM  
UNTUK KLASIFIKASI TEKS PADA MEDIA SOSIAL *TWITTER* (STUDI  
KASUS PEMILIHAN LEGISLATIF 2024)

Disetujui dan disahkan oleh pembimbing :

Pembimbing I



Mochamad Iqbal Ardimansyah, S.T., M.Kom.  
NIP. 920190219910328101

Pembimbing II



Indira Syawanodya S.Kom., M.Kom  
NIP. 92019021992042320

Mengetahui

Ketua Program Studi Rekayasa Perangkat Lunak



Mochamad Iqbal Ardimansyah, S.T., M.Kom.  
NIP. 920190219910328101

## LEMBAR PERNYATAAN

Dengan ini saya menyatakan bahwa skripsi saya yang berjudul "**Perbandingan Metode *Word2vec* Dan TF-IDF Dengan SVM Untuk Klasifikasi Teks Pada Media Sosial *Twitter* (Studi Kasus Pemilihan Legislatif 2024)**", bersama dengan seluruh isi, benar-benar karya saya sendiri. Tidak ada plagiarisme atau pengutipan yang melanggar etika ilmiah yang berlaku. Jika terbukti ada pelanggaran etika ilmiah atau jika ada klaim dari pihak lain tentang kredibilitas karya saya ini, saya siap bertanggung jawab dan menerima konsekuensi atau sanksi yang mungkin timbul.

Bandung, 25 Agustus 2023

Yang membuat pernyataan



Mujtahidul Haq Mahyunda

1909410

## UCAPAN TERIMA KASIH

Segala puji bagi Allah subhanahu wa ta'ala berterima kasih atas karunia dan bimbingan-Nya sehingga penulis dapat menyelesaikan skripsi dengan judul "Perbandingan Metode *Word2vec* Dan TF-IDF Dengan SVM Untuk Klasifikasi Teks Pada Media Sosial *Twitter* (Studi Kasus Pemilihan Legislatif 2024)". Untuk mendapatkan gelar sarjana dan menyelesaikan studi di Universitas Pendidikan Indonesia, pada program studi Rekayasa Perangkat Lunak, skripsi merupakan salah satu syarat.

Dokumen ini mencakup semua materi yang berkaitan dengan penelitian penulis. Selama proses penelitian dan penyusunan skripsi ini, banyak orang membantu dan mendukung saya dengan memberikan saran, bimbingan, kritik, dan dukungan. Penulis ingin mengucapkan terima kasih kepada:

- 1) Prof. Dr. H. M. Solehuddin, M.Pd., M.A., selaku Rektor Universitas Pendidikan Indonesia yang telah memberikan kesempatan kepada penulis untuk mengikuti studi hingga memperoleh gelar sarjana di Universitas Pendidikan Indonesia
- 2) Prod. Dr. Deni Darmawan, S.Pd., M.Si., M.Si., MCE, selaku Direktur Universitas Pendidikan Indonesia yang telah memberikan kesempatan kepada penulis untuk mengikuti studi hingga memperoleh gelar sarjana di Universitas Pendidikan Indonesia
- 3) Bapak Mochamad Iqbal Ardimansyah, S.T., M.Kom., yang merupakan ketua program studi S1 Rekayasa Perangkat Lunak di Universitas Pendidikan Indonesia dan dosen pembimbing skripsi, karena bimbingan, saran, dan kritik beliau sangat berharga bagi penulis.
- 4) Ibu Indira Syawanodya S.Kom., M.Kom., selaku Dosen Pembimbing Akademik dan dosen pembimbing skripsi, karena bimbingan, saran, dan kritik beliau sangat berharga bagi penulis.
- 5) Semua dosen program studi S1 Rekayasa Perangkat Lunak di Universitas Pendidikan Indonesia juga berterima kasih atas bimbingan

- 6) Rekan-rekan mahasiswa rekayasa perangkat lunak yang bekerja sama dan memotivasi satu sama lain selama proses bimbingan dan menyelesaikan skripsi tepat waktu.
- 7) Ayah, ibu, dan kakak sebagai yang telah memberikan nasihat, dorongan, dukungan, motivasi, dan doa sepanjang waktu untuk mendorong penulis untuk menyelesaikan skripsi ini.
- 8) Rekan penulis yang telah memberikan dukungan yaitu: mazaya, ari, ilham, dwi, vivin, krisna, rizki, faris, bimo dan fakhri.
- 9) Banyak pihak lain yang menawarkan bantuan dan dukungan baik langsung maupun tidak langsung selama proses penelitian dan penyusunan skripsi ini, yang tidak dapat disebutkan satu per satu.

Penulis mengakui bahwa skripsi ini memiliki keterbatasan. Oleh karena itu, saran dan kritik yang bermanfaat sangat diharapkan untuk mencegah kesalahan yang sama dalam penelitian serupa di masa depan. Terakhir, penulis memohon maaf sebesar-besarnya jika ada kesalahan dalam penelitian atau tulisan yang disertakan dalam dokumen ini.

Bandung, 25 Agustus 2023

Yang membuat pernyataan



Mujtahidul Haq Mahyunda

1909410

**PERBANDINGAN METODE *WORD2VEC* DAN TF-IDF DENGAN SVM  
UNTUK KLASIFIKASI TEKS PADA MEDIA SOSIAL *TWITTER* (STUDI  
KASUS PEMILIHAN LEGISLATIF 2024)**

MUJTAHIDUL HAQ MAHYUNDA  
1909410

**ABSTRAK**

Media sosial *Twitter* merupakan *platform* terbanyak penyebaran isu berupa teks tentang pemilihan legislatif 2024 di Indonesia. Jumlah teks ada pada media sosial *Twitter* membutuhkan teknik klasifikasi teks yang tepat seperti TF-IDF (*Term Frequency-Inverse Document Frequency*) dan *Word2vec*, teknik klasifikasi teks tersebut dilakukan penelitian yang diperlukan terkait pemilihan legislatif 2024 bertujuan untuk mencari metode terbaik untuk klasifikasi teks dan melihat keunggulan dan kelemahan metode *Word2vec* dan TF-IDF dengan klasifikasi SVM. Metode penelitian menggunakan *Design Science Research Methode (DSRM)* untuk meningkatkan basis pengetahuan teknologi dan ilmu pengetahuan melalui penciptaan benda atau alat inovatif. Hasil evaluasi model TF-IDF pembagian data 90% *Latih* dan 10% *Uji* memiliki hasil tertinggi pada evaluasi *Confusion matrix* dengan *accuracy* 87%, *precision* 83%, *recall* 88% dan *f1-score* 85%. daripada model *Word2vec*. Penggunaan teknik *K-fold cross validation* model TF-IDF hasil tertinggi pada akurasi 86%, *precision* 80%, *recall* 87% dan *f1-score* 84% pada pembagian data 90% *Latih* dan 10% *Uji*, menunjukkan bahwa model TF-IDF memiliki kinerja yang cukup baik untuk menggeneralisasi data yang independen, sehingga menghindari bias dibandingkan model *Word2vec*. Penelitian juga menemukan kelemahan SVM adalah sulit digunakan dalam jumlah sampel berskala besar karena memiliki pengaruh kinerja, sehingga membuat kinerja tidak maksimal. Model TF-IDF menghasilkan vektor fitur yang lebih spesifik, dapat mengatasi masalah kata-kata yang jarang muncul dan dapat vektor fitur yang mudah diinterpretasikan. Penelitian ini diharapkan dapat membantu dalam menentukan metode klasifikasi teks yang tepat dan memberikan kontribusi bagi pengembangan teknik klasifikasi teks dalam pengolahan bahasa alami.

**Kata Kunci:** *Twitter*, Pemilihan legislatif 2024, *Word2vec*, TF-IDF, *Support vector machine*

**PERBANDINGAN METODE WORD2VEC DAN TF-IDF DENGAN SVM  
UNTUK KLASIFIKASI TEKS PADA MEDIA SOSIAL TWITTER (STUDI  
KASUS PEMILIHAN LEGISLATIF 2024)**

MUJTAHIDUL HAQ MAHYUNDA

1909410

**ABSTRACT**

*Media sosial Twitter is the platform with the most spread of issues in the form of texts about the 2024 legislative election in Indonesia. The number of texts on Twitter requires the appropriate text classification techniques such as TF-IDF (Term Frequency-Inverse Document Frequency) and Word2vec. The text classification techniques are carried out in research related to the 2024 legislative election with the aim of finding the best method for text classification and seeing the advantages and disadvantages of the Word2vec and TF-IDF methods with SVM classification. The research method used is the Design Science Research Method (DSRM) to improve the knowledge base of technology and science through the creation of innovative objects or tools. The evaluation results of the TF-IDF model with 90% training data and 10% test data have the highest results in the Confusion matrix evaluation with an accuracy of 87%, precision of 83%, recall of 88%, and f1-score of 85%, compared to the Word2vec model. The use of the K-fold cross validation technique on the TF-IDF model resulted in the highest accuracy of 86%, precision of 80%, recall of 87%, and f1-score of 84% with 90% training data and 10% test data, indicating that the TF-IDF model has a good performance to generalize independent data, thus avoiding bias compared to the Word2vec model. The research also found that the weakness of SVM is that it is difficult to use in large-scale sample sizes because it affects performance, making it less than optimal. The TF-IDF model produces more specific feature vectors, can overcome problems with rare words, and can produce easily interpretable feature vectors. This research is expected to help determine the appropriate text classification method and contribute to the development of text classification techniques in natural language processing.*

**Keywords:** *Twitter, 2024 Legislative Election, Word2vec, TF-IDF, Support Vector Machine.*



## DAFTAR ISI

HALAMAN PENGESAHAN.....	ii
LEMBAR PERNYATAAN.....	iii
UCAPAN TERIMA KASIH.....	iv
ABSTRAK.....	vi
<i>ABSTRACT</i> .....	vii
DAFTAR ISI.....	viii
DAFTAR GAMBAR.....	xi
DAFTAR TABEL.....	xii
BAB I PENDAHULUAN.....	1
I.1    Latar Belakang.....	1
I.2    Rumusan Masalah Penelitian.....	3
I.3    Tujuan Penelitian.....	4
I.4    Manfaat Penelitian.....	4
I.5    Batasan Masalah.....	4
I.6    Sistematika Penulisan Skripsi.....	5
I.7    Hipotesis Penelitian.....	6
BAB II KAJIAN PUSTAKA.....	7
2.1    Penelitian Terkait.....	7
2.2    Klasifikasi Teks.....	10
2.3    Pemilihan Legislatif 2024 di Indonesia.....	11
2.4 <i>Word2vec</i> .....	12
2.5    TF-IDF ( <i>Term Frequency-Inverse Document Frequency</i> ).....	15

2.6	Klasifikasi <i>Support Vector Machine</i> (SVM).....	16
2.7	<i>Confusion Matrix</i> .....	18
2.8	<i>K-fold Cross Validation</i> .....	20
BAB III METODE PENELITIAN.....		21
3.1	<i>Design Science Research Methode</i> (DSRM).....	21
3.2	Tahap Penelitian .....	24
3.3	Objek dan Subjek Penelitian .....	27
3.4	Populasi dan Sampel Penelitian .....	27
3.5	Instrumen Penelitian.....	28
3.6	Metode Pengumpulan Data .....	28
3.7	Jenis dan Sumber Data .....	29
3.8	Metode Analisis Data .....	29
BAB IV HASIL DAN PEMBAHASAN .....		31
4.1	Pengumpulan Data.....	31
4.2	<i>Cleaning Data</i> .....	32
4.3	<i>Preprocessing Data</i> .....	34
4.3.1	<i>Case Folding</i> .....	34
4.3.2	<i>Remove Punctuation</i> .....	35
4.3.3	<i>Tokenization</i> .....	35
4.3.4	<i>Filtering</i> .....	36
4.3.5	<i>Normalization</i> .....	37
4.3.6	<i>Stemming</i> .....	37
4.4	Visualisasi Data .....	38
4.5	<i>Labeling Data</i> .....	39
4.6	<i>Splitting Data</i> .....	41

4.7	<i>Linear SVM</i> dengan TF-IDF.....	41
4.7.1	Pengujian Data Latih dan Data Uji <i>Linear SVM</i> dengan TF-IDF .....	43
4.7.2	Pengujian Validasi pada <i>Linear SVM</i> dengan TF-IDF .....	45
4.8	<i>Linear SVM</i> dengan <i>Word2vec</i> .....	46
4.8.1	Pengujian Data Uji dan Data Latih <i>Linear SVM</i> dengan <i>Word2vec</i> ....	48
4.8.2	Pengujian Validasi pada <i>Linear SVM</i> dengan <i>Word2vec</i> .....	50
4.9	Evaluasi Hasil dan Pembahasan .....	51
BAB V KESIMPULAN DAN SARAN.....		54
5.1	Kesimpulan.....	54
5.2	Saran.....	55
DAFTAR PUSTAKA .....		56
LAMPIRAN.....		60

## DAFTAR GAMBAR

Gambar 2.1 Arsitektur <i>Word2vec</i> Model CBOW dan <i>Skip-gram</i> .....	14
Gambar 2.2 Ilustrasi <i>Window</i> pada <i>Word2vec</i> .....	15
Gambar 2.3 <i>Support Vector Machine</i> .....	17
Gambar 2.4 <i>Hyperplanes</i> dalam Ruang Fitur 2D dan 3D.....	17
Gambar 2.5 <i>Margin</i> pada <i>Support Vector Machine</i> .....	18
Gambar 3.1 Tahapan <i>Design Science Research Methode</i> (DSRM).....	22
Gambar 3.2 Rancangan Tahapan Penelitian .....	24
Gambar 4.1 Diagram Garis Rentang Waktu Pengambilan Data.....	32
Gambar 4.2 Grafik Batang Distribusi Frekuensi Kata .....	38
Gambar 4.3 Visualisasi Frekuensi Kata Menggunakan <i>Wordcloud</i> .....	39
Gambar 4.4 Diagram Batang Jumlah Label Positif dan Negatif.....	40
Gambar 4.5 Persentase Jumlah Label Positif dan Negatif pada Data.....	40
Gambar 4.6 Diagram Batang Perbandingan Hasil Kinerja Data Latih dan Data Uji Pada Model TF-IDF dengan SVM.....	44
Gambar 4.8 Diagram Batang Perbandingan Hasil Kinerja Data Latih dan Data Uji Pada Model <i>Word2vec</i> dengan SVM .....	49

## DAFTAR TABEL

Tabel 2.1 Kajian Penelitian Terkait .....	7
Tabel 2.2 Tahapan Pemilihan Umum 2024.....	12
Tabel 2.3 <i>Confusion Matrix</i> .....	19
Tabel 3.1 Tahapan dan Penjelasan <i>Preprocessing</i> Teks .....	25
Tabel 4.1 <i>Tweet</i> yang Duplikat pada Data <i>Twitter</i> .....	33
Tabel 4.2 <i>Tweet Missing Values</i> pada Data <i>Twitter</i> .....	33
Tabel 4.3 <i>Input</i> dan <i>Output Case Folding</i> .....	34
Tabel 4.4 <i>Input</i> dan <i>Output Remove Punctuation</i> .....	35
Tabel 4.5 <i>Input</i> dan <i>Output Tokenization</i> .....	36
Tabel 4.6 <i>Input</i> dan <i>Output Filtering</i> .....	36
Tabel 4.7 <i>Input</i> dan <i>Output Normalization</i> .....	37
Tabel 4.8 <i>Input</i> dan <i>Output Stemming</i> .....	38
Tabel 4.9 Perbandingan Kinerja Data Latih dan Data Uji Berdasarkan Pembagian Data Menggunakan <i>Confusion Matrix</i> pada TF-IDF dengan SVM.....	43
Tabel 4.10 Perbandingan Kinerja Menggunakan <i>K-fold Cross Validation</i> Berdasarkan Pembagian Data pada TF-IDF dan SVM.....	45
Tabel 4.11 Perbandingan Kinerja Data Latih dan Data Uji Berdasarkan Pembagian Data Menggunakan <i>Confusion Matrix</i> pada <i>Word2vec</i> dengan SVM.....	48
Tabel 4.12 Perbandingan Kinerja Menggunakan <i>K-fold Cross Validation</i> Berdasarkan Pembagian Data pada <i>Word2vec</i> dan SVM .....	50
Tabel 4.13 Hasil Perbandingan Kinerja Data Latih dan Data Uji Berdasarkan Pembagian Data Menggunakan <i>Confusion Matrix</i> pada <i>Tf-idf</i> dan <i>Word2vec</i> dengan SVM .....	51
Tabel 4.14 Hasil Pengujian Menggunakan <i>K-fold Cross Validation</i> pada Model <i>Word2vec</i> dan TF-IDF dengan Klasifikasi SVM.....	52

## DAFTAR PUSTAKA

- Abduh, M. (2023, April). *PDIP Pinrang Target 6 Kursi di Pileg 2024, Setiap Dapil Diisi 1 Perwakilan*. Diakses dari: <https://www.detik.com/sulsel/berita/d-6692283/pdip-pinrang-target-6-kursi-di-pileg-2024-setiap-dapil-diisi-1-perwakilan>
- Aliyah Salsabila, N., Ardhito Winatmoko, Y., Akbar Septiandri, A., & Jamal, A. (2019). *Colloquial Indonesian Lexicon*. Proceedings of the 2018 International Conference on Asian Language Processing, IALP 2018, 226–229.
- Amir, M. (2020). *Keserentakan Pemilu 2024 yang Paling Ideal Berdasarkan Putusan Mahkamah Konstitusi Republik Indonesia*. *Al-Ishlah: Jurnal Ilmiah Hukum*, 23(2), 115–131.
- Andayani, S., & Ryansyah, A. (2017). *Implementasi Algoritma TF-IDF Pada Pengukuran Kesamaan Dokumen*. *JuSiTik: Jurnal Sistem dan Teknologi Informasi Komunikasi*, 1(1), 53. <https://doi.org/10.32524/JUSITIK.V1I1.218>
- Assidyk, A. N., Setiawan, E. B., Si, S., Kurniawan, I., Pd, S., & Si, M. (2020). *Analisis Perbandingan Pembobotan Tf-idf Dan Tf-rf Pada Trending Topic Di Twitter Dengan Menggunakan Klasifikasi K-nearest Neighbor*. *eProceedings of Engineering*, 7(2).
- Azis, H., Purnawansyah, P., Fattah, F., & Putri, I. P. (2020). *Performa Klasifikasi KNN dan Cross Validation Pada Data Pasien Pengidap Penyakit Jantung*. *ILKOM Jurnal Ilmiah*, 12(2), 81–86.
- Bautin, M., Vijayarenu, L., & Skiena, S. (2008). *International Sentiment Analysis for News and Blogs*. Proceedings of the International AAAI Conference on Web and Social Media, 2(1), 19–26. <https://doi.org/10.1609/ICWSM.V2I1.18606>
- Becker, J., Vom Brocke, J., Heddier, M., & Seidel, S. (2015). *In Search of Information Systems (Grand) Challenges: A Community of Inquirers Perspective*. *Business and Information Systems Engineering*, 57(6), 377–390. <https://doi.org/10.1007/S12599-015-0394-0>
- Berrar, D. (2018). *Cross-validation*. *Encyclopedia of Bioinformatics and Computational Biology: ABC of Bioinformatics*, 1–3, 542–545. <https://doi.org/10.1016/B978-0-12-809633-8.20349-X>

- Cahyani, D. E., & Patasik, I. (2021). *Performance comparison of tf-idf and word2vec models for emotion text classification*. *Bulletin of Electrical Engineering and Informatics*, 10(5), 2780–2788. <https://doi.org/10.11591/eei.v10i5.3157>
- Cervantes, J., Garcia-Lamont, F., Rodríguez-Mazahua, L., & Lopez, A. (2020). *A comprehensive survey on support vector machine classification: Applications, challenges and trends*. *Neurocomputing*, 408, 189–215. <https://doi.org/10.1016/J.NEUCOM.2019.10.118>
- Dewi, N. R. (2022). *Analisis Sentimen Tweet Vaksinasi Covid-19 Menggunakan Rnn Dengan Metode Tf-Idf Dan Word2vec*. <http://repository.upnjatim.ac.id/4727/>
- Fajar, R., Program, S., Rekayasa, P., Lunak, N., & Bengkalis, R. (2018). *Implementasi Algoritma Naive Bayes Terhadap Analisis Sentimen Opini Film Pada Twitter*. 3(1).
- Fauzi, R. R. (2022). *Skripsi Analisis Sentimen Dampak Ekonomi Masyarakat Indonesia Akibat Pandemi Covid-19 Pada Media Sosial Twitter Menggunakan Metode Naïve Bayes Classifier, Support Vector Machine Dan Lexicon Based*.
- Felicia, F., & Loisa, R. (2018). *Peran Buzzer Politik dalam Aktivitas Kampanye di Media Sosial Twitter*. *Koneksi*, 2(2), 352–359. <https://doi.org/10.24912/KN.V2I2.3906>
- Firdlous, D. A., & Andrian, R. (2022). *Analisis Sentimen Publik Twitter terhadap Pemilu 2024 menggunakan Model Long Short Term Memory*. *SISTEMASI: Jurnal Sistem Informasi*. <http://sistemasi.ftik.unisi.ac.id>
- Hevner, A. R., March, S. T., Park, J., & Ram, S. (2004). *Design science in information systems research*. *MIS Quarterly: Management Information Systems*, 28(1), 75–105. <https://doi.org/10.2307/25148625>
- Hootsuite. (2022). *DIGITAL 2021 INDONESIA*.
- InfoPublik. (2023, April). *Pendaftaran Bacaleg Pemilu 2024 Dimulai 1 Mei 2023*. Diakses dari: <https://www.infopublik.id/kategori/nasional-politik-hukum/735886/pendaftaran-bacaleg-pemilu-2024-dimulai-1-mei-2023>
- Joachims, T. (1998). *Text categorization with Support Vector Machines: Learning with many relevant features*. 137–142. <https://doi.org/10.1007/BFB0026683>
- Komisi Pemilihan Umum. (2022). *Portal Publikasi Pemilu dan Pemilihan*. Diakses dari: <https://infopemilu.kpu.go.id/>

- Kulkarni, A., Chong, D., & Batarseh, F. A. (2020). *Foundations Of Data Imbalance And Solutions For A Data Democracy*. *Data Democracy: At the Nexus of Artificial Intelligence, Software Development, and Knowledge Engineering*, 83–106. <https://doi.org/10.1016/B978-0-12-818366-3.00005-8>
- Marjaya, D. (2023, April). Pileg 2024, Golkar Targetkan Raya 20 Persen Suara di Bangka Belitung, Pilkada Utamakan Kader. Diakses dari: <https://bangka.tribunnews.com/2023/04/17/pileg-2024-golkar-targetkan-raya-20-persen-suara-di-bangka-belitung-pilkada-utamakan-kader>
- METROTVNEWS.COM. (2023). *Apel Siaga Perubahan Partai NasDem Dipastikan Siap Digelar*. Diakses dari: <https://www.metrotvnews.com/play/bD2CZmPJ-apel-siaga-perubahan-partai-nasdem-dipastikan-siap-digelar>
- Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013). *Efficient Estimation of Word Representations in Vector Space*. 1st International Conference on Learning Representations, ICLR 2013 - Workshop Track Proceedings.
- Mikolov, T., Sutskever, I., Chen, K., Corrado, G. S., & Dean, J. (2013). *Distributed Representations of Words and Phrases and their Compositionality*. *Advances in Neural Information Processing Systems*, 26.
- MTI BINUS UNIVERSITY. (2020). *Word Embedding dengan Word2vec*. Diakses dari: <https://mti.binus.ac.id/2020/11/17/word-embedding-dengan-word2vec/>
- Neupane, P. (2020, Desember). *Text Classification in NLP Explained with Movie Review Example*.
- Nurrahman, A. (2023, April 17). *Twitter Media Sosial Dengan Sebaran Isu Pemilu 2024 Terbanyak*. Diakses dari: <https://validnews.id/nasional/twitter-media-sosial-dengan-sebaran-isu-pemilu-2024-terbanyak>
- Pamungkas, F. S., & Kharisudin, I. (2021). *Analisis Sentimen dengan SVM, NAIVE BAYES dan KNN untuk Studi Tanggapan Masyarakat Indonesia Terhadap Pandemi Covid-19 pada Media Sosial Twitter*. 4, 628–634. <https://journal.unnes.ac.id/sju/index.php/prisma/>
- Pemerintah Indonesia. (2003). *UU No. 12 Tahun 2003*. Diakses dari: <https://peraturan.bpk.go.id/Home/Details/43012>
- Putra, T. D., Utami, E., & P.Kurniawan, M. (2022). *Analisis Sentimen Pemilu 2024 Dengan Naive Bayes Berbasis Particle Swarm Optimization (Pso)*. *EXPLORE*, 13(1), 1–5. <https://ojs.utmmataram.ac.id/index.php/explore/article/view/617>

Mujtahidul Haq Mahyunda, 2023

**PERBANDINGAN METODE WORD2VEC DAN TF-IDF DENGAN SVM UNTUK KLASIFIKASI TEKS PADA MEDIA SOSIAL TWITTER (STUDI KASUS PEMILIHAN LEGISLATIF 2024)**

Universitas Pendidikan Indonesia | [repository.upi.edu](https://repository.upi.edu) | [perpustakaan.upi.edu](https://perpustakaan.upi.edu)



- Rahayu, K. (2023). *Pasca-putusan MK, Parpol dan Caleg Gencarkan Sosialisasi*. Diakses dari: <https://www.kompas.id/baca/polhuk/2023/06/18/pasca-putusan-mk-parpol-dan-caleg-gencarkan-sosialisasi>
- Tempola, F., Muhammad, M., & Khairan, A. (2018). *Perbandingan Klasifikasi Antara KNN dan Naive Bayes pada Penentuan Status Gunung Berapi dengan K-Fold Cross Validation*. *Jurnal Teknologi Informasi dan Ilmu Komputer*, 5(5), 577–584. <https://doi.org/10.25126/jtiik20185983>
- Times.id. (2021). *CSIS: Pemilu Serentak 2024 Wajib Direvisi*. <https://www.times.id/2021/03/csis-pemilu-serentak-2024-wajib-direvisi.html>
- Trstenjak, B., Mikac, S., & Donko, D. (2014). *KNN with TF-IDF based Framework for Text Categorization*. *Procedia Engineering*, 69, 1356–1364. <https://doi.org/10.1016/J.PROENG.2014.03.129>
- Truşcă, M. M. (2019). *Efficiency of SVM classifier with Word2Vec and Doc2Vec models*. *Proceedings of the International Conference on Applied Statistics*, 1(1), 496–503. <https://doi.org/10.2478/icas-2019-0043>
- Vindua, R., & Zailani, A. U. (2023). *Analisis Sentimen Pemilu Indonesia Tahun 2024 Dari Media Sosial Twitter Menggunakan Python*. *JURIKOM (Jurnal Riset Komputer)*, 10(2), 479. <https://doi.org/10.30865/jurikom.v10i2.5945>
- vom Brocke, J., Hevner, A., & Maedche, A. (2020). *Introduction to Design Science Research*. 1–13. [https://doi.org/10.1007/978-3-030-46781-4\\_1](https://doi.org/10.1007/978-3-030-46781-4_1)
- Waskito, L. A. (2019). *Analisis Sentimen Terhadap Pemilihan Presiden Indonesia 2019 Pada Media Sosial Twitter Menggunakan Metode Naive Bayes*.
- Zuhri, K., Adha, N., & Saputri, O. (2020). *Analisis Sentimen Masyarakat Terhadap Pilpres 2019 Berdasarkan Opini Dari Twitter Menggunakan Metode Naive Bayes Classifier*. *Bina Darma Conference on Computer Science (BDCCS)*, 2(5), 259–269.