

**ANALISIS SENTIMEN TERHADAP PEMILIHAN PRESIDEN 2024
MENGUNAKAN METODE *RANDOM FOREST***

SKRIPSI



**UIN SUNAN AMPEL
S U R A B A Y A**

Disusun Oleh
KEVIN BAGUS SAPUTRA
09020220034

**PROGRAM STUDI MATEMATIKA
FAKULTAS SAINS DAN TEKNOLOGI
UNIVERSITAS ISLAM NEGERI SUNAN AMPEL
SURABAYA**

2024

PERNYATAAN KEASLIAN

Saya yang bertanda tangan di bawah ini,

Nama : KEVIN BAGUS SAPUTRA

NIM : 09020220034

Program Studi : Matematika

Angkatan : 2020

Menyatakan bahwa saya tidak melakukan plagiat dalam penulisan skripsi saya yang berjudul "ANALISIS SENTIMEN TERHADAP PEMILIHAN PRESIDEN 2024 MENGGUNAKAN METODE *RANDOM FOREST*". Apabila suatu saat nanti terbukti saya melakukan tindakan plagiat, maka saya bersedia menerima sanksi yang telah ditetapkan.

Demikian pernyataan keaslian ini saya buat dengan sebenar-benarnya.

Surabaya, 03 Januari 2024

Yang menyatakan,



KEVIN BAGUS SAPUTRA
NIM. 09020220034

LEMBAR PERSETUJUAN PEMBIMBING

Skripsi oleh


Nama : KEVIN BAGUS SAPUTRA

NIM : 09020220034


Judul Skripsi : ANALISIS SENTIMEN TERHADAP PEMILIHAN
PRESIDEN 2024 MENGGUNAKAN METODE
RANDOM FOREST

telah diperiksa dan disetujui untuk diujikan.

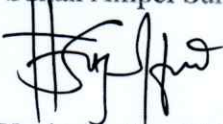
Pembimbing I


Nurissaidah Ulinuha, M. Kom
NIP. 199011022014032004

Pembimbing II


Dr. Lutfi Hakim, M.Ag
NIP. 197312252006041001

Mengetahui,
Ketua Program Studi Matematika
UIN Sunan Ampel Surabaya


Yuniar Farida, M.T.
NIP. 197905272014032002

PENGESAHAN TIM PENGUJI SKRIPSI

Skripsi oleh

Nama : KEVIN BAGUS SAPUTRA
NIM : 09020220034
Judul Skripsi : ANALISIS SENTIMEN TERHADAP PEMILIHAN
PRESIDEN 2024 MENGGUNAKAN METODE
RANDOM FOREST

Telah dipertahankan di depan Tim Penguji
pada tanggal 3 Januari 2024

Mengesahkan,
Tim Penguji

Penguji I



Dr. Dian Candra Rini Novitasari, M.Kom.
NIP. 198511242014032001

Penguji II



Dian Yuliati, M.Si.
NIP. 198707142020122015

Penguji III



Nurissaidah Pinnuha, M. Kom
NIP. 199011022014032004

Penguji IV



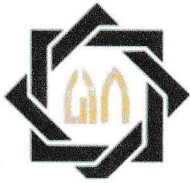
Dr. Lutfi Hakim, M.Ag
NIP. 197312252006041001

Mengetahui,

Dekan Fakultas Sains dan Teknologi
UIN Sunan Ampel Surabaya



Dr. A. Saepul Hamdani, M.Pd.
NIP. 196507312000031002



UIN SUNAN AMPEL
SURABAYA

KEMENTERIAN AGAMA
UNIVERSITAS ISLAM NEGERI SUNAN AMPEL SURABAYA
PERPUSTAKAAN

Jl. Jend. A. Yani 117 Surabaya 60237 Telp. 031-8431972 Fax.031-8413300
E-Mail: perpus@uinsby.ac.id

LEMBAR PERNYATAAN PERSETUJUAN PUBLIKASI
KARYA ILMIAH UNTUK KEPENTINGAN AKADEMIS

Sebagai sivitas akademika UIN Sunan Ampel Surabaya, yang bertanda tangan di bawah ini, saya:

Nama : Kevin Bagus Saputra
NIM : 09020220034
Fakultas/Jurusan : Sains dan Teknologi/Matematika
E-mail address : kecinsaputra@gmail.com

Demi pengembangan ilmu pengetahuan, menyetujui untuk memberikan kepada Perpustakaan UIN Sunan Ampel Surabaya, Hak Bebas Royalti Non-Eksklusif atas karya ilmiah :

Sekripsi Tesis Desertasi Lain-lain (.....)
yang berjudul :

ANALISIS SENTIMEN TERHADAP PEMILIHAN PRESIDEN 2024
MENGUNAKAN METODE RANDOM FOREST

beserta perangkat yang diperlukan (bila ada). Dengan Hak Bebas Royalti Non-Eksklusif ini Perpustakaan UIN Sunan Ampel Surabaya berhak menyimpan, mengalih-media/format-kan, mengelolanya dalam bentuk pangkalan data (database), mendistribusikannya, dan menampilkan/mempublikasikannya di Internet atau media lain secara *fulltext* untuk kepentingan akademis tanpa perlu meminta ijin dari saya selama tetap mencantumkan nama saya sebagai penulis/pencipta dan atau penerbit yang bersangkutan.

Saya bersedia untuk menanggung secara pribadi, tanpa melibatkan pihak Perpustakaan UIN Sunan Ampel Surabaya, segala bentuk tuntutan hukum yang timbul atas pelanggaran Hak Cipta dalam karya ilmiah saya ini.

Demikian pernyataan ini yang saya buat dengan sebenarnya.

Surabaya, 12 Januari 2024

Penulis

Kevin Bagus Saputra

ABSTRAK

ANALISIS SENTIMEN TERHADAP PEMILIHAN PRESIDEN 2024 MENGUNAKAN METODE *RANDOM FOREST*

Pada era modern, informasi banyak tersebar luas di media sosial. Salah satu media sosial yang ramai digunakan oleh masyarakat Indonesia yaitu X. Banyak fenomena yang tersebar di media sosial X, fenomena yang hangat diperbincangkan saat ini mengenai opini publik tentang pemilihan presiden yang diadakan di tahun 2024. Opini negatif maupun positif dapat mempengaruhi pemikiran dari orang-orang yang melihat. Dari kejadian pemilihan presiden 2018, banyak kerusuhan yang terjadi karena opini tanpa didasari argumen yang jelas. Oleh karena itu, penelitian mengenai opini publik untuk melihat seberapa banyak opini negatif maupun positif sangat penting dalam melihat gambaran pemilihan presiden di tahun 2024. Metode *Random Forest* merupakan salah satu metode yang dapat digunakan dalam penelitian sentimen publik. Dasar dari algoritma pembelajaran *Random Forest* yaitu menggunakan konsep *Bagging* dan *Features Randomness*. Model *Random Forest* yang diperoleh terbilang baik dengan menggunakan pembobotan *Term Frequency Invers Document Frequency* berjenis *Unigram* dengan parameter *Mtry* (Peubah Penjelas) berjumlah 70, *Ntry* (Jumlah pohon) yaitu 500, dan menggunakan perhitungan model *Entropy* didapatkan nilai akurasi sebesar 90,75%, *recall* sebesar 90,51%, presisi sebesar 90,75% dan *f-measure* 90,62%. Analisis yang didapat menunjukkan sentimen positif kecenderungan menuju kepada capres, cawapres, dan tokoh masyarakat. Sentimen negatif lebih mengarah terhadap kritikan dan kalimat yang tidak pantas diucapkan. Sentimen netral pada data penelitian kebanyakan menggambarkan suatu berita dan informasi.

Kata kunci: Analisis Sentimen, Random Forest, Pemilihan Presiden

ABSTRACT

SENTIMENT ANALYSIS ON 2024 PRESIDENTIAL ELECTION USING RANDOM FOREST METHOD

In the modern era, information is widely spread on social media. One of the social media that is widely used by Indonesians is X. Many phenomena spread on social media X, a phenomenon that is hotly discussed at this time regarding public opinion about the presidential election held in 2024. Negative and positive opinions can influence the thinking of the people who see. From the events of the 2018 presidential election, many riots occurred because of opinions without being based on clear arguments. Therefore, research on public opinion to see how much negative and positive opinion is very important in seeing the picture of the presidential election in 2024. The Random Forest method is one method that can be used in public sentiment research. The basis of the Random Forest learning algorithm is using the concept of Bagging and Feature Randomness. The Random Forest model obtained is fairly good with using a Term Frequency Invers Document Frequency weighting of the type Unigram with Mtry parameters (Explanatory Variables) totaling 70, Ntry (Number of Trees) is 500, and using the Entropy model calculation, the value of the Entropy model is obtained. trees) which is 500, and using the Entropy model calculation obtained an accuracy of 90,75%, recall of 90,51%, precision of 90,75%, and f-measure of 94.16%. f-measure of 90,62%. The analysis obtained shows that there is a positive sentiment positive sentiment towards presidential candidates, vice presidential candidates, and public figures. Sentiment Negative sentiments are more directed towards criticism and inappropriate sentences. Neutral sentiments in the research data mostly describe news and information. news and information.

Keywords: Sentiment Analysis, Random Forest, Presidential Election.

DAFTAR ISI

HALAMAN JUDUL	i
LEMBAR PERSETUJUAN PEMBIMBING	ii
PENGESAHAN TIM PENGUJI SKRIPSI	iii
HALAMAN PERNYATAAN KEASLIAN	iv
MOTTO	v
HALAMAN PERSEMBAHAN	vi
KATA PENGANTAR	vii
DAFTAR ISI	ix
DAFTAR GAMBAR	xi
ABSTRAK	xiii
ABSTRACT	xiv
I PENDAHULUAN	1
1.1. Latar Belakang Masalah	1
1.2. Rumusan Masalah	7
1.3. Tujuan Penelitian	8
1.4. Manfaat Penelitian	8
1.5. Batasan Masalah	8
1.6. Sistematika Penulisan	9
II TINJAUAN PUSTAKA	12
2.1. Analisis Sentimen	12
2.2. Media Sosial X	13
2.3. Pemilihan Presiden	15
2.4. <i>Text Mining</i>	15
2.5. <i>Text Processing</i>	16
2.6. <i>Outlier Text Data</i>	18
2.7. <i>Term Frequency Inverse Document Frequency</i>	18
2.8. <i>Systematic Sampling</i>	20

2.9. <i>K-Fold Cross Validation</i>	21
2.10. <i>Decision Tree</i>	22
2.11. <i>Random Forest</i>	23
2.12. Evaluasi Sistem	28
2.13. Integrasi Keislaman	31
III METODE PENELITIAN	33
3.1. Jenis Penelitian	33
3.2. Jenis dan Sumber Data	33
3.3. Tahapan Penelitian	34
IV HASIL DAN PEMBAHASAN	39
4.1. <i>Crawling Data</i>	39
4.2. <i>Data sampling</i>	40
4.3. <i>Preprocessing</i>	41
4.4. Pelabelan Data	45
4.5. Pembobotan Kata	46
4.6. Klasifikasi <i>Random Forest</i>	50
4.7. Evaluasi Sistem	70
4.8. Analisis Hasil Penelitian	72
4.9. Hasil Penelitian Berdasarkan Integrasi Islam	84
V PENUTUP	87
5.1. Kesimpulan	87
5.2. Saran	88
DAFTAR PUSTAKA	89
A LAMPIRAN	98

DAFTAR PUSTAKA

- Abah (2023). Sikap presiden dalam politik indonesia. Accessed 1 September 2023.<https://twitter.com/FahmiAgustian/status/1641123914247176192>.
- Afdal, M dan Novita, R. (2024). Sentiment Analysis of PLN Mobile Application Review Using Naïve Bayes Classifier and K-Nearest Neighbor Algorithm Analisis Sentimen Ulasan Aplikasi PLN Mobile Menggunakan Algoritma Naïve Bayes Classifier dan K-Nearest Neighbor. 4(January):10–19.
- Afdhal, I., Kurniawan, R., Iskandar, I., Salambue, R., Budianita, E., and Syafria, F. (2022). Penerapan Algoritma Random Forest Untuk Analisis Sentimen Komentar Di YouTube Tentang Islamofobia. *Jurnal Nasional Komputasi dan Teknologi Informasi*, 5(1):122–130.
- Andreas C. Muller, G. S. (2017). *Introduction to Machine Learning with Python: a guide for data scientist*. O'Reilly Media, Inc., 1005 Gravenstein Highway North, Sebastopol, CA 95472., United States of America.
- Bandung, B. (2023). Sentimen Sampel Positif Kata "Prabowo". Accessed 19 November 2023.<https://twitter.com/bisnisbandung63/status/1713566443328229751>.
- Bardah, D., Afriani, T., and Yatnikasari, A. (2021). Analisis Fishbone dalam Meningkatkan Efektifitas Pelaksanaan Handover Di Rumah Sakit Pemerintah. *Jurnal Kepemimpinan dan Manajemen Keperawatan*, 4(1).

- Breiman, L. (2017). *Classification and Regression Trees*. Chapman Hal, New York.
- Cahyanti, D., Rahmayani, A., and Husniar, S. A. (2020). Analisis performa metode Knn pada Dataset pasien pengidap Kanker Payudara. *Indonesian Journal of Data and Science*, 1(2):39–43.
- Chusnul (2023). Opini publik tentang haram pilih capres partai penista agama. Accessed 1 September 2023. https://twitter.com/ch_chotimah2/status/1707356201212858776.
- Dodo, J. (2023). Kebersamaan Para Calon Presiden 2024. Accessed 05 November 2023. <https://twitter.com/jokowi/status/1718924829020258561>.
- Fahmy Amin, M. (2023). Confusion Matrix in Three-class Classification Problems: A Step-by-Step Tutorial. *Journal of Engineering Research*, 7(1):0–0.
- Fanny, M. (2023). Sentimen Negatif Tertinggi kata "Tidak". Accessed 19 November 2023. <https://twitter.com/Yopie7Yopie/status/1702738058121191524>.
- Firdaus, A. (2021). Text Mining Dan Pola Algoritma Dalam Penyelesaian Masalah Informasi : (Sebuah Ulasan). *Jurnal JUPITER*, 13(1):66.
- Fitriensi, Y. N. (2020). Analisis Hukum Penerapan Dan Bentuk Tindak Pidana Pemilu Menurut Undang-Undang Nomor 7 Tahun 2017. *Journal Of Law*, 2(July):1–10.

- Friastutu, R. (2021). Pencemaran Nama Baik di Media Sosial. *Kumparan.Com*, 6(4):16–32.
- Gustian, L. (2023). Sentimen Positif Teratas Kata "Ganjar". Accessed 19 November 2023. <https://twitter.com/Chepot656529633/status/1713411071351132393>.
- Hasan (2018). Keributan berujung tewasnya warga sampang karena beda pilihan pilpres. Accessed 1 September 2023. https://twitter.com/ZUL_Hasan/status/1067648678448422914.
- Hasibuan (2023). Opini publik tentang partai politik. Accessed 1 September 2023. <https://twitter.com/SuperrGrizz/status/1708678770402660696>.
- Hidayat, A. (2023). Kejanggalaan Putusan Mahkamah Konstitusi Pemilihan Presiden. Accessed 19 November 2023. <https://www.bbc.com/indonesia/articles/c4n60z000ngo>.
- Hidayat, F. (2022). Pertenggaran Beda Pilihan Capres. Accessed 05 November 2023. https://twitter.com/ferry_hdyt13/status/1576073594182201348.
- Himma, Faiqotul (2022). Analisis sentimen adalah: Pengertian, contoh, tipe. Accessed 4 September 2023. <https://majoo.id/solusi/detail/analisis-sentimen-adalah>.
- IndonesiaBelajar (2021). Analisis sentimen adalah: Pengertian, contoh, tipe. Accessed 4 September 2023. <https://www.youtube.com/watch?v=yKovaQ6tyV8>.

Irnanda (2020). Penerapan Klasifikasi C4.5 Dalam Meningkatkan Kecakapan Berbahasa Inggris dalam Masyarakat. *Seminar Nasional Teknologi Komputer Sains (SAINTEKS)*, pages 304–308.

Isnain, A. R., Sakti, A. I., Alita, D., and Marga, N. S. (2021). Sentimen Analisis Publik Terhadap Kebijakan Lockdown Pemerintah Jakarta Menggunakan Algoritma Svm. *Jurnal Data Mining dan Sistem Informasi*, 2(1):31.

Jabar, J. K. (2023). Informasi Sentimen Netral Tertinggi Kata "Wakil". Accessed 19 November 2023. https://twitter.com/JDIHKPU_Jakbar/status/1713455405807116766.

JavaTpoint (2021). Decision tree classification algorithm. Accessed 4 September 2023. <https://www.javatpoint.com/machine-learning-decision-tree-classification-algorithm>.

Jayanti, Q. and Rustiana (2015). Analisis Tingkat Akurasi Model-Model Prediksi Kebangkrutan untuk Memprediksi Voluntary Auditor Switching (Studi Pada Perusahaan Manufaktur Yang Terdaftar di BEI). *Modus*, 27(2):87–108.

Julianto, I. T. (2022). Analisis Sentimen Terhadap Sistem Informasi Akademik Institut Teknologi Garut. *Jurnal Algoritma*, 19(1):449–456.

Kannan, R., Woo, H., Aggarwal, C. C., and Park, H. (2017). Outlier detection for text data. *Proceedings of the 17th SIAM International Conference on Data Mining, SDM 2017*, pages 489–497.

Kpu, L. (2023). Sentimen Netral Teritinggi Kata "Calon". Accessed 19

November 2023. <https://twitter.com/LandakKpu/status/1713944610014470405>.

Larasati, F. A., Ratnawati, D. E., and Hanggara, B. T. (2022). Analisis Sentimen Ulasan Aplikasi Dana dengan Metode Random Forest. ... *Teknologi Informasi dan ...*, 6(9):4305–4313.

Maclean, F., Jones, D., Carin-Levy, G., and Hunter, H. (2013). Understanding twitter. *British Journal of Occupational Therapy*, 76(6):295–298.

Mantrana, A. (2023). sukseskan Pesta Demokrasi Rakyat Melalui Pemilu 2024. Accessed 05 November 2023. <https://twitter.com/abimantrana56/status/1643883908671606791>.

Mays, S. (2019). Other Effective Sampling Methods. Accessed 1 Oktober 2023. <https://stat16.blogspot.com/2015/11/other-effective-sampling-methods.html>.

Miftahusalam, A., Nuraini, A. F., Khoirunisa, A. A., and Pratiwi, H. (2022). Perbandingan Algoritma Random Forest, Naïve Bayes, dan Support Vector Machine Pada Analisis Sentimen Twitter Mengenai Opini Masyarakat Terhadap Penghapusan Tenaga Honorer. *Seminar Nasional Official Statistics*, 2022(1):563–572.

Mikhail Klassen, M. R. (2019). *Mining the social web: data mining Facebook, Twitter, LinkedIn, Instagram, GitHub, and more*.

Mubarok, H. et al. (2018). Demokrasi, politik identitas, dan kohesi sosial: peluang dan tantangan strategi dakwah untuk menghalau provokasi politik di indonesia. *Jurnal Bimas Islam*, 11(2):365–400.

- Musfiroh, D., Khaira, U., Utomo, P. E. P., and Suratno, T. (2021). Analisis Sentimen terhadap Perkuliahan Daring di Indonesia dari Twitter Dataset Menggunakan InSet Lexicon. *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, 1(1):24–33.
- Nishfi, D., Huda, I., Prianto, C., and Awangga, R. M. (2023). Analisis Sentimen Perbandingan Layanan Jasa Pengiriman Kurir Pada Ulasan Play Store Menggunakan Metode Random Forest dan Descision Tree. (54).
- Nugraha, F. A. (2020). *Analisis Sentimen Terhadap Pembatasan Sosial Menggunakan Deep Learning*. Kreatif Industri Nusantara, Indonesia.
- Pratama, M. L., Via, Y. V., and Mandyartha, E. P. (2023). Analisis Performansi Naive Bayes Dan Random Forest Terhadap Sentimen Kenaikan Harga BBM Di Indonesia. XVIII.
- Primajaya, A. and Sari, B. N. (2018). Random Forest Algorithm for Prediction of Precipitation. *Indonesian Journal of Artificial Intelligence and Data Mining*, 1(1):27.
- Rachman, F. F. and Pramana, S. (2020). Analisis Sentimen Pro dan Kontra Masyarakat Indonesia tentang Vaksin COVID-19 pada Media Sosial Twitter. *Health Information Management Journal*, 8(2):100–109.
- Rahmadhany, A., Aldila Safitri, A., and Irwansyah, I. (2021). Fenomena Penyebaran Hoax dan Hate Speech pada Media Sosial. *Jurnal Teknologi Dan Sistem Informasi Bisnis*, 3(1):30–43.
- Ramadhan, R., Sari, Y. A., and Adikara, P. P. (2021). Perbandingan Pembobotan Term Frequency-Inverse Document Frequency dan Term

Frequency-Relevance Frequency terhadap Fitur N-Gram pada Analisis Sentimen. *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, 5(11):5075–5079.

Rauta, U. (2016). Menggagas Pemilihan Presiden yang Demokratis dan Aspiratif. *Jurnal Konstitusi*, 11(3):600.

Riyanto (2022). Hootsuite (we are social): Indonesian digital report 2022. Accessed 4 September 2023.<https://andi.link/hootsuite-we-are-social-indonesian-digital-report-2022/>.

Sabily, A. F., Adikara, P. P., and Fauzi, M. A. (2019). Analisis Sentimen Pemilihan Presiden 2019 pada Twitter menggunakan Metode Maximum Entropy. *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, 3(5):4204–4209.

Saputra, R. (2023). Sentimen Negatif Tertinggi Kata "Aneh". Accessed 19 November 2023.<https://twitter.com/RaymondSap7/status/1713912130532233433>.

Sevinç, E. (2022). An empowered AdaBoost algorithm implementation: A COVID-19 dataset study. *Computers and Industrial Engineering*, 165(January).

Walck, P. E. (2013). BOOK REVIEW: Social Communication in the Twitter Age. *International Journal of Interactive Communication Systems and Technologies*, 3(2):66–69.

Wulandari (2018). Kronologi Pembunuhan di Sampang karena Beda

Pilihan Pilpres. Accessed 19 November 2023.<https://www.bbc.com/indonesia/indonesia-46345977>.

Yang, X.-s. (2019). *Introduction to Algorithms for Data Mining and Machine Learning - Xin-She Yang - Google Books*. Candice Janco, United Kingdom, first edit edition.

Yanti, A. (2023). Perbandingan Metode K-NN Dan Metode Random Forest Untuk Analisis Sentimen pada Tweet Isu Minyak Goreng di Indonesia. *Jurnal Media*, 7(April):756–765.



UIN SUNAN AMPEL
S U R A B A Y A