ANALISIS SENTIMEN BERITA PEMILU 2024 MENGGUNAKAN KLASIFIKASI NAIVE BAYES

SKRIPSI

Diajukan untuk Memenuhi Sebagian dari
Syarat memperoleh Gelar Sarjana Komputer
pada Program Studi Ilmu Komputer

Oleh:

Muhammad Fauzan

1704287

PROGRAM STUDI ILMU KOMPUTER

FAKULTAS PENDIDIKAN MATEMATIKA DAN ILMU PENGETAHUAN ALAM

UNIVERSITAS PENDIDIKAN INDONESIA

2024

ANALISIS SENTIMEN BERITA PEMILU 2024 MENGGUNAKAN KLASIFIKASI NAIVE BAYES

Disusun Oleh:

Muhammad Fauzan

NIM 1704287

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

© Muhammad Fauzan

Universitas Pendidikan Indonesia

Agustus 2024

Hak Cipta Dilindungi Undang-Undang

Skripsi ini tidak boleh diperbanyak seluruhnya atau sebagian,

dengan dicetak ulang, difoto kopi, atau cara lainnya tanpa izin dari penulis

Muhammad Fauzan

1704287

Analisis Sentimen Berita Pemilu 2024 Menggunakan Naive Bayes Classifier

DISETUJUI DAN DISAHKAN OLEH PEMBIMBING:

Pembimbing I,

Digitally signed by EddyPrasetyo N
DN: CN=Eddy Prasetyo N Reason: I am approving this document ocation: your signing location here
Date: 2024-08-03 19:42:41
Foxit Reader Version: 11.2.0

Eddy Prasetyo Nugroho, M.T.

NIP. 197505152008011014

Pembimbing II,

Rizky Rachman J., M.Kom.

NIP. 197711252006041002

Mengetahui,

Ketua Program Studi Ilmu Komputer

Dr. Muhamad Nursalman, M.T.

NIP. 197909292006041002

ANALISIS SENTIMEN BERITA PEMILU 2024 MENGGUNAKAN KLASIFIKASI NAIVE BAYES

Oleh

Muhammad Fauzan – mufauzan30@upi.edu

1704287

ABSTRAK

Pemilihan Umum (Pemilu) tahun 2024 merupakan sarana kedaulatan masyarakat untuk memilih Calon Presiden dan Wakil Presiden. Media daring juga menjadi sumber berita bagi masyarakat dalam mengonsumsi pemberitaan pemilu 2024. Peliputan berita yang dilakukan oleh media mengandung banyak sentimen seperti netral, positif, atau negatif. Penelitian ini bertujuan untuk menganalisis sentimen berita pemilu 2024 menggunakan Naive Bayes Classifier. Penelitian ini memanfaatkan metode web scraping untuk mengumpulkan data berita berdasarkan kata kunci pemilu 2024 dari media daring detik.com. Data berita yang diperoleh kemudian diproses melalui tahap praproses yang meliputi case folding, tokenisasi, stopword removal, dan stemming. Selanjutnya, data tersebut dilabeli sebagai positif, netral, atau negatif menggunakan library NLTK Sentiment Analyzer. Setelah pelabelan, data tersebut dianalisis sentimen menggunakan Algoritma Naive Bayes Classifier dengan dua skenario eksperimen. Skenario pertama adalah analisis sentimen menggunakan Naive Bayes Classifier tanpa tuning. Hasil dari eksperimen pertama menunjukkan akurasi model sebesar 57%. Skenario eksperimen kedua menerapkan hyperparameter alpha pada Algoritma Naive Bayes Classifier. Hasil dari skenario eksperimen kedua menunjukkan akurasi sebesar 65%. Selanjutnya, dilakukan analisis sentimen pada data berita untuk mengetahui sentimen berita pemilu 2024 dengan hasil sentimen netral sebesar 55,2%, sentimen positif sebesar 27,8%, dan sentimen negatif sebesar 17%. Dari hasil penelitian ini, dapat ditarik kesimpulan bahwa analisis sentimen pada berita pemilu 2024 cenderung netral.

Kata Kunci: Analisis Sentimen, Berita Pemilu 2024, Naive Bayes Classifier, Text Mining, Natural Language Processing

SENTIMENT ANALYSIS OF 2024 ELECTION NEWS USING NAIVE BAYES CLASSIFICATION

Arranged by

Muhammad Fauzan – mufauzan30@upi.edu

1704287

ABSTRACT

The 2024 General Election is a means for the public to exercise their right by electing the President and Vice President candidates. Online media also serves as a news source for the public to consume news about the 2024 election. News coverage by media contains various sentiments such as neutral, positive, or negative. This study aims to analyze the sentiment of news articles about the 2024 election using the Naive Bayes Classifier. This research use web scraping methods to collect news data based on the keyword "Pemilu 2024" from the online media site detik.com. The collected news data is then preprocessed, including case folding, tokenization, stopword removal, and stemming. The preprocessed data is then labeled as positive, neutral, or negative using the NLTK Sentiment Analyzer library. Following that labeling, sentiment analysis is performed using the Naive Bayes Classifier algorithm with two experimental scenarios. The first scenario is sentiment analysis using the Naive Bayes Classifier without tuning. The results from the first experiment show a model accuracy of 57%. The second experimental scenario involves applying the alpha hyperparameter to the Naive Bayes Classifier algorithm. Results from the second experiment show an accuracy of 65%. Subsequently, sentiment analysis was conducted on the news data to determine the sentiment of the 2024 election news, resulting in 55.2% neutral sentiment, 27.8% positive sentiment, and 17% negative sentiment. From these findings, it can be concluded that the sentiment analysis of the 2024 election news tends to be neutral.

Keywords: Sentiment Analysis, 2024 Election News, Naive Bayes Classifier, Text Mining, Natural Language Processing

DAFTAR ISI

PERNYATAAN iii
KATA PENGANTAR iv
UCAPAN TERIMA KASIH v
ABSTRAK vii
DAFTAR ISI ix
DAFTAR GAMBAR xi
DAFTAR TABEL xii
BAB I PENDAHULUAN 1
1.1 Latar Belakang 1
1.2 Rumusan Masalah 3
1.3 Tujuan Penelitian 3
1.4 Batasan Penelitian 3
1.5 Sistematika Penulisan 4
BAB II KAJIAN PUSTAKA 6
2.1 Penelitian Terkait 6
2.2 *Text Mining* 7
2.2.1 Metode *Text Mining* 7
2.2.2 Teknik *Text Mining* 8
2.2.3 Proses *Text Mining* 10
2.3 *Natural Language Processing* (NLP) 12
2.3.1 Klasifikasi NLP 13
2.3.2 Metode NLP 15
2.3.3 Alat NLP 16
2.4 Analisis Sentimen 17
2.4.1 Proses Analisis Sentimen 18

2.4.2 Tantangan Analisis Sentimen 21
2.5 Naïve Bayes *Classifier* 22
2.5.1 Tipe-Tipe Naive Bayes Model 23
BAB III METODOLOGI PENELITIAN 25
3.1 Jenis Penelitian 25
3.2 Desain Penelitian 25
3.3 Evaluasi Model 28
3.4 Lingkungan Komputasi 29
3.4.1 Perangkat Keras 29
3.4.2 Perangkat Lunak 29
BAB IV HASIL DAN PEMBAHASAN 30
4.1 Hasil 30
4.1.1 Pembentukan Dataset 30
4.1.2 Pengembangan Model 40
4.1.3 Implementasi Perhitungan Naive Bayes 44
4.1.4 Evaluasi dan Analisis Model 48
4.2 Pembahasan 51
4.2.1 Implementasi Model 54
BAB V KESIMPULAN 56
5.1 Kesimpulan 56
5.2 Saran 56
DAFTAR PUSTAKA 58
LAMPIRAN 61

DAFTAR GAMBAR

Gambar 2. 1 Proses *Information Extraction* 9
Gambar 2. 2 Proses *Visualization* 10
Gambar 2. 3 Proses *Text Mining* 11
Gambar 2. 4 Klasifikasi NLP 13
Gambar 2. 5 Proses Analisis Sentimen 18
Gambar 2. 6 Metode Analisis Sentimen 20

Gambar 3. 1 Desain Penelitian 26
Gambar 3. 2 Proses Web Scraping 27

Gambar 4. 1 Struktur HTML Detik.com 31
Gambar 4. 2 Proses *Data Cleansing* 35
Gambar 4. 3 Persebaran Label Sentimen 40
Gambar 4. 4 Proses *Cross-Validation* 43
Gambar 4. 5 *Confusion Matrix Model Naive Bayes Classifier Baseline* 48
Gambar 4. 6 *Confusion Matrix Tuning Model Naive Bayes* 49
Gambar 4. 7 *Training* dan *Validation Accuracy* 50
Gambar 4. 8 Distribusi Sentimen 51
Gambar 4. 9 TF-IDF Pada Sentimen Netral 52
Gambar 4. 10 TF-IDF Pada Sentimen Negatif 53
Gambar 4. 11 TF-IDF Pada Sentimen Positif 54
Gambar 4. 13 Aktual Label Data 55
Gambar 4. 14 Hasil Implementasi Prediksi Pada Data Baru 55

DAFTAR TABEL

Tabel 3. 1 *Confusion Matrix* 28

Tabel 4. 1 Hasil *Scraping* Data Berita Pemilu 2024 34
Tabel 4. 2 Hasil *Case Folding* 36
Tabel 4. 3 Hasil Tokenisasi 37
Tabel 4. 4 Hasil *Stopword Removal* 37
Tabel 4. 5 Hasil Proses *Stemming* 38
Tabel 4. 6 Hasil Pelabelan Data 39
Tabel 4. 7 Contoh Dataset Hasil TF-IDF 42
Tabel 4. 8 Hasil Perhitungan Probabilitas Kondisional 46

DAFTAR LAMPIRAN

Lampiran 1. Kode Program ‘web_scrap.py’ 61
Lampiran 2. Daftar Kamus Stopword 63
Lampiran 3. Kode Program ‘preprocessing_kampanye.py’ 66
Lampiran 4. Kode Program ‘sentiment_kampanye.py’ 68

DAFTAR PUSTAKA

Abro, A. A., Talpur, M. S. H., & Jumani, A. K. (2023). Natural Language Processing Challenges and Issues: A Literature Review. In Gazi University Journal of Science (Vol. 36, Issue 4, pp. 1522–1536). Gazi Universitesi. https://doi.org/10.35378/gujs.1032517

Andika, L. A., Azizah, P. A. N., & Respatiwulan, R. (2019). Analisis Sentimen Masyarakat terhadap Hasil Quick Count Pemilihan Presiden Indonesia 2019 pada Media Sosial Twitter Menggunakan Metode Naive Bayes Classifier. Indonesian Journal of Applied Statistics, 2(1), 34. https://doi.org/10.13057/ijas.v2i1.29998

Anwar, A. (2003). Komunikasi Politik Paradigma: Teori Aplikasi Strategi Komunikasi Politik Indonesia. Balai Pustaka.

Birjali, M., Kasri, M., & Beni-Hssane, A. (2021). A comprehensive survey on sentiment analysis: Approaches, challenges and trends. Knowledge-Based Systems, 226. https://doi.org/10.1016/j.knosys.2021.107134

Dimitrova, D. V., Shehata, A., Strömbäck, J., & Nord, L. W. (2014). The Effects of Digital Media on Political Knowledge and Participation in Election Campaigns: Evidence From Panel Data. Communication Research, 41(1), 95–118. https://doi.org/10.1177/0093650211426004

Gasparetto, A., Marcuzzo, M., Zangari, A., & Albarelli, A. (2022). Survey on Text Classification Algorithms: From Text to Predictions. Information (Switzerland), 13(2). https://doi.org/10.3390/info13020083

Gupta, V. (2014). A Survey of Natural Language Processing Techniques. International Journal of Computer Science & Engineering Technology, 14–16. http://en.wikipedia.org

Inzalkar, S. M., & Sharma, J. (2015). A Survey on Text Mining-techniques and application. International Journal of Research In Science & Engineering , 488–495. www.ijrise.org|editor@ijrise.org[488-495]

Iqbal, B. M., Lhaksmana, K. M., & Setiawan, E. B. (2023). 2024 Presidential Election Sentiment Analysis in News Media Using Support Vector Machine. Journal of Computer System and Informatics (JoSYC), 4(2), 397–404. https://doi.org/10.47065/josyc.v4i2.3051

Jadhav, S. D., & Channe, H. P. (2013). Comparative Study of K-NN, Naive Bayes and Decision Tree Classification Techniques. In International Journal of Science and Research (IJSR) ISSN (Vol. 5). www.ijsr.net

Khurana, D., Koli, A., Khatter, K., & Singh, S. (2023). Natural language processing: state of the art, current trends and challenges. Multimedia Tools and Applications, 82(3), 3713–3744. https://doi.org/10.1007/s11042-022-13428-4

Kumar, L., & Kalra Bhatia, P. (2013). TEXT MINING: CONCEPTS, PROCESS AND APPLICATIONS. In Journal of Global Research in Computer Science (Vol. 4, Issue 3). www.jgrcs.info

Maheswari, M. U., & Sathiaseelan, Dr. J. G. R. (2017). Text Mining: Survey on Techniques and Applications. International Journal of Science and Research (IJSR), 6(6), 1660–1664.

Medhat, W., Hassan, A., & Korashy, H. (2014). Sentiment analysis algorithms and applications: A survey. Ain Shams Engineering Journal, 5(4), 1093–1113. https://doi.org/10.1016/j.asej.2014.04.011

Newman, N., Fletcher, R., Robertson, C. T., Eddy, K., & Kleis Nielsen, R. (2022). Reuters Institute Digital News Report 2022. https://doi.org/10.60625/risj-x1gn-m549

Prasetyo, H., & Fitrani, A. S. (2023). Sentiment Analysis Before Presidential Election 2024 Using Naïve Bayes Classifier Based On Public Opinion In Twitter. Procedia of Engineering and Life Science, 4. https://doi.org/10.21070/pels.v4i0.1395

Putra, T. D., Utami, E., & Kurniawan, M. P. (2023). Analisis Sentimen Pemilu 2024 dengan Naive Bayes Berbasis Particle Swarm Optimization (PSO). Explore, 13(1), 1–5. https://doi.org/10.35200/ex.v11i2.13

Reshamwala, A., Mishra, D., & Pawar, P. (2013). REVIEW ON NATURAL LANGUAGE PROCESSING. In An International Journal (ESTIJ (Vol. 3, Issue 1).

Shah, S., Akshata Bhat, M., Sumitra Singh, M., Arya Chavan, M., & Aryan Singh, M. (2024). INTERNATIONAL JOURNAL OF PROGRESSIVE RESEARCH IN ENGINEERING MANAGEMENT AND SCIENCE

(IJPREMS) SENTIMENT ANALYSIS. INTERNATIONAL JOURNAL OF PROGRESSIVE RESEARCH IN ENGINEERING MANAGEMENT AND SCIENCE (IJPREMS) , 04(04), 1542–1547. https://doi.org/10.58257/IJPREMS33384

Sun, S., Luo, C., & Chen, J. (2017). A review of natural language processing techniques for opinion mining systems. Information Fusion, 36, 10–25. https://doi.org/10.1016/J.INFFUS.2016.10.004

Vijay Gaikwad, S. (2014). Text Mining Methods and Techniques. In International Journal of Computer Applications (Vol. 85, Issue 17).

Wankhade, M., Rao, A. C. S., & Kulkarni, C. (2022). A survey on sentiment analysis methods, applications, and challenges. Artificial Intelligence Review, 55(7). https://doi.org/10.1007/s10462-022-10144-1

Xu, S. (2018). Bayesian Naïve Bayes classifiers to text classification. Journal of Information Science, 44(1), 48–59. https://doi.org/10.1177/0165551516677946

Alwiah Musdar, I. (2018). METODE NAIVE BAYES DAN LAPLACE SMOOTHING. *JTRISTE*, *5*(2), 8–16. https://www.statista.com/statistics/262205/market-share-held-by-

Khurana, D., Koli, A., Khatter, K., & Singh, S. (2023). Natural language processing: state of the art, current trends and challenges. *Multimedia Tools and Applications*, *82*(3), 3713–3744. https://doi.org/10.1007/s11042-022-13428-4

Lawson, R. (2015). *Web Scraping with Python : scrape data from any website with the power of Python*.

Maheswari, M. U., & Sathiaseelan, Dr. J. G. R. (2017). Text Mining: Survey on Techniques and Applications. *International Journal of Science and Research (IJSR)*, *6*(6), 1660–1664.

Manning, C. D., Raghavan, P., & Schütze, H. (2009). *An Introduction to Information Retrieval*.

Tersiana, A. (2018). *Metode Penelitian*. Anak Hebat Indonesia.