

ABSTRAK

KLASIFIKASI BERITA HOAX PILPRES 2024 MENGGUNAKAN ALGORITMA NAIVEBAYES

Pieter Lomo

Universitas Sanata Dharma

2024

Pada tanggal 28 November 2023 sampai dengan 10 Februari 2024, diumumkan pelaksanaan kampanye terbuka untuk para Capres dan Cawapres terpilih. Dengan tujuan untuk mengenalkan diri, memaparkan visi dan misi, menjelaskan program kerja, dan membangun dukungan dan relawan serta meningkatkan partisipasi pemilih dalam pemilu Capres dan Cawapres 2024. Hal ini dimanfaatkan banyak oknum dalam menyebarkan berita palsu (*Hoaks*) khususnya melalui media sosial. Oleh karena itu akan sangat membantu apabila dilakukan *klasifikasi berita* sebagai salah satu cara untuk membantu para pengguna media sosial dalam mengidentifikasi berita yang tersebar di media sosial. Data yang digunakan adalah data berita diambil menggunakan metode *web scrapping* dengan menggunakan *python* sebanyak 2.164 berita. Berita tersebut berisi informasi-informasi yang dapat digunakan untuk melakukan proses klasifikasi. Penelitian ini menggunakan metode klasifikasi *Naïve Bayes* untuk mendapatkan akurasi yang terbaik. Hasil akurasi paling optimal yang dapat dihasilkan oleh algoritma *Naive Bayes* dengan menggunakan seleksi fitur *Chi-Square* adalah 98,93%, dengan precision 99,90%, recall 97,95%, dan f-1 score 98,92% pada kondisi 3 fold terhadap data hasil oversampling (*Smote*). Sedangkan hasil akurasi paling optimal yang dapat dihasilkan untuk dataset tidak seimbang adalah sebesar 92,32%, dengan precision 100%, recall 46,59%, dan f-1 score 62,70% pada kondisi 11 fold dengan menggunakan seleksi fitur *Chi-Square*. Berdasarkan hasil perbandingan diatas maka dapat disimpulkan bahwa pada percobaan klasifikasi menggunakan metode *Naïve Bayes* dengan seleksi fitur *Chi-Square* terhadap data hasil oversampling (*SMOTE*) lebih baik dibandingkan dengan menggunakan metode *Naive Bayes* terhadap data tidak seimbang dalam mengklasifikasi berita tentang *PILPRES 2024*.

Kata Kunci : *PILPRES 2024, Klasifikasi berita, hoaks, Web Scrapping, Python, Naïve Bayes, Chi-Square, SMOTE, Akurasi, Precision, Recall.*

ABSTRACT

CLASSIFICATION OF HOAX NEWS FOR THE 2024 PRESIDENTIAL ELECTION USING NAIVE BAYES ALGORITHM

Pieter Lomo

Universitas Sanata Dharma

2024

On November 28, 2023 to February 10, 2024, an open campaign was announced for the elected presidential and vice presidential candidates. With the aim of introducing themselves, presenting their vision and mission, explaining work programs, and building support and volunteers as well as increasing voter participation in the 2024 presidential and vice presidential elections. This is utilized by many people in spreading fake news (Hoaks), especially through social media. Therefore, it will be very helpful if news classification is carried out as a way to help social media users identify news spread on social media. The data used is news data taken using the web scrapping method using python as much as 2,164 news. The news contains information that can be used to perform the classification process. This research uses the Naïve Bayes classification method to get the best accuracy. The most optimal accuracy result that can be generated by the Naïve Bayes algorithm using Chi-Square feature selection is 98.93%, with precision 99.90%, recall 97.95%, and f-1 score 98.92% in 3 fold conditions against oversampling data (Smote). While the most optimal accuracy result that can be generated for unbalanced datasets is 92.32%, with 100% precision, 46.59% recall, and f-1 score 62.70% in 11 fold conditions using Chi-Square feature selection. Based on the comparison results above, it can be concluded that the classification experiment using the Naïve Bayes method with Chi-Square feature selection on oversampling data (SMOTE) is better than using the Naïve Bayes method on unbalanced data in classifying news about the 2024 elections.

Keywords: *Election 2024, News classification, hoax, Web Scrapping, Python, Naïve Bayes, Chi-Square, SMOTE, Accuracy, Precision, Recall.*