

ABSTRAK

PERBANDINGAN ALGORITMA NAÏVE BAYES DAN RANDOM FOREST UNTUK
KLASIFIKASI KUNJUNGAN WISATAWAN MANCANEGARA KE INDONESIA

Gloria Stevani Putri Lanyo
Universtias Sanata Dharma
2024

Indonesia merupakan salah satu negara yang memiliki destinasi wisata internasional dengan daya tarik minat wisatawan mancanegara. Namun, jumlah kunjungan wisatawan mancanegara ke Indoensia terjadi penurunan signifikan dalam selama periode pandemi COVID-19. Kemudian tanda-tanda pemulihan terjadi pada tahun 2023 setelah mengalami tantangan pada tahun 2020-2022. Oleh karena itu, studi ini menggunakan data tahun 2017, 2018, 2019, dan 2023 untuk menganalisis pola kunjungan wisatawan, dengan menerapkan metode klasifikasi Gaussian Naïve Bayes dan Random Forest parameter $n_estimators$. Temuan dari penelitian ini menunjukkan bahwa pendekatan machine learning dengan menerapkan kedua metode tersebut dapat digunakan untuk mengklasifikasi kunjungan wisatawan mancanegara ke Indonesia. Dari Implementasi kedua metode tersebut diperoleh hasil Random Forest dengan parameter $n_estimators$ mencapai akurasi sebesar 99.1%, presisi 99.1%, recall 99%, dan f1-score 99.2% dengan $kfold = 7$, 128 tree, dan waktu komputasi 10 detik. Gaussian Naïve Bayes mencapai akurasi sebesar 96.95%, presisi 98.95%, recall 98.88%, dan f1-score 98.90% dengan $k-fold = 17$ dan waktu 3 detik. Dari kedua model tersebut, terbukti bahwa model Random Forest dengan parameter $n_estimators$ mendapatkan akurasi, presisi, recall, dan f-1score lebih tinggi dibandingkan Gaussian Naïve Bayes.

Kata kunci: Wisatawan Mancanegara, Klasifikasi, *Gaussian Naïve Bayes*, *Random Forest*

ABSTRACT**COMPARISON OF NAÏVE BAYES AND RANDOM FOREST ALGORITHMS FOR CLASSIFYING INTERNATIONAL TOURIST VISITS TO INDONESIA**

Gloria Stevani Putri Lanyo
Sanata Dharma University
2024

Indonesia is one of the countries with international tourist destinations that attract foreign tourists. However, the number of foreign tourist arrivals to Indonesia experienced a significant decline during the COVID-19 pandemic period. Signs of recovery began to appear in 2023 after facing challenges in 2020- 2022. Therefore, this study uses data from 2017, 2018, 2019, and 2023 to analyze the pattern of tourist arrivals, applying the Gaussian Naive Bayes and Random Forest classification methods with the parameter $n_estimators$. The findings of this research indicate that the application of machine learning methods, using both methods, can be used to classify foreign tourist arrivals to Indonesia. The implementation of both methods yielded the following results: Random Forest with the parameter $n_estimators$ achieved an accuracy of 99.1%, precision of 99.1%, recall of 99%, and f1-score of 99.2% with $k\text{-fold} = 7$, 128 trees, and a computation time of 10 seconds. Gaussian Naive Bayes achieved an accuracy of 96.95%, precision of 98.95%, recall of 98.88%, and f1-score of 98.90% with $k\text{-fold} = 17$ and a computation time of 3 seconds. From both models, it is evident that the Random Forest model with the parameter $n_estimators$ obtained higher accuracy, precision, recall, and f1-score compared to Gaussian Naive Bayes.

Keyword: Foreign Tourists, Classification, Gaussian Naïve Bayes, Random Forest