

INTISARI

BORUTA DAN LONG SHORT-TERM MEMORY DALAM PERAMALAN JUMLAH KEDATANGAN WISATAWAN DENGAN DATA GOOGLE TRENDS

Nauval Raafi Tanuwijaya
18/430271/PA/17124

Sektor pariwisata memiliki peran penting dalam mendukung perekonomian negara, sehingga dibutuhkan alat bantu untuk membuat kebijakan, salah satunya dengan melakukan peramalan. Berbagai penelitian melakukan peramalan jumlah kedatangan wisatawan dalam mengeksplorasi pemanfaatan data *time series* dari mesin pencari seperti Google Trends karena dianggap lebih *real-time* dibanding data jumlah kedatangan historis tradisional, serta mudah diakses secara terbuka oleh masyarakat umum. Namun, variabel yang terlalu banyak dapat menyebabkan *overfitting* dan memungkinkan adanya variabel tidak prediktif yang akan mengurangi akurasi peramalan.

Penelitian ini akan menggunakan kombinasi metode seleksi fitur Boruta dan model Long Short-Term Memory (LSTM) dalam peramalan jumlah kedatangan wisatawan. Metode Boruta diterapkan untuk memilih fitur yang relevan dari data pencarian Google Trends, dengan tujuan meminimalisasi variabel yang tidak relevan dan meningkatkan akurasi model. Boruta dipilih karena dapat memroses *dataset* non-linier serta dapat bekerja dengan sangat baik tanpa *input* spesifik dari pengguna (Karbasi dkk., 2024). Selain itu, Long Short-Term Memory digunakan karena penggunaan *cell state* yang dapat membawa informasi jangka panjang yang ditentukan dengan struktur *gate* untuk penambahan/pengurangan informasi.

Data yang akan diprediksi adalah data bulanan jumlah kedatangan wisatawan Malaysia ke Indonesia (BPS) periode 2017-2023. Data *time-series* dari Google Trends dengan kata kunci yang beragam akan digunakan variabel prediktor pada peramalan. Hasil menunjukkan kombinasi Boruta-LSTM hanya unggul ketika dibandingkan dengan model LSTM, Boruta-ARIMAX, Pearson-ARIMAX dan ARIMAX. Sementara Pearson-LSTM unggul dibanding yang lain dengan nilai Mean Absolute Error (MAE) 12559.08854, Root Mean Square Error (RMSE) 16855.78685, dan Mean Absolute Percentage Error (MAPE) 7.433952844 persen.

Kata Kunci: peramalan, pariwisata, google trends, boruta, long short-term memory, seleksi fitur

ABSTRACT

BORUTA AND LONG SHORT-TERM MEMORY IN FORECASTING THE NUMBER OF TOURIST ARRIVALS WITH GOOGLE TRENDS DATA

Nauval Raafi Tanuwijaya
18/430271/PA/17124

The tourism sector has an important role in supporting the country's economy, so it needs tools to make policies, one of which is forecasting. Various studies on forecasting the number of tourist arrivals have explored the use of time series data from search engines such as Google Trends because it is considered more real-time than traditional historical arrival data, and is easily accessible to the general public. However, too many variables can lead to overfitting and allow the presence of non-predictive variables that will reduce forecasting accuracy.

This research will use a combination of Boruta feature selection method and Long Short-Term Memory (LSTM) model in forecasting the number of tourist arrivals. The Boruta method is applied to select relevant features from Google Trends search data, with the aim of minimizing irrelevant variables and improving model accuracy. Boruta was chosen because it can process non-linear datasets and can work very well without specific user input (Karbasi et al., 2024). In addition, Long Short-Term Memory was used due to the use of state cells that can carry long-term information defined by a gate structure for adding/subtracting information.

The data to be predicted is the monthly data of the number of Malaysian tourist arrivals to Indonesia (BPS) for the period 2017-2023. Time-series data from Google Trends with various keywords will be used as predictor variables in forecasting. The results show that the Boruta-LSTM combination is only superior when compared to the LSTM, Boruta-ARIMAX, Pearson-ARIMAX and ARIMAX models. While Pearson-LSTM is superior to others with a Mean Absolute Error (MAE) value of 12559.08854, Root Mean Square Error (RMSE) of 16855.78685, and Mean Absolute Percentage Error (MAPE) of 7.433952844 percent.

Keywords: Peramalan, Pariwisata, Google Trends, Boruta, Long Short-Term Memory, Seleksi Fitur