

ABSTRACT

Edbert Anthony (01082180008)

Designing Machine Learning Model Using Logistic Regression, Naive Bayes, and Random Forest Algorithms to Predict Badminton Match Outcomes

(x + 53 pages: 13 figures; 7 tables, 1 appendices)

Badminton is one of the most popular sports worldwide, with the Badminton World Federation (BWF) serving as the international organization overseeing various badminton competitions globally. In this context, this research aims to develop a machine learning model to predict the outcome of a badminton match based on data obtained from the BWF. This model utilizes three main algorithms: Naive Bayes, Random Forest, and Logistic Regression, to identify key factors influencing match results.

Data obtained from the BWF will be processed and prepared. Exploratory Data Analysis (EDA) will be employed to understand data characteristics, such as descriptive statistics, distributions, and correlations between variables. Subsequently, data processing will be carried out, including data cleaning, variable encoding, and the selection of relevant features. Three machine learning models which are Naive Bayes, Random Forest, and Logistic Regression, will be developed and trained using the processed data. After model training, the final results will be evaluated using accuracy assessment and learning curve evaluations. Accuracy will be used to assess the extent to which the models can accurately predict badminton match outcomes, while learning curves will demonstrate the model's performance across various dataset sizes.

The ultimate goal of this research is to provide insights into the performance of each model in predicting badminton match outcomes. Additionally, the study aims to offer insights into the factors influencing the results of a badminton match.

Reference: 16 (2001 – 2023)

ABSTRAK

Edbert Anthony (01082180008)

Perancangan Model Machine Learning Menggunakan Algoritma Logistic Regression, Naive Bayes, dan Random Forest dalam Memprediksi Hasil Pertandingan Badminton

(x + 53 halaman: 13 gambar; 7 tabel; 1 lampiran)

Badminton adalah salah satu olahraga yang sangat populer di seluruh dunia, dengan *Badminton World Federation* (BWF) sebagai organisasi internasional yang mengawasi berbagai pertandingan badminton di seluruh dunia. Dalam konteks ini, penelitian ini bertujuan untuk mengembangkan model machine learning untuk memprediksi kemenangan suatu pertandingan badminton berdasarkan data yang diperoleh dari BWF. Model ini menggunakan tiga algoritma utama, yaitu *Naive Bayes*, *Random Forest*, dan *Logistic Regression*, untuk mengidentifikasi faktor-faktor kunci yang mempengaruhi hasil pertandingan.

Pertama-tama, data yang diperoleh dari BWF akan diolah dan dipersiapkan. Proses *Exploratory Data Analysis* (EDA) akan digunakan untuk memahami karakteristik data, seperti statistik deskriptif, distribusi, dan korelasi antar variabel. Selanjutnya, *data processing* akan dilakukan, termasuk penanganan *data cleaning*, *encoding variable*, dan pemilihan fitur yang relevan. Kemudian, tiga model machine learning, yaitu *Naive Bayes*, *Random Forest*, dan *Logistic Regression*, akan dikembangkan dan dilatih menggunakan data yang telah diproses. Setelah pelatihan model, hasil akhir akan dievaluasi dengan menggunakan evaluasi akurasi dan evaluasi *learning curves*. Akurasi akan digunakan untuk menilai sejauh mana model dapat memprediksi hasil pertandingan dengan tepat, sementara *learning curves* akan memperlihatkan kinerja model pada berbagai ukuran dataset.

Hasil akhir dari penelitian ini diharapkan dapat memberikan gambaran tentang kinerja masing-masing model dalam melakukan prediksi kemenangan pertandingan badminton. Selain itu, penelitian ini juga diharapkan dapat memberikan wawasan tentang faktor - faktor yang memengaruhi hasil suatu pertandingan badminton.

Referensi: 16 (2001 – 2023)