

ABSTRAK

Bola basket adalah olahraga bola berkelompok yang terdiri atas dua tim beranggotakan masing - masing lima orang yang saling bertanding mencetak poin dengan memasukkan bola ke dalam keranjang lawan. Dalam pertandingan bola basket banyak faktor yang menentukan menang atau kalah suatu tim, mulai dari potensi dari pemain itu sendiri, segi strategi maupun penempatan posisi pemain. Kesulitan menentukan komposisi pemain dalam permainan bola basket menjadi topik utama dalam pembahasan ini. Klasifikasi adalah proses menemukan model (fungsi) yang menjelaskan dan membedakan kelas-kelas atau konsep, dengan tujuan agar model yang diperoleh dapat digunakan untuk memprediksi kelas atau objek yang memiliki label kelas tidak diketahui. Salah satu metode dari klasifikasi adalah *random forest* menggunakan struktur pohon. Data yang digunakan berupa data publik diambil dari *website Kaggle.com* dengan judul dataset “*NBA Players stats since 1950*”. Pada penelitian ini digunakan metode *random forest* untuk mengklasifikasi penempatan posisi pemain basket dengan menggunakan atribut *3Par, FTr, ORB%, DRB%, BLK%, AST%, STL%, TOV%, DWS, DBPM, FG%, 3P, 3P%, 2P, FT%, STL, TOV, PF, height, dan weight*. Berdasarkan atribut yang digunakan akan dihasilkan klasifikasi 5 label jenis posisi pemain yaitu *Point Guard (PG), Shooting Guard (SG), Small Forward (SF), Power Forward (PF)* atau *Center (C)*. Proses yang dilakukan dimulai dari proses Preprocessing, terdapat 4 tahap yaitu *Data Cleaning, Data Integration, Data Selection, dan Data Transformation*. Langkah selanjutnya data akan dibagi menjadi *data testing* dan *data training* dengan menggunakan *k-fold cross validation* dalam penelitian ini variasi *cross validation* yang digunakan adalah $k=3$ dan $k=5$. Setelah data dibagi menjadi *data training* dan *data testing* dilakukan permodelan dengan metode *random forest* dengan variasi banyak pohon 2^n , variasi banyak pohon 2, 4, 8, 16, 32, 64, 128, 256, 512, 1024, 2048, dan 4096. Hasil dari percobaan variasi dengan metode *random forest* mendapatkan akurasi tertinggi di angka 80,872483% menggunakan $k = 5$ fold *cross validation* dan variasi banyak pohon sejumlah 2048.

Kata kunci: Klasifikasi, Basket, *Random Forest, Data Mining*

ABSTRACT

Basketball is a group ball sport consisting of two teams of five people each competing to score points by putting the ball into the opponent's basket. In a basketball game, there are many factors that determine whether a team wins or loses, starting from the potential of the players themselves, in terms of strategy and placement of players. The difficulty of determining the composition of players in a basketball game is the main topic in this discussion. Classification is the process of finding a model (function) that describes and distinguishes classes or concepts, with the aim that the obtained model can be used to predict classes or objects that have an unknown class label. One method of classification is random forest using a tree structure. The data used in the form of public data is taken from the Kaggle.com website with the dataset title "NBA Players stats since 1950". In this study, the random forest method was used to classify the placement of basketball players using the attributes 3Par, FTr, ORB%, DRB%, BLK%, AST%, STL%, TOV%, DWS, DBPM, FG%, 3P, 3P%, 2P, FT%, STL, TOV, PF, height, and weight. Based on the attributes used, a classification of 5 types of player position labels will be generated, namely Point Guard (PG), Shooting Guard (SG), Small Forward (SF), Power Forward (PF) or Center (C). The process starts from the Preprocessing process, there are 4 stages, namely Data Cleaning, Data Integration, Data Selection, and Data Transformation. The next step is the data will be divided into testing data and training data using k-fold cross validation. In this study, the variation of cross validation used is k=3 and k=5. After the data is divided into training data and testing data, modeling is carried out using the random forest method with variations in the number of trees 2^n , variations in the number of trees 2, 4, 8, 16, 32, 64, 128, 256, 512, 1024, 2048, and 4096. Results from the variation experiment with the random forest method, the highest accuracy was at 80,872483% using k = 5 fold cross validation and the variation of many trees was 2048.

Keywords: Classification, Basketball, Random Forest, Data Mining