

Random forest untuk mengatasi masalah overfitting pada klasifikasi = Random forest to overcome overfitting problem in classification

Esti Latifah, author

Deskripsi Lengkap: <https://lib.ui.ac.id/detail?id=20475278&lokasi=lokal>

Abstrak

ABSTRAK

Klasifikasi merupakan proses pengelompokan suatu himpunan data ke kelas-kelas yang sudah ada sebelumnya. Pada umumnya, himpunan data dibagi menjadi dua bagian, yaitu training data dan testing data. Dibutuhkan suatu metode klasifikasi yang dapat mengelompokkan training data dan testing data ke dalam suatu kelas dengan tepat. Sering kali metode klasifikasi hanya dapat mengelompokkan training data dengan tepat saja, namun tidak demikian untuk testing data. Artinya, model yang terbentuk tidak cukup stabil atau model tersebut mengalami overfitting. Secara umum, overfitting merupakan kondisi saat akurasi yang dihasilkan pada training data cukup tinggi, namun cenderung tidak mampu memprediksi testing data. Penentuan metode klasifikasi yang rentan terhadap overfitting perlu dipertimbangkan. Random forest merupakan salah satu metode klasifikasi yang rentan terhadap masalah overfitting. Hal tersebut sekaligus menjadi salah satu kelebihan dari metode random forest. Oleh karena itu, pada tugas akhir ini akan dibahas metode random forest serta mengaplikasikannya pada data penderita penyakit Parkinson yang dibagi berdasarkan 2 sub-tipe, yaitu tremor dominant TD dan postural instability gait difficulty PIGD dominant. Selanjutnya, dari data tersebut diperoleh hasil akurasi model yang dihasilkan dalam mengklasifikasi training data, yaitu sekitar 94,25 . Sementara itu, akurasi metode ini dalam melakukan klasifikasi pada data yang tidak terkandung dalam membentuk model sebesar 94,26.

<hr>

ABSTRACT

Classification is the process of grouping a set of data into pre existing classes. In general, the data set is divided into two parts. There are training data and testing data. It takes a classification method that can classify both training data and testing data of its class appropriately. However, some of the classification methods only fit in training data, but it can not apply in testing data. It means that the model is unstable or the model occurs overfitting. In general, overfitting is a condition when the model too fit in training, but unable to predict testing data. In other words, the accuracy of predicting the testing data is decreasing. Therefore, the determination of classification methods that are vulnerable to overfitting need to be considered. Random forest is one of the classification methods that is vulnerable to overfitting. It is also one of the advantages of the random forest method. Therefore, in this final project will be discussed random forest method and applying it to the data of Parkinson 39 s disease patients that is divided by 2 sub types. There are dominant tremor TD and postural instability gait difficulty PIGD dominant. Furthermore, from the data obtained the results of model accuracy in classifying the training data is about 94.25 . Meanwhile, the accuracy of this method in classifying the data not contained in forming a model is about 94.26.