

ABSTRAK

ANDREW (00000019529)

PERBANDINGAN PERFORMA *RANDOM FOREST* DAN *SUPPORT VECTOR MACHINE* PADA DATA BINER DAN *MULTI-CLASS*

Skripsi, Fakultas Sains dan Teknologi (2021)

(xiv + 58 halaman; 18 gambar; 57 tabel; 0 lampiran)

Pemilihan metode dan algoritma yang sesuai dengan data merupakan salah satu langkah yang penting dalam proses pengolahan data. Perkembangan teknologi dan metode pengolahan data memberikan ruang bagi peneliti untuk memilih metode. Tetapi penentuan metode dan parameter yang tepat memerlukan pengamatan dan waktu yang cukup lama dalam sebuah proses penelitian, dikarenakan banyaknya aspek yang dapat menentukan hasil dari pengolahan data. Dalam penelitian klasifikasi, *support vector machine* dan *random forest* adalah dua metode yang sering digunakan. Hasil penelitian menggunakan delapan set data yang berbeda, yaitu empat set data biner dan empat set data *multi-class* dari berbagai bidang yang berbeda (kesehatan, keuangan, klasifikasi benda), menunjukkan klasifikasi data biner memiliki hasil lebih baik menggunakan *support vector machine* karena unggul pada dua set data dan imbang pada satu set data, sedangkan *random forest* lebih baik digunakan pada data *multi-class* karena unggul pada dua set data dan imbang pada satu set data. Namun jika dibandingkan melalui proses pengolahan dan rata-rata performa, dapat dikatakan *random forest* merupakan metode yang lebih baik. Hal itu dikarenakan diperlukannya perlakuan lebih terhadap metode *support vector machine*, yaitu dengan melakukan seleksi atribut.

Kata Kunci : *support vector machine*, *random forest*, klasifikasi, biner, *multi-class*

Referensi : 10 (2001-2019)

ABSTRACT

ANDREW (00000019529)

PERFORMANCE COMPARISON BETWEEN RANDOM FOREST AND SUPPORT VECTOR MACHINE ON BINARY AND MULTI-CLASS DATA

Thesis, Faculty of Science and Technology (2021)

(xiv + 58 pages, 18 figures, 57 tables, 0 appendices)

Selecting the right method and algorithm for the data is one of the important process in data processing. Advancement in technology and data processing method gives researchers an opportunity to choose a method. In the research process, it takes a quite long time to selecting the right method and parameter, this is due to the number of factors that determine the results of data processing. In a classification research, support vector machine and random forest are two of the most commonly used methods. This research use eight datasets, four binary datasets and four multi-class datasets from different fields (health, finance, object classification), show that support vector machine is a better method for the binary data because it excels on two sets of data and draws on one set, and random forest shows a better performance on multi-class data because it excels on two sets of data and draws on one set. However, when compared through processing and average performance, it can be said that random forest is a better method. This is due to the need for more treatment of the support vector machine method, which is selecting attributes.

Keywords : support vector machine, random forest, classification, binary, multi-class

Reference : 10 (2001-2019)