

ABSTRAK

Derrick Devin Lim (03081200023)

PERBANDINGAN TINGKAT AKURASI ALGORITMA RANDOM FOREST DAN K-NEAREST NEIGHBOR (KNN) DALAM MENGLASIFIKASI PENYAKIT KANKER PARU-PARU

(xiii + 76 halaman; 41 gambar; 4 tabel; 1 lampiran)

Kanker paru-paru adalah suatu keadaan dimana terdapat sel kanker yang berkembang secara tidak terkendali di dalam organ paru-paru. Kanker paru-paru merupakan jenis kanker dengan tingkat kematian tertinggi dari total kematian akibat penyakit kanker dan menjadi penyebab utama dari kematian akibat kanker pada pria. Gejala pada kanker paru-paru sulit untuk terdeteksi pada stadium awal, dan hanya dapat terdeteksi ketika memasuki stadium lanjut. Oleh karena itu, penting agar dapat melakukan deteksi dini terhadap penyakit kanker paru-paru agar peluang untuk sembuh lebih besar. Solusi untuk mengatasi hal tersebut adalah memanfaatkan *machine learning* untuk mengklasifikasi tingkat risiko seseorang dalam terkena penyakit kanker paru-paru. Metode yang digunakan dalam penelitian ini adalah *K-Nearest Neighbor* dan *Random Forest*. Algoritma *K-Nearest Neighbor* dan *Random Forest* merupakan algoritma yang sering digunakan untuk melakukan klasifikasi. Penelitian ini bertujuan untuk melakukan perbandingan antara kedua algoritma tersebut dalam melakukan klasifikasi terhadap tingkat risiko seseorang dalam terkena penyakit kanker paru-paru. Dari hasil penelitian ini, dapat disimpulkan bahwa nilai akurasi yang dimiliki oleh *K-Nearest Neighbor* adalah 0.890 dan nilai akurasi *Random Forest* sebesar 0.933. Selain itu, nilai AUC yang dihasilkan oleh *K-Nearest Neighbor* adalah 0.973 dan untuk algoritma *Random Forest* adalah 0.991. Yang dapat disimpulkan bahwa *Random Forest* lebih unggul dari *K-Nearest Neighbor*.

Kata kunci: Kanker paru-paru, *K-Nearest Neighbor*, *Random Forest*

Referensi: 38 (2018-2023).

ABSTRACT

Derrick Devin Lim (03081200023)

COMPARISON OF THE ACCURACY OF RANDOM FOREST AND K-NEAREST NEIGHBOR (KNN) ALGORITHMS IN CLASSIFYING LUNG CANCER DISEASES

(xiii + 76 pages; 41 figures; 4 tables; 1 appendixes)

Lung cancer is a condition where there are cancer cells that develop uncontrollably in the lung organ. Lung cancer has the highest mortality rate of all cancer deaths and is the leading cause of cancer deaths in men. Symptoms of lung cancer are difficult to detect in the early stages, and can only be detected when it enters the advanced stages. Therefore, it is important to be able to do early detection of lung cancer so that the chances of recovery are greater. The solution to overcome this is to utilize machine learning to classify a person's risk level for lung cancer. The methods used in this research are K-Nearest Neighbor and Random Forest. K-Nearest Neighbor and Random Forest algorithms are algorithms that are often used for classification. This research aims to make a comparison between the two algorithms in classifying a person's risk level for lung cancer. From the results of this study, it can be concluded that the accuracy value owned by K-Nearest Neighbor is 0.890 and the accuracy value of Random Forest is 0.933. In addition, the AUC value generated by K-Nearest Neighbor is 0.973 and for the Random Forest algorithm is 0.991. Which can be concluded that Random Forest is superior to K-Nearest Neighbor.

Keywords: Lung cancer, K-Nearest Neighbor, Random Forest

References: 38 (2018-2023).