

Data Mining Untuk Memprediksi Lama Perawatan Pasien Covid-19 Di DIY

1. Agung Supoyo, 17121020, agungsupoyo@gmail.com
2. Putri Taqwa Prasetyaningrum, S.T., M.T., MCE. 0527108603,
putri@mercubuana-yogya.ac.id

ABSTRAK

Covid-19 merupakan penyakit berbahaya yang dapat menyebabkan kematian pada pasien dengan komorbid. Pada awal tahun 2021 kasus penularan Covid-19 di DIY yang masih tinggi dan ditambah dengan sulitnya mencari ruang perawatan rumah sakit, sehingga diperlukan penelitian berbasis data untuk memprediksi waktu perawatan pasien. Selain itu juga diperlukan analisis terhadap faktor-faktor yang paling mempengaruhi lama perawatan pasien.

Penelitian menggunakan *dataset* penyelidikan epidemiologi covid positif DIY sejumlah 3.029 baris yang terdiri dari 55 atribut dan dataset laporan hasil pemeriksaan lab sejumlah 2.823 baris data dengan 15 atribut. Dataset diperoleh dari Dinas Kominfo DIY untuk kasus periode Maret sampai dengan September 2020. Diperlukan *preprocessing* (*data reduction, data cleaning data integration*) sebelum dilakukan analisis *data mining*. *Preprocessing* menghasilkan *dataset* sejumlah 271 baris dengan 32 atribut.

Analisis *data mining* menggunakan algoritma *Random Forest*, k-NN dan *Deep Learning* menghasilkan *performance model* dengan RMSE masing-masing sebesar 4,949; 6,359 dan 5,436. Setelah dilakukan seleksi atribut untuk optimasi dihasilkan nilai RMSE sebesar 4.853 pada algoritma *Random Forest* dengan menggunakan 25 atribut.

Hasil *modeling* belum cukup baik untuk dapat digunakan dalam memprediksi lama perawatan karena jika dibandingkan dengan rata-rata lama perawatan sebesar 15.339 hari dihasilkan NRMSE sebesar 31,63%. Nilai *performance model* dipengaruhi oleh pemilihan atribut yang digunakan. Lima atribut yang paling berpengaruh terhadap lama perawatan pasien adalah usia, jenis kelamin, kecamatan, apakah pasien merupakan petugas kesehatan dan batuk. Untuk meningkatkan *performance model* diperlukan penelitian lanjutan menggunakan lebih banyak *record data* dengan tambahan atribut lain yang mempengaruhi lama perawatan.

Kata kunci: prediksi, lama perawatan Covid-19, *Random Forest*, k-NN, *Deep Learning*

DATA MINING FOR PREDICTING DAYS TO RECOVERY OF COVID-19 PATIENTS IN DIY

1. Agung Supoyo, 17121020, agungsupoyo@gmail.com
2. Putri Taqwa Prasetyaningrum, S.T., M.T., MCE. 0527108603,
putri@mercubuana-yogya.ac.id

ABSTRACT

At the beginning of 2021, the number of Covid-19 cases in DIY is very high and it is difficult to find the treatment room of the hospital, so it is necessary to analyze the prediction of treatment time. The results of the analysis will support the government's decision to formulate ward availability policies and implement The Community Activities Restrictions Enforcement. In addition, it is necessary to analyze the attributes that have the greatest impact on the length of patient treatment.

The study used a data set obtained from Dinas Kominfo DIY for the case from March to September 2020. Preprocessing (data reduction, data cleaning and data integration) is required before data mining analysis. The preprocessing will generate a data set containing 271 data records in 32 columns.

Data mining analysis using Random Forest, k-NN and Deep Learning algorithms resulted in a performance model with RMSE of 4.949; 6,359 and 5,436. After selecting the attributes for optimization, an RMSE value of 4.853 is generated in the Random Forest algorithm using 25 attributes.

The results of the modeling are not good enough to be used in predicting the length of treatment because when compared with the average length of treatment of 15,339 days, the NRMSE is 31.63%. The value of the performance model is influenced by the selection of the attributes used. The five attributes that most influence the length of patient care are age, gender, sub-district, whether the patient is a health worker and cough. To improve the performance of the model, further research is needed using more data records with the addition of other attributes that affect the length of treatment.

Keywords: regression, covid-19 recovery time, Random Forest, k-NN, Deep Learning