

# Klasifikasi Tingkat Risiko Pasien COVID-19 Dengan Pemilihan Model Machine Learning Didukung Oleh Reliabilitas Conformal = Classification Of COVID-19 Patients Degree Of Severity With Machine Learning Model Selection Supported By Conformal Reliability

Abyoso Hapsoro Nurhadi, author

Deskripsi Lengkap: <https://lib.ui.ac.id/detail?id=9999920521037&lokasi=lokal>

---

## Abstrak

Berdasarkan data World Health Organization (WHO) hingga 6 Januari 2022, terdapat 4.264.669 kasus terkonfirmasi COVID-19 dengan jumlah kematian sebanyak 144.116 pasien di Indonesia. Figur ini meningkat drastis saat dibandingkan data WHO hingga 25 April 2021 sebesar 1.636.792 kasus terkonfirmasi dengan jumlah kematian sebanyak 44.500 pasien. Varian B.1.617.2 atau lebih umum dikenal sebagai Delta dinyatakan hadir di Indonesia pada 3 Mei 2021 dengan dua kasus positif terdeteksi di Jakarta. Varian ini memiliki daya tular yang lebih tinggi dan mengakibatkan gejala COVID-19 lebih parah sehingga menjadi varian yang mendominasi persebaran COVID-19 di Indonesia. Menurut revisi protokol tatalaksana COVID-19 edisi ketiga, seorang pasien COVID-19 dapat dibedakan menjadi lima kategori berdasarkan severitas kasus yang diderita dengan tingkat risiko tertinggi yaitu kritis. Pasien COVID-19 yang digolongkan kategori kritis menunjukkan gejala Acute Respiratory Distress Syndrome (ARDS), sepsis, dan syok sepsis. Dengan menganalisis berbagai faktor yang terkait dengan gejala-gejala tersebut, dapat dibangun sebuah pemahaman berbentuk model Machine Learning untuk mengestimasi tingkat risiko kasus seorang pasien COVID-19. Model Machine Learning yang dibangun mencakup berbagai model, seperti model berbasis tree maupun berbasis ensemble. Dalam penelitian ini, tingkat risiko disimplikasi menjadi dua, yaitu severe dan non-severe berdasarkan urgensi perawatan khusus di rumah sakit. Untuk menentukan model optimal, digunakan metrik evaluasi Recall guna memberi perhatian kepentingan pasien tergolong kasus severe berhasil dideteksi severe dengan benar. Digunakan data pasien positif COVID-19 pada salah satu rumah sakit di Jakarta dari Januari 2020 hingga Agustus 2021 yang dibagi menjadi dua periode, sebelum dan sesudah adanya varian Delta. Dengan pembagian data ini, dapat dibangun tiga model Machine Learning yaitu model sebelum Delta, model setelah Delta, dan model keseluruhan. Dari ketiga model yang terbangun, akan diperiksa apakah ada perbedaan yang signifikan. Lebih lanjut, model-model Machine Learning yang terbentuk akan diuji tingkat kepercayaan terhadap prediksinya menggunakan metode Conformal. Diperoleh model Random Forest berhasil mengklasifikasikan data COVID-19 dengan lebih baik dibanding model-model lainnya. Model Random Forest pada seluruh variabel respon mencapai Recall 86,49%. Dengan identifikasi 4 variabel terpenting, model mencapai Recall 80,18%. Mendukung hasil ini, model percaya 90% dengan prediksi yang dihasilkan.

.....According to World Health Organization (WHO) data to 6 January 2022, there have been 4.264.669 confirmed cases of COVID-19 with 144.116 patient deaths in Indonesia. This figure has significantly increased when compared with WHO data to 25 April 2021, where there were 1.636.792 confirmed cases with 44.500 patient deaths. The B.1.617.2 variant or more commonly known as Delta was announced to be present in Indonesia on 3 May 2021 with two positive cases detected in Jakarta. This variant is more contagious and causes worse COVID-19 symptoms which made it the dominating variant of COVID-19 distribution in Indonesia. According to the revision of COVID-19 governance protocol third edition, a

patient of COVID-19 can be differentiated to five categories depending on the severity of their case with the highest risk being critical. A patient of COVID-19 that is categorized as critical will show symptoms of Acute Respiratory Distress Syndrome (ARDS), sepsis, and sepsis shock. Through analyzing the factors that are related to these symptoms, we can build an understanding in the form of Machine Learning to estimate a COVID-19 patient's degree of severity. The Machine Learning model that will be built encompasses many models, such as tree-based models and ensemble-based models. In this research, this degree is simplified into two, which are severe and non-severe with accordance to the urgency of special care in hospitals. To determine optimal models, the Recall evaluation metric is used as a means to give better attention to making sure severe patient cases are properly classified as severe. The data used will be positive COVID-19 patients in a Jakarta-based hospital from January 2020 until August 2021 which is split into two periods of before and after the presence of Delta variant. With this division, we can build three Machine Learning models which have learned before Delta, after Delta, and overall. From each of these built models, we will then determine if there exists a significant difference between them. Furthermore, the Machine Learning models that are built will be tested in their confidence on their own prediction using the Conformal method. We procure that Random Forest model classifies COVID-19 data better than all other models. Random Forest built on all response variables achieve 86,49% Recall. With the identification of 4 most important variables, the model achieves 80,18%. Supporting this result, the model has 90% faith in its prediction.