

# Prediksi Kematian Pasien COVID-19 Berdasarkan Skoring Variabel Menggunakan Regresi Logistik dan Machine learning = Mortality Prediction of Corona Virus Disease (COVID)-19 Patients Based on Scoring Developed with Logistic Regression and Machine Learning

Raden Rara Diah Handayani, author

Deskripsi Lengkap: <https://lib.ui.ac.id/detail?id=9999920536884&lokasi=lokal>

---

## Abstrak

Dalam dua tahun terakhir pandemi corona virus disease 2019 (COVID-19) telah menginfeksi > 220 juta orang dan 5 juta orang meninggal. Di Indonesia > 4 juta orang terinfeksi dan > 140.000 orang meninggal. Pada puncak pandemi, kebutuhan perawatan tidak seimbang dengan sarana rumah sakit sehingga WHO menganjurkan untuk memprioritaskan pasien secara ekual. Untuk itu diperlukan prediktor luaran pasien COVID-19. Penelitian ini bertujuan menyusun prediktor luaran pasien COVID-19 menggunakan regresi logistik dan machine learning.

Penelitian terdiri atas 2 tahap. Tahap pertama adalah kohort retrospektif untuk menyusun prediktor kematian di rumah sakit dengan regresi logistik dan machine learning (decision tree, random forest, support vectore machine, gradient boost and extreme gradient boost). Pasien terkonfirmasi COVID-19 diinput di data registri REG-COVID-19 pada bulan Maret–Juli 2020 di RS Persahabatan (RSP) dan RS Universitas Indonesia (RSUI). Tahap kedua adalah kohort prospektif pada pasien COVID-19 di RSP, RSUI dan RSPI Suliati Saroso pada bulan Maret–Mei 2021. Data yang diinput adalah data demografi, gejala klinis, komorbid, laboratorium, skor Brixia dari radiografi toraks, luaran pasien dari perawatan dan lama rawat. Pada tahap penyusunan diperoleh 271 subjek untuk analisis machine learning, 239 subjek untuk model 1, sebanyak 180 subjek model 2, dan 152 subjek model 3 dan model 4. Hasil analisis regresi logistik model 1 terdiri atas 7 variabel yaitu demam, diabetes melitus, frekuensi napas, saturasi O<sub>2</sub>, leukosit, SGOT dan CRP dengan AUC 0,930. Model 2 memberikan hasil hampir sama tetapi SGOT menjadi SGPT dengan AUC 0,926. Model 3 memiliki AUC 0,919 dan model 4 memberikan AUC 0,924 dengan variabel D dimer > 2000 menjadi salah satu prediktor. Validasi semua model regresi logistik dan machine learning menunjukkan penurunan AUC, tetapi tidak berbeda bermakna (uji perbandingan AUC,  $p = 0,683-0,736$ ). Perbandingan model regresi logistik dan machine learning juga tidak berbeda bermakna (uji perbandingan AUC dengan rumus Hanley,  $p = 0,492-0,923$ ).

Disimpulkan prediksi kematian pasien COVID-19 menggunakan regresi logistik dan machine learning memiliki akurasi yang baik sehingga regresi logistik dan machine learning dapat dijadikan prediktor luaran pasien COVID-19.

.....Corona virus disease 2019 (COVID-19) pandemic has lasted almost 2 years worldwide with more than two hundred million world population were infected and almost 5 million (2%) death. In Indonesia, there have been more than 4 million people were infected with more than 140.000 (3.5%) death. At the peak of the outbreak there were discrepancy between health care facilities and demands. WHO recommended to prioritize patient equally, to avoid patient discrimination by social class, race, and gender. The best prediction tool should be valid, reliable and feasible. Many studies develop assessment with logistic regression and machine learning with the goal to improve accuracy. Some study showed variety of predictors in outcome prediction, in this study we developed and validated assessment tool to predict

hospital mortality comparing logistic regression and machine learning, included support vector machine (SVM), decision tree (DT), random forest (RF), gradient boost (GB) and extreme gradient boost (XGB). Our study was conducted in 2 stages. The first stage study was cohort retrospective to develop assessment tool to predict hospital mortality by comparing logistic regression and machine learning among hospitalized COVID-19 patients from March to July 2020. The second was cohort prospective study among the same population, to validate the tools. The development data were collected from Persahabatan hospital and Universitas Indonesia hospital who registered in REG-COVID-19, 271 subjects were eligible for machine learning analysis and 239 subjects for logistic regression data set 1; 180 subjects for data set 2; 152 for data set 3 and 4. Analysis of data set 1 resulted in 8 variables as mortality prediction include fever, DM, respiratory rate (RR), oxygen saturation, leucocyte, ALT > 42, CRP > 88, with AUC 0,930. Data set 2 resulted in similar variables except AST, with AUC 0,926. Data set 3 resulted in 6 variables with AUC 0,919 and Data set 4 resulted in 7 variables included fever, HR, RR, leucocyte, age above 52, CRP > 86 and D-dimer > 2000 with AUC 0,924. Validation of all models showed decreasing AUC. Machine learning analysis resulted in 5 models with the best was XGB among all set data with AUC between 0,8–0,9. There were decreasing of AUC of all models, but not statistically different (p 0.683–0.736). Comparing developed models with logistic regression and machine learning showed there were differences but not statistically significant. (p 0.492-0.923)