

ABSTRAK

Salah satu indikator kesejahteraan rakyat dalam suatu negara adalah faktor kesehatan dan gizi yang tercukupi (Badan Pusat Statistik, 2022). Termasuk di dalamnya juga keselamatan persalinan bagi ibu hamil dan anak (bayi) yang dilahirkan. Namun data milik Kementerian Kesehatan di Indonesia menunjukkan bahwa terdapat kenaikan jumlah kematian pada ibu tahun 2021 sebanyak 6.856 kasus, meningkat dari jumlah kasus yang sama pada tahun 2019 yang sebanyak 4.197 (Kemen PPPA, 2022). Hal itu dapat dicegah dengan melakukan pemeriksaan antenatal agar selanjutnya dapat ditentukan jenis persalinan yang akan dilakukan saat melahirkan.

Berangkat dari hal tersebut, penelitian ini bertujuan untuk membangun sebuah sistem yang dapat mengklasifikasi (*supervised learning*) jenis persalinan pada ibu hamil dengan metode *ensemble learning* (pembelajaran ensambel) menggunakan algoritma *categorical boosting* (*CatBoost*). Untuk dapat memberikan kontribusi yang berarti, penelitian sejak persiapan data (*preparation*) hingga pemilihan skenario pengujian untuk pemodelan dipilih secara cermat agar diperoleh hasil yang seoptimal mungkin. Digunakan skenario penerapan teknik *balancing* dengan *SMOTE-N* dan variasi pembelajaran *k-fold* (5 dan 10) pada *cross-validation*. Evaluasi metrik menggunakan *confusion matrix* untuk mengukur kinerja pemodelan dari *CatBoost*.

Pemodelan terbaik diperoleh pada pembelajaran *CatBoost* dengan *balancing SMOTE-N* yang dilatih *10-fold cross-validation*, dimana akurasi tertinggi untuk pelatihan dengan rata-rata 0.93 dan pengujian bernilai 0.95 (95%).

Kata kunci: *balancing, CatBoost, categorical boosting, confusion matrix, data preparation, ensambel, ensemble learning, parameter CatBoost, k-fold cross-validation, klasifikasi, persalinan, SMOTE-N, supervised learning.*

ABSTRACT

One of the indicators of people's welfare in a country is adequate health and nutrition (Badan Pusat Statistik, 2022). That also includes the safe delivery for pregnant women and babies who are born. But, data from the Indonesian Ministry of Health shows the death of maternal was increase in 2021 with 6.856 cases, that was an increase from the same case in 2019 with just 4.197 (Kemen PPPA, 2022). This can be prevented with conducting antenatal check-up, so then the type of delivery can be determined when labor was about to perform.

Departing from that information, this research aims to build a system that could perform classification (supervised learning) the type of delivery on pregnant women with the ensemble learning method using the categorical boosting algorithm (CatBoost). To be able make a meaningful contribution, the research from data preparation to selection of test scenarios for modeling was carefully chosen in order to obtain the most optimal results as possible. The scenario of applying balancing technique with SMOTE-N and variation of k-fold learning with 5 and 10 folds for cross-validation. Metric evaluation uses confusion matrix to measure the modeling performance of CatBoost.

The best modeling was obtained in CatBoost learning with SMOTE-N balancing trained by 10-fold cross-validation, where the highest accuracy was for training with average of 0.93 and testing is 0.95 (95%).

Keywords: balancing, CatBoost, categorical boosting, confusion matrix, data preparation, ensemble learning, CatBoost parameter, k-fold cross-validation, classification, labor, SMOTE-N, supervised learning

