Universitas Indonesia Library >> UI - Tesis Membership

Rancang Bangun Model Prediksi Risiko Stunting di Indonesia dengan Penerapan Data Science = Development of Stunting Risk Prediction Model in Indonesia with the Application of Data Science

Tarigan, Ibrahim Amyas Aksar, author

Deskripsi Lengkap: https://lib.ui.ac.id/detail?id=20523571&lokasi=lokal

---

Abstrak

Stunting adalah kondisi di mana balita memiliki panjang atau tinggi badan di bawah rata-rata yang diakibatkan oleh buruknya gizi yang didapatkan. Stunting berpotensi memperlambat perkembangan otak, dengan dampak jangka panjang berupa keterbelakangan mental, rendahnya kemampuan belajar, dan risiko serangan penyakit kronis (diabetes, hipertensi dan obesitas). Salah satu upaya untuk memitigasi dan mengurangi angka kejadian stunting adalah dengan menerapkan ilmu Data Science dalam mengevaluasi dan membuat model prediksi risiko dari faktor yang dianggap dominan dalam terjadinya stunting. Kerangka kerja CRISP-DM digunakan sebagai metodologi penelitian ini. Data yang digunakan adalah data sekunder dari Indonesia Family Life Survey (IFLS) tahun 2014-2015, yang dilakukan di 13 provinsi di Indonesia. Pemodelan dilakukan menggunakan bahasa pemrograman Python yang dikerjakan pada Jupyter Notebook. Dua metode sampling (Random Sampling dan Oversampling) digunakan untuk mengatasi masalah data imbalance. Empat algoritma classifier diujikan untuk mengetahui performa model. Hasil penelitian menunjukkan keempat algoritma dinyatakan model yang baik dengan performa rata-rata lebih dari 70%. Pada teknik undersampling, skor akurasi yang didapatkan Logistic Regression sebesar 98,31%, KNN sebesar 77,53%, SVC sebesar 97,48%, dan Decision Tree Classifier sebesar 80,76%. Sedangkan pada teknik oversampling, skor akurasi yang didapatkan Logistic Regression adalah 95%, KNN sebesar 71%, SVC sebesar 93%, dan Decision Tree Classifier sebesar 84%. Logistic Regression adalah algoritma klasifikasi terbaik yang dapat digunakan pada dataset ini dan hasil perbandingan skor Logistic Regression pada teknik undersampling (95,09%) dan oversampling (91,98%) setelah 10 kali percobaan menunjukkan bahwa meskipun teknik undersampling memiliki performa yang lebih baik, namun teknik oversampling memiliki hasil yang lebih konsisten.

.....Stunting is a condition in which a toddler has a length or height below the average caused by poor nutrition. Stunting has the potential to slow down brain development, with long-term effects in the form of intellectual disability, low learning abilities, and the risk of developing chronic diseases (diabetes, hypertension and obesity). One of the efforts to mitigate and reduce the incidence of stunting is to apply Data Science in evaluating and making risk prediction models from factors that are considered dominant in the occurrence of stunting. The CRISP-DM framework was used as the methodology of this study. The data used is secondary data from the Indonesia Family Life Survey (IFLS) 2014-2015, which was conducted in 13 provinces in Indonesia. The modelling is made using the Python programming language which is done on Jupyter Notebook. Two sampling methods (Random Sampling and Oversampling) were used to overcome the problem of data imbalance. Four classifier algorithms were tested to determine the performance of the model. The research results showed four algorithms were declared good models with an average performance of more than 70%. In the undersampling technique, the accuracy score obtained by Logistic Regression is 98.31%, KNN is 77.53%, SVC is 97.48%, and Decision Tree Classifier is 80.76%. While in the oversampling technique, the accuracy score obtained by Logistics Regression is 95%, KNN is 71%,

SVC is 93%, and Decision Tree Classifier is 84%. Logistic Regression is the best classification algorithm that can be used in this dataset and the results of the comparison of Logistic Regression scores on the undersampling (95.09%) and oversampling (91.98%) techniques after 10 experiments show that although the undersampling technique has better performance, oversampling technique has more consistent results.