

MARWA HALIZA, 22.240.0070

STUDI KOMPARASI MODEL PREDIKSI RISIKO STUNTING MENGGUNAKAN METODE RANDOM FOREST DAN XGBOOST: STUDI KASUS DI KECAMATAN PEKALONGAN BARAT, dibawah bimbingan Bapak Dr. P.A. CHRISTIANTO S.Kom., M.Kom. dan Ibu TRI PUDJI W., S.E., M.Si.

84 + lxxxvi halaman / 9 gambar / 12 tabel / 3 lampiran / 14 pustaka (2022-2025)

ABSTRAK

Stunting merupakan salah satu permasalahan gizi kronis yang masih menjadi perhatian serius di Indonesia, termasuk di Kota Pekalongan. Upaya penanganan stunting dilakukan oleh Dinas Sosial Pengendalian Penduduk dan Keluarga Berencana (DINSOS P2KB) bersama Dinas Kesehatan melalui pendataan demografi dan identifikasi keluarga berisiko. Namun, kompleksitas data menyebabkan identifikasi keluarga berisiko stunting belum optimal apabila dianalisis menggunakan metode konvensional. Penelitian ini bertujuan membandingkan performa algoritma Random Forest dan XGBoost dalam memprediksi risiko stunting pada keluarga di Kecamatan Pekalongan Barat menggunakan pendekatan CRISP-DM. Data yang digunakan merupakan data sekunder dari sistem SIGA berupa laporan data stunting DINSOS P2KB Kota Pekalongan periode 2023–2024 sebanyak 26.094 KK. Data terdiri atas lima atribut fitur biner, yaitu Baduta, Balita, Air Tidak Layak, Jamban Tidak Layak, dan Bukan KB Modern, serta satu atribut target Risiko Stunting. Tahapan pengolahan data meliputi pembersihan duplikat, penanganan missing value, validasi domain nilai, seleksi atribut, serta pembagian data latih dan data uji dengan proporsi 80:20 menggunakan stratified sampling. Setelah proses pembersihan, diperoleh 25.822 data yang terdiri atas 20.657 data latih dan 5.165 data uji. Hasil pengujian menunjukkan bahwa Random Forest dan XGBoost menghasilkan performa klasifikasi yang identik, dengan accuracy 98,72%, precision 96,15%, recall 84,75%, dan F1-Score 90,09%. Meskipun memiliki performa klasifikasi yang sama, XGBoost lebih efisien secara komputasi dengan waktu pelatihan 0,4441 detik dan prediksi 0,0299 detik, sedangkan Random Forest membutuhkan 3,5585 detik dan 0,0396 detik. Berdasarkan hasil tersebut, XGBoost lebih efisien untuk klasifikasi risiko stunting berbasis data administratif di lingkungan DINSOS P2KB Kota Pekalongan.

Kata kunci : Stunting, Random Forest, XGBoost, klasifikasi, CRIPS-DM.

MARWA HALIZA, 22.240.0070

COMPARATIVE STUDY OF STUNTING RISK PREDICTION MODELS USING RANDOM FOREST AND XGBOOST METHODS: A CASE STUDY IN WEST PEKALONGAN DISTRICT, under the guidance of Mr. Dr. P.A. CHRISTIANTO S.Kom., M.Kom. dan Mrs TRI PUDJI W., S.E., M.Si.

84 + lxxxvi pages / 9 figures / 12 tables / 3 appendices / 14 references (2022–2025)

ABSTRACT

Stunting is one of the chronic nutritional problems that remains a serious concern in Indonesia, including in Pekalongan City. Efforts to address stunting are carried out by the Social Affairs, Population Control, and Family Planning Office (DINSOS P2KB) together with the Health Office through demographic data collection and identification of at-risk families. However, the complexity of the data causes the identification of families at risk of stunting to be less than optimal when analyzed using conventional methods. This study aims to compare the performance of the Random Forest and XGBoost algorithms in predicting stunting risk among families in West Pekalongan District using the CRISP-DM approach. The data used are secondary data from the SIGA system in the form of stunting data reports from DINSOS P2KB of Pekalongan City for the 2023–2024 period, consisting of 26,094 households. The data consist of five binary feature attributes, namely Under-Two Children (Baduta), Toddlers (Balita), Inadequate Water, Inadequate Sanitation Facilities, and Non-Modern Family Planning, as well as one target attribute, namely Stunting Risk. The data processing stages include duplicate removal, missing value handling, domain value validation, attribute selection, and division of training and testing data at a ratio of 80:20 using stratified sampling. After the data cleaning process, 25,822 data records were obtained, consisting of 20,657 training data and 5,165 testing data. The test results show that Random Forest and XGBoost achieved identical classification performance, with an accuracy of 98.72%, precision of 96.15%, recall of 84.75%, and F1-Score of 90.09%. Although they achieved the same classification performance, XGBoost was more computationally efficient, with a training time of 0.4441 seconds and a prediction time of 0.0299 seconds, while Random Forest required 3.5585 seconds and 0.0396 seconds, respectively. Based on these results, XGBoost is more efficient for stunting risk classification based on administrative data within DINSOS P2KB of Pekalongan City.

Keywords: Stunting, Random Forest, XGBoost, classification, CRIPS-DM.