

ABSTRAK

Della Andini

Perancangan Sistem Prediksi Penyakit Diabetes Berbasis Web Menggunakan Algoritma *Random Forest* Dengan Teknik SMOTE

11 lampiran depan + 76 halaman + 12 tabel + 30 gambar + 7 lampiran akhir

Latar Belakang: Diabetes merupakan penyakit kronis dengan prevalensi tinggi di dunia maupun di Indonesia, menempati peringkat kelima global menurut *International Diabetes Federation* (IDF). Keterlambatan deteksi dini menyebabkan komplikasi serius, sehingga dibutuhkan sistem prediksi berbasis teknologi yang mudah diakses.

Metode: Penelitian ini merancang sistem prediksi diabetes berbasis web menggunakan algoritma *Random Forest* dengan metode pengembangan sistem *Waterfall*. Dataset yang digunakan adalah *Pima Indians Diabetes Database* dari *Kaggle*. Untuk mengatasi ketidakseimbangan data digunakan teknik SMOTE, serta dilakukan *hyperparameter tuning*. Sistem dikembangkan menggunakan *Python* dan *framework Streamlit*, kemudian di-deploy melalui *Streamlit Cloud*.

Hasil Penelitian: Model prediksi mencapai akurasi 78,35% dengan nilai *precision*, *recall*, dan *F1-score* yang cukup baik. Sistem menampilkan hasil berupa “Risiko Rendah” atau “Risiko Tinggi” serta dilengkapi fitur edukasi kesehatan yang memberikan saran pencegahan, informasi komplikasi, dan manajemen gaya hidup.

Kesimpulan: Sistem yang dirancang tidak hanya memberikan prediksi akurat, tetapi juga meningkatkan literasi kesehatan. Dengan akses mudah berbasis web, sistem ini dapat menjadi alat bantu deteksi dini dan edukasi mandiri bagi masyarakat maupun tenaga kesehatan.

Kata Kunci : Prediksi diabetes, *Random Forest*, SMOTE, *Streamlit*, *Machine Learning*, berbasis web.

Referensi : 52 (2017-2025)

ABSTRACT

Della Andini

Design of a Web-Based Diabetes Prediction System Using the Random Forest Algorithm with SMOTE Technique

11 preliminary pages + 76 pages + 12 tables + 30 figures + 7 final appendices

Background: Diabetes is a chronic disease with a high prevalence both globally and in Indonesia, ranking fifth worldwide according to the International Diabetes Federation (IDF). Delays in early detection often lead to serious complications therefore, a technology-based prediction system that is easily accessible is required.

Method: This study designs a web-based diabetes prediction system using the Random Forest algorithm with the Waterfall system development method. The dataset used is the Pima Indians Diabetes Database from Kaggle. To address class imbalance, the SMOTE technique was applied, and hyperparameter tuning was performed. The system was developed using Python and the Streamlit framework, then deployed through Streamlit Cloud.

Research Results: The prediction model achieved an accuracy of 78.35% with satisfactory precision, recall, and F1-score values. The system provides outputs in the form of “Low Risk” or “High Risk” and is equipped with health education features that offer prevention advice, complication information, and lifestyle management.

Conclusion: The designed system not only provides accurate predictions but also enhances health literacy. With easy access through the web, this system can serve as an early detection and self-education tool for both the community and healthcare professionals.

Keywords: Diabetes prediction, Random Forest, SMOTE, Streamlit, Machine Learning, web-based.

References: 52 (2017–2025)