

IMPLEMENTASI MODEL *CATBOOST* DAN XAI DALAM APLIKASI PREDIKSI RETENSI KARYAWAN BERBASIS *STREAMLIT*

Jesselyn Mu

ABSTRAK

Perusahaan merupakan entitas yang bertujuan memperoleh keuntungan melalui produksi atau jasa, dengan karyawan sebagai aset terpenting. Namun, saat ini retensi karyawan menjadi tantangan besar, khususnya di Indonesia. Data menunjukkan terjadi peningkatan *turnover* sejak tahun 2020 hingga 2024, terutama di kalangan usia produktif. Faktor internal dan eksternal turut mempengaruhi keputusan karyawan untuk bertahan atau mengundurkan diri. Oleh karena itu, penelitian ini bertujuan untuk mengembangkan aplikasi berbasis *Web* dengan menggunakan *Streamlit* yang dapat memprediksi retensi karyawan dan mengidentifikasi faktor hasil prediksi tersebut. Metodologi yang digunakan melibatkan algoritma *CatBoost* untuk klasifikasi (retensi atau tidak) dan regresi (lama masa kerja), serta *Explainable AI* dengan *SHAP Values* untuk interpretasi hasil prediksi. Hasil penelitian menunjukkan bahwa model klasifikasi terbaik mencapai akurasi 0,93 dan *F1-score* 0,96 pada data uji. Model regresi menghasilkan *R² Score* sebesar 0,99 dengan MAPE 0,01, menunjukkan performa prediksi yang sangat baik. Selain itu, penerapan *SHAP Values* berhasil mengidentifikasi faktor hasil prediksi klasifikasi tersebut. Aplikasi ini berhasil divalidasi melalui *BlackBox Testing* dengan skor sempurna 100%. Penelitian ini diharapkan dapat membantu perusahaan mengurangi *turnover* dan mempertahankan karyawan potensial.

Kata Kunci: *CatBoost*, *Explainable AI*, Retensi, *Streamlit*

IMPLEMENTATION OF CATBOOST AND XAI MODELS IN STREAMLIT-BASED EMPLOYEE RETENTION PREDICTION APPLICATION

Jesselyn Mu

ABSTRACT

A company is an entity that aims to generate profit through the production of goods or services, with employees as its most valuable asset. However, employee retention has become a major challenge today, especially in Indonesia. Data shows a continuous increase in turnover from 2020 to 2024, particularly among the productive age group. Both internal and external factors influence employees' decisions to stay or resign. Therefore, this study aims to develop a Web-based application using Streamlit to predict employee retention and identify the contributing factors behind the predictions. The methodology involves the CatBoost algorithm for classification (retained or not) and regression (length of employment), as well as Explainable AI with SHAP Values for interpreting the prediction results. The findings show that the best classification model achieved an accuracy of 0.93 and an F1-score of 0.96 on the test data. The regression model achieved an R^2 score of 0.99 with a MAPE of 0.01, indicating excellent predictive performance. Additionally, the implementation of SHAP Values successfully identified the contributing factors in the classification predictions. The application was successfully validated through BlackBox Testing with a perfect score of 100%. This research is expected to help companies reduce turnover and retain valuable employees.

Keyword: *CatBoost, Explainable AI, Retention, Streamlit*