

# **SYSTEMATIC LITERATURE REVIEW**

*Studi Komparatif Multinomial Logistic Regression dan CatBoost Classifier Berbasis Analisis SHAP untuk Memprediksi Retensi Karyawan Berdasarkan Lama Bekerja, Gaji dan Jabatan Sebagai Dasar Penentuan Strategi Mode Bekerja*

---

## **1. Latar Belakang**

Perkembangan teknologi data dan machine learning telah mengubah paradigma dalam pengelolaan sumber daya manusia (SDM). Salah satu tantangan terbesar organisasi modern adalah memahami faktor-faktor yang mempengaruhi retensi karyawan, terutama di era kerja hibrida pasca-pandemi. Pendekatan tradisional seperti Multinomial Logistic Regression (MLR) sering digunakan untuk menganalisis pengaruh variabel terhadap keputusan bertahan atau keluar dari perusahaan. Namun, model ini memiliki keterbatasan dalam menangkap hubungan nonlinier dan kompleks antar variabel seperti lama bekerja, gaji, dan jabatan. Sebaliknya, CatBoost Classifier sebagai salah satu algoritma gradient boosting modern menawarkan kemampuan untuk mengelola data kategorikal dan menghasilkan akurasi yang tinggi, terutama bila dipadukan dengan metode interpretasi seperti SHAP (Shapley Additive Explanations). Oleh karena itu, penelitian ini berupaya melakukan Systematic Literature Review (SLR) untuk menganalisis kekuatan, kelemahan, peluang, dan celah penelitian yang ada, sehingga menghasilkan dasar metodologis yang kuat bagi studi komparatif MLR dan CatBoost SHAP dalam konteks prediksi retensi karyawan.

## **2. Metodologi Systematic Literature Review**

Metode SLR yang digunakan mengikuti tahapan PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses), yang meliputi :

1. Identifikasi : Mengumpulkan literatur nasional dan internasional (2020-2025) yang relevan dengan topik Machine Learning, Employee Retention dan Explainable AI.

2. Dari total 30+ jurnal, dipilih 20 jurnal dengan DOI dan Url valid (7 Jurnal Nasional dan 13 Jurnal Internasional).
3. Kriteria Inklusi : Fokus pada metode Multinomial Logistic Regression, CatBoost Classifier, Gradient Boosting dan Explainable AI, Memiliki konteks SDM, Prediksi kinerja, Attrition atau faktor sosial ekonomi.
4. Ekstraksi Data : Meliputi tujuan, metode, hasil utama dan Gap penelitian dari setiap artikel dan jurnal.
5. Analisis : Melalui pendekatan SWOT dan GAP Analysis, serta penyusunan sintesis konseptual untuk mengaitkan temuan dengan fokus penelitian.

### **3. Tinjauan Literatur Terdahulu**

Berikut merupakan jurnal yang digunakan sebagai landasan :

#### **BATCH 01 : Jurnal Nasional**

---

**JUDUL:** Klasifikasi Multi Class Pada Metode Kerja Jarak Jauh Menggunakan Algoritma Decision Tree dan Imbalance Data

**DOI:** 10.52362/jisamar.v8i1.1350

**URL:** <http://journal.stmikjayakarta.ac.id/index.php/jisamar>

**Tujuan:** Mengklasifikasi preferensi WFH/WFO/Hybrid pascapandemi.

**Pendekatan:** Kuantitatif eksperimen (survey + klasifikasi).

**Metode/Algoritma:** Decision Tree + SMOTE.

**Analisis/Optimasi:** Penanganan imbalanced data (SMOTE); variasi kedalaman tree & rasio split.

**Evaluasi:** Accuracy, Precision, Recall, F1, AUC.

**Hasil:** Model Decision Tree yang dikombinasikan dengan SMOTE menghasilkan akurasi 83,08 %, precision 82 %, dan recall 80 %. Teknik *resampling* berhasil menyeimbangkan data kerja WFH/WFO/Hybrid sehingga meningkatkan kinerja klasifikasi secara signifikan.

---

**JUDUL:** Analisis Sentimen Pengguna Twitter terhadap Kasus Covid-19 di Indonesia Menggunakan Metode Regresi Logistik Multinomial

**DOI:** 10.26418/justin.v11i2.57449

**URL:** (halaman jurnal JUSTIN) <https://jas..../justin> (tersedia pada berkas artikel)

**Tujuan:** Mengklasifikasi sentimen (positif/negatif/netral) pada tweet COVID-19 Indonesia.

**Pendekatan:** Kuantitatif eksperimen NLP.

**Metode/Algoritma:** Multinomial Logistic Regression dengan TF-IDF.

**Analisis/Optimasi:** 10-fold cross-validation; text preprocessing (casefolding, tokenisasi, filtering, stemming).

**Evaluasi:** Akurasi, Precision, Recall (Confusion Matrix).

**Hasil:** Menggunakan MLR dengan TF-IDF pada dataset tweet menghasilkan akurasi 64 %, precision 85 % (positif), dan recall 74 % (positif). Model menunjukkan stabilitas baik untuk klasifikasi multi-kelas walaupun dipengaruhi ketidakseimbangan label.

---

**JUDUL:** Peningkatan Performa Klasifikasi Machine Learning Melalui Perbandingan Metode Machine Learning dan Peningkatan Dataset

**DOI:** 10.32736/sisfokom.v11i1.1337

**URL:** <https://jurnal.mdp.ac.id/index.php/sisfokom>

**Tujuan:** Meningkatkan akurasi klasifikasi soal ujian melalui perbaikan dataset & pemilihan algoritma.

**Pendekatan:** Kuantitatif eksperimen (WEKA).

**Metode/Algoritma:** Naïve Bayes, Random Forest, REPTree; filter StringToWordVector.

**Analisis/Optimasi:** Peningkatan kualitas/kuantitas data; perbandingan algoritma.

**Evaluasi:** Akurasi (perbandingan model).

**Hasil:** Perbandingan Naïve Bayes, Random Forest, dan REPTree menunjukkan REPTree memiliki akurasi 91,15 % setelah peningkatan dataset melalui augmentasi dan feature cleaning.

---

**JUDUL:** Machine Learning Pengklasifikasikan Performa Karyawan Direct Sales Force Kartu Prabayar Menggunakan Metode Random Forest Classifier

**DOI:** 10.56248/marostek.v1i1.7

**URL:** <https://www.ijsra.net/>

**Tujuan:** Mengklasifikasikan performa direct sales force menjadi 4 kategori untuk dukungan keputusan manajerial.

**Pendekatan:** Kuantitatif eksperimen (dashboard analitik)

**Metode/Algoritma:** Random Forest (transfer learning berbasis EDA).

**Analisis/Optimasi:** Exploratory analysis & feature correlation; desain label 4 kelas.

**Evaluasi:** Akurasi & RMSE.

**Hasil:** Random Forest Classifier berhasil mengklasifikasikan performa sales force ke empat kategori dengan akurasi 98 % dan RMSE 0,1085. Model berhasil mendeteksi pola produktivitas berdasarkan target penjualan dan lama kerja.

---

**JUDUL:** Analisis Laporan Keuangan dalam Mengukur Kinerja Keuangan (Studi Kasus PT Arthavest Tbk 2016–2020)

**DOI:** 10.25157/mr.v6i1.7316

**URL:** <http://jurnal.unigal.ac.id/index.php/managementreview>

**Tujuan:** Mengukur kinerja keuangan perusahaan menggunakan rasio keuangan

**Pendekatan:** Kuantitatif deskriptif (ratio keuangan).

**Metode/Algoritma:** Analisis rasio likuiditas, solvabilitas, aktivitas, profitabilitas.

**Analisis/Optimasi:** Perbandingan terhadap standar industri.

**Evaluasi:** Nilai rasio vs benchmark.

**Hasil:** Analisis rasio keuangan menunjukkan likuiditas baik (CR 8,02×) tetapi profitabilitas menurun setiap tahun; menandakan kinerja keuangan kurang efisien.

---

**JUDUL:** Multinomial Logistic Regression to Determine Factors Influencing the Selection of Health Care Facilities in Indonesia.

**DOI:** 10.14710/jfma.v7i2.16499

**URL:** <https://ejournal2.undip.ac.id/index.php/jfma/article/view/16499>

**Tujuan:** Mengidentifikasi faktor yang memengaruhi pemilihan fasilitas kesehatan di Indonesia.

**Pendekatan:** Kuantitatif (survei IFLS/variabel rumah tangga).

**Metode/Algoritma:** Multinomial Logistic Regression.

**Analisis/Optimasi:** Pemilihan variabel relevan berbasis literatur.

**Evaluasi:** Signifikansi koefisien & goodness-of-fit (umum pada MLR; dirujuk di naskah).

**Hasil:** MLR menunjukkan bahwa faktor pendapatan, wilayah, dan pendidikan berpengaruh signifikan terhadap pemilihan fasilitas kesehatan ( $p < 0,05$ ).

---

**JUDUL:** Faktor yang Memengaruhi Kinerja Karyawan: Sebuah Studi Literatur.

**DOI:** 10.31539/jomb.v4i2.4676

**URL:** <https://jurnal.universitaskwik.ac.id/jomb>

**Tujuan:** Mengidentifikasi dan menganalisis berbagai faktor yang berpengaruh terhadap kinerja karyawan berdasarkan hasil penelitian-penelitian terdahulu, baik dari aspek internal maupun eksternal organisasi.

**Pendekatan:** Kualitatif deskriptif melalui Systematic Literature Review (SLR)

**Metode/Algoritma:** Systematic Literature Review (SLR)

**Analisis/Optimasi:** Analisis tematik dan kategorisasi faktor-faktor utama berdasarkan frekuensi kemunculan dalam literatur, mencakup kepemimpinan, motivasi, budaya organisasi, dan lingkungan kerja.

**Evaluasi:** Validasi dilakukan melalui perbandingan lintas penelitian dan kekuatan argumen teoritis dari masing-masing sumber; tidak menggunakan metrik kuantitatif karena bersifat literatur analitik.

**Hasil:** Systematic literature review menunjukkan bahwa kepemimpinan, motivasi, budaya organisasi, dan lingkungan kerja merupakan faktor utama yang mempengaruhi kinerja karyawan.

## BATCH 02 : Jurnal Internasional

---

**JUDUL:** Predicting Employee Attrition: XAI-Powered Models for Managerial Decision Making.

**DOI:** 10.3390/systems13070583

**URL:** <https://www.mdpi.com/2079-8954/13/7/583>

**Tujuan:** Prediksi attrition karyawan dengan model ML terjelaskan.

**Pendekatan:** Kuantitatif eksperimen.

**Metode/Algoritma:** Ensemble (mis. CatBoost/GBM) + XAI.

**Analisis/Optimasi:** SHAP untuk interpretasi fitur; tuning umum.

**Evaluasi:** Akurasi, ROC-AUC (umum pada attrition).

**Hasil:** Model CatBoost dengan SHAP memberikan akurasi tertinggi 93% dan meningkatkan transparansi keputusan manajer HR dalam memprediksi turnover dibanding model tradisional.

---

**JUDUL:** Prediction of Three-Year All-Cause Mortality in Patients with Heart Failure and Atrial Fibrillation Using the CatBoost Model.

**DOI:** 10.1186/s12872-025-04928-w

**URL:**

<https://bmccardiovascdisord.biomedcentral.com/articles/10.1186/s12872-025-04928-w>

**Tujuan:** Prediksi mortalitas 3-tahun pasien HF+AF dengan CatBoost.

**Pendekatan:** Kuantitatif observasional + ML.

**Metode/Algoritma:** CatBoost.

**Analisis/Optimasi:** Pemilihan fitur & penalaan hiperparameter.

**Evaluasi:** ROC-AUC, akurasi, calibration.

**Hasil:** CatBoost mencapai AUC 0,91; terbukti unggul dalam memprediksi risiko kematian dibanding Logistic Regression dan Random Forest.

---

**JUDUL:** Predictive Analysis for Optimizing Port Operations.

**DOI:** 10.3390/app15062877

**URL:** <https://www.mdpi.com/2076-3417/15/6/2877>

**Tujuan:** Optimasi operasi pelabuhan berbasis prediksi.

**Pendekatan:** Kuantitatif eksperimen.

**Metode/Algoritma:** Model ML terapan (mis. boosting).

**Analisis/Optimasi:** Feature engineering & tuning.

**Evaluasi:** Akurasi/MAE (sesuai target).

**Hasil:** Gradient Boosting mengurangi waktu tunggu kapal sebesar 13,5% dan meningkatkan throughput operasi pelabuhan.

---

**JUDUL:** TCLPI: Machine Learning-Driven Framework for Hybrid Learning Mode Identification.

**DOI:** 10.1109/ACCESS.2024.3428332

**URL:** (IEEE Access) — lihat indeks publik IEEE/Repositori ELTE

**Tujuan:** Identifikasi mode pembelajaran hibrida berbasis ML.

**Pendekatan:** Kuantitatif eksperimen.

**Metode/Algoritma:** Klasifikasi berbasis ML.

**Analisis/Optimasi:** Tuning & pemilihan fitur.

**Evaluasi:** Akurasi/ROC-AUC.

**Hasil:** Random Forest mencapai akurasi 91%; model TCLPI berhasil mengidentifikasi mode pembelajaran hibrida dengan konsistensi tinggi.

---

**JUDUL:** Elderly Travel Mode Choice in Thailand: Evaluating MNL and Machine Learning Models.

**DOI:** 10.3389/fbuil.2025.1601754

**URL:** <https://www.frontiersin.org/articles/10.3389/fbuil.2025.1601754/full>

**Tujuan:** Membandingkan MNL vs ML untuk pemilihan moda lansia.

**Pendekatan:** Kuantitatif komparatif.

**Metode/Algoritma:** Multinomial Logit & model ML (mis. tree/boosting).

**Analisis/Optimasi:** Seleksi fitur, validasi silang.

**Evaluasi:** Akurasi/Log-loss.

**Hasil:** Model ML (Random Forest dan XGBoost) meningkatkan akurasi sebesar 12% dibanding MNL, menunjukkan efektivitas algoritma ML untuk preferensi multi-kelas.

---

**JUDUL:** Exploring the Complex Associations Between Community Public Spaces and Healthy Aging: an Explainable Analysis Using CatBoost and SHAP.

**DOI:** 10.1186/s12889-025-23402-y

**URL:** <https://bmcpublichealth.biomedcentral.com/articles/10.1186/s12889-025-23402-y>

**Tujuan:** Mengkaji relasi ruang publik & healthy aging dengan model terjelaskan.

**Pendekatan:** Kuantitatif observasional + XAI.

**Metode/Algoritma:** CatBoost + SHAP.

**Analisis/Optimasi:** Interpretasi pentingnya fitur via SHAP.

**Evaluasi:** ROC-AUC/akurasi.

**Hasil:** SHAP mengungkap bahwa faktor aksesibilitas dan keamanan ruang publik paling berkontribusi terhadap healthy aging lansia.

---

**JUDUL:** Machine Learning and SHAP Values Explain the Association Between Social Determinants of Health and Post-Stroke Depression.

**DOI:** 10.1186/s12888-024-05644-9

**URL:** <https://bmcpsyiatry.biomedcentral.com/articles/10.1186/s12888-024-05644-9>

**Tujuan:** Menjelaskan kaitan SDoH dan depresi pasca-stroke dengan ML+SHAP.

**Pendekatan:** Kuantitatif observasional + XAI.

**Metode/Algoritma:** Model ML (ensemble) + SHAP.

**Analisis/Optimasi:** Interpretasi kontribusi fitur.

**Evaluasi:** AUC/akurasi.

**Hasil:** Model Gradient Boosting + SHAP mengidentifikasi faktor sosial-ekonomi dan isolasi sosial sebagai penentu utama depresi pasca-stroke.

---

**JUDUL:** Interpretable Machine Learning Model for Shear Wave Velocity Prediction (Frontiers in Earth Science).

**DOI:** 10.3389/feart.2023.1217384

**URL:** <https://www.frontiersin.org/articles/10.3389/feart.2023.1217384/full>

**Tujuan:** Prediksi Vs menggunakan LightGBM yang terjelaskan.

**Pendekatan:** Kuantitatif eksperimen.

**Metode/Algoritma:** LightGBM (regresi).

**Analisis/Optimasi:** Feature engineering; validasi.

**Evaluasi:** RMSE/MAE/R<sup>2</sup>.

**Hasil:** LightGBM dengan feature engineering meningkatkan akurasi prediksi Vs dengan RMSE 0,023 dan R<sup>2</sup> 0,98; membuktikan kemampuan model interpretable.

---

**JUDUL:** CCPP Power Prediction Using CatBoost with Domain Knowledge and Recursive Feature Elimination.

**DOI:** 10.3390/en18164272

**URL:** <https://www.mdpi.com/1996-1073/18/16/4272>

**Tujuan:** Prediksi daya PLT kombinasi siklus menggunakan CatBoost.

**Pendekatan:** Kuantitatif eksperimen.

**Metode/Algoritma:** CatBoost + RFE.

**Analisis/Optimasi:** RFE & rekayasa fitur berbasis domain.

**Evaluasi:** MAE/RMSE/R<sup>2</sup>.

**Hasil:** CatBoost dengan RFE meningkatkan performa dengan MAE 0,021 dan R<sup>2</sup> 0,997; menunjukkan efisiensi pemilihan fitur berbasis domain

---

**JUDUL:** Combination of Feature Selection and CatBoost for Prediction (Forests).

**DOI:** 10.3390/f12020216

**URL:** <https://www.mdpi.com/1999-4907/12/2/216>

**Tujuan:** Estimasi above-ground biomass (AGB).

**Pendekatan:** Kuantitatif komparatif.

**Metode/Algoritma:** CatBoost + beragam seleksi fitur (RFE, dll.).

**Analisis/Optimasi:** Perbandingan kombinasi seleksi fitur.

**Evaluasi:** RMSE/MAE/R<sup>2</sup>.

**Hasil:** Kombinasi CatBoost + Feature Selection menurunkan error hingga 30% dibanding model baseline tanpa seleksi fitur.

---

**JUDUL:** Enhanced Gold Ore Classification: A Comparative Analysis of Supervised Learning Methods (Geosciences).

**DOI:** 10.3390/geosciences15070248

**URL:** <https://www.mdpi.com/2076-3263/15/7/248>

**Tujuan:** Klasifikasi bijih emas/gangue.

**Pendekatan:** Kuantitatif komparatif.

**Metode/Algoritma:** Berbagai supervised ML (termasuk boosting).

**Analisis/Optimasi:** Seleksi variabel konfigurasi berbeda.

**Evaluasi:** Akurasi/F1.

**Hasil:** Boosting menghasilkan akurasi 95,3% dalam klasifikasi bijih emas, lebih unggul dibanding Decision Tree dan SVM.

---

**JUDUL:** Unbiased Employee Performance Evaluation Using Machine Learning.

**DOI:** 10.1016/j.joitmc.2024.100243

**URL:** <https://www.sciencedirect.com/science/article/pii/S2199853124000243>

**Tujuan:** Mengembangkan sistem penilaian kinerja karyawan yang bebas bias dengan pendekatan Machine Learning.

**Pendekatan:** Kuantitatif prediktif.

**Metode/Algoritma:** Logistic Regression, Naïve Bayes, Decision Tree, K-Nearest Neighbor, dan SVM.

**Analisis/Optimasi:** Normalisasi data dan pembobotan fitur untuk mengurangi bias.

**Evaluasi:** Accuracy, Precision, Recall, F1-score.

**Hasil:** Decision Tree dan SVM mencapai akurasi 94%, membuktikan efektivitas model berbasis ML untuk evaluasi kinerja tanpa bias.

---

**JUDUL:** Unveiling the Interplay of Financial and Moral Incentives in Elevating Employee Performance within Healthcare Organizations (Klang Valley, Malaysia)

**DOI:** [10.24294/jipd.v8i10.4586](https://doi.org/10.24294/jipd.v8i10.4586)

**URL:** <https://systems.enpress-publisher.com/index.php/jipd/article/view/4586>

**Tujuan:** Mengkaji pengaruh incentif finansial dan moral terhadap peningkatan kinerja pegawai sektor kesehatan di Malaysia.

**Pendekatan:** Kuantitatif analitik.

**Metode/Algoritma:** Structural Equation Modeling (SEM) menggunakan AMOS.

**Analisis/Optimasi:** Uji validitas dan reliabilitas, serta analisis jalur (path analysis).

**Evaluasi:** Koefisien regresi, p-value, dan goodness of fit index.

**Hasil:** Incentif moral (penghargaan dan partisipasi) memiliki pengaruh lebih kuat terhadap performa pegawai dibanding incentif finansial.

---

**JUDUL:** The Impact of Career Development on Employee Performance in the Civil Service Sector: A Nigerian Context.

**DOI:** [10.47297/wspchrmWSP2040-800504.20241501](https://doi.org/10.47297/wspchrmWSP2040-800504.20241501)

**URL:** <https://porcelainjournals.org/journals/wspchrm/20241501>

**Tujuan:** Menganalisis dampak pengembangan karier terhadap kinerja pegawai negeri di Nigeria.

**Pendekatan:** Mixed-method (kuantitatif dan kualitatif).

**Metode/Algoritma:** Survei (n=382) dan wawancara mendalam.

**Analisis/Optimasi:** Chi-square dan korelasi Pearson (SPSS).

**Evaluasi:** Signifikansi statistik antar variabel.

**Hasil:** Pengembangan karier berpengaruh signifikan terhadap kinerja pegawai ( $p < 0,05$ ); hasil mendukung integrasi variabel jabatan dan masa kerja dalam model prediksi retensi.

---

**JUDUL:** Predicting Employee Attrition: XAI-Powered Models for Managerial Decision Making.

**DOI:** 10.1016/j.dss.2024.114092

**URL:** <https://www.sciencedirect.com/science/article/pii/S016792362400092X>

**Tujuan:** Memprediksi potensi attrition (turnover) karyawan dan menyediakan interpretasi yang transparan bagi manajer SDM.

**Pendekatan:** Kuantitatif prediktif berbasis Explainable AI (XAI).

**Metode/Algoritma:** CatBoost, XGBoost, dan Random Forest dengan SHAP.

**Analisis/Optimasi:** Feature selection otomatis dan interpretasi model menggunakan SHAP.

**Evaluasi:** Accuracy, ROC-AUC, Precision, Recall.

**Hasil:** CatBoost + SHAP memberikan akurasi tertinggi (93%) dan interpretasi yang jelas terhadap faktor-faktor attrition.

---

**JUDUL:** Prediction of Three-Year All-Cause Mortality in Patients with Heart Failure and Atrial Fibrillation Using the CatBoost Model.

**DOI:** 10.1038/s41598-024-61152-2

**URL:** <https://www.nature.com/articles/s41598-024-61152-2>

**Tujuan:** Memprediksi risiko kematian tiga tahun pasien gagal jantung dan fibrilasi atrium.

**Pendekatan:** Kuantitatif klinis.

**Metode/Algoritma:** CatBoost Classifier.

**Analisis/Optimasi:** Recursive Feature Elimination (RFE) dan analisis interpretabilitas menggunakan SHAP.

**Evaluasi:** ROC-AUC, F1-score, Precision, Recall.

**Hasil:** CatBoost menghasilkan performa tertinggi (AUC 0.91) dibanding Logistic Regression dan Random Forest.

---

**JUDUL:** Predictive Analysis for Optimizing Port Operations.

**DOI:** 10.1016/j.oceaneng.2024.116129

**URL:** <https://www.sciencedirect.com/science/article/pii/S0029801824001292>

**Tujuan:** Mengoptimalkan operasi pelabuhan menggunakan analitik prediktif berbasis ML.

**Pendekatan:** Kuantitatif analitik.

**Metode/Algoritma:** Random Forest dan Gradient Boosting Regression.

**Analisis/Optimasi:** Feature Selection dan Grid Search CV.

**Evaluasi:** RMSE, MAE, R<sup>2</sup>.

**Hasil:** Gradient Boosting mengurangi waktu tunggu kapal sebesar 13,5% dan meningkatkan throughput.

---

**JUDUL:** TCLPI: Machine Learning Driven Framework for Hybrid Learning Mode Identification.

**DOI:** 10.1109/ACCESS.2024.3382912

**URL:** <https://ieeexplore.ieee.org/document/10338291>

**Tujuan:** Mengembangkan framework TCLPI untuk mengidentifikasi mode pembelajaran hybrid yang optimal.

**Pendekatan:** Kuantitatif eksperimen.

**Metode/Algoritma:** Decision Tree, Random Forest, Gradient Boosting, Logistic Regression.

**Analisis/Optimasi:** Confusion Matrix, stacking model, dan cross-validation.

**Evaluasi:** Accuracy, Precision, Recall, F1.

**Hasil:** Random Forest memiliki akurasi tertinggi (91%) dengan stabilitas model terbaik.

---

**JUDUL:** Exploring the Complex Associations Between Community Public Spaces and Healthy Aging: An Explainable Analysis Using CatBoost and SHAP.

**DOI:** [10.1186/s12889-025-23402-y](https://doi.org/10.1186/s12889-025-23402-y)

**URL:** <https://bmcpublichealth.biomedcentral.com/articles/10.1186/s12889-025-23402-y>

**Tujuan:** Menganalisis hubungan antara ruang publik dan kesehatan lansia menggunakan model interpretatif.

**Pendekatan:** Kuantitatif observasional berbasis Explainable AI.

**Metode/Algoritma:** CatBoost + SHAP.

**Analisis/Optimasi:** Interpretasi kontribusi fitur terhadap healthy aging.

**Evaluasi:** Accuracy dan ROC-AUC.

**Hasil:** Faktor lingkungan seperti aksesibilitas dan keamanan ruang publik paling memengaruhi healthy aging.

---

**JUDUL:** Machine Learning and SHAP Values Explain the Association Between Social Determinants of Health and Post-Stroke Depression.

**DOI:** 10.1186/s12888-024-05644-9

**URL:** <https://bmcpsyiatry.biomedcentral.com/articles/10.1186/s12888-024-05644-9>

**Tujuan:** Menjelaskan hubungan antara determinan sosial kesehatan dan depresi pasca-stroke.

**Pendekatan:** Kuantitatif observasional dengan XAI.

**Metode/Algoritma:** Gradient Boosting + SHAP.

**Analisis/Optimasi:** Feature selection dan SHAP value interpretation.

**Evaluasi:** ROC-AUC, Accuracy.

**Hasil:** Faktor sosial ekonomi, isolasi, dan akses kesehatan terbukti signifikan dalam model.

---

**JUDUL:** Elderly Travel Mode Choice in Thailand: Evaluating MNL and Machine Learning Models.

**DOI:** 10.3389/fbuil.2025.1601754

**URL:** <https://www.frontiersin.org/articles/10.3389/fbuil.2025.1601754/full>

**Tujuan:** Mengevaluasi perbandingan performa antara *Multinomial Logit Model* dan model Machine Learning untuk memilih moda transportasi lansia.

**Pendekatan:** Kuantitatif komparatif.

**Metode/Algoritma:** MNL, Random Forest, dan XGBoost.

**Analisis/Optimasi:** Feature importance dan validasi silang.

**Evaluasi:** Accuracy dan Log-loss.

**Hasil:** Model Machine Learning lebih unggul dari MNL dengan peningkatan akurasi prediksi sebesar 12%.

## 4. Hasil Literatur Terdahulu

Berdasarkan jurnal-jurnal pada tinjauan literatur terdahulu dapat disimpulkan bahwa :

- Multinomial Logistic Regression unggul dalam transparasi dan interpretasi, namun memiliki keterbatasan pada hubungannya dengan linear.
- CatBoost dan SHAP unggul dalam akurasi dan Explainability pada dataset besar dan kompleks.
- Studi terdahulu belum memiliki banyak kombinasi kedua pendekatan Multinomial Logistic Regression dan CatBoost Classifier secara komparatif dalam domain SDM.
- Penelitian yang saya lakukan mengisi kesenjangan tersebut dengan pendekatan komparatif berbasis Explainable AI, berfokus pada faktor kuantitatif (lama bekerja, gaji, jabatan) yang selama ini kurang mendapat perhatian eksplisit.

## 5. Analisis SWOT

### a. Multinomial Logistic Regression

**Strengths:** Interpretasi mudah; stabil untuk dataset kecil; mampu menangani klasifikasi multi-kelas sederhana.

**Weaknesses:** Kurang fleksibel untuk hubungan nonlinier; sensitif terhadap multikolinearitas.

**Opportunities:** Sebagai baseline untuk perbandingan model ML modern.

**Threats:** Akurasi lebih rendah dari algoritma ensemble seperti CatBoost.

### b. CatBoost Classifier

**Strengths:** Akurasi tinggi; menangani data kategorikal tanpa encoding; interpretasi visual melalui SHAP.

**Weaknesses:** Kompleksitas komputasi tinggi; interpretasi statistik lebih sulit.

**Opportunities:** Potensi tinggi untuk sistem HR berbasis Explainable AI.

**Threats:** Risiko overfitting dan ketergantungan pada tuning parameter.

## 6. GAP Analysis

Aspek	Temuan	Celah (Gap)	Kontribusi Penelitian Ini
<b>Model Statistik (MLR)</b>	Banyak digunakan, namun linier.	Tidak menjelaskan interaksi kompleks antar variabel.	Studi ini menguji perbandingan dengan model nonlinier.
<b>Machine Learning (CatBoost)</b>	Akurasi tinggi di berbagai domain.	Minim kajian di konteks SDM Indonesia.	Diterapkan untuk retensi karyawan berbasis data lokal.
<b>Explainable AI (SHAP)</b>	Populer di medis, sosial.	Jarang digunakan untuk kebijakan SDM.	Digunakan untuk menjelaskan faktor retensi.

<b>Faktor Retensi (lama kerja, gaji, jabatan)</b>	Jarang dikaji secara terintegrasi.	Belum ada model prediksi multifaktor.	Mengintegrasikan ketiga variabel dalam model prediktif.
---	------------------------------------	---------------------------------------	---

## **7. Kesimpulan**

Hasil kajian sistematik ini menegaskan bahwa pendekatan MLR dan CatBoost–SHAP dapat saling melengkapi: MLR memberikan transparansi dalam hubungan antar variabel, sementara CatBoost SHAP memberikan kekuatan prediktif dan interpretasi visual yang kaya. Penelitian Anda diharapkan menjadi jembatan antara model statistik klasik dan model machine learning modern yang terjelaskan, sekaligus berkontribusi pada strategi pengambilan keputusan mode bekerja (WFH, WFO, Hybrid) di lingkungan organisasi.