



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

SKRIPSI

JESSELYN MU

NIM. 2110511008

S1 INFORMATIKA

FAKULTAS ILMU KOMPUTER

UNIVERSITAS PEMBANGUNAN NASIONAL “VETERAN” JAKARTA

JAKARTA

2025



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

SKRIPSI

**SEBAGAI SALAH SATU SYARAT UNTUK MEMPEROLEH GELAR
SARJANA KOMPUTER**

JESSELYN MU

NIM. 2110511008

S1 INFORMATIKA

FAKULTAS ILMU KOMPUTER

UNIVERSITAS PEMBANGUNAN NASIONAL “VETERAN” JAKARTA

JAKARTA

2025

PERNYATAAN ORISINALITAS

Tugas Akhir ini adalah hasil karya sendiri dan semua sumber baik yang dikutip maupun yang dirujuk telah saya nyatakan dengan benar.

Nama : Jesselyn Mu

NIM : 2110511008

Tanggal : 27 Mei 2025

Bilamana di kemudian hari ditemukan ketidaksesuaian dengan pernyataan saya ini, maka saya bersedia dituntut dan diproses sesuai dengan ketentuan yang berlaku.

Jakarta, 27 Mei 2025

Yang Menyatakan



Jesselyn Mu

PERNYATAAN PERSETUJUAN PUBLIKASI TUGAS AKHIR UNTUK KEPENTINGAN AKADEMIS

Sebagai civitas akademika Universitas Pembangunan Nasional Veteran Jakarta, saya yang bertanda tangan di bawah ini:

Nama : Jesselyn Mu
NIM : 2110511008
Fakultas : Ilmu Komputer
Program Studi : S-1 Informatika

Demi pengembangan ilmu pengetahuan, menyetujui untuk memberikan kepada Universitas Pembangunan Nasional Veteran Jakarta Hak Bebas Royalti Non eksklusif (Non – exclusive Royalty Free Right) atas skripsi saya yang berjudul:

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

Dengan Hak Bebas Royalti ini Universitas Pembangunan Nasional Veteran Jakarta berhak menyimpan, mengalih media/memformatkan, mengelola dalam bentuk pangkalan data (basis data), merawat dan mempublikasikan skripsi saya selama tetap mencantumkan nama saya sebagai penulis dan sebagai pemilik Hak Cipta. Demikian pernyataan ini saya buat dengan sebenarnya.

Dibuat di: Jakarta

Pada tanggal: 27 Mei 2025

Yang Menyatakan



Jesselyn Mu

LEMBAR PENGESAHAN

Judul : Implementasi Model CatBoost dan xAI dalam Aplikasi Prediksi Retensi Karyawan Berbasis Streamlit
Nama : Jesselyn Mu
NIM : 2110511008
Program Studi : SI Informatika

Disetujui oleh :

Penguji 1:

Indra Permana Solihin, S.Kom, M.Kom.

Penguji 2:

I Wayan Ranga Pinastawa, M.Kom.

Pembimbing 1:

Neny Rosmawarni, M.Kom.

Pembimbing 2:

Kharisma Wiati Gusti, M.T.



Diketahui oleh:

Koordinator Program Studi:

Dr. Widya Cholil, M.I.T

NIP. 221112080

Dekan Fakultas Ilmu Komputer:

Prof. Dr. Ir. Supriyanto, M.Sc., IPM.

NIP. 197605082003121002



Tanggal Ujian Tugas Akhir:

15 April 2025

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 R^2 *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*

KATA PENGANTAR

Segala puji dan syukur peneliti panjatkan ke hadirat Tuhan Yang Maha Esa atas limpahan rahmat dan karunia-Nya sehingga penulis dapat menyelesaikan skripsi ini dengan judul “Implementasi Model *CatBoost* dan xAI dalam Aplikasi Prediksi Retensi Karyawan Berbasis *Streamlit*” sebagai salah satu syarat untuk menyelesaikan program Sarjana di Universitas Pembangunan Nasional “Veteran” Jakarta pada Fakultas Ilmu Komputer, Jurusan Informatika.

Skripsi ini disusun dengan tujuan untuk menguraikan latar belakang, permasalahan, tujuan, metodologi penelitian, hingga hasil penelitian yang sudah dilakukan. Dalam proses penyusunan skripsi ini, penulis menyadari bahwa banyak bantuan, bimbingan, serta dukungan dari berbagai pihak.

Penulis ingin menyampaikan penghargaan dan terima kasih yang sebesar-besarnya kepada:

1. Ayahanda dan Ibunda, Jiu Sen dan Martha Nindia Cristin, atas cinta, dukungan, dan kepercayaan yang tiada henti.
2. Bapak Prof. Dr. Ir. Supriyanto, ST., M.Sc., IPM selaku Dekan Fakultas Ilmu Komputer, Universitas Pembangunan Nasional “Veteran” Jakarta.
3. Ibu Dr. Widya Cholil, S.Kom., M.I.T selaku Kepala Program Studi S1-Informatika, Fakultas Ilmu Komputer, Universitas Pembangunan Nasional “Veteran” Jakarta.
4. Ibu Neny Rosmawarni, S.Kom., M.Kom dan Ibu Kharisma Wiati Gusti, M.T. selaku dosen pembimbing 1 dan 2 yang selalu membimbing, memberikan arahan, dan masukan kepada penulis dalam menyusun skripsi.
5. Bapak Bayu Hananto, S.Kom, M.Kom selaku dosen pembimbing akademik.
6. Tim Dosen Fakultas Ilmu Komputer UPNVJ yang sudah memberikan bimbingan serta ilmu yang berharga selama penulis menjadi mahasiswa.
7. Kakak dan Adik, Zefanya Priskilla dan Michael Vincent Mu, atas dukungan tenaga dan semangat yang tak henti-hentinya serta kesediaannya mendengarkan keluh kesah penulis.
8. Nenek dan Tante, Tjioe Boei Ngo dan Irlen Ceciliana, atas dukungan materi dan tenaga setiap hari, serta perhatiannya terhadap kebutuhan keluarga.
9. Sahabat dan teman-teman seperjuangan, atas dukungan, tukar pikiran, dan kebersamaan sepanjang masa perkuliahan.

Penulis menyadari bahwa skripsi ini masih jauh dari kesempurnaan. Oleh karena itu, saran dan kritik yang membangun sangat penulis harapkan demi perbaikan dan penyempurnaan penelitian ini di masa mendatang. Akhir kata, penulis berharap semoga proposal ini dapat memberikan manfaat dan kontribusi positif, baik bagi penulis sendiri maupun bagi pembaca dan pengembangan ilmu pengetahuan.

Jakarta, 15 April 2025



Jesselyn Mu

NIM. 2110511008

DAFTAR ISI

PERNYATAAN ORISINALITAS.....	i
PERNYATAAN PERSETUJUAN PUBLIKASI TUGAS AKHIR UNTUK KEPENTINGAN AKADEMIS.....	ii
LEMBAR PENGESAHAN	iii
ABSTRAK.....	iv
ABSTRACT.....	v
KATA PENGANTAR.....	vi
DAFTAR ISI	vii
DAFTAR TABEL.....	x
DAFTAR GAMBAR	xiii
DAFTAR RUMUS	xvi
DAFTAR SIMBOL	xvii
DAFTAR LAMPIRAN	xxi
BAB 1 PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Rumusah Masalah	3
1.3 Batasan Masalah.....	3
1.4 Tujuan dan Manfaat Penelitian	4
1.5 Sistematika Penulisan	5
BAB 2 TINJAUAN PUSTAKA.....	7
2.1 Kajian Teoritis	7
2.1.1 Retensi.....	7
2.1.2 Karyawan	7
2.1.3 Data Mining	7
2.1.4 Exploratory Data Analysis (EDA)	8
2.1.5 Pembersihan Data.....	8
2.1.5.1 Handling Missing Value	9
2.1.5.2 Handling Outliers Data.....	9
2.1.6 Feature Engineering	9
2.1.7 Data Imbalanced	10
2.1.7.1 SMOTE -NC	10
2.1.7.2 Random Under Sampling	10
2.1.8 Machine Learning	11
2.1.9 Ensemble Learning	11
2.1.9.1 Boosting	12
2.1.10 Gradient Boosting Decision Tree	13
2.1.11 CatBoost Algorithm.....	14
2.1.12 Hyperparameter Optuna	15
2.1.13 xAI (Explainable AI).....	17
2.1.13.1 SHAP (Shapley Additive Explanations)	17
2.1.14 Entity Relationship Diagram.....	18
2.1.15 Unified Modeling Language (UML).....	18
2.1.15.1 Use Case Diagram	18
2.1.15.2 Activity Diagram	19
2.1.16 Streamlit	19
2.1.17 Metrik Evaluasi Klasifikasi.....	19

2.1.17.1	<i>Area Under the Curve (AUC)</i>	19
2.1.17.2	<i>LogLoss</i>	20
2.1.17.3	<i>Confusion Matrix</i>	20
2.1.17.4	<i>Accuracy</i>	21
2.1.17.5	<i>Precision</i>	21
2.1.17.6	<i>Recall</i>	22
2.1.17.7	<i>F1-Score</i>	22
2.1.18	Metrik Evaluasi Regresi	22
2.1.18.1	MAE (<i>Mean Absolute Error</i>)	22
2.1.18.2	MAPE (<i>Mean Absolute Percentage Error</i>).....	23
2.1.18.3	RMSE (<i>Root Mean Squared Error</i>)	23
2.1.18.4	R^2 <i>Score</i>	23
2.1.19	<i>Cross Validation</i>	24
2.1.19.1	<i>StratifiedKFold</i>	24
2.1.19.2	<i>Time Series Cross Validation</i>	24
2.1.20	<i>Black Box Testing</i>	25
2.2	Penelitian Terdahulu.....	25
BAB 3 METODE PENELITIAN.....		31
3.1	Metode Penelitian.....	31
3.2	Rancangan Solusi / Metode yang Diusulkan	35
3.3	Teknik Pengumpulan Data (Eksperimen Desain)	51
3.4	Metode Analisis.....	53
3.4.1	Simulasi <i>Exploratory Data Analysis</i>	54
3.4.2	Simulasi <i>Preprocessing Data</i>	56
3.4.3	Simulasi <i>Feature Engineering</i>	57
3.4.4	Simulasi Perubahan Fitur Kategorikal Menjadi Numerik dengan <i>CatBoost</i>	61
3.4.5	Simulasi Prediksi Menggunakan <i>CatBoost Classification</i>	65
3.4.6	Simulasi Evaluasi Model <i>CatBoost Classification</i>	70
3.4.7	Simulasi Integrasi xAI dengan Model <i>CatBoost Classification</i>	72
3.4.8	Simulasi Pembuatan Model <i>CatBoost Regression</i>	76
3.4.9	Simulasi Evaluasi Model <i>CatBoost Regression</i>	80
3.5	Alat dan Bahan Penelitian.....	84
3.6	Jadwal Penelitian.....	85
BAB 4 HASIL DAN PEMBAHASAN.....		87
4.1	Analisis Hasil	87
4.1.1	Hasil Pengumpulan Data.....	87
4.1.2	Hasil <i>Exploratory Data Analysis (EDA)</i>	88
4.1.3	Hasil <i>Preprocessing Data</i>	97
4.1.3.1	Imputasi <i>Missing Value</i>	97
4.1.3.2	Menghapus <i>Outlier</i> pada Data Numerik	97
4.1.4	Hasil <i>Future Engineering</i>	98
4.1.4.1	Hasil Pembuatan fitur umur (<i>age_years</i>).....	99
4.1.4.2	Pembuatan fitur lama kerja (<i>active_work & active_work_months</i>)	99
4.1.4.3	Pembuatan fitur pendapatan dalam periode tertentu (<i>income_3_months & income_6_months</i>).....	99

4.1.4.4	Pembuatan fitur total pendapatan selama masa kerja (<i>total_income_work</i>)	100
4.1.4.5	Pembuatan fitur rasio pendapatan terhadap jumlah tanggungan (<i>income_dependant_ratio</i>)	100
4.1.4.6	Pembuatan fitur efisiensi kerja (<i>work_efficiency</i>).....	100
4.1.5	Hasil Pembuatan Model <i>Classification</i>	101
4.1.5.1	Hasil Pembuatan Model <i>CatBoost Classification</i> dengan data latih tanpa <i>sampling</i>	101
4.1.5.2	Hasil Pembuatan Model <i>CatBoost Classification</i> dengan data latih yang di-sampling menggunakan SMOTE-NC	104
4.1.5.3	Hasil Pembuatan Model <i>CatBoost Classification</i> dengan data latih yang di-sampling menggunakan <i>Random Under Sampling</i>	108
4.1.6	Evaluasi Model <i>Classification</i>	111
4.1.6.1	Hasil Evaluasi Model <i>CatBoost Classification</i> dengan data latih yang tidak di-sampling	112
4.1.6.2	Hasil Evaluasi Model <i>CatBoost Classification</i> dengan data latih yang di-sampling menggunakan SMOTE-NC	115
4.1.6.3	Hasil Evaluasi Model <i>CatBoost Classification</i> dengan data latih yang di-sampling menggunakan <i>Random Under Sampling</i>	118
4.1.6.4	Hasil pengujian pada data uji dan pemilihan model	121
4.1.7	Pembuatan Model <i>Regression</i>	124
4.1.7	Evaluasi Model <i>Regression</i>	128
4.1.7.1	Pengujian dan Pemilihan Model	131
4.1.8	Integrasi Aplikasi <i>Streamlit</i>	133
4.1.9.1	Hasil Perancangan Desain Aplikasi	133
4.1.9.2	Pengembangan Aplikasi <i>Streamlit</i>	143
4.1.9.3	Integrasi Model dan xAI dalam Aplikasi	150
4.1.9	Hasil Pengujian Aplikasi	153
4.1.9.1	Hasil Pengujian Aplikasi <i>Role</i> HR	153
4.1.9.2	Hasil Pengujian Aplikasi <i>Role</i> Karyawan	160
4.1.9.3	Hasil Pengujian Aplikasi <i>Role</i> Pimpinan	164
4.2	Hasil dan Rekomendasi.....	167
BAB 5	PENUTUP	169
5.1	Kesimpulan	169
5.2	Saran.....	169
DAFTAR	PUSTAKA	171
LAMPIRAN		179
DAFTAR	RIWAYAT HIDUP.....	192

DAFTAR TABEL

Tabel 2. 1 Tabel <i>Literature Review</i>	25
Tabel 2. 2 Keterangan <i>Confusion Matrix</i> (Sumber: (Nurhayati et al. 2024))	21
Tabel 3.1. Rancangan <i>Feature Engineering</i> yang Akan Dibuat.....	37
Tabel 3.2. Rancangan Model <i>CatBoost Classification</i>	40
Tabel 3.3. Rancangan Pencarian <i>Hyperparameter</i> Terbaik Menggunakan <i>Optuna</i>	41
Tabel 3.4. Rancangan <i>Hyperparameter</i> untuk <i>CatBoost Regression</i>	45
Tabel 3.5. Rancangan <i>Hyperparameter</i> Terbaik Menggunakan <i>Optuna</i>	46
Tabel 3.6 Distribusi dan Kriteria Data Penelitian	51
Tabel 3.7. Contoh Data Percobaan I.....	53
Tabel 3.8. Contoh Data Percobaan II	54
Tabel 3.9. Contoh Data Percobaan III.....	54
Tabel 3.10 Contoh Hasil Imputasi Data	57
Tabel 3.11. Contoh Hasil Pembuatan Fitur I.....	60
Tabel 3.12. Contoh Hasil Pembuatan Fitur II	60
Tabel 3.13. Contoh Hasil Pembuatan Fitur III	60
Tabel 3.14. Fitur-Fitur Kategorikal I.....	61
Tabel 3.15. Fitur-Fitur Kategorikal II.....	61
Tabel 3.16. Data Sesudah dan Sebelum Perubahan Kategorikal Menjadi Numerikal I	64
Tabel 3.17. Data Sesudah dan Sebelum Perubahan Kategorikal Menjadi Numerikal II.....	64
Tabel 3.18. Data Sesudah dan Sebelum Perubahan Kategorikal Menjadi Numerikal III.....	65
Tabel 3.19 Hasil Perhitungan Residu Tiap Observasi	65
Tabel 3.20. Hasil Perhitungan Residu Tiap Fitur I.....	66
Tabel 3.21. Hasil Perhitungan Residu Tiap Fitur II	67
Tabel 3.22. Hasil Perhitungan Residu Tiap Fitur III	67
Tabel 3.23. Hasil Perhitungan Residu Tiap Fitur IV	67
Tabel 3.24 Hasil Prediksi Klasifikasi Setiap Observasi untuk <i>Depth</i> = 1	68
Tabel 3.25. Hasil Prediksi Klasifikasi Setiap Observasi untuk <i>Depth</i> = 2	69
Tabel 3.26 Contoh Hasil Prediksi.....	69
Tabel 3.27 Contoh Hasil Prediksi untuk Evaluasi Klasifikasi	70
Tabel 3.28 <i>Confusion Matrix</i>	70
Tabel 3.29 Hasil Akhir Perhitungan Setiap <i>Subset</i>	72
Tabel 3.30. Contoh Hasil Akhir Perhitungan <i>Subset</i> Kontribusi Setiap Fitur I.....	73
Tabel 3.31. Contoh Hasil Akhir Perhitungan <i>Subset</i> Kontribusi Setiap Fitur II	74
Tabel 3.32. Contoh Hasil Akhir Perhitungan <i>Subset</i> Kontribusi Setiap Fitur III ...	74
Tabel 3.33. Contoh Hasil Akhir Perhitungan <i>Subset</i> Kontribusi Setiap Fitur IV ...	75
Tabel 3.34 Hasil Perhitungan Kontribusi Setiap Fitur	75
Tabel 3.35 Hasil Perhitungan Residu Untuk Setiap Observasi	76
Tabel 3.36. Contoh Hasil Perhitungan Residu Setiap Fitur untuk Regresi I.....	77
Tabel 3.37. Contoh Hasil Perhitungan Residu Setiap Fitur untuk Regresi II	77

Tabel 3.38. Contoh Hasil Perhitungan Residu Setiap Fitur untuk Regresi III	78
Tabel 3.39. Contoh Hasil Perhitungan Residu Setiap Fitur untuk Regresi III	78
Tabel 3.40 Contoh Hasil Final Prediksi dengan <i>Depth</i> = 1	79
Tabel 3.41. Contoh Hasil Prediksi dengan <i>Depth</i> = 2	80
Tabel 3.42. Contoh Hasil Prediksi dengan <i>Depth</i> = 6	80
Tabel 3.43 Contoh Hasil Prediksi untuk Evaluasi Regresi	80
Tabel 3.44 Hasil Perhitungan Mencari Nilai <i>Error</i>	81
Tabel 3.45 Hasil Konversi Nilai Mutlak	82
Tabel 3.46 Hasil Konversi Setiap Observasi	82
Tabel 3.47. Hasil Perhitungan <i>Error</i> Kuadrat	83
Tabel 3.48. Jadwal Penelitian	85
Tabel 4.1. Hasil Pelatihan Model Klasifikasi Tanpa <i>Optuna</i>	101
Tabel 4.2. Hasil Pencarian <i>Hyperparameter</i> Terbaik untuk Model Klasifikasi Tanpa Data <i>Sampling</i>	102
Tabel 4.3. Hasil Pelatihan Model Klasifikasi Menggunakan <i>Optuna</i>	103
Tabel 4.4. Hasil Pelatihan Model Klasifikasi dengan SMOTE-NC dan Tanpa Menggunakan <i>Optuna</i>	105
Tabel 4.5. Hasil Pencarian <i>Hyperparameter</i> Terbaik untuk Model Klasifikasi dengan SMOTE-NC	106
Tabel 4.6. Hasil Pelatihan Model Klasifikasi dengan SMOTE-NC dengan Menggunakan <i>Optuna</i>	107
Tabel 4.7. Hasil Pelatihan Model Klasifikasi dengan <i>Random Under Sampling</i> Tanpa Menggunakan <i>Optuna</i>	108
Tabel 4.8. Hasil Pencarian <i>Hyperparameter</i> Terbaik untuk Model Klasifikasi dengan <i>Random Under Sampling</i>	109
Tabel 4.9. Hasil Pelatihan Model Klasifikasi dengan <i>Random Under Sampling</i> dan <i>Optuna</i>	110
Tabel 4.10. Hasil Akurasi <i>Cross Validation</i> untuk Model Klasifikasi Tanpa <i>Optuna</i>	113
Tabel 4.11. Hasil Akurasi <i>Cross Validation</i> untuk Model Klasifikasi Dengan <i>Optuna</i>	114
Tabel 4.12. Hasil Akurasi <i>Cross Validation</i> untuk Model Klasifikasi dengan SMOTE-NC	116
Tabel 4.13. Hasil Akurasi <i>Cross Validation</i> untuk Model Klasifikasi dengan SMOTE-NC dan <i>Optuna</i>	118
Tabel 4.14. Hasil Akurasi <i>Cross Validation</i> untuk Model Klasifikasi dengan <i>Random Under Sampling</i>	119
Tabel 4.15. Hasil Akurasi <i>Cross Validation</i> untuk Model Klasifikasi Menggunakan <i>Random Under Sampling</i> dan <i>Optuna</i>	121
Tabel 4.16. Hasil Pengujian pada Data Uji Menggunakan Model Klasifikasi Tanpa Data <i>Sampling</i>	121
Tabel 4.17. Hasil Pengujian pada Data Uji Menggunakan Model Klasifikasi dengan Data SMOTE-NC	122
Tabel 4.18. Hasil Pengujian pada Data Uji Menggunakan Model Klasifikasi dengan Data <i>Random Under Sampling</i>	122

Tabel 4.19. Ringkasan Keseluruhan Hasil untuk Model Klasifikasi	123
Tabel 4.20. Hasil Pelatihan Model Regresi Tanpa <i>Optuna</i>	125
Tabel 4.21. Hasil Pencarian <i>Hyperparameter</i> untuk Model Regresi	126
Tabel 4.22. Hasil Pelatihan Model Regresi dengan Menggunakan <i>Optuna</i>	127
Tabel 4.23. Hasil R^2 <i>Score Cross Validation</i> untuk Model Tanpa <i>Optuna</i>	129
Tabel 4.24. Hasil R^2 <i>Score Cross Validation</i> untuk Model Regresi dengan <i>Optuna</i>	131
Tabel 4.25. Hasil Pengujian pada Data Uji Menggunakan Model <i>CatBoost</i> Regresi.....	131
Tabel 4.26. Hasil Evaluasi Keseluruhan Model Regresi.....	132
Tabel 4.27. Hasil Pengujian <i>Form Login</i> untuk <i>Role</i> HR	153
Tabel 4.28. Hasil Pengujian Halaman Prediksi untuk <i>Role</i> HR.....	154
Tabel 4.29. Hasil Pengujian Halaman <i>Dashboard</i> untuk <i>Role</i> HR	155
Tabel 4.30. Hasil Pengujian Halaman Laporan untuk <i>Role</i> HR.....	156
Tabel 4.31. Hasil Pertanyaan Tambahan untuk <i>Role</i> HR	160
Tabel 4.32. Hasil Pengujian <i>Form Login</i> untuk <i>Role</i> Karyawan	161
Tabel 4.33. Hasil Pengujian <i>Form Kepuasan Kerja</i> untuk <i>Role</i> Karyawan	162
Tabel 4.34. Hasil Pengujian <i>Form Keluhan</i> untuk <i>Role</i> Karyawan	163
Tabel 4.35. Hasil Pengujian <i>Form Login</i> untuk <i>Role</i> Pimpinan	164
Tabel 4.36. Hasil Pengujian Halaman <i>Form</i> Penilaian untuk <i>Role</i> Pimpinan	165
Tabel 4.37. Hasil Pengujian Halaman <i>Dashboard</i> untuk <i>Role</i> Pimpinan	166

DAFTAR GAMBAR

Gambar 2. 1 Cara Kerja <i>Boosting</i> (Sumber: (Maretva Cendani dan Wibowo 2022))	12
Gambar 2. 2 Langkah Kerja <i>Optuna</i> (Sumber: (Zhuo 2024)).....	16
Gambar 2. 3 Ilustrasi <i>Confusion Matrix</i> (Sumber: (Nurhayati et al. 2024))	20
Gambar 3.1 Kerangka Penelitian	31
Gambar 3.2. Rancangan Solusi <i>Exploratory Data Analysis</i> (EDA)	36
Gambar 3.3. Rancangan Solusi Tahapan <i>Preprocessing</i> Data	37
Gambar 3.4 Rancangan Solusi untuk Model <i>CatBoost</i> Classification.....	39
Gambar 3.5 Rancangan Solusi untuk Model <i>CatBoost</i> Regression	44
Gambar 3.6. Rancangan Solusi Pembuatan Aplikasi	49
Gambar 3.7. <i>Sitemap</i> Rancangan Aplikasi	50
Gambar 4.1. Hasil Pengumpulan Data untuk Data Pelatihan	87
Gambar 4.2. Hasil Pengumpulan Data untuk Data Pengujian	88
Gambar 4.3. Hasil Deskripsi Statistik Dasar.....	88
Gambar 4.4. Hasil Pengecekan Outlier	89
Gambar 4.5. Hasil Pengecekan Data Kosong	89
Gambar 4.6. Distribusi Status Churn Berdasarkan Status Pernikahan.....	90
Gambar 4.7. Rata-Rata Jumlah Tanggungan Berdasarkan Status Churn	91
Gambar 4.8. Distribusi Status Churn Berdasarkan Pendidikan	91
Gambar 4.9. Distribusi Status Churn Berdasarkan Domisili	92
Gambar 4.10. Distribusi Status Churn Berdasarkan Departemen.....	92
Gambar 4.11. Distribusi Status Churn Berdasarkan Posisi	93
Gambar 4.12. Distribusi Kepuasan Kerja Berdasarkan Status Churn.....	94
Gambar 4.13. Sebaran Rating Kinerja Berdasarkan Status Churn	94
Gambar 4.14. Sebaran Ketidakhadiran Berdasarkan Status Churn.....	95
Gambar 4.15. Sebaran Pendapatan Berdasarkan Status Churn.....	96
Gambar 4.16. Jumlah Target	96
Gambar 4.17. Hasil Pengisian Nilai Kosong	97
Gambar 4.18. Hasil Penghapusan Data Outlier	97
Gambar 4.19. Data dengan <i>active_work_months</i> di bawah 3 Bulan	98
Gambar 4.20. Hasil Filtering Data	98
Gambar 4.21. Hasil Pembuatan Fitur Umur.....	99
Gambar 4.22. Hasil Pembuatan Fitur Lama Kerja.....	99
Gambar 4.23. Hasil Pembuatan Fitur Pendapatan dalam Periode Tertentu	99
Gambar 4.24. Hasil Pembuatan Fitur Total Pendapatan Selama Masa Kerja	100
Gambar 4.25. Hasil Pembuatan Fitur Rasio Pendapatan Terhadap Jumlah Tanggungan	100
Gambar 4.26. Hasil Pembuatan Fitur Efisiensi Kerja	100
Gambar 4.27. Grafik Learning Curve Pelatihan Model Klasifikasi Tanpa Optuna	102
Gambar 4.28. Grafik Learning Curve Pelatihan Model Klasifikasi Dengan Optuna	104

Gambar 4.29. Grafik Learning Curve Pelatihan Model Klasifikasi Dengan SMOTE-NC dan tanpa menggunakan Optuna.....	105
Gambar 4.30. Grafik Learning Curve Pelatihan Model Klasifikasi Dengan SMOTE-NC dan Menggunakan Optuna.....	107
Gambar 4. 31. Grafik Learning Curve Pelatihan Model Klasifikasi Dengan Random Under Sampling Tanpa Menggunakan Optuna.....	109
Gambar 4.32. Grafik Learning Curve Pelatihan Model Klasifikasi Dengan Random Under Sampling dan Optuna	111
Gambar 4.33. Confusion Matrix untuk Data Latih Menggunakan Model Klasifikasi Tanpa Data Sampling	112
Gambar 4.34. Confusion Matrik untuk Data Validasi Menggunakan Model Klasifikasi Tanpa Data Sampling	112
Gambar 4.35. Grafik Cross Validation untuk Model Klasifikasi Tanpa Optuna..	112
Gambar 4.36. Confusion Matrixx untuk Data Latih Menggunakan Model Klasifikasi Tanpa Sampling dan Optuna.....	114
Gambar 4.37. Confusion Matrixx untuk Data Validasi Menggunakan Model Klasifikasi Tanpa Sampling dan Optuna.....	114
Gambar 4.38. Grafik Cross Validation untuk Model Klasifikasi Dengan Optuna	114
Gambar 4.39. Confusion Matrix untuk Data Latih Menggunakan Model Klasifikasi dengan SMOTE-NC	115
Gambar 4.40. Confusion Matrix untuk Data Validasi Menggunakan Model Klasifikasi dengan SMOTE-NC	115
Gambar 4.41 Grafik Cross Validation untuk Model Klasifikasi dengan SMOTE-NC	116
Gambar 4.42. Confusion Matrix untuk Data Latih Menggunakan Model Klasifikasi dengan SMOTE-NC dan Optuna.....	117
Gambar 4.43. Confusion Matrix untuk Data Validasi Menggunakan Model Klasifikasi dengan SMOTE-NC dan Optuna.....	117
Gambar 4.44. Grafik Cross Validation untuk Model Klasifikasi dengan SMOTE-NC dan Optuna.....	117
Gambar 4.45. Confusion Matrix untuk Data Latih Menggunakan Model Klasifikasi dengan Random Under Sampling.....	118
Gambar 4.46. Confusion Matrix untuk Data Validasi Menggunakan Model Klasifikasi dengan Random Under Sampling.....	118
Gambar 4.47. Grafik Cross Validation untuk Model Klasifikasi dengan Random Under Sampling	119
Gambar 4.48. Confusion Matrix untuk Data Latih Menggunakan Model Klasifikasi dengan Random Under Sampling dan Optuna	120
Gambar 4.49. Confusion Matrix untuk Data Latih Menggunakan Model Klasifikasi dengan Random Under Sampling dan Optuna	120
Gambar 4.50. Grafik Cross Validation untuk Model Klasifikasi Menggunakan Random Under Sampling dan Optuna	120
Gambar 4.51. Grafik Learning Curve pada Model Regresi Tanpa Optuna.....	125

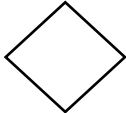
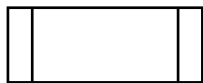
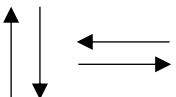
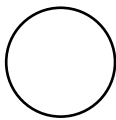
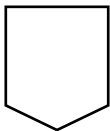
Gambar 4.52. Grafik Learning Curve pada Model Regresi dengan Menggunakan Optuna	127
Gambar 4.53. Hasil Prediksi pada 20 Indeks Pertama Menggunakan Model Regresi Tanpa Optuna	128
Gambar 4.54. Grafik Cross Validation untuk Model Regresi Tanpa Optuna.....	129
Gambar 4.55. Hasil Prediksi pada 20 Indeks Pertama Menggunakan Model Regresi dengan Optuna	130
Gambar 4.56. Grafik Cross Validation untuk Model Regresi dengan Optuna	130
Gambar 4.57. Use Case Diagram Aplikasi TALENTRA.....	133
Gambar 4.58. Activity Diagram Masuk Aplikasi.....	134
Gambar 4.59. Activity Diagram Melakukan Prediksi.....	135
Gambar 4.60. Activity Diagram Melihat Data & Tren Karyawan	135
Gambar 4.61. Activity Diagram Melihat Data Histori Prediksi & SHAP Value..	136
Gambar 4.62. Activity Diagram Mengunduh Data Histori Prediksi & SHAP Value	137
Gambar 4.63. Activity Diagram Mengisi Form Kepuasan Kerja.....	138
Gambar 4.64. Activity Diagram Mengisi Form Komplain	139
Gambar 4.65. Activity Diagram Mengisi Form Kinerja Karyawan.....	140
Gambar 4.66. Activity Diagram Melihat Histori Prediksi atau Data Karyawan..	140
Gambar 4.67. Activity Diagram Mengunduh Data	141
Gambar 4.68. Rancangan Database Aplikasi TALENTRA	142
Gambar 4.69. Halaman Utama Aplikasi TALENTRA.....	143
Gambar 4.70. Halaman Login Aplikasi TALENTRA.....	144
Gambar 4.71. Halaman Prediksi Aplikasi TALENTRA	145
Gambar 4.72. Halaman Dashboard Role HR Aplikasi TALENTRA	146
Gambar 4.73. Halaman Laporan Aplikasi TALENTRA	147
Gambar 4.74. Halaman Form Penilaian Kinerja Karyawan	147
Gambar 4.75. Halaman Dashboard Role Pimpinan Aplikasi TALENTRA	148
Gambar 4.76. Halaman Form Kepuasan Kerja Aplikasi TALENTRA	149
Gambar 4.77. Halaman Form Keluhan Aplikasi TALENTRA	149
Gambar 4.78. Tahapan Integrasi Model dan xAI pada Aplikasi	150
Gambar 4.79. Tampilan untuk Hasil Prediksi Retensi	151
Gambar 4.80. Tampilan untuk Hasil Prediksi Tidak Retensi	151
Gambar 4.81. Hasil SHAP Diagram	151
Gambar 4.82. Tampilan untuk Hasil Prediksi Retensi dan Plot xAI.....	152
Gambar 4.83. Tampilan untuk Hasil Prediksi Tidak Retensi dan xAI	153

DAFTAR RUMUS

Rumus 2.1	14
Rumus 2.2	14
Rumus 2.3	17
Rumus 2.4	20
Rumus 2.5	21
Rumus 2.6	22
Rumus 2.7	22
Rumus 2.8	22
Rumus 2.9	22
Rumus 2.10	23
Rumus 2.11	23
Rumus 2.12	23
Rumus 2.13	25
Rumus 3.1	55
Rumus 3.2	55
Rumus 3.3	55
Rumus 3.4	56
Rumus 3.5	56
Rumus 3.6	65
Rumus 3.7	68
Rumus 3.8	81
Rumus 4.1	152

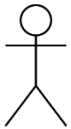
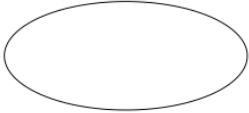
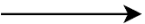
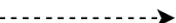
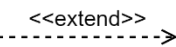
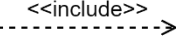
DAFTAR SIMBOL

Simbol 1. Simbol-Simbol *Flowchart*

No.	Simbol	Nama Simbol	Fungsi
1.		Terminator	Untuk menyatakan awal atau akhir dalam suatu proses.
2.		<i>Input/Output</i>	Untuk mewakili proses menerima <i>input</i> atau menghasilkan <i>output</i> .
3.		<i>Process</i>	Untuk menunjukkan langkah atau operasi yang dilakukan oleh sistem atau komputer untuk menjalankan suatu proses.
4.		<i>Decision</i>	Untuk menggambarkan titik pengambilan keputusan, di mana proses dapat bercabang berdasarkan jawaban "ya" atau "tidak."
5.		<i>Predefine Process</i>	Untuk menggambarkan bagian dari proses yang lebih rinci atau sub-prosedur yang sudah ditetapkan.
6.		<i>Flow</i>	Untuk menghubungkan atau mengarahkan antar simbol dalam sebuah diagram.
7.		<i>On-Page Reference</i>	Untuk menunjukkan proses masuk atau keluar di dalam satu lembar kerja yang sama.
8.		<i>Off-Page Reference</i>	Untuk menghubungkan proses yang terjadi di antara lembar kerja yang berbeda dan menunjukkan penyambungan proses yang berlanjut di halaman lain.
9.		<i>Manual Operation</i>	Untuk menandakan adanya proses yang dilakukan secara manual, tidak otomatis oleh komputer.

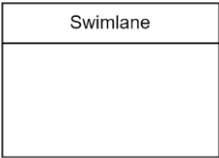
No.	Simbol	Nama Simbol	Fungsi
10.		<i>Document</i>	Untuk menunjukkan bahwa proses input atau output terkait dengan dokumen.
11.		<i>Preparation</i>	Untuk menunjukkan tempat penyimpanan atau persiapan data sebelum digunakan dalam pengolahan lebih lanjut.

Simbol 2. Simbol-Simbol *Use Case Diagram*

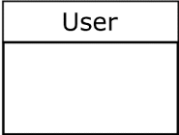
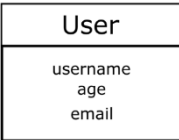
No	Simbol	Nama Simbol	Arti
1.		Aktor	Untuk menggambarkan peran seorang, sistem, atau perangkat yang berinteraksi dengan <i>use case</i> .
2.		<i>Use Case</i>	Untuk menunjukkan fungsi atau layanan yang diberikan oleh sistem kepada aktor.
3.		<i>Association</i>	Untuk menjelaskan bahwa entitas dapat memiliki hubungan atau tidak memiliki hubungan sama sekali dengan entitas lain.
4.		Generalisasi	Untuk merepresentasikan hubungan pewarisan antara aktor atau antar <i>use case</i> .
5.		<i>Extend</i>	Untuk menambahkan fungsionalitas tambahan pada <i>use case</i> lain dengan syarat tertentu.
6.		<i>Include</i>	Untuk melengkapi <i>use case</i> dengan fungsionalitas tambahan yang berasal dari <i>use case</i> lain.

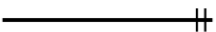
Simbol 3. Simbol-Simbol *Activity Diagram*

No.	Simbol	Nama Simbol	Arti
1.		Status Awal	Untuk menunjukkan titik awal dalam sebuah diagram aktivitas.

No.	Simbol	Nama Simbol	Arti
2.		Aktivitas	Untuk menggambarkan suatu aktivitas sistem yang dimulai dengan kata kerja.
3.		Percabangan	Untuk merepresentasikan pilihan cabang dalam suatu aktivitas yang lebih dari satu.
4.		Penggabungan	Untuk menggambarkan penggabungan dua atau lebih aktivitas menjadi satu.
5.		Status Akhir	Untuk menunjukkan titik akhir dalam suatu diagram aktivitas.
6.		<i>Swimlane</i>	Untuk menggambarkan pemisahan suatu entitas yang bertanggung jawab terhadap aktivitas yang terjadi.

Simbol 4. Simbol-Simbol *Entity Relationship Diagram* (ERD)

No	Simbol	Nama Simbol	Arti
1.		Entitas	Untuk merepresentasikan suatu objek di dalam sebuah <i>database</i> .
2.		Atribut	Untuk menggambarkan isi yang ada di dalam sebuah entitas.
3.		Nol	Untuk menunjukkan bahwa entitas dapat memiliki hubungan atau tidak memiliki hubungan dengan entitas lain.
4.		Satu	Untuk menyatakan bahwa satu entitas hanya bisa terhubung dengan satu entitas lainnya.
5.		<i>Crow's Foot</i>	Untuk menjelaskan bahwa satu entitas dapat berhubungan dengan banyak entitas lain.

No	Simbol	Nama Simbol	Arti
6.		Nol + <i>Crow's Foot</i>	Untuk menunjukkan bahwa suatu entitas bisa tidak memiliki hubungan atau memiliki banyak hubungan dengan entitas lain.
7.		Satu + <i>Crow's Foot</i>	Untuk menegaskan bahwa entitas harus memiliki setidaknya satu hubungan dengan entitas lain, tetapi bisa memiliki lebih dari satu.
8.		Satu + Satu	Untuk menyatakan bahwa setiap instansi dari satu entitas hanya dapat berhubungan dengan satu instansi dari entitas lainnya.

DAFTAR LAMPIRAN

Lampiran 1. Surat Perizinan Riset	179
Lampiran 2. Surat Balasan dari PT. XYZ	180
Lampiran 3. Bukti Pertanyaan Untuk Fitur <i>job_satisfaction</i> dan <i>performance_rating</i>	182
Lampiran 4. Bukti Pengujian <i>BlackBox Testing</i>	183
Lampiran 5. Sumber-Sumber <i>Feature Engineering</i>	184
Lampiran 6. Hasil Pembagian Data Latih Menggunakan SMOTE-NC.....	185
Lampiran 7. Hasil Pembagian Data Latih Menggunakan <i>Random Under Sampling</i>	188
Lampiran 8. Keterangan Fitur-Fitur Tambahan dalam xAI	189
Lampiran 9. Surat Pernyataan Validasi	190
Lampiran 10. Keterangan Hasil Pengecekan Plagiasi (<i>TurnItIn</i>).....	191