



SKRIPSI

PREDIKSI INDEKS EKONOMI HIJAU DI PROVINSI JAWA TIMUR MENGGUNAKAN STACKING ENSEMBLE: XGBOOST, LIGHTGBM, DAN CATBOOST

ANDINI FITRIYAH SALSABILAH

NPM 21081010217

DOSEN PEMBIMBING

Dr. Basuki Rahmat, S.Si., MT.

Eva Yulia Puspaningrum, S.Kom., M.Kom.

**KEMENTERIAN PENDIDIKAN TINGGI, SAINS, DAN TEKNOLOGI
UNIVERSITAS PEMBANGUNAN NASIONAL VETERAN JAWA TIMUR
FAKULTAS ILMU KOMPUTER
PROGRAM STUDI INFORMATIKA
SURABAYA
2025**



SKRIPSI

PREDIKSI INDEKS EKONOMI HIJAU DI PROVINSI JAWA TIMUR MENGGUNAKAN STACKING ENSEMBLE: XGBOOST, LIGHTGBM DAN CATBOOT

ANDINI FITRIYAH SALSABILAH

NPM 21081010217

DOSEN PEMBIMBING

Dr. Basuki Rahmat, S.Si., MT.

Eva Yulia Puspaningrum, S.Kom., M.Kom.

**KEMENTERIAN PENDIDIKAN TINGGI, SAINS, DAN TEKNOLOGI
UNIVERSITAS PEMBANGUNAN NASIONAL VETERAN JAWA TIMUR
FAKULTAS ILMU KOMPUTER
PROGRAM STUDI INFORMATIKA
SURABAYA
2025**

Halaman ini sengaja dikosongkan

LEMBAR PENGESAHAN

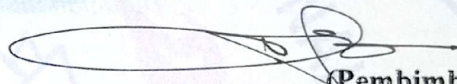
PREDIKSI INDEKS EKONOMI HIJAU DI PROVINSI JAWA TIMUR MENGUNAKAN STACKING ENSEMBLE: XGBOOST, LIGHTGBM, DAN CATBOOST

Oleh :
ANDINI FITRIYAH SALSABILAH
NPM. 21081010217

Telah dipertahankan di hadapan dan diterima oleh Tim Penguji Skripsi Prodi Informatika Fakultas Ilmu Komputer Universitas Pembangunan Nasional Veteran Jawa Timur Pada tanggal 4 Juni 2025

Menyetujui

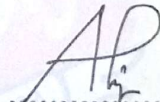
Dr. Basuki Rahmat, S.Si., MT.
NIP. 19690723 2021211 002


..... (Pembimbing I)

Eva Yulia Puspaningrum, S.Kom., M.Kom.
NIP. 19890705 202121 2 002


..... (Pembimbing II)

Dr. Eng. Ir. Anggraini Puspita Sari, ST., MT.
NPT. 222198 60 816400


..... (Ketua Penguji)

Fawwaz Ali Akbar, S.Kom., M.Kom.
NIP. 19920317 201803 1 002


..... (Anggota Penguji)

Mengetahui,
Dekan Fakultas Ilmu Komputer


Prof. Dr. Ir. Novirina Hendrasarie, MT
NIP. 19681126 199403 2 001

Halaman ini sengaja dikosongkan

LEMBAR PERSETUJUAN

**PREDIKSI INDEKS EKONOMI HIJAU DI PROVINSI JAWA TIMUR
MENGUNAKAN STACKING ENSEMBLE: XGBOOST, LIGHTGBM,
DAN CATBOOST**

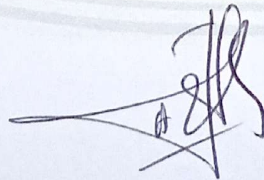
Oleh :

ANDINI FITRIYAH SALSABILAH

NPM. 21081010217

Menyetujui,

**Koordinator Program Studi Informatika
Fakultas Ilmu Komputer**



Fetty Tri Anggraeny, S.Kom., M.Kom.

NIP. 19820211 2021212 005

Halaman ini sengaja dikosongkan

SURAT PERNYATAAN BEBAS PLAGIASI

Saya yang bertanda tangan di bawah ini :

Nama : Andini Fitriyah Salsabilah
NPM : 21081010217
Program : Sarjana (S1)
Program Studi : Informatika
Fakultas : Ilmu Komputer

Menyatakan bahwa dalam dokumen ilmiah Skripsi ini tidak terdapat bagian dari karya ilmiah lain yang telah diajukan untuk memperoleh gelar akademik di suatu lembaga Pendidikan Tinggi, dan juga tidak terdapat karya atau pendapat yang pernah ditulis atau diterbitkan oleh orang/lembaga lain, kecuali yang secara tertulis disitasi dalam dokumen ini dan disebutkan secara lengkap dalam daftar pustaka.

Dan saya yang menyatakan bahwa dokumen ilmiah ini bebas dari unsur-unsur plagiasi. Apabila dikemudian hari ditemukan indikasi plagiat pada Skripsi ini, saya bersedia menerima sanksi sesuai dengan peraturan perundang-undangan yang berlaku.

Demikian surat pernyataan ini saya buat dengan sesungguhnya tanpa ada paksaan dari siapapun juga dan untuk dipergunakan sebagaimana mestinya.



Surabaya, 19 Juni 2025

Yang Membuat Pernyataan



Andini Fitriyah Salsabilah

NPM. 21081010217

Halaman ini sengaja dikosongkan

ABSTRAK

Nama Mahasiswa / NPM : Andini Fitriyah Salsabilah / 21081010217
Judul Skripsi : Prediksi Indeks Ekonomi Hijau di Provinsi Jawa Timur Menggunakan Stacking Ensemble: XGBoost, LightGBM, dan CatBoost
Dosen Pembimbing : 1. Dr. Basuki Rahmat, S.Si., MT.
2. Eva Yulia Puspaningrum, S.Kom., M.Kom.

Indonesia tengah menghadapi tantangan dalam mencapai pembangunan berkelanjutan, khususnya dalam menyeimbangkan pertumbuhan ekonomi dengan kelestarian lingkungan. Ketimpangan antara aspek ekonomi, sosial, dan lingkungan menuntut hadirnya alat ukur yang mampu menilai pencapaian ekonomi hijau secara menyeluruh. *Green Economy Index* (GEI), yang diperkenalkan oleh Kementerian PPN/Bappenas, menjadi instrumen penting dalam mengukur kinerja ekonomi hijau. Namun, keterbatasan data di tingkat provinsi, seperti di Jawa Timur, menyulitkan proses evaluasi dan perumusan kebijakan daerah yang tepat. Penelitian ini mengusulkan pendekatan prediksi GEI berbasis *machine learning* menggunakan algoritma stacking ensemble yang menggabungkan tiga model andal: XGBoost, LightGBM, dan CatBoost. Model dibangun menggunakan data indikator ekonomi, sosial, dan lingkungan, serta diuji melalui *holdout* set untuk mengukur akurasi dan kemampuan generalisasi. Hasil menunjukkan bahwa stacking ensemble menghasilkan performa terbaik dengan RMSE 0.0302, MAE 0.0220, dan R^2 0.9767, mengungguli model tunggal lainnya. Sebagai pembandingan, CatBoost mencatat RMSE 0.0317, MAE 0.0229, dan R^2 0.9743. LightGBM mencatat RMSE 0.0348, MAE 0.0260, dan R^2 0.9691. Sementara XGBoost mencatat RMSE 0.0413, MAE 0.0315, dan R^2 0.9564. Hasil ini membuktikan bahwa stacking ensemble mampu mengungguli model tunggal dalam memprediksi GEI secara akurat. Dengan demikian, pendekatan ini dapat digunakan sebagai alat bantu dalam mendukung pengambilan keputusan berbasis data untuk pembangunan ekonomi hijau yang berkelanjutan di tingkat regional.

Kata kunci : CatBoost, Indeks Ekonomi Hijau, LightGBM, Regresi, Stacking Ensemble, XGBoost

Halaman ini sengaja dikosongkan

ABSTRACT

Student Name / NPM : Andini Fitriyah Salsabilah / 21081010217
Thesis Title : Prediction of the Green Economy Index in East Java Province Using Stacking Ensemble: XGBoost, LightGBM, and CatBoost
Advisor : 1. Dr. Basuki Rahmat, S.Si., MT.
2. Eva Yulia Puspaningrum, S.Kom., M.Kom.

ABSTRACT

Indonesia is currently facing significant challenges in achieving sustainable development, particularly in balancing economic growth with environmental sustainability. Disparities among economic, social, and environmental aspects necessitate the presence of a comprehensive measurement tool to evaluate green economy performance. The Green Economy Index (GEI), introduced by the Ministry of National Development Planning (Bappenas), serves as a crucial instrument in assessing green economic achievements. However, data limitations at the provincial level, such as in East Java, hinder effective evaluation and policy formulation. This study proposes a machine learning-based approach to predicting GEI using a stacking ensemble algorithm that integrates three robust base models: XGBoost, LightGBM, and CatBoost. The model was developed using economic, social, and environmental indicators and evaluated using a holdout set to assess its accuracy and generalization capability. The results demonstrate that the stacking ensemble achieved the best performance with an RMSE of 0.0302, MAE of 0.0220, and R^2 of 0.9767, outperforming all individual models. For comparison, CatBoost recorded RMSE of 0.0317, MAE of 0.0229, and R^2 of 0.9743. LightGBM achieved RMSE of 0.0348, MAE of 0.0260, and R^2 of 0.9691. While XGBoost showed RMSE of 0.0413, MAE of 0.0315, and R^2 of 0.9564. These findings confirm that the stacking ensemble approach surpasses single models in accurately predicting the GEI. Thus, this approach can be a valuable data-driven decision support tool for advancing sustainable green economic development at the regional level.

Keywords: CatBoost, Green Economy Index, LightGBM, Regression, Stacking Ensemble, XGBoost

Halaman ini sengaja dikosongkan

KATA PENGANTAR

Puji syukur kehadirat Allah SWT atas segala rahmat, hidayah dan karunia-Nya kepada penulis sehingga skripsi dengan judul **“Prediksi Indeks Ekonomi Hijau di Provinsi Jawa Timur Menggunakan Stacking Ensemble: XGBoost, LightGBM, dan CatBoost”** dapat terselesaikan dengan baik.

Adapun tujuan dari penyusunan skripsi ini adalah untuk memenuhi salah satu persyaratan dalam menyelesaikan program studi S1 Informatika di Fakultas Ilmu Komputer, Universitas Pembangunan Nasional "Veteran" Jawa Timur.

Pada kesempatan ini, penulis ingin menyampaikan penghargaan dan rasa terima kasih yang sebesar-besarnya kepada pihak-pihak yang telah memberikan dukungan dalam penyelesaian skripsi ini:

1. Bapak Prof. Dr. Ir. Akhmad Fauzi, M.MT., selaku Rektor Universitas Pembangunan Nasional "Veteran" Jawa Timur.
2. Ibu Dr. Ir. Novirina Hendrasarie, MT., selaku Dekan Fakultas Ilmu Komputer, Universitas Pembangunan Nasional "Veteran" Jawa Timur.
3. Ibu Fetty Anggraeny, S.Kom, M.Kom., selaku Koordinator Program Studi Informatika, Fakultas Ilmu Komputer, Universitas Pembangunan Nasional "Veteran" Jawa Timur.
4. Bapak Andreas Nugroho Sihananto, S.Kom., M.Kom., selaku koordinator skripsi yang membantu dalam proses administrasi
5. Bapak Dr. Basuki Rahmat, S.Si., MT dan Ibu Eva Yulia Puspaningrum, S.Kom., M.Kom., selaku dosen pembimbing yang selalu memberikan dukungan dan bimbingan selama proses penyusunan laporan skripsi.
6. Ibu Dr. Eng. Ir. Anggraini Puspita Sari, ST., MT dan Bapak Fawwaz Ali Akbar, S.Kom., M.Kom., selaku dosen penguji seminar hasil skripsi yang telah memberikan masukan dan arahan yang sangat berharga dalam penyempurnaan skripsi ini.
7. Bapak Eka Prakarsa Mandyartha, S.T., M.Kom, selaku dosen wali yang senantiasa memberikan arahan, nasihat, dan bimbingan selama masa perkuliahan.

8. Terima kasih kepada keluarga saya, ayah Seniman, ibu Umaroh, saudara kembar saya Andina, dan nenek saya Satri atas doa, kasih sayang, dan dukungan tanpa henti yang menjadi sumber kekuatan penulis.
9. Kepada teman-teman seperjuangan saya Hanun, Arifa, Nadhia, Rina dari Informatika angkatan 2021 yang telah menjadi sumber semangat, diskusi, sepanjang perjalanan studi ini. Terima kasih atas kebersamaan, kerja sama, dan dukungan yang telah diberikan, baik dalam menghadapi tantangan akademik maupun non-akademik.
10. Rekan-rekan di UKM Penalaran dan Kreativitas khususnya Lely, serta keluarga besar GenBI khususnya Allan, Puput dan Serba Bisa Tim, yang telah menjadi bagian penting dalam perjalanan intelektual dan pengembangan diri penulis.
11. Kepada sahabat-sahabat terdekat saya Umroh, Robi, Nasa, Putri, Fiya, Diya, Nabila, dan Anti yang telah menjadi tempat berbagi cerita, keluh kesah, tawa, dan semangat. Terima kasih atas dukungan yang tulus, kebersamaan yang hangat, serta semua momen berharga yang telah dilalui bersama.
12. Kepada teman-teman seperjuangan saya Adel, Amar, Zila, Oca, dan Mas Alvi yang menjadi rekan satu penguji pada masa seminar hasil skripsi. Terima kasih atas semangat, kerja sama, dan saling dukung selama proses revisi hingga penyelesaian tugas akhir ini.

Surabaya, 19 Juni 2025

Penulis

DAFTAR ISI

LEMBAR PENGESAHAN	iii
LEMBAR PERSETUJUAN	v
SURAT PERNYATAAN BEBAS PLAGIASI.....	vii
ABSTRAK	ix
ABSTRACT	xi
KATA PENGANTAR.....	xiii
DAFTAR ISI.....	xv
DAFTAR GAMBAR.....	xix
DAFTAR TABEL	xxi
BAB I PENDAHULUAN.....	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	4
1.3 Tujuan Penelitian	4
1.4 Manfaat Penelitian	4
1.5 Batasan Penelitian	4
BAB II TINJAUAN PUSTAKA.....	7
2.1 Penelitian Terdahulu	7
2.2 Landasan Teori.....	10
2.2.1 Konsep Ekonomi Hijau	10
2.2.2 Indeks Ekonomi Hijau.....	10
2.2.3 Prediksi Indeks Ekonomi Hijau	12
2.2.4 <i>Machine Learning</i>	13
2.2.5 Prediksi.....	13
2.2.6 Model Regresi	14
2.2.7 Pohon Keputusan (<i>Decision Tree</i>)	15
2.2.8 <i>Nodes</i>	16
2.2.9 <i>Threshold</i>	16
2.2.10 Nilai <i>Output</i>	17
2.2.11 Perhitungan Prediksi	17
2.2.12 Residual.....	18

2.2.13 <i>Extreme Gradient Boosting (XGBoost)</i>	18
2.2.14 <i>Light Gradient Boosting Machine (LightGBM)</i>	21
2.2.15 <i>Categorical Boosting (CatBoost)</i>	25
2.2.16 <i>Stacking Ensemble</i>	27
2.2.17 <i>Mean Absolute Error (MAE)</i>	28
2.2.18 <i>Root Mean Squared Error (RMSE)</i>	29
2.2.19 Koefisien Determinasi (R^2 atau <i>R-squared</i>)	29
2.2.20 <i>Paired t-test</i>	30
2.2.21 <i>Min Max Scaling</i>	31
2.2.22 <i>Pearson Correlation</i>	31
2.2.23 <i>Hyperparameter</i>	31
BAB III METODOLOGI PENELITIAN	33
3.1 Tahapan Penelitian	33
3.2 Identifikasi Masalah	34
3.3 Pemahaman Data	34
3.4 Pra-Pemrosesan Data	36
3.4.1 Muat Data	36
3.4.2 Eksplorasi Data Awal	36
3.4.3 Normalisasi Data	36
3.4.4 Deteksi dan Penghapusan Multikolinearitas	37
3.5 Pengembangan Model	39
3.5.1 Dataset Normalisasi	40
3.5.2 Memisahkan Data Fitur dan Target	40
3.5.3 Membagi Data Menjadi <i>Base</i> , <i>Meta</i> , dan <i>Holdout set</i>	40
3.5.4 Set <i>Hyperparameter</i>	41
3.5.5 Pembuatan <i>Base Model</i> XGBoost	41
3.5.6 Pembuatan <i>Base Model</i> LightGBM	43
3.5.7 Pembuatan <i>Base Model</i> CatBoost	44
3.5.8 Prediksi <i>Base Learner</i> untuk Dataset <i>Meta</i>	45
3.5.9 Membuat Dataset Baru untuk <i>Meta Learner</i>	45
3.5.10 Pelatihan <i>Meta Model</i>	46
3.5.11 Prediksi Akhir pada <i>Holdout Set</i>	47

3.6 Evaluasi Model.....	47
3.7 Visualisasi Hasil Prediksi.....	48
3.8 Skenario Pengujian.....	50
BAB IV HASIL DAN PEMBAHASAN	53
4.1 Impor Pustaka dan Modul	53
4.2 Prapemrosesan Data	54
4.2.1 Pemuatan Data	54
4.2.2 Eksplorasi Data	55
4.2.3 Hasil Normalisasi Data	59
4.2.4 Deteksi dan Penghapusan Multikolinearitas	60
4.3 Pengembangan Model.....	61
4.3.1 Pisah Data Fitur dan Target.....	61
4.3.2 Pembagian Data	62
4.3.3 Set Hiperparameter.....	63
4.3.4 Pelatihan Model Dasar	64
4.3.5 Membuat Dataset Baru untuk <i>Meta Learner</i>	64
4.3.6 Pelatihan <i>Meta Learner</i>	65
4.3.7 Prediksi Akhir pada Data <i>Holdout</i>	66
4.4 Evaluasi Model.....	69
4.5 Visualisasi Hasil Prediksi.....	71
4.6 Analisis dan Hasil Skenario Pengujian	76
4.6.1 Skenario Uji dengan <i>Learning Rate</i>	76
4.6.2 Skenario Uji dengan <i>N_estimators</i>	81
4.6.3 Skenario Uji dengan <i>Max_Depth</i>	85
4.6.4 Pelatihan dan Pengujian Model.....	90
4.6.5 Analisis Hasil Perbandingan Model.....	95
BAB V PENUTUP	99
5.1 Kesimpulan	99
5.2 Saran.....	100
DAFTAR PUSTAKA	101

Halaman ini sengaja dikosongkan

DAFTAR GAMBAR

Gambar 2. 1 Jenis <i>Nodes</i> pada <i>Decision Tree</i>	16
Gambar 2. 2 Skema Algoritma XGBoost	19
Gambar 2. 3 Histogram Berdasarkan Algoritma <i>Decision Tree</i>	22
Gambar 2. 4 Kontruksi <i>Leaf Wise</i> dan <i>Level Wise</i>	23
Gambar 2. 5 Skema Diagram Algoritma CatBoost.....	26
Gambar 2. 6 Alur Kerja Teknik <i>Stacking</i>	27
Gambar 3. 1 Tahapan Penelitian	33
Gambar 3. 2 Alur Pra-Pemrosesan Data	36
Gambar 3. 3 Diagram Alur Proses <i>Stacking Ensemble</i>	40
Gambar 3. 4 Contoh Plot Diagram Hasil Prediksi	48
Gambar 3. 5 Contoh <i>Plotting</i> Residual	49
Gambar 4. 1 Preview Dataset Indeks Ekonomi Hijau.....	55
Gambar 4. 2 Tipe Data Setiap Kolom.....	57
Gambar 4. 3 Hasil Cek Nilai Hilang	57
Gambar 4. 4 Heatmap Korelasi Variabel (GEI).....	58
Gambar 4. 5 Hasil Normalisasi Data	60
Gambar 4. 6 Hasil Pembagaian Dataset	63
Gambar 4. 7 Hasil Prediksi Model LightGBM	68
Gambar 4. 8 Hasil Prediksi Model XGBoost.....	68
Gambar 4. 9 Hasil Prediksi Model CatBoost.....	68
Gambar 4. 10 Hasil Prediksi Model <i>Stacking</i>	69
Gambar 4. 11 Plotting Hasil Prediksi Model <i>Stacking</i>	73
Gambar 4. 12 Plotting Residual Model <i>Stacking</i>	75
Gambar 4. 13 Evaluasi Metrik dengan <i>Learning Rate</i> 0.001.....	77
Gambar 4. 14 Evaluasi Metrik dengan <i>Learning Rate</i> 0.01.....	78
Gambar 4. 15 Evaluasi Metrik dengan <i>Learning Rate</i> 0.1.....	79
Gambar 4. 16 Evaluasi Metrik dengan <i>N Estimators</i> 50.....	81
Gambar 4. 17 Evaluasi Metrik dengan <i>N Estimators</i> 100.....	82
Gambar 4. 18 Evaluasi Metrik dengan <i>N Estimators</i> 1000.....	83

Gambar 4. 19 Evaluasi Metrik dengan <i>Max_depth</i> 3.....	86
Gambar 4. 20 Evaluasi Metrik dengan <i>Max_depth</i> 6.....	87
Gambar 4. 21 Evaluasi Metrik dengan <i>Max_depth</i> 8.....	88
Gambar 4. 22 Evaluasi Metrik Untuk Model LightGBM.....	91
Gambar 4. 23 Plot Diagram Hasil Prediksi LightGBM	91
Gambar 4. 24 Evaluasi Metrik Untuk Model XGBoost.....	92
Gambar 4. 25 Plot Diagram Hasil Prediksi LightGBM	92
Gambar 4. 26 Evaluasi Metrik Untuk Model CatBoost.....	93
Gambar 4. 27 Plot Diagram Hasil Prediksi CatBoost	93
Gambar 4. 28 Evaluasi Metrik Untuk Model <i>Stacking</i>	94
Gambar 4. 29 Plot Diagram Hasil Prediksi <i>Stacking</i>	94
Gambar 4. 30 Perbandingan Hasil Plot Residu Semua Model	96

DAFTAR TABEL

Tabel 2. 1 Penjelasan Kolom Uji <i>Paired t-test</i>	30
Tabel 2. 2 Hiperparameter Model	32
Tabel 3. 1 Dataset Penelitian.....	35
Tabel 3. 2 Contoh Dataset Sebelum Normalisasi.....	37
Tabel 3. 3 Contoh Dataset Hasil Normalisasi	37
Tabel 3. 4 Matriks Korelasi Pearson Antar Fitur	38
Tabel 3. 5 Contoh Dataset Setelah Penghapusan Multikolinearitas	39
Tabel 3. 6 Contoh Hasil Prediksi Model XGBoost.....	42
Tabel 3. 7 Contoh Hasil Prediksi Model LightGBM	44
Tabel 3. 8 Contoh Hasil Prediksi Model CatBoost.....	45
Tabel 3. 9 Contoh Dataset Baru untuk Proses <i>Stacking</i>	46
Tabel 3. 10 Contoh Hasil Prediksi Model <i>Stacking</i>	47
Tabel 3. 11 Penyetelan Parameter dan Rentang Nilai.....	50
Tabel 4. 1 Analisis Kuantitatif Variabel Indeks Ekonomi Hijau	55
Tabel 4. 2 Hasil Perbandingan Evaluasi Metrik.....	71
Tabel 4. 3 Hasil Pengujian <i>Learning Rate</i>	80
Tabel 4. 4 Hasil Pengujian <i>n_estimators</i>	84
Tabel 4. 5 Hasil Pengujian <i>max_depth</i>	89
Tabel 4. 6 Perbandingan Hasil Evaluasi Metrik Semua Model	95
Tabel 4. 7 Hasil Uji T-test antara <i>Stacking</i> vs Algoritma Dasar	97

Halaman ini sengaja dikosongkan