

SKRIPSI

**PREDIKSI HARGA KENDARAAN BEKAS DI
PLATFORM OLX MENGGUNAKAN METODE
*GRADIENT BOOSTING REGRESSOR***



POLITEKNIK NEGERI BALI

Oleh :

Pandu Anggi Purwanto

NIM. 2115354010

**PROGRAM STUDI SARJANA TERAPAN
TEKNOLOGI REKAYASA PERANGKAT LUNAK
JURUSAN TEKNOLOGI INFORMASI
POLITEKNIK NEGERI BALI
2024**

ABSTRAK

Pesatnya pertumbuhan industri otomotif di Indonesia, sehingga jumlah kendaraan sepeda motor mencapai 132 juta unit pada tahun 2023, hal tersebut memicu tingkat transaksi kendaraan pada pasar kendaraan bekas dengan harga yang bervariasi. Penelitian ini bertujuan untuk membangun model prediksi harga motor bekas menggunakan algoritma *Gradient Boosting Regressor* (GBR) yang diterapkan pada data OLX Indonesia. Model dievaluasi dengan membandingkan performa pada dua rasio pembagian data, yaitu 80:20 dan 70:30. Evaluasi dilakukan menggunakan metrik RMSE, MAE, MAPE, dan R^2 baik pada skala logaritmik maupun skala asli, serta waktu yang dibutuhkan untuk tuning hyperparameter. Hasil menunjukkan bahwa model dengan split 80:20 memiliki performa sedikit lebih baik dengan nilai RMSE log sebesar 0.1095 dan R^2 sebesar 0.891, dibandingkan dengan RMSE log 0.1114 dan R^2 sebesar 0.888 pada split 70:30. Pada skala asli, split 80:20 menunjukkan RMSE sekitar 2,2 juta dan R^2 sebesar 0.915, sementara split 70:30 menghasilkan RMSE 2,24 juta dan R^2 sebesar 0.912. MAPE pada kedua rasio hampir setara, yaitu sekitar 8%, menandakan bahwa model mampu memprediksi harga dengan tingkat kesalahan rata-rata yang rendah. Visualisasi performa model menunjukkan distribusi error yang simetris dan residual yang tersebar secara acak di sekitar nol, menandakan tidak adanya bias signifikan. Validasi manual data dan penanganan outlier turut berkontribusi dalam meningkatkan akurasi model. Dengan transformasi logaritmik dan teknik pra-pemrosesan yang tepat, model mampu menangkap pola harga dengan baik. Secara keseluruhan, model yang dibangun dapat dijadikan alat bantu untuk memprediksi harga motor bekas secara akurat, serta memberikan nilai tambah bagi pengguna dalam menentukan harga jual atau beli secara objektif dan data-driven.

Kata Kunci: Prediksi Harga, Gradient Boosting Regressor, Platform Online, RMSE, R^2 , MAPE, Motor Bekas.

ABSTRACT

The rapid growth of the automotive industry in Indonesia, so that the number of motorcycle vehicles reached 132 million units in 2023, this triggered the level of vehicle transactions in the used vehicle market with varying prices. This study aims to build a used motorcycle price prediction model using the Gradient Boosting Regressor (GBR) algorithm applied to OLX Indonesia data. The model is evaluated by comparing performance at two data sharing ratios, namely 80:20 and 70:30. The evaluation was conducted using RMSE, MAE, MAPE, and R^2 metrics on both logarithmic and original scales, as well as the time required for hyperparameter tuning. Results show that the model with an 80:20 split performs slightly better with a log RMSE of 0.1095 and R^2 of 0.891, compared to a log RMSE of 0.1114 and R^2 of 0.888 for the 70:30 split. At the original scale, the 80:20 split showed an RMSE of about 2.2 million and an R^2 of 0.915, while the 70:30 split resulted in an RMSE of 2.24 million and an R^2 of 0.912. The MAPE for both ratios is almost equal at around 8%, indicating that the model is able to predict prices with a low average error rate. Visualization of the model's performance showed symmetrical error distributions and residuals randomly scattered around zero, indicating the absence of significant bias. Manual validation of the data and handling of outliers contributed to improving the accuracy of the model. With logarithmic transformation and proper pre-processing techniques, the model is able to capture price patterns well. Overall, the model built can be used as a tool to accurately predict the price of used motorcycles, as well as provide added value for users in determining the selling or buying price objectively and data-driven.

Keywords: *Price Prediction, Gradient Boosting Regressor, Online Platform, RMSE, R^2 , MAPE, Used Motorcycle.*

DAFTAR ISI

HALAMAN PERNYATAAN KEASLIAN KARYA SKRIPSI	iv
ABSTRAK.....	v
ABSTRACT.....	vi
KATA PENGANTAR.....	vi
DAFTAR ISI.....	ix
DAFTAR TABEL	xii
DAFTAR GAMBAR	xiii
DAFTAR LAMPIRAN	xv
BAB I.....	1
PENDAHULUAN.....	1
1.1. Latar Belakang	1
1.2. Rumusan Masalah	2
1.3. Batasan Masalah	2
1.4. Tujuan Penelitian	3
1.5. Manfaat Penelitian	3
1.6. Sistematika Penulisan.....	3
BAB II TINJAUAN PUSTAKA.....	5
2.1. Penelitian Sebelumnya	5
2.2. Landasan Teori.....	6
2.2.1. Python.....	6
2.2.2. Machine Learning	6
2.2.3. Data Non Linear.....	7
2.2.4. Web Scraping.....	7
2.2.5. Flowchart	7
2.2.6. Gradient Boosting	8
2.2.7. Metriks Evaluasi	12
2.2.8. Logaritmik	15
2.2.9. RandomizedSearchCV	15
2.2.10. Cross-Validation	16
2.2.11. Residual Plot.....	16
2.2.12. Interquartile Range.....	17

2.2.13. Box Plot	18
2.2.14. Flask	18
2.2.15. Black-box Testing	19
BAB III METODE PENELITIAN.....	20
3.1. Objek dan Metode Penelitian.....	20
3.1.1. Objek Penelitian.....	20
3.1.2. Metode Penelitian	20
3.2. Analisis Kondisi Eksisting.....	20
3.3. Rancangan Penelitian	21
3.3.1. Kebutuhan Sistem	21
3.3.2. Flowchart Model.....	21
3.3.3. Flowchart Sistem.....	27
3.4. Pengujian Penelitian.....	28
3.4.1. Pengujian Model	28
3.4.2. Pengujian Sistem.....	28
BAB IV HASIL DAN PEMBAHASAN	29
4.1. Implementasi Pembuatan Model.....	29
4.1.1. Pengumpulan Data	29
4.1.2. Validasi Data Manual.....	29
4.1.3. Pra-pemrosesan Data.....	29
4.1.4. Pembagian Data (Train-Test Split)	31
4.1.5. Penskalaan Fitur	31
4.1.6. Pemodelan Machine Learning	31
4.1.7. Evaluasi Model	33
4.2. Implementasi Pembuatan Sistem	33
4.2.1. Antarmuka Sistem.....	33
4.2.2. Proses Prediksi	34
4.2.3. Contoh Input dan Output	34
4.3. Hasil	35
4.3.1. Hasil Tuning Hyperparameter	36
4.3.2. Perbandingan Metrik Kinerja Model	37
4.3.3. Visualisasi Performa Model	37

4.3.4. Visualisasi Perbandingan Split	42
4.3.5. Hasil Prediksi.....	45
4.4. Pembahasan	54
BAB V PENUTUP.....	57
5.	57
5.1. Kesimpulan.....	57
5.2. Saran.....	57
DAFTAR PUSTAKA.....	59
LAMPIRAN.....	62

DAFTAR TABEL

Tabel 2.1 Simbol Flowchart.....	8
Tabel 2.2 Penjelasan Simbol Rumus Model Awal	9
Tabel 2.3 Penjelasan Simbol Rumus Gradient.....	10
Tabel 2.4 Penjelasan Simbol Rumus Weak Learner	10
Tabel 2.5 Penjelasan Simbol Rumus Koefisien Optimal	11
Tabel 2.6 Penjelasan Simbol Rumus Model Diperbarui.....	12
Tabel 2.7 Penjelasan Simbol Rumus MAE	13
Tabel 2.8 Penjelasan Simbol Rumus MSE	13
Tabel 2.9 Penjelasan Simbol Rumus RMSE.....	14
Tabel 2.10 Penjelasan Simbol Rumus MAPE.....	14
Tabel 2.11 Penjelasan Simbol Rumus R2.....	15
Tabel 3.1 Contoh Data Perhitungan Evaluasi.....	23
Tabel 4.1 Contoh Input Data	34
Tabel 4.2 Contoh Hasil Output	34

DAFTAR GAMBAR

Gambar 2.1 Jenis Residu	17
Gambar 2.2 Box Plox	18
Gambar 3.1 Flowchart Model	21
Gambar 3.2 Alur Preprocessing	22
Gambar 3.3 Alur Pemodelan Algoritma GBR	23
Gambar 3.4 Flowchart Sistem	27
Gambar 4.1 Hyperparameter	32
Gambar 4.2 Antarmuka Sistem	34
Gambar 4.3 Hasil Prediksi	35
Gambar 4.4 Parameter Terbaik 80:20	36
Gambar 4.5 Parameter Terbaik 70:30	36
Gambar 4.6 Perbandingan Metrik Kinerja Model	37
Gambar 4.7 Perbandingan Metrik Kinerja Model Lanjutan	37
Gambar 4.8 Performa Model 80:20	38
Gambar 4.9 Distribusi Error 80:20	40
Gambar 4.10 Performa Model 70:30	41
Gambar 4.11 Distribusi Error 70:30	42
Gambar 4.12 Perbandingan Split RMSE, R ² , dan MAPE	43
Gambar 4.13 Perbandingan Waktu Split	44
Gambar 4.14 Hasil Prediksi Honda Vario 2020	45
Gambar 4.15 Harga Tertinggi Honda Vario 2020 Pada Platform OLX	46
Gambar 4.16 Harga Terendah Honda Vario 2020 Pada Platform OLX	46
Gambar 4.17 Harga Rata-Rata Honda Vario 2020 Berdasar Data Pemodelan	47
Gambar 4.18 Hasil Prediksi Honda Beat 2023	48
Gambar 4.19 Harga Tertinggi Honda Beat 2023 Pada Platform OLX	48
Gambar 4.20 Harga Terendah Honda Beat 2023 Pada Platform OLX	49

Gambar 4.21 Harga Rata-Rata Honda Beat 2023 Berdasar Data Pemodelan	49
Gambar 4.22 Hasil Prediksi Yamaha Mio 2022	50
Gambar 4.23 Harga Tertinggi Yamaha Mio 2022 Pada Platform OLX	50
Gambar 4.24 Harga Terendah Yamaha Mio 2022 Pada Platform OLX	51
Gambar 4.25 Harga Rata-Rata Yamaha Mio 2022 Berdasar Data Pemodelan	51
Gambar 4.26 Hasil Prediksi Yamaha Xmax 2021	52
Gambar 4.27 Harga Tertinggi Yamaha Xmax 2021 Pada Platform OLX.....	52
Gambar 4.28 Harga Terendah Yamaha Xmax 2021 Pada Platform OLX	53
Gambar 4.29 Harga Rata-Rata Yamaha Xmax 2021 Berdasar Data Pemodelan	53

DAFTAR LAMPIRAN

Lampiran 1: Pernyataan Telah Menyelesaikan Bimbingan Skripsi	62
Lampiran 2: Lembar Perbaikan Penguji 1	63
Lampiran 3: Lembar Perbaikan Penguji 2	64
Lampiran 4: Lembar Perbaikan Penguji 3	65

BAB I

PENDAHULUAN

1.1. Latar Belakang

Indonesia mengalami pertumbuhan pesat terutama di industri otomotif pada beberapa dekade terakhir. Badan Pusat Statistik (BPS) menyebutkan bahwa kendaraan sepeda motor yang terdata di Indonesia sebanyak 132 Juta pada tahun 2023. Pertumbuhan ini didukung oleh minat masyarakat karena kendaraan sepeda motor dipandang efektif dan efisien pada kondisi jalan macet maupun pada jalan kecil, selain itu harganya yang cukup terjangkau oleh kalangan menengah kebawah [1]. Akibat jumlah kendaraan yang terus bertambah membuat pasar kendaraan bekas mengalami perkembangan yang pesat karena banyaknya pemilik kendaraan yang ingin menjual kendaraan mereka baik itu karena ingin mengganti yang baru maupun karena alasan ekonomi. Fenomena ini membuat pasar kendaraan mengalami persaingan, dengan banyaknya transaksi yang terjadi setiap hari. Platform *online* memiliki peranan penting untuk memfasilitasi transaksi dalam memudahkan penjual menjangkau pembeli di seluruh Indonesia.

Melakukan penjualan kendaraan bekas bukanlah suatu masalah yang mudah. Salah satu tantangannya yaitu mencari informasi terlebih dahulu terkait harga kendaraan bekas melalui platform *online*. Hal ini terkadang membuat penjual banyak menghabiskan waktu untuk mencari harga yg sesuai dengan kendaraan yang ingin dijual [3].

Untuk mengatasi tantangan tersebut, maka perlu pendekatan berbasis teknologi seperti *machine learning*. *Machine learning* merupakan sebuah teknologi yang mampu mengolah data dengan otomatis yang sering digunakan pada sistem untuk mempelajari bagaimana cara mengenali pola maupun untuk menarik kesimpulan baru dari pola sebelumnya, contoh penerapannya seperti deteksi wajah, *filter spam*, serta melakukan segmentasi pelanggan [4]. *Machine learning* memiliki dua pendekatan yaitu *supervised learning* yang memanfaatkan data yang terklasifikasi untuk membangun model prediksi dan *unsupervised learning* yang dapat melakukan identifikasi pola tanpa menggunakan data yang terklasifikasi [4]. Maka *machine learning* cocok digunakan untuk memprediksi harga kendaraan yang bersifat *non-linear* menggunakan algoritma *gradient boosting regressor (GBR)*. GBR merupakan algoritma yang efektif untuk menangani data *non-linear* yang kompleks serta mampu mengolah variabel dengan bobot yg berbeda-beda [2]. Sebuah penelitian mencoba untuk

meneliti kinerja GBR dengan mengimplementasikan fungsi polinomial. Hasil yang diberikan yaitu terjadi peningkatan nilai R^2 dari 0,9988 menjadi 0,9998 serta terjadi sebuah penurunan RMSE hingga 0,008 [5]. Selain meneliti kinerja GBR, juga ada penelitian yang berfokus dalam estimasi harga jual mobil dan memperoleh nilai prediksi kesalahan 2,65% dari harga real yaitu sebesar Rp. 114.650.000,- dan setelah dilakukan prediksi menggunakan regresi linier berganda hasilnya sebesar Rp. 112.000.000,- [3]. Penelitian lainnya menemukan bahwa *gradient boosting regression tree* (GBRT) menghasilkan nilai MAE, *root mean squared error* (RMSE), dan *mean absolute percentage error* (MAPE) lebih kecil dibandingkan menggunakan algoritma *support vector machine* (SVM) maupun *Recurrent neural network* (RNN) dalam memprediksi konsumsi energi rumah tangga [6]. Selain itu juga ada penelitian yang memprediksi angka harapan hidup menggunakan beberapa model seperti *decision tree regression*, *random forest regression*, *gradient boosting regression*, serta *XGBoost regression*, pada penelitian tersebut didapatkan hasil bahwa kinerja *random forest regression* lebih unggul dimana RMSE menunjukkan nilai sebesar 2,04 saat sebelum dilakukan tuning, dan setelah dilakukan tuning RMSE menunjukkan nilai sebesar 1,94 [17]. Berdasarkan penelitian-penelitian sebelumnya maka penelitian ini bertujuan untuk memprediksi harga kendaraan motor dengan GBR, dikarenakan GBR merupakan algoritma *supervised learning* yang mampu menangani data *non-linear* karena harga kendaraan dapat dipengaruhi banyak faktor seperti merek, model, tahun, serta kondisi kendaraan tersebut. Maka dengan penerapan algoritma GBR untuk memprediksi harga kendaraan diharapkan dapat membantu penjual menentukan harga yang kompetitif, dan juga membantu pembeli untuk menentukan keputusan sebelum membeli sebuah kendaraan motor.

1.2. Rumusan Masalah

Dari latar belakang di atas, maka didapatkan permasalahan yaitu bagaimana penerapan algoritma *Gradient Boosting Regressor* dalam melakukan prediksi harga kendaraan bekas berdasarkan data yang didapat dari platform *online*?

1.3. Batasan Masalah

Untuk membuat penelitian terfokus, maka beberapa batasan masalah yang dapat ditentukan sebagai berikut :

- a. Data yang digunakan bersifat *static* melalui hasil *scraping* pada platform OLX.

- b. Penelitian hanya menggunakan data kendaraan sepeda motor.
- c. Prediksi harga kendaraan bekas menggunakan atribut seperti merek, model, tahun produksi, kondisi kendaraan, serta lokasi penjualan.
- d. Algoritma yang digunakan untuk melakukan prediksi adalah *Gradient Boosting Regressor* tanpa membandingkan algoritma lain.

1.4. Tujuan Penelitian

Penelitian ini bertujuan untuk membangun model prediksi menggunakan *Gradient Boosting Regressor* untuk menentukan harga kendaraan bekas berdasarkan data yang didapat dari platform *online*.

1.5. Manfaat Penelitian

Penelitian ini diharapkan memberikan manfaat sebagai berikut :

- a. Bagi Penjual : Dapat membantu menentukan harga kendaraan bekas supaya lebih cepat terjualnya.
- b. Bagi Pembeli : Dapat membantu pembeli untuk menentukan keputusan sebelum membeli kendaraan bekas.
- c. Bagi Akademisi : Dapat memberikan kontribusi dalam penerapan *machine learning* menggunakan *gradient boosting regressor* pada prediksi jual beli kendaraan bekas.

1.6. Sistematika Penulisan

Untuk mempermudah dalam mengetahui pembahasan yang terdapat pada skripsi ini, maka diperlukan sistematika yang berfungsi sebagai kerangka dan pedoman penulisan skripsi. Berikut adalah sistematika penulisannya:

- a. Bagian Awal Skripsi

Pada bagian awal memuat halaman sampul depan, halaman persetujuan, halaman pengesahan, halaman pernyataan, halaman abstrak, halaman abstrak dalam bahasa inggris, halaman kata pengantar, halaman daftar isi, halaman daftar tabel, dan halaman daftar gambar.

- b. Bagian Utama Skripsi

Pada bagian utama terbagi menjadi beberapa bab dan sub bab sebagai berikut:

BAB I PENDAHULUAN

Bab ini terdiri dari latar belakang, rumusan masalah, batasan masalah, tujuan penelitian, manfaat penelitian, dan sistematika penulisan.

BAB II TINJAUAN PUSTAKA

Bab ini terdapat penelitian sebelumnya yang berisikan tentang hasil-hasil penelitian yang sudah dilakukan, dan terdapat juga landasan teori yang berisikan tentang *python*, *machine learning*, *data non linear*, *web scraping*, *flowchart*, *gradient boosting*, metrik evaluasi, *flask*, *black-box testing*.

BAB III METODE PENELITIAN

Bab ini merupakan tahapan tentang metode penelitian yang akan dilakukan oleh penulis meliputi objek dan metode penelitian, analisis kondisi eksisting, rancangan penelitian, serta pengujian penelitian.

BAB IV HASIL DAN PEMBAHASAN

Bab ini berisikan gambaran hasil penelitian dan analisa yang tersusun dengan baik diklasifikasikan ke dalam hasil penelitian, dan pembahasan.

BAB V PENUTUP

Bab ini berisikan kesimpulan dan saran dari penelitian yang telah dilakukan. Kesimpulan dikemukakan dari hasil penelitian sedangkan saran berisikan pengembangan yang bisa dilakukan terhadap masalah yang terdapat pada hasil penelitian.

c. Bagian Akhir Skripsi

Pada bagian akhir dari skripsi ini berisikan tentang daftar pustaka dan daftar lampiran.

BAB V

PENUTUP

5.1. Kesimpulan

Berdasarkan hasil dan pembahasan pada penelitian ini mengenai prediksi harga kendaraan bekas menggunakan *Gradient Boosting Regressor* dapat disimpulkan bahwa penelitian ini berhasil dalam menerapkan algoritma *Gradient Boosting Regressor* untuk membangun model prediksi dan GBR ini terbukti efektif dalam menangani kompleksitas dan sifat *non-linear* pada data harga kendaraan yang dipengaruhi berbagai faktor seperti ‘Merek’, ‘Model’, serta ‘Tahun Produksi’. Rasio data split 80:20 memberikan hasil yang sedikit lebih baik dibandingkan 70:30, namun perbedaan tersebut tidak signifikan, yang menunjukkan bahwa model ini cukup stabil terhadap variasi data pelatihan.

Keberhasilan model ini juga dipengaruhi beberapa tahap seperti proses validasi manual data yang memastikan kualitas dataset, penanganan outlier yang menjaga kestabilan distribusi harga, serta penggunaan transformasi logaritmik yang membantu normalisasi target. Seluruh tahapan metodologi yang diterapkan dari pengambilan data, pembersihan, pra-pemrosesan, hingga pemodelan.

5.2. Saran

Berdasarkan hasil penelitian dan kesimpulan yang telah dicapai, ada beberapa saran untuk pengembangan penelitian ini seperti:

1. Penambahan fitur: pada penelitian selanjutnya sangat diharapkan untuk melakukan penambahan fitur seperti jumlah kilometer, warna kendaraan, serta kondisi kendaraan. Hal ini mungkin dapat meningkatkan akurasi model prediksi.
2. Melakukan perbandingan algoritma: Meskipun penelitian ini berfokus pada GBR, diharapkan penelitian selanjutnya dapat membandingkan kinerja GBR dengan algoritma *machine learning* lainnya (misalnya, *Random Forest*, *XGBoost*, *LightGBM*, atau bahkan model *deep learning*) untuk mengidentifikasi algoritma yang paling optimal untuk kasus prediksi harga kendaraan bekas.

3. Melakukan validasi eksternal: Melakukan validasi model dengan data dari platform online lain atau sumber data eksternal lainnya dapat meningkatkan generalisasi dan *robustness* model.
4. Melakukan pemodelan secara *realtime*: Dengan melakukan pemodelan secara *realtime*, model dapat belajar dan beradaptasi dengan perubahan harga yang terjadi pada platform *online*, sehingga memungkinkan hasil prediksi lebih akurat.

DAFTAR PUSTAKA

- [1] A. Acuviarta and A. M. P. Permana, "Analisis Faktor Yang Mempengaruhi Permintaan Sepeda Motor di Kota-Kota Besar Jawa Barat," *Jurnal Rekayasa Industri dan Ekonomi*, vol. 2, no. 3, pp. 171-180, 2022.
- [2] S. Suryana, B. Warsito, and S. Suparti, "Penerapan Gradient Boosting dengan Hyperopt untuk Memprediksi Keberhasilan Telemarketing Bank," *Jurnal Gaussian*, vol. 10, no. 4, pp. 617-623, 2021
- [3] E. D. S. Mulyani, F. Mulady, D. Ramadhan, A. Ariyantono, D. Ramdani, R. Wahyundana, and M. Gilang, "Estimasi Harga Jual Mobil Bekas Menggunakan Metode Regresi Linier Berganda," *STMIK Tasikmalaya, Program Studi Teknik Informatika*, vol. 9, no. 1, 2020.
- [4] A. Fathurohman, "Machine Learning untuk Pendidikan: Mengapa dan Bagaimana," *Jurnal Informatika dan Teknologi Komputer*, vol. 1, no. 3, pp. 57–62, 2021.
- [5] N. V. Putranto, M. Akrom, G. A. Trinapradika, and A. History, "Implementasi Fungsi Polinomial pada Algoritma Gradient Boosting Regressor: Studi Regresi pada Dataset Obat-Obatan Kadaluarsa sebagai Material Antikorosi," *Jurnal Teknologi dan Manajemen Informatika*, vol. 9, no. 2, pp. 172–182, 2023
- [6] P. Nie, M. Roccotelli, M. P. Fanti, Z. Ming, and Z. Li, "Prediction of home energy consumption based on gradient boosting regression tree," *School of Electro-Mechanical Engineering, Xidian University, Xi'an, China; Polytechnic University of Bari, Bari, Italy; Institute of Systems Engineering, Macau University of Science and Technology, Taipa, Macao SAR*, 2024.
- [7] N. Fedorov, Y. Petrichenko, "Gradient Boosting–Based Machine Learning Methods in Real Estate Market Forecasting," *Institute of Information Technologies, Chelyabinsk State University*, 2021.
- [8] P. D. Pitroda, B. C. Donga, H. B. Domadiya, and D. H. Domadiya, "Beyond The Basics: A Detailed Survey of Advanced Python Applications and Innovations," *International Journal for Research in Applied Science and Engineering Technology (IJRASET)*, vol. 8, no. 5, pp. 457-463, 2020.

- [9] D. D. Ayani, H. S. Pratiwi, and H. Muhardi, "Implementasi Web Scraping untuk Pengambilan Data pada Situs Marketplace," *Jurnal Sistem dan Teknologi Informasi (JUSTIN)*, vol. 7, no. 4, pp. 257, 2019.
- [10] Z. He, D. Lin, T. Lau, and M. Wu, "Gradient Boosting Machine: A Survey," *arXiv preprint arXiv:1908.06951*, 2019.
- [11] S. B. Koduri, L. Guniseti, C. R. Ramesh, K. Mutyalu, and D. Ganesh, "Prediction of crop production using adaboost regression method Prediction of crop production using adaboost regression method," *J. Phys. Conf. Ser.*, 2019.
- [12] A. Zalukhu *et al.*, "PERANGKAT LUNAK APLIKASI PEMBELAJARAN FLOWCHART," *Jurnal Teknologi Informasi dan Industri*, vol. 4, no. 1, 2023.
- [13] D. F. Ningtyas and N. Setiyawati, "Implementasi Flask Framework pada Pembangunan Aplikasi Purchasing Approval Request," *J. Janitra Inform. dan Sist. Inf.*, vol. 1, no. 1, pp. 19–34, 2021.
- [14] C. Padmavathya dan S.S. Sugantha Mallika, "A Survey on Applications of Nonlinear Data Structures," *Turkish Journal of Computer and Mathematics Education*, vol. 12, no. 10, pp. 6336–6350, 2021.
- [15] S. Handayanto dan I. Nuryasin, "Pengujian Blackbox Decision Table pada Sistem Aplikasi Mobile Sharing Story App," *Jurnal Ilmu Komputer dan Informatika*, vol. 12, no. 10, pp. 6336–6350, 2021.
- [16] M. N. Fietroh and Suhaedi, "Keputusan Pembelian Sepeda Motor Bekas di Kalangan Mahasiswa Universitas Teknologi Sumbawa," *Jurnal Ekonomi dan Bisnis*, vol. 9, no. 3, pp. 236-242, Des. 2021.
- [17] G. Chairunisa, M. K. Najib, S. Nurdianti, S. F. Imni, W. Sanjaya, R. D. Andriani, H. Henriyansah, R. S. P. Putri, dan D. Ekaputri, "Life Expectancy Prediction Using Decision Tree, Random Forest, Gradient Boosting, and XGBoost Regressions," *Jurnal Sintak*, vol. 2, no. 2, pp. 71-82, 2024.
- [18] N. H. Cahyana, Y. Fauziah, W. Wisnalmawati, dan A. S. Aribowo, "Hyperparameter Optimization of Semi-Supervised Sentiment Annotation Model on Marketplace Dataset" in *Proceedings of the 2nd International Conference on Advance Research in Social and Economic Science (ICARSE 2023)*, vol. 842, pp. 161–170, 2024.

- [19] J. Allgaier and R. Pryss, “Cross-Validation Visualized: A Narrative Guide to *Advanced Methods*” *Machine Learning and Knowledge Extraction*, vol. 6, no. 2, pp. 1378–1388, 2024.
- [20] W. Li, D. Cook, E. Tanaka, and S. VanderPlas, “A Plot is Worth a Thousand Tests: Assessing Residual Diagnostics with the Lineup Protocol” , *Journal of Computational and Graphical Statistics.*, vol. 33, no. 4, pp. 1497–1511, 2024.
- [21] H.-Y. Kim, “Statistical notes for clinical researchers: simple linear regression 3 – residual analysis” *Restor. Dent. Endod.*, vol. 44, no. 1, p. e11, 2019.
- [22] R. Aznar-Gimeno, et al. “Incorporating a New Summary Statistic into the Min–Max Approach: A Min–Max–Median, Min–Max–IQR Combination of Biomarkers for Maximising the Youden Index,” *Mathematics*, vol. 9, no. 19, art. 2497, 2021.
- [23] M. Baumeister, M. Munko, K.-P. Gladow, M. Ditzhaus, N. Chakarov, and M. Pauly, “Early and Late Buzzards: Comparing Different Approaches for Quantile-based Multiple Testing in Heavy-Tailed Wildlife Research Data,” arXiv preprint arXiv:2409.14926, 2024.
- [24] A. Mazarei, R. Sousa, J. Mendes-Moreira, and J. Gama, “Online boxplot derived outlier detection,” *International Journal of Data Science and Analytics*, vol. 19, pp. 83–97, 2025.