



SKRIPSI

IMPLEMENTASI METODE *XGBOOST* DALAM PREDIKSI PENJUALAN PRODUK PAWONKOE DAN ANALISIS ULASAN PELANGGAN MENGUNAKAN METODE *MBERT*

SELENA NURMANINA AFANDY
NPM 21083010076

DOSEN PEMBIMBING
Kartika Maulida Hindrayani, S.Kom., M.Kom
Aviolla Terza Damaliana, S.Si, M.Stat

**KEMENTERIAN PENDIDIKAN TINGGI, SAINS, DAN TEKNOLOGI
UNIVERSITAS PEMBANGUNAN NASIONAL VETERAN JAWA TIMUR
FAKULTAS ILMU KOMPUTER
PROGRAM STUDI SAINS DATA
SURABAYA
2026**

LEMBAR PENGESAHAN

IMPLEMENTASI METODE *XGBOOST* DALAM PREDIKSI PENJUALAN PRODUK PAWONKOE DAN ANALISIS ULASAN PELANGGAN MENGUNAKAN METODE *MBERT*

Oleh:
SELENA NURMANINA AFANDY
NPM. 21083010076

Telah dipertahankan di hadapan dan diterima oleh Tim Penguji Sidang Skripsi Program Studi Sains Data Fakultas Ilmu Komputer Universitas Pembangunan Nasional Veteran Jawa Timur pada Tanggal 24 Februari 2026:

Menyetujui,

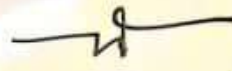
Kartika Maulida Hindrayani, S.Kom., M.Kom
NIP. 199209092022032009


..... (Pembimbing I)

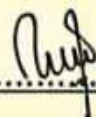
Aviolla Terza Damaliana, S.Si, M.Stat
NIP. 199408022022032015


..... (Pembimbing II)

Dr. Ir. Mohammad Idhom, S.P., S.Kom., M.T.
NIP. 198303102021211006


..... (Penguji I)

Muhammad Nasrudin, M. Stat.
NIP. 199609092024061002


..... (Penguji II)

Mengetahui,
Dekan Fakultas Ilmu Komputer


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

LEMBAR PERSETUJUAN

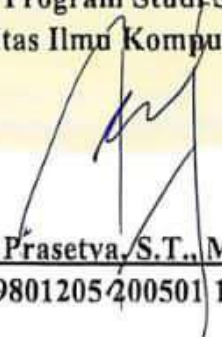
IMPLEMENTASI METODE *XGBOOST* DALAM PREDIKSI PENJUALAN PRODUK PAWONKOE DAN ANALISIS ULASAN PELANGGAN MENGUNAKAN METODE *MBERT*

Oleh:
SELENA NURMANINA AFANDY
NPM. 21083010076

Telah disetujui untuk mengikuti Ujian Skripsi

Menyetujui,

Koordinator Program Studi Sains Data
Fakultas Ilmu Komputer


Dr. Eng. Ir. Dwi Arman Prasetya, S.T., M.T., IPU., Asean Eng.
NIP. 19801205200501 1 002

SURAT PERNYATAAN BEBAS PLAGIASI

Saya yang bertanda tangan di bawah ini:

Nama Mahasiswa : Selena Nurmanina Afandy
NPM : 21083010076
Program : Sarjana (S1)
Program Studi : Sains Data
Fakultas : 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 menyatakan bahwa dokumen ilmiah ini bebas dari unsur-unsur plagiasi. Apabila di kemudian 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, 23 Februari 2026
Yang Membuat Pernyataan,



Selena Nurmanina Afandy

NPM. 21083010076

ABSTRAK

Nama Mahasiswa / NPM : Selena Nurmanina Afandy / 21083010076
Judul Skripsi : Implementasi Metode *XGBOOST* Dalam Prediksi Penjualan Produk Pawonkoe Dan Analisis Ulasan Pelanggan Menggunakan Metode *MBERT*
Dosen Pembimbing : 1. Kartika Maulida Hindrayani, S.Kom., M.Kom
2. Aviolla Terza Damaliana, S.Si, M.Stat

Fluktuasi penjualan harian dan dinamika persepsi pelanggan menjadi tantangan utama dalam pengelolaan UMKM Pawonkoe Banyuwangi selama periode Januari–Desember 2024. Variasi penjualan yang signifikan berpotensi menimbulkan ketidakefisienan dalam perencanaan produksi apabila tidak didukung sistem prediksi yang akurat. Selain itu, terdapat 365 data ulasan pelanggan yang belum dimanfaatkan secara optimal sebagai dasar evaluasi strategis. Penelitian ini menggunakan 366 data penjualan harian dan 365 ulasan pelanggan melalui pendekatan terintegrasi berbasis *machine learning*. Model prediksi penjualan dibangun menggunakan algoritma *XGBOOST* dengan fitur *lag*, *rolling mean* tujuh hari, dan variabel hari dalam minggu serta pembagian data 80% pelatihan dan 20% pengujian. Hasil evaluasi menunjukkan MAPE sebesar 20,44% pada skema *1-step supervised* dan rata-rata 19,66% pada *Recursive 7-step Forecasting* sebagai hasil terbaik secara numerik. Namun, pada skema *rolling autoregressive* yang merepresentasikan kondisi peramalan riil diperoleh MAPE sebesar 46,27%, menunjukkan peningkatan *error* akibat akumulasi prediksi beruntun. Analisis sentimen menggunakan *Multilingual BERT (MBERT)* menghasilkan *accuracy* sebesar 0,91 dengan distribusi 245 sentimen positif (67,1%), 117 netral (32,1%), dan 3 negatif (0,8%). Hasil ini menunjukkan persepsi pelanggan didominasi sentimen positif. Penelitian ini mengintegrasikan prediksi penjualan dan analisis sentimen dalam sistem berbasis web menggunakan Streamlit sebagai pendukung pengambilan keputusan berbasis data bagi UMKM Pawonkoe.

Kata Kunci: *XGBOOST*, *Multilingual BERT (MBERT)*, Prediksi Penjualan, Analisis Sentimen, *Machine learning*

ABSTRACT

Student Name / NPM : Selena Nurmanina Afandy / 21083010076
Undergraduate thesis title : *Implementation Of XGBOOST For Pawonkoe Product Sales Prediction And Customer Review Analysis Using MBERT*
Advisors : 1. Kartika Maulida Hindrayani, S.Kom., M.Kom
2. Aviolla Terza Damaliana, S.Si, M.Stat

Daily sales fluctuations and dynamic customer perceptions became major challenges in managing Pawonkoe MSME in Banyuwangi during the January–December 2024 period. Significant sales variations may lead to inefficiencies in production planning if not supported by an accurate Forecasting system. In addition, 365 customer reviews were available but had not been optimally utilized as a basis for strategic evaluation. This study employed 366 daily sales records and 365 customer reviews using an integrated machine learning approach. The sales Forecasting model was developed using the XGBOOST algorithm with Lag features, a seven-day rolling mean, and day-of-week variables, with an 80% Training and 20% testing data split. The evaluation results showed a Mean Absolute Percentage Error (MAPE) of 20.44% in the 1-step supervised scheme and an average MAPE of 19.66% in Recursive 7-step Forecasting as the best numerical result. However, in the rolling autoregressive scheme, which represents real-world Forecasting conditions, a MAPE of 46.27% was obtained, indicating increased error due to accumulated sequential predictions. Sentiment analysis using Multilingual BERT (MBERT) achieved an accuracy of 0.91, with a distribution of 245 Positive sentiments (67.1%), 117 Neutral (32.1%), and 3 Negative (0.8%). These results indicate that customer perception is predominantly Positive. This study integrates sales Forecasting and sentiment analysis into a web-based system using Streamlit to support data-driven decision-making for Pawonkoe MSME.

Keywords: *XGBOOST, Multilingual BERT (MBERT), Sales Forecasting, Sentiment Analysis, Machine learning*

KATA PENGANTAR

Puji syukur kehadirat Allah SWT atas segala rahmat, hidayah dan karunia-Nya kepada penulis sehingga proposal skripsi dengan judul **“Implementasi Metode *XGBOOST* Dalam Prediksi Penjualan Produk Pawonkoe Dan Analisis Ulasan Pelanggan Menggunakan Metode *MBERT*”** dapat terselesaikan dengan baik.

Penulis mengucapkan terima kasih kepada Kartika Maulida Hindrayani, S.Kom., M.Kom selaku Dosen Pembimbing I dan Aviolla Terza Damaliana, S.Si, M.Stat selaku Dosen Pembimbing 2 yang telah meluangkan waktunya untuk memberikan bimbingan, nasihat serta motivasi kepada penulis. Selain itu, selama penyusunan proposal skripsi penulis juga banyak menerima bantuan dari berbagai pihak. Untuk itu penulis mengucapkan terima kasih kepada:

1. Kepada kedua orang tua tercinta serta keluarga besar, penulis mengucapkan terima kasih yang sebesar-besarnya atas doa, kasih sayang, serta dukungan moral dan materi yang senantiasa mengiringi setiap langkah penulis.
2. Ibu Prof. Dr. Ir. Novirina Hendrasarie, M.T., selaku Dekan Fakultas Ilmu Komputer UPN “Veteran” Jawa Timur.
3. Bapak Dr. Eng. Ir. Dwi Arman Prasetya, S.T., M.T., IPU., Asean, Eng., selaku Koordinator Prodi Sains Data UPN “Veteran” Jawa Timur.
4. Kepada Dosen – Dosen Prodi Sains Data.
5. Rekan penulis dengan Nim 202110120311078, penulis menyampaikan terima kasih yang sedalam-dalamnya atas bantuan dan perhatian yang diberikan selama proses penyusunan skripsi ini. Terima kasih telah menjadi tempat penulis berkeluh, bersedia mendengarkan saat penulis merasa lelah dan bersedih, bahkan ketika penulis berada pada titik ingin menyerah. Kehadiranmu melalui waktu yang diluangkan, kesabaran dalam menemani diskusi, serta dukungan yang tenang namun konsisten membantu penulis menata kembali pikiran ketika semuanya terasa berat. Terima kasih pula telah membantu penulis bagian-bagian penting, memberi masukan, dan menguatkan tanpa menghakimi. Bantuan yang diberikan bukan hanya memudahkan secara teknis,

tetapi juga menjadi penopang emosional yang berarti, sehingga penulis mampu bertahan dan menyelesaikan skripsi ini hingga akhir.

6. Rekan - rekan penulis Mareta Putri Wardhana, Wenny Maria Tampubolon, Anissa Andiar Bhalqis, serta Naomi Dwi Anggraini, yang telah menemani penyusunan skripsi melalui dukungan, diskusi, dan semangat yang konsisten.
7. Teman - teman terbaik penulis yang penulis sebut sebagai tim 073, 072, 068, dan 82, terima kasih atas kebersamaan, dukungan, serta energi positif yang menguatkan penulis selama menjalani perkuliahan hingga tahap penyelesaian skripsi ini.
8. penulis ingin menyampaikan terima kasih kepada diri sendiri yang telah bertahan hingga titik ini. Terima kasih karena tetap berusaha melangkah meskipun sering merasa lelah, ragu, bahkan ingin menyerah. Terima kasih telah melalui hari-hari yang tidak selalu mudah, menghadapi tekanan, kebingungan, dan berbagai proses panjang dengan tetap berusaha bangkit setiap kali terjatuh. Setiap usaha kecil, setiap malam yang dihabiskan untuk berpikir dan memperbaiki, serta setiap doa yang dipanjatkan akhirnya membawa penulis sampai pada tahap ini. Skripsi ini bukan hanya tentang hasil akhir, tetapi juga tentang perjalanan panjang yang penuh pembelajaran. Semoga langkah ini menjadi pengingat bahwa segala usaha, ketekunan, dan kesabaran tidak pernah sia-sia.
9. Last but not least, I wanna thank me. I wanna thank me for believing in me, I wanna thank me for doing all this hard work. I wanna thank me for having no days off. I wanna thank me for never quitting.

Penulis menyadari bahwa di dalam penyusunan skripsi ini banyak terdapat kekurangan. Untuk itu kritik dan saran yang membangun dari semua pihak sangat diharapkan demi kesempurnaan penulisan skripsi ini. Akhirnya, dengan segala keterbatasan yang penulis miliki semoga laporan ini dapat bermanfaat bagi semua pihak umumnya dan penulis pada khususnya.

Surabaya, 24 Februari 2026

Penulis

DAFTAR ISI

LEMBAR PENGESAHAN	ii
LEMBAR PERSETUJUAN	iii
SURAT PERNYATAAN BEBAS PLAGIASI.....	iv
ABSTRAK	v
<i>ABSTRACT</i>	vi
KATA PENGANTAR.....	vii
DAFTAR ISI.....	viii
DAFTAR GAMBAR.....	xi
DAFTAR TABEL	xii
DAFTAR LAMPIRAN	xii
DAFTAR NOTASI.....	xiv
BAB I PENDAHULUAN.....	1
1.1. Latar Belakang	1
1.2. Rumusan Masalah	4
1.3. Batasan Masalah.....	4
1.4. Tujuan Penelitian	5
1.5. Manfaat Penelitian	6
BAB II TINJAUAN PUSTAKA.....	9
2.1. Penelitian Terdahulu	9
2.2. Kerangka Teori.....	14
2.2.1 <i>Arsitektur Extreme Gradient Boosting (XGBOOST)</i>	14
2.2.2 <i>Bidirectional Encoder Representations from Transformers (BERT)</i> ...	24
2.2.3 Evaluasi Model	28
2.2.4 <i>Streamlit</i>	30
BAB III METODOLOGI PENELITIAN	33
3.1. Variabel Penelitian dan Sumber Data	33
3.1.1. Variabel prediksi	33
3.1.2. Variabel sentimen.....	34
3.2. Langkah Analisis.....	34

3.2.1. Langkah Analisis Penjualan	37
3.2.2. Langkah Analisis Sentimen	41
3.3. Desain Sistem	45
3.3.1. Fitur <i>Input</i> dan Pengelolaan Data	46
3.3.2. Fitur Pemrosesan Otomatis Berbasis Program	46
3.3.3. Fitur Analisis Gabungan	47
3.3.4. Fitur <i>Dashboard Monitoring</i>	47
3.3.5. Fitur Pendukung Pengambilan Keputusan	47
3.3.6. Kegunaan bagi Pengguna (Manager/Pemilik UMKM)	47
BAB IV HASIL DAN PEMBAHASAN	49
4.1. Hasil Penelitian	49
4.1.1. Hasil Analisis Prediksi Penjualan	49
4.1.1.1. <i>Load Data</i>	50
4.1.1.2. <i>Preprocessing data</i>	50
4.1.1.3. <i>Exploratory Data Analysis (EDA)</i>	56
4.1.1.4. Statistika deskriptif	57
4.1.1.5. Modeling Data dan Evaluasi	64
4.1.1.6. Pembagian Data	64
4.1.1.7. <i>Training Model XGBOOSTRegressor</i>	66
4.1.1.8. Evaluasi Model	68
4.1.1.9. Prediksi Masa Depan (<i>Time series</i>)	73
4.1.2. Hasil Analisis Sentimen Pelanggan	78
4.1.2.1. <i>Load data</i>	78
4.1.2.2. Data Pre-Processing	79
4.1.2.3. <i>Missing Value</i>	79
4.1.2.4. <i>Handling Missing Value</i>	80
4.1.2.5. <i>Text Cleaning</i> dan Tokenisasi Awal	81
4.1.2.6. <i>Feature engineering</i>	83
4.1.2.7. <i>Smart Sentiment Labeling</i>	87
4.1.2.8. <i>Confidence Score</i>	92
4.1.2.9. Persiapan Model MBERT	94

4.1.2.10. Proses <i>Training</i> dan Optimisasi.....	96
4.1.2.11. Validasi pada Data <i>Test</i>	97
4.1.2.12. Evaluasi Dan Visualisasi Model MBERT	99
4.2. Tampilan antar muka (Streamlit)	109
BAB V PENUTUP	121
5.1. Kesimpulan	121
5.2. Saran Pengembangan	123
DAFTAR PUSTAKA	125
LAMPIRAN.....	129

DAFTAR GAMBAR

Gambar 3.1 Diagram alir analisis	34
Gambar 3.2 Langkah Analisis Penjualan	37
Gambar 3.3 Langkah Analisis Sentimen	41
Gambar 3.4 Alir Desain Sistem.....	46
Gambar 4.2 <i>Time series</i> Total Penjualan Setelah <i>Preprocessing</i>	57
Gambar 4.3 Distribusi Total Penjualan	60
Gambar 4.4 Hasil <i>Forecasting</i> 7 hari	77
Gambar 4.6 <i>Word Cloud</i> Positif	102
Gambar 4.7 <i>Word Cloud</i> Negatif.....	104
Gambar 4.8 <i>Word Cloud</i> Netral.....	106
Gambar 4.9 Halaman depan <i>Dashboard</i>	111
Gambar 4.10 Fitur <i>Input</i> Data Penjualan Dan <i>Sentiment</i>	113
Gambar 4.11 Halaman grafik prediksi penjualan dan <i>Forecasting</i> 7 hari.....	116
Gambar 4.12 Fitur analisis Sentimen	119

DAFTAR TABEL

Tabel 2.1 Penelitian Terdahulu	9
Tabel 2.2 kriteria MAPE.....	23
Tabel 3.1 Data Penjualan Produk Pawonkoe.....	33
Tabel 3.2 Data ulasan pelanggan	34
Tabel 4.1 Hasil <i>Load Data</i> Penjualan	50
Tabel 4.2 Hasil <i>Missing Value</i>	51
Tabel 4.3 Daftar <i>Lag features</i>	53
Tabel 4.4 Hasil <i>Feature engineering</i> Untuk Pemodelan	54
Tabel 4.5 Statistika deskriptif Total Penjualan.....	59
Tabel 4.6 Tabel Statistik Penjualan per Hari dalam Seminggu	61
Tabel 4.7 Tabel Statistik Penjualan Bulanan Pada Tahun 2024	62
Tabel 4.8 Tabel Korelasi Fitur	63
Tabel 4.9 Tabel hasil Pembagian Data	66
Tabel 4.10 Hasil Evaluasi Model 1-Step (t+1) pada Data Uji	69
Tabel 4.11 Hasil Evaluasi Model <i>Recursive 7-Step</i> per <i>Horizon</i>	70
Tabel 4.12 Hasil Evaluasi 1-Step <i>rolling autoregressive</i>	71
Tabel 4.13 Rata-rata Evaluasi Model <i>Recursive 7-Step (Horizon 1–7)</i>	71
Tabel 4.14 Ringkasan Perbandingan Evaluasi.....	72
Tabel 4.15 Hasil Prediksi Penjualan 7 Hari ke Depan.....	74
Tabel 4.16 <i>Load data sentiment</i>	78
Tabel 4.17 <i>feature engineering</i> Sentimen Berdasarkan <i>Rating</i> Pelanggan	84
Tabel 4.19 Distribusi Sentimen Awal (<i>Auto Labeling</i>)	88
Tabel 4.20 Hasil Manual <i>Review</i>	89
Tabel 4.21 Perbandingan Sebelum dan Sesudah <i>Re-Labeling</i>	89
Tabel 4.22 Hasil <i>Relabeling</i> Sentimen Berdasarkan Skor.....	91
Tabel 4.23 Hasil analisis <i>confidence score</i>	93
Tabel 4.24 Distribusi kelas	94
Tabel 4.25 Hasil <i>Split Data</i>	95
Tabel 4.26 Hasil Performa setiap <i>Epoch</i>	97
Tabel 4.27 <i>Classification Report</i>	100

DAFTAR LAMPIRAN

Lampiran 1. Hasil uji plagiasi	129
Lampiran 2. Bukti sentimen	130

DAFTAR NOTASI

I	: Intensitas
W_R	: <i>Weight factor</i>
H	: <i>Hue</i>
S	: <i>Saturation</i>
V	: <i>Value</i>
dst	: Gambar akumulator
scr	: Gambar <i>input</i>
F	: <i>Foreground</i>
B	: <i>Background</i>
f	: <i>Frame</i>
SE	: <i>Structuring element</i>
ECD	: <i>Equivalent circular diameter</i>
b	: Bias
$f(x)m$	: Fungsi vektor masukan
$d(x', x)$	: Jarak di antara data uji z ke setiap vector data latih
$K(x, y)$	: Fungsi kernel linear
$G(x)$	: Fungsi gaussian satu dimensi
σ	: Standard deviasi dari distribusi
$G(x, y)$	: Fungsi gaussian dua dimensi