

# **PREDIKSI HARGA BAHAN BANGUNAN DENGAN MENGUNAKAN GRADIENT BOOSTING REGRESSOR**

**SKRIPSI**

**OLEH:**

**ABDUL LATIF**

**218160019**



**PROGRAM STUDI TEKNIK INFORMATIKA  
FAKULTAS TEKNIK  
UNIVERSITAS MEDAN AREA  
MEDAN  
2026**

**UNIVERSITAS MEDAN AREA**

© Hak Cipta Di Lindungi Undang-Undang

1. Dilarang Mengutip sebagian atau seluruh dokumen ini tanpa mencantumkan sumber
2. Pengutipan hanya untuk keperluan pendidikan, penelitian dan penulisan karya ilmiah
3. Dilarang memperbanyak sebagian atau seluruh karya ini dalam bentuk apapun tanpa izin Universitas Medan Area

Document Accepted 1/7/26

Access From ([repository.uma.ac.id](http://repository.uma.ac.id))1/7/26

# **PREDIKSI HARGA BAHAN BANGUNAN DENGAN MENGUNAKAN GRADIENT BOOSTING REGRESSOR**

## **SKRIPSI**

**Diajukan sebagai Salah Satu Syarat untuk Memperoleh  
Gelar Sarjana di Fakultas Teknik  
Universitas Medan Area**



**Oleh:  
ABDUL LATIF  
218160019**

**PROGRAM STUDI TEKNIK INFORMATIKA  
FAKULTAS TEKNIK  
UNIVERSITAS MEDAN AREA  
MEDAN  
2026**

UNIVERSITAS MEDAN AREA

© Hak Cipta Di Lindungi Undang-Undang

1. Dilarang Mengutip sebagian atau seluruh dokumen ini tanpa mencantumkan sumber
2. Pengutipan hanya untuk keperluan pendidikan, penelitian dan penulisan karya ilmiah
3. Dilarang memperbanyak sebagian atau seluruh karya ini dalam bentuk apapun tanpa izin Universitas Medan Area

Document Accepted 1/7/26

Access From (repository.uma.ac.id)1/7/26

## HALAMAN PENGESAHAN

Judul Skripsi : Prediksi Harga Bahan Bangunan Menggunakan Gradient Boosting Regressor

Nama : Abdul Latif  
NPM : 218160019  
Fakultas : Teknik

Disetujui Oleh  
Komisi Pembimbing



Andre Hasudungan Lubis S.Ti, M.Sc  
Pembimbing I



Tanggal Lulus: 12 Maret 2026



## HALAMAN PERTANYAAN

Saya menyatakan bahwa skripsi yang saya susun, sebagai syarat memperoleh gelar sarjana merupakan hasil karya tulis saya sendiri. Adapun bagian-bagian tertentu dalam penulisan skripsi ini yang saya kutip dari hasil karya orang lain telah dituliskan sumbernya secara jelas sesuai dengan norma, kaidah, dan etika penulisan ilmiah.

Saya bersedia menerima sanksi pencabutan gelar akademik yang saya peroleh dan sanksi-sanksi lainnya dengan peraturan yang berlaku, apabila di kemudian hari ditemukan adanya plagiat dalam skripsi ini.

Medan, 12 Maret 2026



Abdul Latif  
218160019

**HALAMAN PERNYATAAN PERSETUJUAN PUBLIKASI  
TUGAS AKHIR/SKRIPSI/TESIS UNTUK KEPENTINGAN AKADEMIS**

Sebagai sivitas akademik Universitas Medan Area, saya yang bertanda tangan di bawah ini:

Nama : Abdul Latif  
NPM : 218160019  
Program Studi : Teknik Informatika  
Fakultas : Teknik  
Jenis Karya : Skripsi

Demi pengembangan ilmu pengetahuan, menyetujui untuk memberikan kepada Universitas Medan Area **Hak Bebas Royalti Noneksklusif (*Non-exclusive Royalty-Free Right*)** atas karya ilmiah saya yang berjudul :

**Prediksi Harga Bahan Bangunan Menggunakan Gradient Boosting Regressor**

Beserta perangkat yang ada (jika diperlukan). Dengan Hak Bebas Royalti Noneksklusif ini Universitas Medan Area berhak menyimpan, mengalihmedia/format-kan, mengelola dalam bentuk pangkalan data (*database*), merawat, dan memublikasikan Skripsi saya selama tetap mencantumkan nama saya sebagai penulis/pencipta dan sebagai pemilik Hak Cipta.

Demikian pernyataan ini saya buat dengan sebenarnya.

Dibuat di : Medan  
Pada tanggal : 12 Maret 2026  
Yang menyatakan



(Abdul Latif)

## RIWAYAT HIDUP

Abdul Latif lahir pada tanggal 01 Desember 2003 di Desa Sei Rotan Tembung , Kecamatan Percut Sei Tuan, Kabupaten Deli Serdang, Provinsi Sumatera Utara dari Ayah Alm. Sariman dan Ibunda Sumiati yang luar biasa hebat. Penulis merupakan anak terakhir dari delapan bersaudara. Pada tahun 2021 penulis lulus dari SMKN 1 Percut Sei Tuan kemudian mendaftar pada tahun 2021 menjadi mahasiswa di Fakultas Teknik Universitas Medan Area dengan jurusan Teknik Informatika.



## KATA PENGANTAR

Dengan penuh rasa syukur kepada Tuhan Yang Maha Esa atas berkat dan rahmat-Nya, penulis akhirnya dapat menyelesaikan skripsi dengan judul “Prediksi Harga Bahan Bangunan Menggunakan Gradient Boosting Regressor”. Skripsi ini disusun sebagai salah satu syarat untuk menyelesaikan pendidikan Program Sarjana pada Fakultas Teknik Program Studi Teknik Informatika di Universitas Medan Area. Penulis menyadari bahwa meskipun skripsi ini telah disusun dengan sebaik – baiknya, masih terdapat banyak kekurangan dan jauh dari kata kesempurnaan. Oleh karena itu, penulis sangat mengharapkan kritik dan saran yang bersifat konstruktif dari para pembaca untuk perbaikan di masa yang akan datang. Penulis juga ingin menyampaikan rasa terima kasih yang sebesar – besarnya kepada :

1. Allah SWT. yang selalu memberikan kenikmatan dan kesehatan sehingga penulis masih bisa menjalankan hidup dengan baik dan selalu memberikan penulis perlindungan dan kekuatan.
2. Yang sangat istimewa kedua orang tua penulis, Ayahanda Alm.Sariman dan ibunda Sumiati yang selalu memenuhi kebutuhan penulis mendukung, dan mendoakan penulis sehingga penulis menyelesaikan pendidikan hingga perguruan tinggi.
3. Bapak Prof. Dr. Dadan Ramdan, M.Eng., M.Sc., selaku Rektor Universitas Medan Area.
4. Bapak Dr. Eng. Supriatno, S.T., M.T. selaku Dekan Fakultas Teknik Universitas Medan Area.
5. Bapak Andre Hasudungan Lubis S.Ti.,M.Sc, selaku Dosen Pembimbing penulis yang mengarahkan penulis dalam proses menyelesaikan skripsi ini.
6. Bapak Dr. Dian Noviandri S.T Selaku Ketua Panitia kemudian Ibu Susilawati, S.Kom., M.Kom, Selaku Sekretaris Panitia, Dan Bapak Muhathir, S.T.,M.Kom., Selaku Penguji Sidang Skripsi dalam pelaksanaan pengujian penelitian penulis, yang memberikan saran dan masukan kepada penulis.

7. Seluruh Dosen Teknik Informatika Universitas Medan Area yang telah membekali penulis dengan ilmu yang sangat bermanfaat.
8. Bang Robby Kurniawan Sari Damanik, ST., selaku IT Support Teknik Informatika yang selalu mengarahkan penulis dalam administrasi.
9. Kepada saudara kandung penulis yang turut selalu mendukung segala kegiatan penulis hingga berhasil menyelesaikan pendidikan ini.
10. Kepada teman teman terdekat penulis yang selalu merayakan setiap pencapaian penulis sampai penulis bisa menyelesaikan pendidikan ini dengan semangat yang tinggi. Last but not least, terkhusus dan paling tersayang Bambang Saktiawan S.kom, Mhd. Ramadhan Saputra S.kom, Evindo Amanda S.kom, Khairul Amali S.kom, Mhd. Imam Ali On The Way S.kom, Lutfi Chaidar Ikhrum yang sampai sekarang lost cantac.

Medan, 12 Maret 2026



Abdul Latif

218160019

## ABSTRAK

Harga bahan bangunan merupakan salah satu faktor penting yang memengaruhi perencanaan biaya konstruksi. Penelitian ini bertujuan membangun model prediksi harga bahan bangunan menggunakan algoritma Gradient Boosting Regressor berdasarkan atribut jenis bahan, merek, jumlah stok, lokasi distribusi, dan harga jual. Data penelitian terdiri dari 1000 catatan transaksi yang diperoleh dari Toko Bangunan Risky Mandiri di Medan pada tahun 2025. Proses penelitian diawali dengan pengumpulan data, dilanjutkan dengan tahap praproses berupa pembersihan data, pengkodean kategori, dan normalisasi. Analisis data eksploratori dilakukan melalui statistik deskriptif dan visualisasi untuk memahami distribusi serta hubungan antar variabel. Model Gradient Boosting Regressor kemudian diterapkan dengan pembagian data menjadi 80% data latih dan 20% data uji. Hasil penelitian menunjukkan bahwa model menghasilkan nilai koefisien determinasi ( $R^2$ ) sebesar 0,945 pada data latih dan 0,918 pada data uji, dengan tingkat kesalahan prediksi yang relatif kecil. Temuan ini membuktikan bahwa Gradient Boosting Regressor mampu memberikan performa prediksi yang akurat terhadap harga bahan bangunan. Penelitian ini diharapkan dapat menjadi dasar dalam pengambilan keputusan terkait pengadaan dan perencanaan biaya konstruksi agar lebih efisien dan tepat sasaran.

**Kata Kunci: Harga Bahan Bangunan, Prediksi Harga, Analisis Regresi, Data Eksploratori, Gradient Boosting Regressor.**

## ABSTRACT

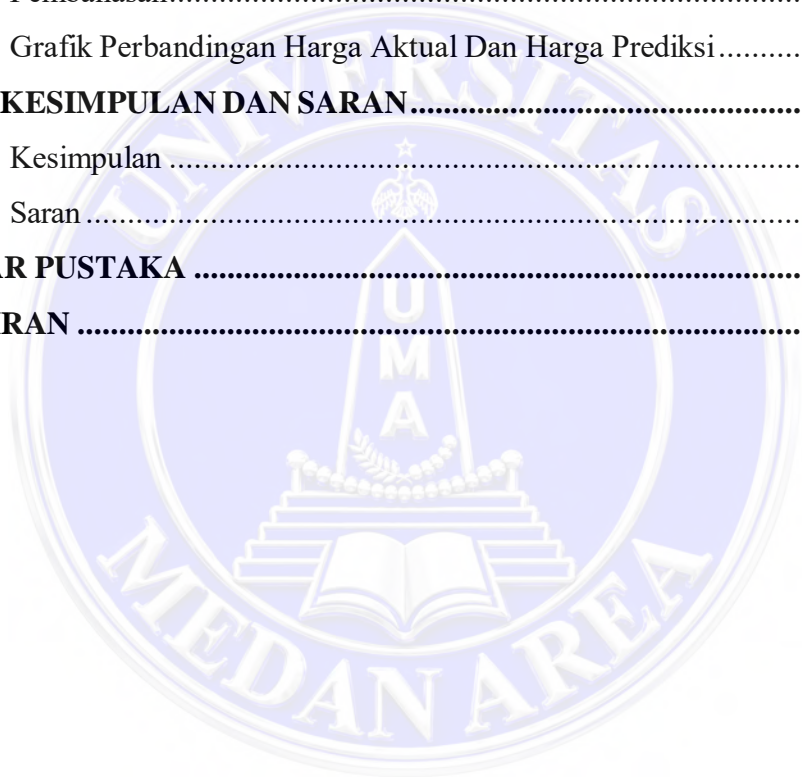
*The price of building materials is one of the crucial factors influencing construction cost planning. This study aims to develop a predictive model for building material prices using the Gradient Boosting Regressor algorithm, based on attributes such as material type, brand, stock availability, distribution location, and selling price. The dataset consists of 1,000 transaction records collected from Toko Bangunan Risky Mandiri in Medan in 2025. The research process began with data collection, followed by preprocessing steps including data cleaning, categorical encoding, and normalization. Exploratory Data Analysis (EDA) was conducted through descriptive statistics and visualization to examine data distribution and relationships between variables. The Gradient Boosting Regressor model was then applied by splitting the dataset into 80% training data and 20% testing data. The results show that the model achieved a coefficient of determination ( $R^2$ ) of 0.945 on the training set and 0.918 on the testing set, with relatively small prediction errors. These findings demonstrate that the Gradient Boosting Regressor provides accurate performance in predicting building material prices. This study is expected to serve as a foundation for decision-making in procurement and construction cost planning, enabling more efficient and targeted resource allocation.*

**Keywords:** *Building Material Prices, Price Prediction, Regression Analysis, Exploratory Data, Gradient Boosting Regressor.*

## DAFTAR ISI

<b>HALAMAN PENGESAHAN .....</b>	<b>i</b>
<b>HALAMAN PERTANYAAN .....</b>	<b>ii</b>
<b>HALAMAN PERNYATAAN PERSETUJUAN PUBLIKASI .....</b>	<b>iii</b>
<b>RIWAYAT HIDUP .....</b>	<b>iv</b>
<b>KATA PENGANTAR .....</b>	<b>v</b>
<b>ABSTRAK .....</b>	<b>vii</b>
<b>ABSTRACT .....</b>	<b>viii</b>
<b>DAFTAR ISI .....</b>	<b>ix</b>
<b>DAFTAR TABEL .....</b>	<b>xii</b>
<b>DAFTAR GAMBAR .....</b>	<b>xiii</b>
<b>BAB I PENDAHULUAN .....</b>	<b>14</b>
1.1. Latar Belakang .....	14
1.2. Rumusan Masalah .....	15
1.3. Batasan Masalah .....	15
1.4. Tujuan Penelitian .....	16
1.5. Manfaat Penelitian .....	16
<b>BAB II TINJAUAN PUSTAKA .....</b>	<b>18</b>
2.1. Harga Bahan Bangunan .....	18
2.2. Regresi Machine Learning .....	20
2.2.1. Evaluasi Performa Regresi .....	20
2.3. Gradient Boosting Regressor .....	21
2.3.1. Flowchart Algoritma Gradient Boosting Regressor .....	22
2.5. Penelitian Terdahulu .....	24
<b>BAB III METODOLOGI PENELITIAN .....</b>	<b>27</b>
3.1. Tahapan Penelitian .....	27
3.1.1. Pengumpulan Data .....	28
3.1.2. Prapemrosesan Data .....	18
3.1.3. EDA(Exploratory Data Analysis) .....	18
3.1.4. Proses Implementasi .....	19
3.1.5. Evaluasi .....	19

3.2. Perhitungan Manual .....	22
<b>BAB IV.....</b>	<b>25</b>
4.1. Karakteristik Data EDA (Exploratory Data Analysis) .....	25
4.1.1. Distribusi Jenis Bahan Bangunan .....	25
4.1.2. Boxplot Per Jenis Bahan.....	26
4.1.3. Heatmap Korelasi Antar Variabel Numerik .....	26
4.1.4. Histogram Distribusi Stok Bahan Bangunan .....	27
4.1.5. Histogram Distribusi Harga Bahan Bangunan .....	28
4.2. Implementasi .....	29
4.3. Pembahasan.....	30
4.4. Grafik Perbandingan Harga Aktual Dan Harga Prediksi.....	30
<b>BAB V KESIMPULAN DAN SARAN.....</b>	<b>32</b>
5.1. Kesimpulan .....	32
5.2. Saran .....	32
<b>DAFTAR PUSTAKA .....</b>	<b>34</b>
<b>LAMPIRAN .....</b>	<b>38</b>



## DAFTAR TABEL

	Halaman
Tabel 2.1. Penelitian Terdahulu .....	13
Tabel 3.1. Rangkuman Data Primer Penelitian .....	17
Tabel 4.1. Statistik Deskriptif Atribut Numerik.....	25
Tabel 4.2. Distribusi Jenis Bahan Bangunan .....	26
Tabel 4.3. Hasil Evaluasi Model Gradient Boosting Regressor.....	30
Tabel 4.4. Perbandingan Harga Aktual Dan Prediksi.....	31



## DAFTAR GAMBAR

	Halaman
Gambar 2.1. <i>Flowchart</i> .....	11
Gambar 3.1. Tahapan Penelitian .....	15
Gambar 4.1. Boxplot Per Jenis Bahan.....	27
Gambar 4.2. Heatmap Korelasi Antar Variabel Numerik.....	27
Gambar 4.3. Histogram Distribusi Stok Bahan Bangunan .....	28
Gambar 4.4. Histogram Distribusi Harga Bahan Bangunan.....	29
Gambar 4.5. Grafik Perbandingan Harga Aktual Dan Harga Prediksi.....	32



# BAB I

## PENDAHULUAN

### 1.1. Latar Belakang

Harga bahan bangunan merupakan salah satu komponen utama yang mempengaruhi besarnya biaya konstruksi. Fluktuasi harga yang terjadi di pasar dipengaruhi oleh berbagai faktor, seperti jenis material, merek, ketersediaan stok, lokasi distribusi, biaya logistik, serta dinamika permintaan dan penawaran. Ketidakpastian harga ini dapat mengakibatkan penyusunan anggaran menjadi kurang akurat, memengaruhi efisiensi pelaksanaan proyek, serta menimbulkan risiko pembengkakan biaya (*cost overrun*) (Sari & Prabowo, 2025). Metode estimasi harga yang dilakukan secara konvensional, baik melalui perkiraan berdasarkan pengalaman maupun analisis historis sederhana, memiliki keterbatasan dalam menangkap pola hubungan yang kompleks antar variabel. Pendekatan tersebut memiliki keterbatasan dalam penerapannya pada data berukuran besar dengan hubungan yang bersifat linier. Oleh karena itu, dibutuhkan metode prediksi yang mampu mengolah data secara sistematis, mengidentifikasi keterkaitan antar atribut, serta memberikan hasil estimasi dengan tingkat akurasi yang tinggi (Andriyani dkk., 2024). Dalam beberapa tahun terakhir, pendekatan berbasis kecerdasan buatan dan pembelajaran mesin telah semakin digunakan untuk memprediksi harga dan biaya konstruksi dengan cara yang lebih adaptif dan presisi. Sebagai contoh, penelitian “*Applying Hybrid Machine Learning for Construction Material Price Prediction and Procurement Cost Optimization*” (Gebregiorgis Atnafie, 2025) menyajikan metode hibrida yang mengombinasikan model multilayer perceptron dan jaringan penguatan (*Deep Q Network*) untuk memprediksi harga material sekaligus menentukan waktu pengadaan secara ekonomis. Hasilnya menunjukkan bahwa model prediksi berbasis pembelajaran mesin lebih unggul dibanding metode regresi linier tradisional dalam menangani fluktuasi harga konstruksi. Penelitian terdahulu menunjukkan efektivitas Gradient Boosting dalam berbagai konteks prediksi harga. Sari & Prabowo (2023) membandingkan kinerja Random Forest, Gradient Boosting, dan XGBoost untuk

prediksi harga rumah, dan menemukan bahwa Gradient Boosting memberikan hasil kompetitif dengan tingkat kesalahan rendah. Andriyani et al. (2024) juga berhasil memanfaatkan Gradient Boosted Trees Regression untuk memprediksi harga beras premium dengan nilai R-Squared sebesar 0,9047 dan RMSE sebesar 0,0473, menandakan tingkat akurasi yang tinggi. Berdasarkan temuan-temuan tersebut, penerapan Gradient Boosting Regressor pada prediksi harga bahan bangunan diharapkan dapat menghasilkan model yang akurat dan reliabel. Model ini berpotensi menjadi alat bantu dalam pengambilan keputusan bagi kontraktor, pengembang, maupun konsumen, sehingga perencanaan biaya dapat dilakukan secara lebih tepat dan efisien.

## 1.2. Rumusan Masalah

Bagaimana memprediksi harga bahan bangunan dengan menggunakan model Gradient Boosting Regressor?

## 1.3. Batasan Masalah

Batasan masalah diperlukan untuk memperjelas ruang lingkup penelitian dan menjaga agar pembahasan tetap fokus serta tidak menyimpang dari tujuan utama penelitian. Adapun batasan dalam penelitian ini adalah sebagai berikut:

1. Penelitian ini menggunakan data primer berupa data bahan bangunan yang dikumpulkan secara langsung dari sumber distribusi atau toko bahan bangunan di Kota Medan, dengan atribut yang meliputi jenis material, merek, stok, dan lokasi distribusi.
2. Faktor eksternal seperti kondisi ekonomi makro, kebijakan pemerintah, dan dinamika pasar yang lebih luas tidak dimasukkan dalam analisis prediksi harga.
3. Model prediksi yang dikembangkan hanya menggunakan algoritma Gradient Boosting Regressor.
4. Data primer yang digunakan telah melewati tahap prapemrosesan meliputi pembersihan, normalisasi, dan encoding, serta dibatasi pada rentang waktu pengumpulan tertentu.

5. Prediksi harga bahan bangunan difokuskan pada harga jual di tingkat distributor atau toko bahan bangunan, tidak mencakup harga jual ke konsumen akhir.
6. Evaluasi performa model dilakukan menggunakan metrik kesalahan seperti *Mean Absolute Percentage Error* (MAPE), *Mean Absolute Error* (MAE), dan *Root Mean Squared Error* (RMSE), tanpa menggunakan koefisien determinasi ( $R^2$ ).

#### 1.4. Tujuan Penelitian

Penelitian ini memiliki tujuan yaitu:

1. Memprediksi harga bahan bangunan menggunakan algoritma Gradient Boosting Regressor berbasis data primer yang dikumpulkan langsung dari sumber distribusi atau toko bahan bangunan.
2. Mengidentifikasi pengaruh atribut seperti jenis material, merek, stok, dan lokasi distribusi terhadap hasil prediksi harga bahan bangunan.
3. Mengevaluasi kinerja model Gradient Boosting Regressor dalam memprediksi harga bahan bangunan menggunakan metrik *Mean Absolute Percentage Error* (MAPE), *Mean Absolute Error* (MAE), dan *Root Mean Squared Error* (RMSE).
4. Menyediakan model prediksi harga yang dapat digunakan sebagai alat bantu pengambilan keputusan dalam perencanaan biaya konstruksi secara lebih tepat dan efisien.

#### 1.5. Manfaat Penelitian

Manfaat penelitian ini sebagai berikut:

1. Manfaat Teoretis

Memberikan kontribusi terhadap pengembangan ilmu pengetahuan, khususnya di bidang penerapan algoritma machine learning dalam pemodelan prediksi harga berbasis data primer. Menjadi referensi bagi penelitian selanjutnya yang membahas penerapan Gradient Boosting Regressor atau algoritma sejenis dalam konteks prediksi harga bahan bangunan.

## 2. Manfaat Praktis

Memberikan solusi berbasis data yang dapat membantu kontraktor, pengembang, dan pelaku usaha bahan bangunan dalam memperkirakan harga secara lebih akurat. Membantu proses perencanaan biaya konstruksi agar lebih tepat, efisien, dan meminimalkan risiko pembengkakan biaya. Menyediakan model prediksi yang dapat dijadikan dasar pengambilan keputusan strategis, baik dalam pengadaan bahan maupun pengaturan stok.



## **BAB II**

### **TINJAUAN PUSTAKA**

#### **2.1. Harga Bahan Bangunan**

Harga bahan bangunan merupakan nilai tukar yang harus dibayarkan oleh konsumen untuk memperoleh suatu produk bahan bangunan yang dibutuhkan dalam kegiatan konstruksi maupun renovasi. Menurut teori ekonomi, harga adalah salah satu faktor penting yang memengaruhi keputusan pembelian konsumen karena mencerminkan nilai dan daya beli masyarakat. Dalam konteks pembangunan, menurut Sari & Prabowo (2025), fluktuasi harga bahan bangunan memiliki dampak yang cukup signifikan terhadap perencanaan anggaran dan keputusan pembelian konsumen. Fluktuasi harga bahan bangunan yang terjadi di pasar berdampak langsung pada biaya pembangunan. Apabila harga bahan bangunan mengalami kenaikan, maka biaya keseluruhan pembangunan juga akan meningkat, sehingga dapat menunda pelaksanaan proyek atau mengurangi volume pembelian oleh konsumen. Sebaliknya, ketika harga bahan bangunan mengalami penurunan, hal ini mendorong peningkatan permintaan karena konsumen berusaha memanfaatkan kondisi harga yang lebih rendah untuk memenuhi kebutuhan mereka.

Naik turunnya harga bahan bangunan disebabkan oleh berbagai faktor yang saling berkaitan. Menurut hasil kajian yang terdapat dalam dokumen, faktor-faktor tersebut antara lain:

1. **Permintaan dan Penawaran**

Ketidakseimbangan antara permintaan dan penawaran menjadi penyebab utama fluktuasi harga bahan bangunan.

“Fluktuasi harga dipengaruhi oleh permintaan dan penawaran yang tidak stabil di pasar.”

2. **Biaya Produksi**

Perubahan pada biaya produksi, baik akibat kenaikan harga bahan baku maupun tenaga kerja, akan berdampak pada harga jual bahan bangunan.

“Kenaikan biaya produksi bahan bangunan akan berimbas pada kenaikan

harga yang harus dibayar oleh konsumen.”

3. Distribusi dan Transportasi

Proses distribusi yang terhambat atau kenaikan ongkos transportasi juga memengaruhi harga bahan bangunan di pasaran.

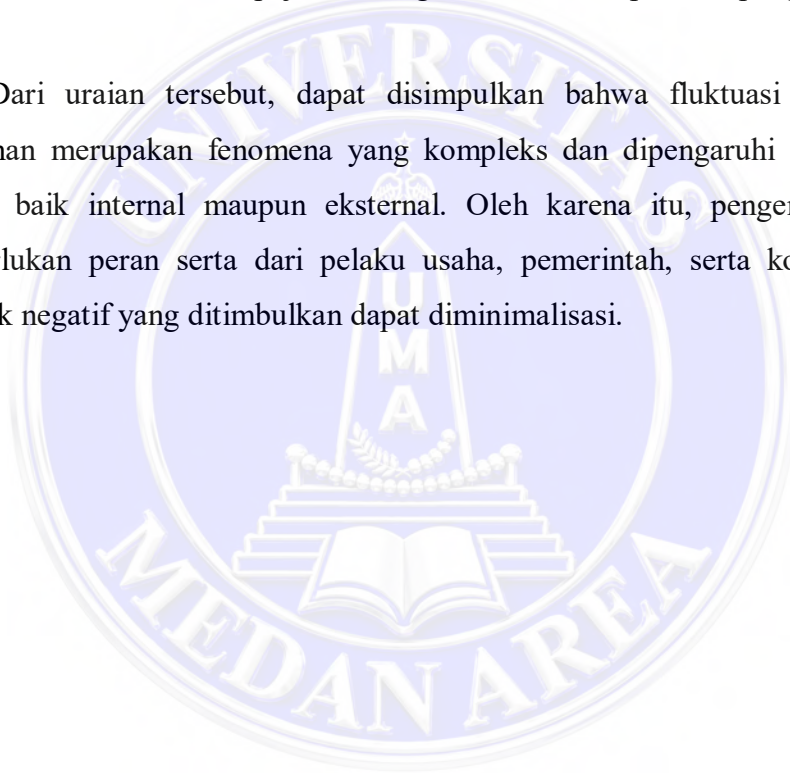
“Keterlambatan distribusi atau kenaikan ongkos transportasi juga mempengaruhi harga bahan bangunan.”

4. Kebijakan Pemerintah

Regulasi atau kebijakan pemerintah, seperti pengenaan pajak dan aturan impor, dapat memengaruhi tingkat harga.

“Pemerintah melalui pajak atau regulasi tertentu dapat mempengaruhi harga.”

Dari uraian tersebut, dapat disimpulkan bahwa fluktuasi harga bahan bangunan merupakan fenomena yang kompleks dan dipengaruhi oleh berbagai faktor, baik internal maupun eksternal. Oleh karena itu, pengendalian harga memerlukan peran serta dari pelaku usaha, pemerintah, serta konsumen agar dampak negatif yang ditimbulkan dapat diminimalisasi.



## 2.2. Regresi Machine Learning

Regresi merupakan salah satu metode statistik yang paling banyak digunakan dalam ilmu data dan pembelajaran mesin. Tujuan utama regresi adalah memodelkan hubungan antara satu variabel terikat (*dependent variable*) dengan satu atau lebih variabel bebas (*independent variables*) berdasarkan data historis (He dkk., 2019). Dalam konteks pembelajaran mesin, regresi digunakan untuk memprediksi nilai numerik berkelanjutan berdasarkan pola yang diperoleh dari data historis. Model regresi tradisional seperti Regresi Linear berasumsi bahwa hubungan antara variabel prediktor dengan variabel target bersifat linier. Namun, dalam banyak kasus dunia nyata, hubungan antarvariabel sering kali kompleks dan tidak mengikuti pola linier sederhana (Li & Science, 2022). Karakteristik utama regresi dalam pembelajaran mesin adalah kemampuannya mempelajari pola secara otomatis dari data tanpa harus mendefinisikan persamaan matematis yang kaku. Dengan demikian, regresi tidak hanya sekadar metode analisis hubungan variabel, tetapi juga alat prediksi yang dapat diterapkan pada berbagai bidang, termasuk ekonomi, kesehatan, rekayasa, maupun analisis harga komoditas.

### 2.2.1. Evaluasi Performa Regresi

Evaluasi performa regresi bertujuan menilai sejauh mana hasil prediksi mendekati nilai sebenarnya. Beberapa ukuran yang umum digunakan adalah:

1. *Mean Absolute Error (MAE)*

Mengukur rata-rata selisih absolut antara nilai aktual dan prediksi. Semakin kecil MAE, semakin akurat prediksi.

2. *Mean Squared Error (MSE)*

Rata-rata kesalahan kuadrat antara nilai aktual dan prediksi. MSE menekankan penalti lebih besar pada kesalahan yang ekstrem.

3. Root Mean Squared Error (RMSE)

Akar dari MSE, sehingga satuannya sama dengan variabel target. RMSE memudahkan interpretasi karena menunjukkan rata-rata besar kesalahan.

4. Mean Absolute Percentage Error (MAPE)

Mengukur kesalahan dalam bentuk persentase terhadap nilai aktual. MAPE banyak digunakan pada analisis harga karena mudah dipahami dalam konteks ekonomi.

5. Koefisien Determinasi ( $R^2$ )

Menunjukkan proporsi variasi nilai aktual yang dapat dijelaskan oleh model. Nilai mendekati 1 menandakan model semakin baik.

Secara keseluruhan, penggunaan lebih dari satu metrik diperlukan agar penilaian model lebih objektif. MAPE bermanfaat untuk melihat kesalahan relatif, sementara MAE, MSE, dan RMSE menggambarkan besar kesalahan absolut. Dengan demikian, kombinasi metrik ini memberikan gambaran menyeluruh mengenai keakuratan dan ketepatan model regresi (Andriyani dkk., 2024).

### 2.3. Gradient Boosting Regressor

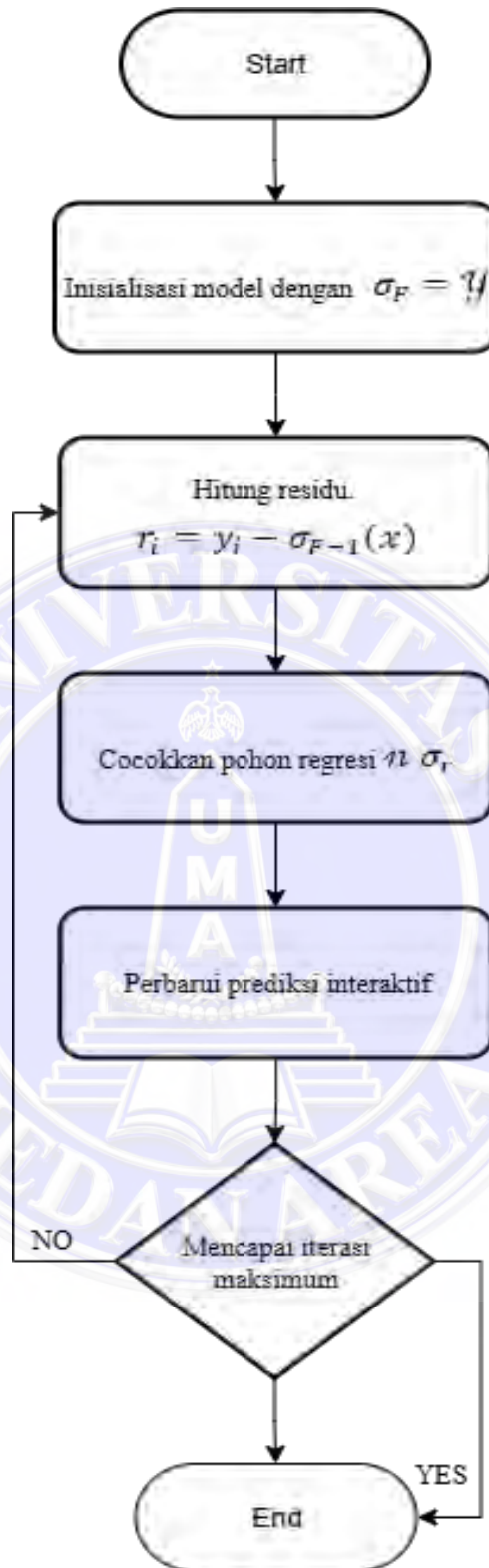
Gradient Boosting Regressor (GBR) merupakan salah satu algoritma pembelajaran mesin berbasis ensemble yang dikembangkan dari konsep boosting. Algoritma ini bekerja dengan membangun model secara bertahap melalui penggabungan sejumlah decision tree sederhana (weak learners) untuk menghasilkan model prediksi yang lebih kuat dan akurat (He dkk., 2019). Prinsip dasarnya adalah meminimalkan fungsi kesalahan (loss function) dengan pendekatan gradient descent secara iteratif, sehingga setiap model baru berfokus memperbaiki kesalahan dari model sebelumnya. Kelebihan Gradient Boosting memberikan, Akurasi tinggi menangkap pola kompleks, Fleksibel terhadap berbagai jenis fungsi loss, dan Memberikan feature importance yang memudahkan interpretasi. Adapun kelemahannya adalah rentan overfitting jika parameter tidak dikendalikan dengan baik, serta

membutuhkan waktu komputasi yang relatif lebih lama dibandingkan metode lain.

### 2.3.1. Flowchart Algorithma Gradient Bossting Regressor

Pada gambar 2.1 di bawah ini merupakan proses membangun model prediksi.





Gambar 2.1. Flowchart

Flowchart di atas menjelaskan alur berpikir algoritma Gradient Boosting Regressor dalam membangun sebuah model prediksi. Proses diawali dengan pembentukan model awal yang sangat sederhana, biasanya berupa nilai rata-rata dari data yang diteliti. Model ini masih jauh dari sempurna, namun berfungsi sebagai titik tolak untuk perbaikan selanjutnya. Langkah berikutnya adalah menghitung perbedaan antara nilai sebenarnya dengan nilai yang diperkirakan oleh model awal. Perbedaan ini menunjukkan kesalahan yang perlu diperbaiki. Dari kesalahan tersebut, dibangunlah suatu pohon keputusan sederhana yang berfungsi untuk menangkap pola yang belum mampu dijelaskan oleh model sebelumnya. Hasil dari pohon tersebut kemudian ditambahkan ke dalam model awal dengan bobot tertentu agar perbaikannya lebih terkontrol. Setelah itu, model yang diperbarui kembali dievaluasi untuk melihat apakah tingkat kesalahannya sudah cukup kecil atau masih perlu diperbaiki lagi. Jika hasil evaluasi menunjukkan bahwa kesalahan masih cukup besar, maka proses perhitungan kesalahan dan pembangunan pohon baru diulang kembali. Proses ini dilakukan berulang kali sampai tingkat kesalahan dianggap memadai atau jumlah perulangan yang telah ditentukan tercapai. Pada akhirnya, seluruh rangkaian perbaikan menghasilkan sebuah model akhir. Model ini bukan hanya sekadar hasil dari satu kali perkiraan, melainkan gabungan dari banyak tahapan perbaikan yang saling melengkapi. Dengan cara demikian, model akhir memiliki kemampuan prediksi yang jauh lebih baik dibandingkan model awal yang sederhana.

#### **2.4. EDA(*Exploratory Data Analysis*)**

Exploratory Data Analysis (EDA) adalah langkah awal penting dalam analisis data yang berperan sebagai proses menemukan pola, karakteristik, dan masalah dalam data sebelum dilakukan pemodelan. Menurut (Komorowski dkk., 2016) EDA bertujuan untuk menelaah distribusi data, mendeteksi outlier atau anomali, serta memahami korelasi antar variabel tanpa asumsi awal yang terlalu kuat (misalnya model yang sudah ditetapkan).

#### **2.5. Penelitian Terdahulu**

Berikut penelitian terdahulu yang pernah dilakukan sebelumnya yang menggunakan metode Gradient Boosting Regressor:

Tabel 2.1. Penelitian Terdahulu

Peneliti (Tahun)	Judul Penelitian	Metode	Hasil Utama	Relevansi
(Li & Science, 2022 )	<i>A Gentle Introduction to Gradient Boosting</i>	Gradient Boosting Regressor	Penjabaran matematis fit ke gradien negatif dan struktur update GBR.	Memberikan pemahaman teoretis yang mendukung implementasi model prediksi harga bahan bangunan
(Gebregiorgis Atnafie, 2025)	<i>Applying Hybrid Machine Learning for Construction Material Price Prediction and Procurement Cost Optimization</i>	Gradient Boosting	Model hibrida yang menggabungkan multilayer perceptron (MLP) dengan Deep Q Network (DQN) mampu memprediksi harga material konstruksi dengan lebih akurat dibandingkan model regresi linier tradisional.	Sama-sama berfokus pada prediksi harga bahan bangunan sebagai dasar perencanaan biaya konstruksi.
(He dkk., 2019)	<i>Gradient Boosting Machine: A Survey</i>	Gradient Boosting	Survei tentang metode Gradient Boosting, varian, dan aplikasinya di berbagai domain prediksi.	Memberikan tinjauan komprehensif tentang perkembangan algoritma Gradient Boosting.
(Sari & Prabowo, 2025)	Analisis Perbandingan Prediksi Harga Rumah Dengan Random Forest, Gradient Boosting, dan XGBoost	Gradient Boosting, XGBoost	Menemukan bahwa Gradient Boosting memiliki performa prediksi harga rumah yang kompetitif dibandingkan metode lain.	Relevan karena sama-sama memprediksi harga properti berbasis data historis.
(Andriyani dkk., 2024)	Prediksi Harga Beras Premium Tahun 2024 Menggunakan Metode Gradient Boosted Trees Regression	Gradient Boosted Trees (GBR)	Menghasilkan nilai $R^2$ sebesar 0,9047 dan RMSE 0,0473, menunjukkan akurasi tinggi pada prediksi harga komoditas.	Relevan karena penelitian ini menunjukkan efektivitas GBR dalam memprediksi harga berbasis data

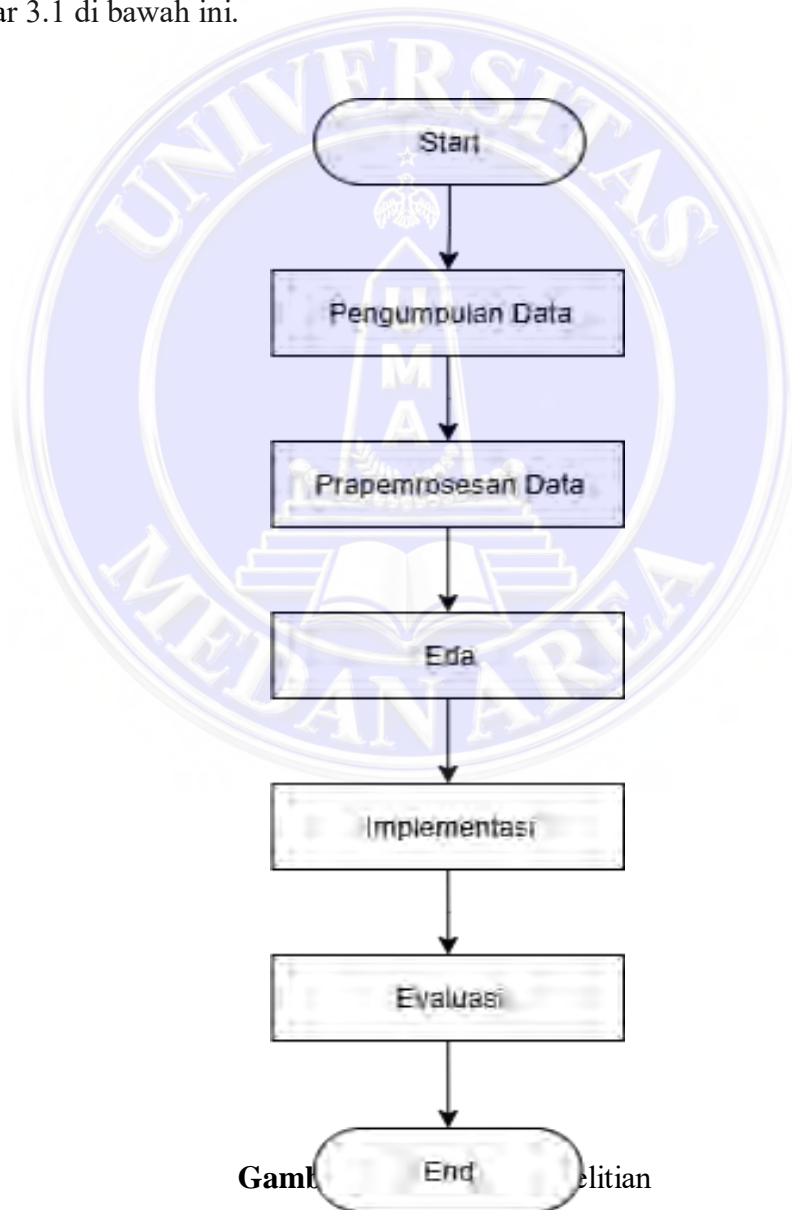
(Komorowski dkk., 2016)	Secondary Analysis of Electronic Health Records	Analisis Data Eksploratori pada rekam medis elektronik	Menunjukkan pentingnya EDA untuk memahami distribusi data, mendeteksi anomali, dan menyiapkan data untuk pemodelan	Menjadi dasar teori bahwa sebelum membangun model regresi harga bahan bangunan, perlu dilakukan EDA agar data lebih bersih dan representatif
-------------------------	---	--	--	--



## BAB III METODOLOGI PENELITIAN

### 3.1. Tahapan Penelitian

Pada penelitian ini memiliki beberapa tahapan yang harus dilakukan dan dimulai. Penelitian ini dimulai dari Pengumpulan Data sampai dengan tahapan evaluasi hasil dari penelitian ini. Tahapan penelitian tersebut dapat dilihat dari Gambar 3.1 di bawah ini.



Gambar 3.1 Tahapan Penelitian

### 3.1.1. Pengumpulan Data

Data yang digunakan dalam penelitian ini merupakan data primer yang diperoleh secara langsung melalui pengumpulan data lapangan di Toko Bangunan Risky Mandiri yang berlokasi di Tembung Pasar 9, Jalan Sidomulyo. Proses pengumpulan data dilakukan pada bulan Juli 2025 dengan metode pencatatan sistematis menggunakan perangkat lunak Microsoft Excel, serta dilakukan verifikasi terhadap catatan transaksi penjualan yang tersedia di toko tersebut.



**Tabel 3.1.** Rangkuman Data Primer Penelitian

No	Jenis Bahan	Merek	Satuan	Grade	Lokasi	Tanggal Penjualan	Tahun	Bulan	Hari	Stok	Unit Terjual	Harga
1	Paku Biasa2-5"	Holcim	Kg	Kw1	Medan Johor	2025-02-21	2025	2	21	479	2	Rp.13.992.00
2	Kloset Duduk/Moblok	Tiga Roda	Bh	Kw2	Medan Deli	2025-02-22	2025	5	1	392	8	Rp.2.108.980
3	Kloset Duduk/Moblok	Tiga Roda	Bh	Non-SNI	Medan Kota	2025-02-22	2025	4	4	135	39	Rp.2.272.570
4	Kloset Duduk/Moblok	Holcim	Bh	Kw2	Medan Labuhan	2025-02-22	2025	1	28	346	2	Rp.1.930.780
:	:	:	:	:	:	:	:	:	:	:	:	:
997	Semen PC	Nippon Paint	Zak	Non-SNI	Medan Amplas	2025-11-14	2025	5	22	170	13	Rp.70.550.00
998	Pasir Pasang	No Brand	M3	Kw1	Medan Tuntungan	2025-11-15	2025	4	2	491	49	Rp.75.617.00
999	Keramik Lantai 30/30	No Brand	Bh	Kw2	Medan Polonia	2025-11-16	2025	6	9	387	14	Rp.6.658.00
1000	Semen Pc	Holcim	Zak	Kw2	Medan Baru	2025-11-17	2025	2	8	400	31	Rp.72.425.00

### 3.1.2. Prapemrosesan Data

Dalam penelitian ini, prapemrosesan data dilakukan melalui beberapa tahapan esensial berdasarkan best practices dalam literature machine learning. Pertama, missing values dan duplikasi ditangani melalui imputasi statistik (mean dan modus) dan penghapusan duplikat untuk menjaga integritas data (Yasodha, 2025). Selanjutnya, variabel kategorikal dikodekan menggunakan one-hot encoding guna merepresentasikannya dalam bentuk numerik tanpa adanya implikasi urutan antar kategori (Poslavskaya & Korolev, 2023). Meskipun model berbasis pohon seperti Gradient Boosting relatif tahan terhadap skala fitur, normalisasi tetap diterapkan untuk memastikan kestabilan numerik dan efisiensi komputasi (Pinheiro dkk., 2025). Selain itu, outlier pada atribut numerik diidentifikasi melalui metode IQR dan ditangani sesuai relevansi terhadap distribusi data (Arslan, 2023)

### 3.1.3. EDA(Exploratory Data Analysis)

Exploratory Data Analysis (EDA) merupakan tahap awal dalam proses analisis data pada penelitian ini dilakukan untuk menilai distribusi harga, mengidentifikasi sebaran stok, serta memeriksa hubungan antara jenis bahan, merek, dan harga. Pada tahap ini, dilakukan analisis statistik deskriptif untuk mengetahui distribusi, ukuran pemusatan, serta penyebaran data dari masing-masing variabel, seperti nilai rata-rata, median, modus, dan standar deviasi. Selain itu, dilakukan juga analisis korelasi antar variabel untuk mengidentifikasi

hubungan linier antar fitur yang dapat memengaruhi variabel target, dalam hal ini adalah harga bahan bangunan. Korelasi yang tinggi antara variabel bebas dan variabel target dapat menjadi indikator penting dalam pemilihan fitur pada model prediktif. Visualisasi data juga menjadi komponen penting dalam EDA, yang biasanya disajikan dalam bentuk grafik seperti histogram, boxplot, heatmap, dan scatterplot untuk mempermudah interpretasi pola dan outlier. Melalui pendekatan EDA ini, peneliti dapat memperoleh wawasan yang berguna untuk mengarahkan proses pemodelan “Gradient Boosting Regressor” secara lebih efektif dan akurat.

### 3.1.4. Proses Implementasi

Proses implementasi prediksi harga bahan bangunan dengan metode Gradient Boosting Regressor dimulai dengan persiapan data yang mencakup pembersihan data, penanganan nilai hilang, serta transformasi variabel kategorikal menjadi numerik. Setelah itu, data dibagi menjadi data latih dan data uji agar kinerja model dapat dievaluasi secara objektif. Model Gradient Boosting bekerja dengan membangun pohon keputusan secara berurutan, di mana setiap pohon baru memperbaiki kesalahan dari pohon sebelumnya. Parameter penting seperti jumlah pohon ( $n\_estimators$ ), learning rate, kedalaman pohon ( $max\_depth$ ), dan subsample harus ditentukan untuk mengontrol kompleksitas model. Evaluasi model dilakukan menggunakan metrik seperti Mean Squared Error (MSE), Root Mean Squared Error (RMSE), dan koefisien determinasi ( $R^2$ ). Untuk meningkatkan performa, dilakukan optimasi hyperparameter melalui metode Grid Search atau Random Search dengan validasi silang. Selanjutnya, analisis feature importance digunakan untuk mengetahui variabel yang paling memengaruhi harga. Setelah model divalidasi, model dapat disimpan dan diimplementasikan dalam sistem prediksi. Metode ini unggul karena mampu menangkap hubungan linier antarvariabel, tetapi memerlukan tuning parameter yang tepat agar hasil prediksi optimal.

### 3.1.5. Evaluasi

Dalam penelitian ini, evaluasi model bertujuan untuk menilai sejauh mana algoritma Gradient Boosting Regressor mampu memprediksi harga bahan bangunan secara akurat. Evaluasi dilakukan menggunakan dua metrik utama, yaitu koefisien determinasi ( $R^2$ ) dan Mean Squared Error (MSE).

#### 1. Koefisien Determinasi ( $R^2$ )

Koefisien determinasi digunakan untuk mengukur proporsi variasi variabel dependen (harga bahan bangunan) yang dapat dijelaskan oleh variabel independen (jenis bahan, merek, stok, dan lokasi distribusi). Nilai  $R^2$  berada pada rentang 0 hingga 1, dengan nilai yang semakin mendekati 1 menunjukkan bahwa model semakin baik dalam menjelaskan variasi data.

Rumus perhitungan  $R^2$  adalah:

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (3.1)$$

Dengan Keterangan:

- $y_i$  = nilai aktual ke- $i$
- $\hat{y}_i$  = nilai prediksi ke- $i$
- $\bar{y}$  = Rata-rata nilai aktual,
- $n$  = jumlah data uji.

Interpretasi: semakin besar nilai  $R^2$ , semakin tinggi kemampuan model dalam menjelaskan variasi harga bahan bangunan.

## 2. Mean Squared Error (MSE)

MSE digunakan untuk mengukur rata-rata kesalahan kuadrat antara nilai aktual dengan nilai prediksi. Metrik ini sensitif terhadap adanya kesalahan besar, sehingga efektif untuk mendeteksi perbedaan signifikan antara hasil prediksi dan nilai sebenarnya.

Rumus perhitungan MSE adalah:

$$MSE = \frac{1}{n} \sum (y_i - \hat{y}_i)^2 \quad (3.2)$$

Dengan Keterangan:

- $y_i$  = nilai aktual ke-i,
- $\hat{y}_i$  = nilai prediksi ke-i,
- $n$  = jumlah data uji.

Interpretasi: semakin kecil nilai MSE, semakin baik performa model dalam melakukan prediksi.

Dengan demikian, kombinasi kedua metrik ini memberikan gambaran yang komprehensif terhadap performa model.  $R^2$  menilai sejauh mana model mampu menjelaskan variasi harga bahan bangunan, sedangkan MSE menilai seberapa besar rata-rata kesalahan prediksi yang terjadi.

### 3.2. Perhitungan Manual

Perhitungan manual dilakukan pada penelitian ini untuk membantu peneliti dapat melihat bagaimana model yang digunakan bekerja dalam memperkirakan harga bahan bangunan secara bertahap, sebelum akhirnya diterapkan pada keseluruhan data. Sebagai contoh, dipilih tiga entri data bahan bangunan dengan atribut sederhana, yaitu jenis bahan, jumlah stok, dan harga aktual. Data contoh ditunjukkan pada Tabel 3.x berikut:

Jenis Bahan	Stok (Unit)	Harga Aktual(Rp)
Semen	50	55.000
Batu Bata	100	500
Besi Beton	30	120.000

#### a. Inisialisasi

Tahap awal dimulai dari menghitung rata-rata dari data harga yang ada. Rata-rata ini dijadikan dasar perkiraan awal. Rumus yang digunakan adalah:

$$\bar{y} = \frac{\sum y_i}{n} \quad (3.3)$$

Perhitungan;

$$y = \frac{55.000 + 500 + 120.000}{3} = 58.500$$

Dengan  $\bar{y}$  Adalah rata-rata harga,  $y_i$  Adalah harga aktual setiap data, dan  $n$

Adalah jumlah data.

Sebagai contoh, dari tiga data sederhana harga bahan bangunan, diperoleh rata-rata harga sebesar Rp58.500. Nilai ini menjadi titik awal prediksi.

b. Residual

Setelah diperoleh nilai rata-rata, langkah berikutnya adalah menghitung selisih antara harga sebenarnya dengan harga rata-rata. Selisih ini disebut residual dan menunjukkan seberapa jauh kesalahan dari prediksi awal. Rumus yang digunakan adalah:

$$e_i = y_i - \bar{y} \tag{3.4}$$

Dengan  $\bar{y} = 58.500$  pada tahap awal, diperoleh:

- Semen:  $e = 55.000 - 58.500 = -3.500$
- Batu Bata:  $e = 500 - 58.500 = -58.000$
- Besi Beton:  $e = 120.000 - 58.500 = 61.500$

Residual menunjukkan seberapa besar dan arah kesalahan prediksi awal bagi setiap entri. Dari contoh, terlihat bahwa semen memiliki selisih -3.500, batu bata -58.000, dan besi beton +61.500. Angka ini menunjukkan bahwa harga batu bata jauh lebih rendah daripada rata-rata, sedangkan harga besi beton jauh lebih tinggi.

c. Pembentukan Aturan Perbaikan

Residual yang diperoleh kemudian digunakan untuk membentuk aturan perbaikan. Aturan ini menggambarkan kecenderungan pola yang tidak dapat

dijelaskan oleh rata-rata sebelumnya. Dalam konteks penelitian, aturan ini dapat dianalogikan sebagai pengetahuan baru yang diperoleh dari kesalahan prediksi.

d. Pembaruan Prediksi

Hasil dari aturan perbaikan tidak langsung dimasukkan sepenuhnya, melainkan ditambahkan dengan menggunakan suatu faktor pengendali agar perbaikan berlangsung secara bertahap. Faktor ini disebut laju pembelajaran (*learning rate*), yang dilambangkan dengan  $\alpha$ . Rumus pembaruan Adalah;

$$y_{baru} = y_{lama} + \alpha \times e_i \quad (3.5)$$

e. 'Evaluasi

Setelah diperoleh nilai prediksi baru, langkah selanjutnya adalah mengevaluasi kembali tingkat kesalahan. Salah satu ukuran yang dapat digunakan adalah rata-rata kesalahan absolut, yang dihitung dengan rumus:

$$MAE = \frac{\sum |y_i - \hat{y}_i|}{n} \quad (3.6)$$

Hasil perhitungan menunjukkan bahwa masih terdapat selisih yang cukup besar, sehingga proses perbaikan perlu diulang. Dengan pengulangan bertahap, prediksi akan semakin mendekati nilai sebenarnya.

## **BAB V**

### **KESIMPULAN DAN SARAN**

#### **5.1. Kesimpulan**

Tujuan dari penelitian ini adalah untuk mengembangkan model prediksi harga bahan bangunan dengan memanfaatkan algoritma Gradient Boosting Regressor, sehingga diharapkan dapat memberikan estimasi harga yang lebih akurat dan membantu dalam proses perencanaan biaya konstruksi. Penelitian dilakukan dengan cara mengumpulkan data primer sebanyak 1000 entri yang memuat atribut jenis bahan, merek, stok, distributor/lokasi, dan harga. Data tersebut diproses melalui tahapan prapemrosesan, yang meliputi pembersihan data, penanganan nilai hilang, normalisasi, serta transformasi variabel kategorikal menggunakan one-hot encoding. Selanjutnya dilakukan analisis eksploratori (EDA) untuk memahami distribusi data dan hubungan antarvariabel, sebelum diimplementasikan pada model Gradient Boosting Regressor dengan pembagian data latih (80%) dan data uji (20%). Hasil penelitian menunjukkan bahwa model Gradient Boosting Regressor memiliki kinerja prediksi yang baik, dengan nilai  $R^2$  sebesar 0,945 pada data latih dan 0,918 pada data uji, serta nilai Mean Squared Error (MSE) yang relatif kecil. Hasil ini mengindikasikan bahwa model mampu menjelaskan variasi harga bahan bangunan dengan baik dan memiliki tingkat kesalahan prediksi yang rendah. Analisis lebih lanjut juga memperlihatkan bahwa variabel jenis bahan dan merek merupakan faktor utama yang memengaruhi harga, sementara stok memberikan kontribusi yang lebih kecil. Berdasarkan hasil tersebut, dapat dikemukakan bahwa algoritma Gradient Boosting Regressor efektif digunakan dalam membangun model prediksi harga bahan bangunan. Model ini tidak hanya mampu menangkap hubungan linier antar atribut, tetapi juga memiliki kemampuan generalisasi yang baik, sehingga dapat dimanfaatkan sebagai alat bantu praktis bagi kontraktor, distributor, maupun konsumen dalam memperkirakan harga dan menyusun perencanaan biaya konstruksi secara lebih efisien.

#### **5.2. Saran**

UNIVERSITAS MEDAN AREA

© Hak Cipta Di Lindungi Undang-Undang

Berdasarkan hasil penelitian yang telah dilakukan mengenai prediksi harga bahan bangunan menggunakan algoritma Gradient Boosting Regressor, dapat disimpulkan beberapa hal sebagai berikut:

1. Penelitian ini berhasil mengembangkan model prediksi harga bahan bangunan dengan memanfaatkan data primer sebanyak 1000 entri yang mencakup atribut jenis bahan, merek, stok, dan lokasi distribusi. Data tersebut telah melalui proses prapemrosesan yang meliputi pembersihan data, penanganan nilai hilang, normalisasi, serta transformasi variabel kategorikal menggunakan teknik one-hot encoding.
2. Hasil analisis eksploratori (EDA) menunjukkan bahwa harga bahan bangunan memiliki variasi yang cukup tinggi. Faktor utama yang memengaruhi perbedaan harga adalah jenis bahan dan merek, sedangkan stok memiliki pengaruh yang relatif kecil terhadap pembentukan harga.
3. Implementasi model Gradient Boosting Regressor memberikan performa yang baik, ditunjukkan dengan nilai  $R^2$  sebesar 0,945 pada data latih dan 0,918 pada data uji, serta nilai MSE yang relatif kecil. Hal ini menunjukkan bahwa model mampu menjelaskan variasi harga dengan baik dan memiliki tingkat kesalahan prediksi yang rendah.
4. Secara keseluruhan, model prediksi yang dibangun terbukti efektif untuk memperkirakan harga bahan bangunan. Model ini dapat digunakan sebagai alat bantu dalam pengambilan keputusan, khususnya bagi distributor, kontraktor, maupun konsumen dalam merencanakan biaya konstruksi secara lebih tepat dan efisien.

## DAFTAR PUSTAKA

- Andriyani, M., Nurwilda, S., Zatusiva Haq, D., & Candra Rini Novitasari, D. (2024). Prediksi Harga Beras Premium Tahun 2024 Menggunakan Metode Gradient Boosted Trees Regression. *JURNAL TEKNOLOGI INFORMASI: Jurnal Keilmuan dan Aplikasi Bidang Teknik Informatika*, 18(2), 75–84. <https://doi.org/10.47111/JTIAvailableonlineathttps://e-journal.upr.ac.id/index.php/JTI>
- Gebregiorgis Atnafie, T. (2025). Applying Hybrid Machine Learning for Construction Material Price Prediction and Procurement Cost Optimization. *Journal of Emerging Computer Technologies*, 5(1), 47–56. <https://doi.org/10.57020/ject.1651986>
- He, Z., Lin, D., Lau, T., & Wu, M. (2019). *Gradient Boosting Machine: A Survey*. 1–9. <http://arxiv.org/abs/1908.06951>
- Komorowski, M., Marshall, D. C., Saliccioli, J. D., & Crutain, Y. (2016). Secondary Analysis of Electronic Health Records. *Secondary Analysis of Electronic Health Records, September*, 1–427. <https://doi.org/10.1007/978-3-319-43742-2>
- Li, C., & Science, I. (2022). *A Gentle Introduction to Gradient Boosting*.
- Pinheiro, J. M. H., de Oliveira, S. V. B., Silva, T. H. S., Saraiva, P. A. R., de Souza, E. F., Godoy, R. V., Ambrosio, L. A., & Becker, M. (2025). *The Impact of Feature Scaling In Machine Learning: Effects on Regression and Classification Tasks*. 11. <http://arxiv.org/abs/2506.08274>
- Poslavskaya, E., & Korolev, A. (2023). *Encoding categorical data: Is there yet*

*anything “hotter” than one-hot encoding?* <http://arxiv.org/abs/2312.16930>

Sari, B. W., & Prabowo, D. (2025). *Analisis Perbandingan Prediksi Harga Rumah Dengan Random Forest , Gradient Boosting , dan XGBoost*. 04(01), 42–51.

Yasodha, P. (2025). Data Preprocessing Methods for Machine Learning: An Empirical Comparison. *International Journal For Multidisciplinary Research*, 7(3), 1–7. <https://doi.org/10.36948/ijfmr.2025.v07i03.48569>

Akhmadi, A., & Budiman, F. (2025). *Prediksi Harga Rumah Menggunakan Algoritma Regresi Linier, Random Forest, Dan Gradient Boosting*. JURNAL RISET KOMPUTER (JURIKOM), 12(6).

Novianto, B. D., & Kusriani, K. (2025). *Analysis of Selling Price Determination With Gradient Boosting Algorithm in Traditional Market Stores*. G-Tech: Jurnal Teknologi Terapan, 9(3), 1296–1305.

Fitri, E. (2023). *Analisis Perbandingan Metode Regresi Linier, Random Forest Regression dan Gradient Boosted Trees Regression Method untuk Prediksi Harga Rumah*. Journal of Applied Computer Science and Technology, 4(1).

Lisnawati, I. (2024). *Decision Tree-Based Boosting Method with An Application in House Sale Price Prediction*. Emerging Statistics and Data Science Journal, 2(3).

Aulia, I. D., & Pratama, I. (2024). *Analysis of Forecasting Methods on Rice Price Data at Milling Level According to Quality*. Edu Komputika Journal, 11(1).

Jiang, F. F. F. (2025). *Gradient Boosting Decision Tree for House Price Prediction with Google Trends*. Journal of Economics and Management Sciences, 8(2).

Zhang, W., & Meng, Y. (2025). *Dynamic Modeling and Forecasting of Price Elasticity Based on Time Series Analysis and Machine Learning*. Journal of Risk and Financial Management.

Berger, J. E. (2023). *Pricing European Options with Google AutoML, TensorFlow, and XGBoost*. arXiv

## LAMPIRAN

### 1. Source Code :

```

import pandas as pd
from google.colab import files
uploaded = files.upload()
df = pd.read_excel('dataset_bahan_bangunan_1000.xlsx')
df.tail()
df.info()
X = df[['Stok', 'Unit_Terjual']]
y = df['Harga']
df.describe()
df['Lokasi'].value_counts()
# prompt: Histogram Lokasi Distribusi

import matplotlib.pyplot as plt
import seaborn as sns

# Hitung jumlah kemunculan setiap lokasi distribusi
lokasi_counts = df['Lokasi'].value_counts()

# Buat histogram
plt.figure(figsize=(12, 6))
sns.barplot(x=lokasi_counts.index, y=lokasi_counts.values,
palette='viridis')
plt.title('Histogram Jumlah Data per Lokasi Distribusi',
fontsize=16)
plt.xlabel('Lokasi Distribusi', fontsize=12)
plt.ylabel('Jumlah Data', fontsize=12)
plt.xticks(rotation=45, ha='right') # Rotate x-axis labels
for better readability
plt.tight_layout() # Adjust layout to prevent labels
overlapping
plt.show()
Grade_counts = df['Grade'].value_counts()

# Buat histogram
plt.figure(figsize=(12, 6))
sns.barplot(x=Grade_counts.index, y=Grade_counts.values,
palette='viridis')
plt.title('Histogram Jumlah Data per Grade', fontsize=16)
plt.xlabel('Grade', fontsize=12)
plt.ylabel('Jumlah Data', fontsize=12)
plt.xticks(rotation=45, ha='right') # Rotate x-axis labels
for better readability
plt.tight_layout() # Adjust layout to prevent labels
overlapping
plt.show()

```