

**IMPLEMENTASI *HYBRID MACHINE LEARNING*
ALGORITHM UNTUK PREDIKSI HARGA RUMAH DI
BANDUNG**



UMN

UNIVERSITAS
MULTIMEDIA
NUSANTARA

Skripsi

**RYAN GUNAWAN
0000043095**

**PROGRAM STUDI SISTEM INFORMASI
FAKULTAS TEKNIK DAN INFORMATIKA
UNIVERSITAS MULTIMEDIA NUSANTARA
TANGERANG SELATAN**

2024

**IMPLEMENTASI *HYBRID MACHINE LEARNING*
ALGORITHM UNTUK PREDIKSI HARGA RUMAH DI
BANDUNG**



Skripsi

Diajukan sebagai Salah Satu Syarat untuk Memperoleh

Gelar Sarjana S.Kom

RYAN GUNAWAN
00000043095

PROGRAM STUDI SISTEM INFORMASI
FAKULTAS TEKNIK DAN INFORMATIKA
UNIVERSITAS MULTIMEDIA NUSANTARA
TANGERANG SELATAN

2024

i

HALAMAN PERNYATAAN TIDAK PLAGIAT

Dengan ini saya,

Nama : Ryan Gunawan

Nomor Induk Mahasiswa : 00000043095

Program studi : Sistem Informasi

Skripsi dengan judul:

IMPLEMENTASI *HYBRID MACHINE LEARNING ALGORITHM* UNTUK PREDIKSI HARGA RUMAH DI BANDUNG

merupakan hasil karya saya sendiri bukan plagiat dari karya ilmiah yang ditulis oleh orang lain, dan semua sumber, baik yang dikutip maupun dirujuk, telah saya nyatakan dengan benar serta dicantumkan di Daftar Pustaka.

Jika di kemudian hari terbukti ditemukan kecurangan/penyimpangan, baik dalam pelaksanaan skripsi maupun dalam penulisan laporan skripsi, saya bersedia menerima konsekuensi dinyatakan TIDAK LULUS untuk Tugas Akhir yang telah saya tempuh.

Tangerang, 13 Mei 2024



Ryan Gunawan

HALAMAN PENGESAHAN

Skripsi dengan judul

IMPLEMENTASI *HYBRID MACHINE LEARNING* UNTUK PREDIKSI HARGA RUMAH DI BANDUNG

Oleh

Nama : Ryan Gunawan
NIM : 00000043095
Program Studi : Sistem Informasi
Fakultas : Teknik dan Informatika

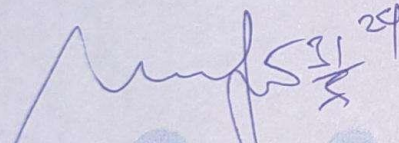
Telah diujikan pada hari Selasa, 21 Mei 2024

Pukul 10.00 s.d 12.00 dan dinyatakan

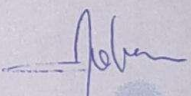
LULUS

Dengan susunan penguji sebagai berikut.

Ketua Sidang


Dr. Santo Fernandi Wijaya
0310016902

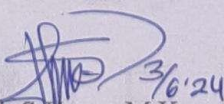
Penguji


Melissa Indah Fianty, S.Kom., MMSI.
0313019201

Pembimbing


Ir. Raymond Sunardi Oetama, M.C.I.S.,
0328046803/023897

Ketua Program Studi Sistem Informasi


Ririn Ikana Desanti, S.Kom., M.Kom.
313058001

NUSANTARA

**LEMBAR PERSETUJUAN PUBLIKASI
KARYA ILMIAH MAHASISWA**

Yang bertanda tangan dibawah ini:

Nama : Ryan Gunawan
Nomor Induk Mahasiswa : 00000043095
Program Studi : Sistem Informasi
Jenjang : S1
Judul Karya Ilmiah :

**IMPLEMENTASI *HYBRID MACHINE LEARNING ALGORITHM* UNTUK
PREDIKSI HARGA RUMAH DI BANDUNG**

Menyatakan dengan sesungguhnya bahwa saya bersedia:

Memberikan izin sepenuhnya kepada Universitas Multimedia Nusantara untuk mempublikasikan hasil karya ilmiah saya di repositori Knowledge Center, sehingga dapat diakses oleh Civitas Akademika/Publik. Saya menyatakan bahwa karya ilmiah yang saya buat tidak mengandung data yang bersifat konfidensial dan saya juga tidak akan mencabut kembali izin yang telah saya berikan dengan alasan apapun.

Saya tidak bersedia, dikarenakan:

Dalam proses pengajuan untuk diterbitkan ke jurnal/konferensi nasional/internasional (dibuktikan dengan *letter of acceptance*)*.

Tangerang, 21 Mei 2024



Ryan Gunawan

* Jika tidak bisa membuktikan LoA jurnal/HKI selama 6 bulan kedepan, saya bersedia mengizinkan penuh karya ilmiah saya untuk diunggah ke KC UMN dan menjadi hak institusi UMN.

KATA PENGANTAR

Puji dan syukur saya panjatkan kepada Tuhan yang Maha ESA atas berkat yang dilimpahkan kepada saya sehingga berkesempatan untuk membuat penulisan skripsi tentang “IMPLEMENTASI *HYBRID MACHINE LEARNING ALGORITHM* UNTUK PREDIKSI HARGA RUMAH DI BANDUNG” hingga selesai tepat pada waktunya. Penulisan skripsi ini diajukan dengan tujuan memenuhi persyaratan kelulusan Program Strata 1 untuk mendapatkan gelar sarjana pada Program Studi Sistem Informasi, Fakultas Teknik dan Informatika di Universitas Multimedia Nusantara. Penyelesaian proses pembuatan laporan skripsi ini dibantu oleh berbagai pihak yang memberikan motivasi, semangat, dan bimbingan untuk menyelesaikan program studi ini. Dengan demikian saya dengan tulus hati mengapresiasi pihak- pihak yang telah membantu dan membimbing saya selamat proses penyusunan skripsi.

Saya mengucapkan terima kasih

1. Bapak Dr. Ninok Leksono, M.A, selaku Rektor Universitas Multimedia Nusantara.
2. Bapak Dr. Eng. Niki Prastomo, St.T., M.Sc., selaku Dekan Fakultas Universitas Multimedia Nusantara.
3. Ibu Ririn Ikana Desanti, S.Kom, M.Kom, selaku Ketua Program Studi Universitas Multimedia Nusantara.
4. Bapak Ir. Raymond Sunardi Oetama, M. CIS, sebagai Pembimbing pertama yang telah memberikan bimbingan, arahan, dan motivasi atas terselesainya tugas akhir ini.
5. Keluarga saya yang telah memberikan bantuan dukungan material dan moral, sehingga penulis dapat menyelesaikan tugas akhir ini.

Tangerang, 13 Mei 2024



Ryan Gunawan

IMPLEMENTASI *HYBRID MACHINE LEARNING* ALGORITHM UNTUK PREDIKSI HARGA RUMAH DI BANDUNG

Ryan Gunawan

ABSTRAK

Pemahaman terhadap faktor-faktor yang memengaruhi nilai properti esensial bagi penjual dan pembeli. Bagi penjual, pemahaman tentang dinamika pasar, karakteristik properti, dan faktor eksternal seperti perkembangan infrastruktur dan tren ekonomi lokal memungkinkan penetapan harga properti yang akurat dan kompetitif. Bagi pembeli, pengetahuan tentang variabel yang mempengaruhi nilai properti memungkinkan analisis lebih teliti, mempertimbangkan potensi apresiasi nilai properti, dan membuat keputusan investasi strategis. Penelitian ini menganalisis variabel yang mempengaruhi harga rumah dan mengembangkan sistem prediksi harga rumah untuk membantu pengambilan keputusan bagi penjual dan pembeli properti. Penelitian ini menganalisis dataset katalog rumah menggunakan Big Data Analytics. Setelah analisis, algoritma prediksi Support Vector Regression, XGBoost, dan Gradient Boost diimplementasikan dan dibandingkan untuk menentukan model terbaik dan dikombinasikan dengan algoritma clustering. Tujuan model prediksi dan clustering adalah memprediksi nilai properti suatu daerah berdasarkan faktor yang ada dan memberikan gambaran harga pasar properti. Perbandingan terhadap ketiga kombinasi algoritma regresi dan clustering menunjukkan XGBoost sebagai model terbaik dengan RMSE terendah yaitu 182.569.784,85 untuk Cluster 0, 238.994.002,58 untuk Cluster 1, dan 317.144.761,10 untuk Cluster 2. Nilai MAPE dari XGBoost juga lebih rendah dibandingkan algoritma lain dengan MAPE 11,38% untuk Cluster 0, 5,11% untuk Cluster 1, dan 3% untuk Cluster 2. Model XGBoost digunakan untuk diimplementasikan ke sistem prediksi berbasis website.

Kata kunci: *Big Data Analytics, Gradient Boost, XGBoost, Machine learning, Support Vector Regression.*

THE IMPLEMENTATION OF HYBRID MACHINE LEARNING ALGORITHM FOR HOUSE PRICE PREDICTION IN BANDUNG

Ryan Gunawan

ABSTRACT (English)

A thorough understanding of the factors influencing property valuation is essential for both sellers and buyers. For sellers, a deep understanding of market dynamics, property characteristics, and external factors such as infrastructure development and local economic trends enables them to set more accurate and competitive property prices. For buyers, comprehensive knowledge of variables affecting property value allows for more detailed analysis, consideration of potential appreciation in property value, and strategic investment decision-making. Therefore, this research will analyze variables affecting house prices and develop a house price prediction system to aid decision-making for both property sellers and buyers. To identify the factors influencing property value, this study will analyze a house catalog dataset using Big Data Analytics. After the analysis, Support Vector Regression, XGBoost, and Gradient Boost prediction algorithms will be implemented and compared to determine the best model, combined with clustering algorithms. The purpose of creating prediction and clustering models is to predict property values in an area based on existing factors and provide an overview of property market prices. The comparison of the three algorithm combinations for regression and clustering shows that XGBoost is the best model with the lowest RMSE of 182,569,784.85 for Cluster 0, 238,994,002.58 for Cluster 1, and 317,144,761.10 for Cluster 2. The MAPE value of XGBoost is also lower than other algorithms with a MAPE value of 11.38% for Cluster 0, 5.11% for Cluster 1, and 3% for Cluster 2. The XGBoost model is selected for implementation in the website-based prediction system.

Keywords: *Big Data Analytics, Gradient Boost, XGBoost, Machine learning, Support Vector Regression.*

DAFTAR ISI

HALAMAN PERNYATAAN TIDAK PLAGIAT	ii
HALAMAN PERSETUJUAN PUBLIKASI KARYA ILMIAH UNTUK KEPENTINGAN AKADEMIS	iv
KATA PENGANTAR.....	iv
ABSTRAK	vi
DAFTAR ISI.....	viii
DAFTAR TABEL	x
DAFTAR GAMBAR.....	xi
DAFTAR RUMUS	xiii
BAB I PENDAHULUAN.....	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah.....	6
1.3 Batasan Masalah	6
1.4 Tujuan dan Manfaat Penelitian	7
1.5 Sistematika Penulisan	7
BAB II LANDASAN TEORI	9
2.1 Penelitian Terdahulu	9
2.2 Teori Penelitian.....	13
2.3 <i>Framework</i> dan Algoritma.....	14
2.3.1 CRISP-DM.....	14
2.3.2 <i>Support Vector Regression</i>	16
2.3.3 <i>Gradient Boosting</i>	17
2.3.4 <i>XGBoost</i>	18
2.3.5 K-means	19
2.4 Tools	19
BAB III METODOLOGI PENELITIAN	22
3.1 Objek Penelitian.....	22
3.2 Alur Penelitian	22
3.3 Metode Penelitian	23
3.3.1 Business Understanding.....	25
3.3.2 Data Understanding.....	25
3.4 Teknik Pengumpulan Data.....	26
3.5 Teknik Analisis Data	26

3.5.1	Data Preparation.....	26
3.5.2	Data Modeling.....	26
3.5.3	Evaluation	27
3.5.4	Deployment.....	27
BAB IV ANALISIS DAN HASIL PENELITIAN		29
4.1	Business Understanding.....	29
4.2	Data Understanding	29
4.2.1	Data Collection.....	31
4.2.2	Exploratory Data Analysis (EDA)	34
4.3	Data Preparation	42
4.3.1	Data Cleaning.....	42
4.3.2	Data Standardization	43
4.3.3	One Hot Encoding.....	44
4.4	Modelling.....	46
4.4.1	Kmeans Clustering	47
4.4.2	<i>Support Vector Regression</i>	51
4.4.3	<i>Gradient Boost</i>	53
4.4.4	<i>XGBoost</i>	56
4.5	Evaluation	58
4.6	Deployment.....	62
4.7	Result & Discussion.....	69
BAB V SIMPULAN DAN SARAN.....		73
5.1	Simpulan.....	73
5.2	Saran	74
DAFTAR PUSTAKA.....		75
LAMPIRAN.....		78

DAFTAR TABEL

Tabel 2.1 Penelitian Terdahulu	9
Tabel 3.1 Perbandingan Metodologi Penelitian.....	23
Tabel 3.2 Software <i>Requirement</i>	28
Tabel 4.1 Fitur Hasil Web Scrapping	34
Tabel 4.2 Summary Variabel Numerik	35
Tabel 4.3 Summary Variabel Kategorikal	35
Tabel 4.4 Metric Evaluasi SVR	52
Tabel 4.5 Metric Evaluasi Gradient Boost.....	55
Tabel 4.6 Metric Evaluasi <i>XGBoost</i>	57
Tabel 4.7 Perbandingan Performa Model Prediksi	58

UMMN

UNIVERSITAS

MULTIMEDIA

NUSANTARA

DAFTAR GAMBAR

Gambar 1.1 Grafik Perkembangan Indeks Harga Properti Residensial.....	2
Gambar 1.2 Grafik Tingkat Penjualan Properti Residensial.....	3
Gambar 4.1 Landing Page Lamudi.com.....	30
Gambar 4.2 Halaman Katalog Lamudi.....	30
Gambar 4.3 Halaman Detail Produk Lamudi.com.....	31
Gambar 4.4 Daftar Import Package Web Scrapping.....	32
Gambar 4.5 Kode untuk Memulai Sesi Chrome.....	32
Gambar 4.6 Kode Pengambilan Link Detail Produk.....	32
Gambar 4.7 Kode Web Scrapping.....	33
Gambar 4.8 Bar Plot Distribusi Harga Rumah.....	36
Gambar 4.9 Bar Plot Distribusi Jumlah Bedroom.....	37
Gambar 4.10 Bar Plot Distribusi Jumlah Bathrooms.....	37
Gambar 4.11 Bar Plot Distribusi Jumlah Car_spaces.....	38
Gambar 4.12 Bar Plot Distribusi Jumlah Multiple_floor.....	39
Gambar 4.13 Pie Chart Distribusi Fasilitas Security.....	39
Gambar 4.14 Bar Plot Top 10 Distribusi Lokasi.....	40
Gambar 4.15 Pie Chart Distribusi Fasilitas Terdekat.....	41
Gambar 4.16 Bar Plot Top 10 Lokasi dengan Harga Rata-rata Tertinggi.....	42
Gambar 4.17 Kode Penanganan Missing Value.....	43
Gambar 4.18 Standarisasi Kolom Harga.....	43
Gambar 4.19 Standarisasi Kolom Lokasi.....	44
Gambar 4.20 Dataframe Setelah Data Cleaning.....	45
Gambar 4.21 Implementasi One Hot Encoding pada Kolom Lokasi.....	45
Gambar 4.22 Dataframe setelah One Hot Encoding.....	46
Gambar 4.23 Kode Implementasi Elbow Method.....	47
Gambar 4.24 Plot Elbow Method.....	47
Gambar 4.25 Implementasi Clustering Kmeans.....	48
Gambar 4.26 Scatter Plot Clustering Berdasarkan Land_size.....	48
Gambar 4.27 Cluster 0 Kelas Harga Rumah.....	49
Gambar 4.28 Cluster 1 Kelas Harga Rumah.....	50
Gambar 4.29 Cluster 2 Kelas Harga Rumah.....	51
Gambar 4.30 Kode Import Package SVR dan Fitur Model SVR.....	51
Gambar 4.31 Implementasi One Hot Encoding Kolom Lokasi SVR.....	52
Gambar 4.32 Implementasi Model SVR untuk Setiap Cluster.....	52
Gambar 4.33 Kode Import Package SVR dan Fitur Model Gradient Boost.....	53
Gambar 4.34 Implementasi One Hot Encoding Kolom Lokasi Gradient Boost.....	54
Gambar 4.35 Implementasi Gradient Boost pada Setiap Cluster.....	54
Gambar 4.36 Import Package SVR dan Inisialisasi Fitur.....	56
Gambar 4.37 Implementasi One Hot Encoding Kolom Lokasi XGBoost.....	56
Gambar 4.38 Implementasi model XGBoost untuk Setiap Cluster.....	57
Gambar 4.39 Top 10 Feature SVR.....	59
Gambar 4.40 Top 10 Feature Gradient Boost.....	60
Gambar 4.41 Top 10 Feature XGBoost.....	61

Gambar 4.42 File Tsx Utama.....	62
Gambar 4.43 Landing Page <i>website</i>	63
Gambar 4.44 Tampilan Form Fitur Prediksi.....	63
Gambar 4.45 Contoh Pengisian Form.....	64
Gambar 4.46 Tampilan Review Form.....	65
Gambar 4.47 Tampilan Hasil Prediksi.....	66
Gambar 4.48 Kode Pemanggilan API.....	67
Gambar 4.49 Waktu Pemrosesan HTTP Request	68
Gambar 4.50 Kode API endpoint /predict	68



UMMN

UNIVERSITAS
MULTIMEDIA
NUSANTARA

DAFTAR RUMUS

Rumus 2.1 Rumus Support Vector Regression Linear	17
Rumus 2.2 Rumus Support Vector Regression Non-Linear	17
Rumus 2.3 Rumus Gradient Boost.....	18
Rumus 2.4 Rumus Extreme Gradient Boost	18
Rumus 2.5 Rumus K-Means	19



UMMN

UNIVERSITAS
MULTIMEDIA
NUSANTARA