



**PENERAPAN *WEB SCRAPING* DAN *MACHINE LEARNING*
UNTUK PREDIKSI HARGA RUMAH DI JAKARTA
MENGUNAKAN ALGORITMA *RANDOM FOREST***



LAPORAN TUGAS AKHIR

UNIVERSITAS
MERCU BUANA
ALIFIANDY NUGRAHA
41519110163

**PROGRAM STUDI TEKNIK INFORMATIKA
FAKULTAS ILMU KOMPUTER
UNIVERSITAS MERCU BUANA
JAKARTA
2023**



**PENERAPAN *WEB SCRAPING* DAN *MACHINE LEARNING*
UNTUK PREDIKSI HARGA RUMAH DI JAKARTA
MENGUNAKAN ALGORITMA *RANDOM FOREST***

LAPORAN TUGAS AKHIR

**ALIFIANDY NUGRAHA
41519110163**

**UNIVERSITAS
MERCU BUANA**

Diajukan sebagai salah satu syarat untuk memperoleh gelar sarjana

**PROGRAM STUDI TEKNIK INFORMATIKA
FAKULTAS ILMU KOMPUTER
UNIVERSITAS MERCU BUANA
JAKARTA
2023**

HALAMAN PERNYATAAN KARYA SENDIRI

Saya yang bertanda tangan dibawah ini:

Nama : Alifiandy Nugraha
NIM : 41519110163
Program Studi : Teknik Informatika
Judul Laporan Skripsi : Penerapan *Web Scraping* Dan *Machine Learning* Untuk Prediksi Harga Rumah Di Jakarta Menggunakan Algoritma *Random Forest*

Menyatakan bahwa Laporan Tugas Akhir ini adalah hasil karya saya sendiri dan bukan plagiat, serta semua sumber baik yang dikutip maupun dirujuk telah saya nyatakan dengan benar. Apabila ternyata ditemukan di dalam Laporan Tugas Akhir saya terdapat unsur plagiat, maka saya siap mendapatkan sanksi akademis yang berlaku di Universitas Mercu Buana.

UNIVERSITAS Jakarta, 1 Desember 2023
MERCU BUANA



Alifiandy Nugraha

HALAMAN PENGESAHAN

Laporan Skripsi ini diajukan oleh:

Nama : Alifiandy Nugraha
NIM : 41519110163
Program Studi : Teknik Informatika
Judul Laporan Skripsi : Penerapan *Web Scraping* Dan *Machine Learning* Untuk Prediksi Harga Rumah Di Jakarta Menggunakan Algoritma *Random Forest*

Telah berhasil dipertahankan pada sidang di hadapan Dewan Penguji dan diterima sebagai bagian persyaratan yang diperlukan untuk memperoleh gelar Sarjana Strata I pada Program Studi Teknik Informatika, Fakultas Ilmu Komputer, Universitas Mercu Buana.

Disahkan oleh:

Pembimbing : Sabar Rudiarto, S.Kom., M.Kom.
NIDN : 0309036902
Ketua Penguji : Afiyati, S.Si, MT
NIDN : 0316106908
Penguji 1 : Umniy Salamah, S.Kom, MMSI
NIDN : 0306098104
Penguji 2 : Yustika Erliani, SE, MMSI
NIDN : 0305057809

(Sabar Rudiarto)

(Afiyati)

(Umi Salamah)

(Yustika Erliani) 12/1/23

Jakarta, 27 Desember 2023

Mengetahui,

Dekan

Ketua Program Studi



Dr. Bambang Jokonowo, S.Si., M.T.I



Dr. Hadi Santoso, S.Kom, M.Kom

KATA PENGANTAR

Puji syukur saya panjatkan kepada Allah SWT, Tuhan Yang Maha Esa, karena atas berkat dan rahmat-Nya, saya dapat menyelesaikan Laporan Skripsi ini. Penulisan Laporan Skripsi ini dilakukan dalam rangka memenuhi salah satu syarat untuk mencapai gelar Sarjana Komputer pada Fakultas Ilmu Komputer Universitas Mercu Buana. Saya menyadari bahwa, tanpa bantuan dan bimbingan dari berbagai pihak, dari masa perkuliahan sampai pada penyusunan skripsi ini, sangatlah sulit bagi saya untuk menyelesaikan Laporan Skripsi ini. Oleh karena itu, saya mengucapkan terima kasih kepada:

1. Prof. Dr. Ir. Andi Adriansyah, M.Eng selaku Rektor Universitas Mercu Buana.
2. Bapak Dr. Bambang Jokonowo, S.Si, M.T.I selaku Dekan Fakultas Ilmu Komputer.
3. Bapak Dr. Hadi Santoso, S.Kom, M.Kom selaku Ketua Program Studi Teknik Informatika
4. Bapak Sabar Rudiarto, S.Kom, M.Kom selaku Dosen Pembimbing yang telah menyediakan waktu, tenaga, dan pikiran untuk mengarahkan saya dalam penyusunan skripsi ini.
5. Ibu Afiyati, S.Si, MT selaku Ketua Dosen Penguji Tugas Akhir atas koreksi dan arahan serta masukannya.
6. Ibu Umniy Salamah, S.Kom, MMSI selaku Dosen Penguji 1 Tugas Akhir atas koreksi dan arahan serta masukannya.
7. Ibu Yustika Erliani, SE, MMSI selaku Dosen Penguji 2 Tugas Akhir atas koreksi dan arahan serta masukannya.

Akhir kata, saya berharap Allah SWT, Tuhan Yang Maha Esa berkenan membalas segala kebaikan semua pihak yang telah membantu. Semoga Laporan Skripsi ini membawa manfaat bagi pengembangan ilmu.

Jakarta, 1 Desember 2023

Alifiandy Nugraha

**HALAMAN PERNYATAAN PERSETUJUAN PUBLIKASI TUGAS
AKHIR UNTUK KEPENTINGAN AKADEMIS**

Sebagai sivitas akademik Universitas Mercu Buana, saya yang bertanda tangan di bawah ini:

Nama : Alifiandy Nugraha
NIM : 41519110163
Program Studi : Teknik Informatika
Judul Laporan Skripsi : Penerapan *Web Scraping* Dan *Machine Learning* Untuk Prediksi Harga Rumah Di Jakarta Menggunakan Algoritma *Random Forest*

Demi pengembangan ilmu pengetahuan, dengan ini memberikan izin dan menyetujui untuk memberikan kepada Universitas Mercu Buana **Hak Bebas Royalti Non-Eksklusif (*Non-exclusive Royalty-Free Right*)** atas karya ilmiah saya yang berjudul di atas beserta perangkat yang ada (jika diperlukan).

Dengan Hak Bebas Royalti Non-Eksklusif ini Universitas Mercu Buana berhak menyimpan, mengalihmedia/format-kan, mengelola dalam bentuk pangkalan data (*database*), merawat, dan mempublikasikan Laporan Skripsi saya selama tetap mencantumkan nama saya sebagai penulis/pencipta dan sebagai pemilik Hak Cipta.

Demikian pernyataan ini saya buat dengan sebenarnya.

Jakarta, 1 Desember 2023

Yang menyatakan,



Alifiandy Nugraha

ABSTRAK

Nama : Alifiandy Nugraha
NIM : 41519110163
Program Studi : Teknik Informatika
Judul Laporan Skripsi : Penerapan *Web Scraping* Dan *Machine Learning* Untuk Prediksi Harga Rumah Di Jakarta Menggunakan Algoritma *Random Forest*
Pembimbing : Sabar Rudiarto, S.Kom., M.Kom

Penelitian ini mengkaji tentang teknik *web scraping* dan *machine learning* untuk memprediksi harga rumah di Jakarta menggunakan algoritma *random forest*. Teknik *web scraping* digunakan untuk melakukan pengumpulan data dan algoritma *random forest* adalah model untuk prediksi *machine learning*. Fokus utama penelitian ini adalah mengidentifikasi parameter-parameter yang mempengaruhi harga rumah di Jakarta serta memberikan panduan hasil prediksi harga rumah yang dapat disesuaikan berdasarkan spesifikasi rumah yang diinginkan. Metode yang digunakan adalah observasi pada situs jual beli properti dan teknik *web scraping* untuk mengumpulkan data rumah di Jakarta. Data yang dikumpulkan kemudian digunakan untuk melatih model algoritma *random forest*. Hasil penelitian menunjukkan bahwa teknik *web scraping* sangat efektif untuk mengumpulkan data dengan berhasil mengumpulkan 1272 data rumah di Jakarta pada dataset 1 dan 1492 data rumah di Jakarta pada dataset 2. Hasil evaluasi algoritma *random forest* menggunakan *R-squared* sebesar 0.59 pada dataset 1 dan pada dataset 2 sebesar 0.90. Kesimpulan penelitian ini adalah teknik *web scraping* dan *machine learning* adalah kombinasi yang dapat digunakan untuk membuat suatu sistem prediksi. *Web scraping* digunakan untuk mengambil data di internet dan *machine learning* untuk membuat model prediksi. Namun membuat suatu sistem dengan kombinasi teknik tersebut harus selektif pada pengumpulan data dan pemilihan model *machine learning* karena pengumpulan data akan mempengaruhi kinerja model *machine learning*. Terbukti pada dataset 1 dan dataset 2 terdapat perbedaan variasi data yang cukup signifikan, sehingga pada metrik *R-Squared* terdapat selisih hasil akurasi yang juga signifikan.

Kata Kunci : *web scraping, machine learning, random forest, sistem prediksi*

ABSTRACT

Name : Alifiandy Nugraha
NIM : 41519110163
Study Program : Informatics Engineering
Title Paper : Application of Web Scraping and Machine Learning
for House Price Prediction in Jakarta Using Random
Forest Algorithm
Counsellor : Sabar Rudiarto, S.Kom., M.Kom

This research examines web scraping and machine learning techniques to predict house prices in Jakarta using the random forest algorithm. Web scraping technique is used to collect data and random forest algorithm is a model for machine learning prediction. The main focus of this research is to identify parameters that affect house prices in Jakarta and provide guidance on house price prediction results that can be adjusted based on the desired house specifications. The method used is observation on property buying and selling sites and web scraping techniques to collect data on houses in Jakarta. The data collected is then used to train the random forest algorithm model. The results showed that the web scraping technique was very effective for collecting data by successfully collecting 1272 house data in Jakarta in dataset 1 and 1492 house data in Jakarta in dataset 2. The evaluation results of the random forest algorithm using R-squared are 0.59 on dataset 1 and 0.90 on dataset 2. The conclusion of this research is that web scraping and machine learning techniques are a combination that can be used to create a prediction system. Web scraping is used to retrieve data on the internet and machine learning to create a prediction model. However, making a system with a combination of these techniques must be selective in data collection and machine learning model selection because data collection will affect the performance of the machine learning model. It is proven that in dataset 1 and dataset 2 there is a significant difference in data variation, so that in the R-Squared metric there is a significant difference in accuracy results.

Keywords: *web scraping, machine learning, random forest, prediction system*

DAFTAR ISI

HALAMAN JUDUL	i
HALAMAN PERNYATAAN KARYA SENDIRI	ii
HALAMAN PENGESAHAN.....	iii
KATA PENGANTAR.....	iv
HALAMAN PERNYATAAN PERSETUJUAN PUBLIKASI TUGAS AKHIR UNTUK KEPENTINGAN AKADEMIS	v
ABSTRAK	vi
ABSTRACT	vii
DAFTAR ISI.....	viii
DAFTAR GAMBAR.....	xi
DAFTAR LAMPIRAN	xii
BAB I PENDAHULUAN.....	1
1.1 Latar Belakang	1
1.2 Perumusan Masalah.....	2
1.3 Tujuan Penelitian.....	2
1.4 Manfaat Penelitian.....	2
1.5 Batasan Penelitian	3
BAB II TINJAUAN PUSTAKA.....	4
2.1 Penelitian Terkait	4
2.2 Teori Pendukung	11
BAB III METODE PENELITIAN	16
3.1 Jenis Penelitian	16
3.2 Tahapan Penelitian	16
BAB IV PEMODELAN	22
4.1 Dataset	22
4.1.1 Pencarian Situs Web	22
4.1.2 Analisis Struktur Halaman Web	24
4.1.3 Implementasi Kode	26
4.1.4 Pengujian Kode.....	27
4.2 Eksplorasi Data Analisis.....	29
4.2.1 Dataset 1	29

4.2.2 Dataset 2	33
4.3 Splitting Data.....	35
4.4 Preprocessing Data	36
4.5 Implementasi Model.....	36
4.6 Evaluasi Algoritma <i>Random Forest</i>	38
BAB V_KESIMPULAN DAN SARAN	40
5.1 Kesimpulan.....	40
5.2 Saran	40
DAFTAR PUSTAKA	42
LAMPIRAN.....	45



DAFTAR TABEL

Table 2.1 Daftar Penelitian Terkait.....	4
Table 4.1 Hasil Evaluasi	38



DAFTAR GAMBAR

Gambar 3.1 Tahapan Penelitian	16
Gambar 3.2 Langkah-langkah <i>Web Scraping</i>	17
Gambar 3.3 Algoritma <i>Decision Tree</i>	19
Gambar 3.4 Algoritma <i>Random Forest</i>	20
Gambar 4.1 Halaman List Rumah.....	23
Gambar 4.2 Halaman Detail Rumah	24
Gambar 4.3 Elemen List Rumah Yang Menampilkan Alamat Detail Rumah.....	25
Gambar 4.4 Elemen Detail Rumah Yang Berisi Informasi Properti.....	26
Gambar 4.5 Struktur <i>Dataset 1</i>	30
Gambar 4.6 <i>Corelation Heatmap Dataset 1</i>	32
Gambar 4.7 Struktur <i>Dataset 2</i>	33
Gambar 4.8 <i>Correlation Heatmap Dataset 2</i>	34



DAFTAR LAMPIRAN

Lampiran 1 Bimbingan	45
Lampiran 2 Halaman Persetujuan	46
Lampiran 3 <i>Curriculum Vitae</i>	47
Lampiran 4 Surat Pernyataan HKI.....	48

