

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*