

42537 MAXIMILIANUS RAYMOND

by Maximilianus Raymond

Submission date: 15-Jun-2022 03:38PM (UTC+0700)

Submission ID: 1857217274

File name: 42537 MAXIMILIANUS RAYMOND.pdf (297.78K)

Word count: 1157

Character count: 7484

Prediksi Harga Rumah menggunakan Linear Regression

²
Jarot S. Suroso
Information Systems Management
Department
Bina Nusantara University
Jakarta 11480, Indonesia
jsembodo@binus.edu

Maximilianus Raymond Kusnadi
CyberSecurity
Department
Bina Nusantara University
Jakarta 11480, Indonesia
maximilianus.kusnadi@binus.ac.id

Justin Orlean
Master Track Teknik Informatika
Department
Bina Nusantara University
Jakarta 11480, Indonesia
justn.orlean@binus.ac.id

Abstract— Rumah merupakan tempat tinggal serta berlingkungnya manusia dari segala ancaman alam liar. Rumah adalah salah satu kebutuhan primer yang cukup penting dikehidupan saat ini. Oleh karena itu, untuk membeli dan mempersiapkan kebutuhan primer tersebut manusia perlu merencanakan kebutuhan biaya yang akan dikeluarkan untuk membeli atau membangun sebuah rumah. Maka dari itu, tujuan penelitian ini adalah untuk membuat prediksi harga rumah yang menggunakan metode machine learning yang menggunakan algoritma linear regression.

Keywords— linear regression, house prices, prediction, machine learning, harga rumah

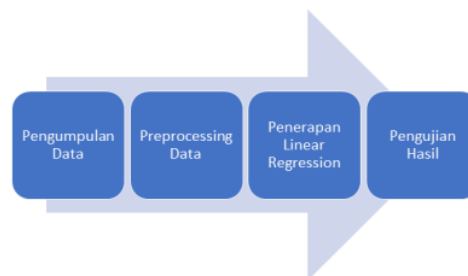
I. INTRODUCTION

Rumah merupakan bangunan yang terletak di atas tanah yang merupakan bagian dari kebutuhan primer banyak orang. Rumah digunakan sebagai tempat berlindungnya seseorang dari cuaca luar dan merupakan tempat untuk berkumpul bagi sebagian orang. Rumah dapat digunakan menjadi tempat istirahat dan juga tempat sebagian orang untuk melakukan kegiatan sehari-hari didalam rumah. Tidak hanya itu, rumah juga seringkali digunakan oleh banyak orang sebagai alat untuk investasi, dikarenakan pergerakan harga rumah yang semakin melonjak terkait banyaknya jumlah orang yang memiliki kebutuhan untuk memiliki rumah [1]. Dikarenakan oleh hal itu, maka prediksi harga rumah sangat bermanfaat untuk digunakan sebagai alat bantu untuk mempertimbangkan keputusan dalam pembelian rumah.

Pada penelitian ini, kami akan menerapkan metode linear regression terhadap suatu dataset untuk memprediksi harga rumah. Pada akhir penelitian, prediksi dari hasil machine learning akan dibandingkan dengan harga asli dari dataset untuk mengetahui akurasi dari model. Untuk membuat prediksi harga rumah tersebut kita menggunakan algoritma linear regression pada machine learning dalam penelitian ini, yang pernah dikembangkan juga oleh penelitian Sefto P. [2] untuk memprediksi harga tanah yang menggunakan variabel seperti banyaknya kamar tidur, tempat parkir, bahkan banyaknya kamar mandi. Manfaat dari penelitian ini sendiri adalah untuk mempelajari penerapan machine learning yang digunakan untuk memprediksi harga rumah dengan algoritma linear regression.

II. METODE PENELITIAN

Penelitian ini akan menggunakan metode linear regression terhadap suatu dataset untuk memprediksi harga rumah. Setelah dataset sudah terkumpul, maka akan dilakukan preprocessing terhadap dataset tersebut sebelum diterapkan algoritma linear regression. Tahap akhir dari penelitian ini adalah menguji prediksi yang dikeluarkan oleh model linear regression terhadap nilai asli untuk mengetahui tingkat keakuratan model. Alur penelitian ditunjukkan oleh gambar 1 di bawah.



Gambar 1. Alur Penelitian

A. Pengumpulan Data

Dataset untuk penelitian ini bernama "House price prediction" diambil dari website Kaggle dan dibuat oleh user bernama Shree. Dataset berisi 4600 data rumah yang masing-masing memiliki 18 atribut. Data diambil pada tahun 2014 dan lokasi rumah berada di berbagai kota di Amerika Serikat.

B. Data Preprocessing

Sebelum linear regression diterapkan pada dataset, perlu dilakukan preprocessing agar atribut-atribut pada dataset dapat sesuai dengan model linear regression. Selain itu preprocessing berguna untuk meningkatkan akurasi dari prediksi dengan cara menghapus, mengubah, atau menggabungkan beberapa atribut. Beberapa hal yang dilakukan pada data preprocessing adalah menghapus atribut yang redundan atau irelevan, mengubah atribut kategorik menjadi numerik, dan membagi dataset menjadi data untuk training dan testing.

C. Linear Regression

Linear regression adalah sebuah metode linear untuk menemukan relasi antara independent dan dependent variable [3]. Jika diberikan sebuah dataset $\{y_1, x_1, x_2, \dots, x_n\}$, dengan y sebagai dependent variable, setiap x sebagai independent variable, dan n sebagai jumlah data dalam dataset, maka model linear regression akan menganggap relasi antar variabel-variabel itu linear. Oleh karena itu, model linear regression akan mengikuti fungsi persamaan linear yang ditunjukkan oleh (1).

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_p x_p + \varepsilon \quad (1)$$

D. Pengujian Hasil

Setelah linear regression diterapkan menggunakan data training, hasil prediksi akan dibandingkan dengan data aktual untuk mengetahui akurasi dari model linear regression. Prediksi akan dilakukan menggunakan data testing yang berbeda dengan data training. Hasil akan dievaluasi menggunakan tiga macam metrik, yaitu R-Squared (R^2), Root Mean Squared Error (RMSE), dan Mean Absolute Error (MAE). R-Squared digunakan untuk mencari proporsi dari variasi hasil dependent variable yang dapat diprediksi dari independent variable. R-Squared mengukur seberapa baik model dapat memprediksi hasil dari independent variable yang diberikan. R-Squared didapat dari persamaan (2).

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

Variabel y_i adalah hasil prediksi model dan y adalah rata-rata dependent variable. Root Mean Squared Error digunakan untuk mencari jumlah seluruh kuadrat selisih antara nilai asli dengan prediksi, kemudian diakarkan. RMSE berfungsi sebagai indikator seberapa jauh prediksi model dengan data asli. RMSE didapat dari persamaan (3).

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n}} \quad (3)$$

Mean Absolute Error digunakan untuk mencari rata-rata penyimpangan antara data asli dengan hasil prediksi. MAE didapatkan dari persamaan (4).

$$MAE = \frac{\sum_{i=1}^n |y_i - \hat{y}_i|}{n} \quad (4)$$

III. HASIL DAN PEMBAHASAN

Sebelum dilakukan training dan testing, dilakukan feature selection untuk memilih feature yang penting dari dataset. Dari proses ini, feature *date*, *country*, *street*, dan *statezip* dikeluarkan dari dataset karena setiap data memiliki nilai *statezip*, *street*, dan *date* yang berbeda dan memiliki *country* yang sama.

Dataset dibagi ke dalam data training dan data testing dengan perbandingan 4:1. Sebelum dilakukan prediksi, model dilatih menggunakan data training agar dapat digunakan untuk memprediksi harga rumah. Setelah dilakukan training, maka dapat dilakukan testing menggunakan data testing.

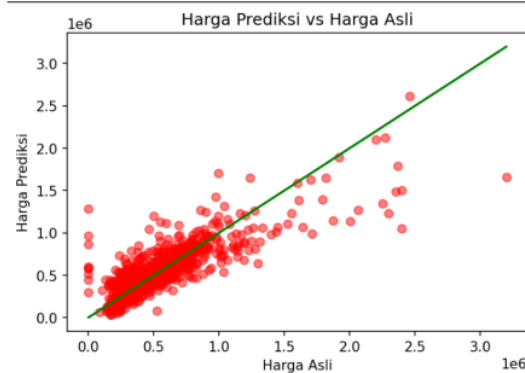
Kami melakukan dua kali training dan testing, sekali menggunakan atribut *city* yang diubah menjadi numerik (0 atau 1) dan sekali tanpa menggunakan atribut *city*. Hasil prediksi kemudian dievaluasi dengan hasil aktualnya menggunakan R-Squared, Root Mean Squared Error, dan Mean Absolute Error. Tabel berikut menunjukkan hasil evaluasi dari kedua testing.

Tabel 1. Evaluasi Hasil Prediksi

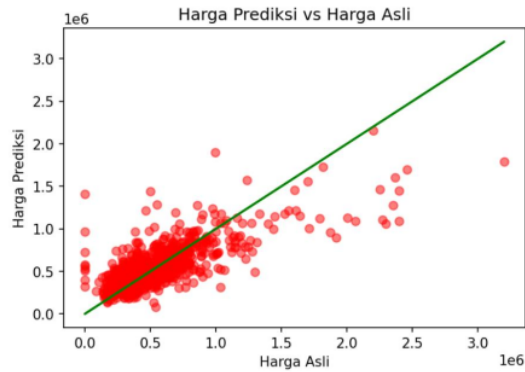
	R2	RMSE	MAE
Tanpa Atribut City	0.552554923	231978.5209	154702.3051
Dengan Atribut City	0.667652488	199928.2312	128156.0059

Dari hasil evaluasi, didapati bahwa model yang dilatih menggunakan dataset yang memiliki atribut *city* lebih akurat dibandingkan dataset yang tidak memiliki atribut *city*. Nilai R-Squared prediksi model yang menggunakan atribut *city* lebih tinggi, yang menandakan bahwa prediksi lebih akurat dibandingkan yang tidak menggunakan atribut *city*. Nilai RMSE dan MAE yang lebih rendah mengindikasikan rata-rata kesalahan yang lebih sedikit.

Berikut plot yang menggambarkan data asli yang dibandingkan dengan hasil prediksi.



Gambar 2. Harga Prediksi vs Harga Asli (Dataset menggunakan atribut *city*)



Gambar 3. Harga Prediksi vs Harga Asli (Dataset tanpa atribut *city*)

IV. KESIMPULAN

Kami melakukan prediksi harga rumah menggunakan dataset yang berisi 4600 data berbeda yang masing-masing memiliki 18 atribut. Dilakukan preprocessing data berupa pemilihan feature untuk mengoptimasi kinerja model.

Kemudian dilakukan pembagian data untuk training dan testing dengan proporsi 4:1. Digunakan model linear regression untuk memprediksi harga rumah berdasarkan dataset.

Model dilatih menggunakan dua macam training data, yaitu data yang memiliki atribut *city* dan yang tidak memiliki atribut *city*. Setelah dilakukan training dan prediksi menggunakan data testing, ditemukan bahwa hasil prediksi model yang dilatih menggunakan data training yang memiliki atribut *city* memiliki akurasi yang lebih baik dibandingkan model yang dilatih menggunakan dataset tanpa atribut *city*. Dapat disimpulkan bahwa semakin banyak feature relevan yang digunakan dalam training, maka hasil prediksi akan semakin baik.

REFERENCES

- [1] Utama, C. (2011). Rumah, Layakkah Menjadi Aset Untuk Investasi?. *Bina Ekonomi*, 15(2).
- [2] Sefto, P. (2016). Prediksi Harga Tanah Menggunakan Algoritma Linear Regression. *Technologia*, 7(2), pp. 122-129.
- [3] Yan, Xin (2009). Linear Regression Analysis: Theory and Computing, *World Scientific*, pp. 1-2, ISBN 9789812834119.

42537 MAXIMILIANUS RAYMOND

ORIGINALITY REPORT

12%

SIMILARITY INDEX

11%

INTERNET SOURCES

7%

PUBLICATIONS

0%

STUDENT PAPERS

PRIMARY SOURCES

1	www.aasmr.org Internet Source	2%
2	Jarot S. Suroso, Sri Mumpuni Ngesti Rahaju, Kusnadi. "Evaluation Of IS Risk Management Using Octave Allegro In Education Division", 2018 International Conference on Orange Technologies (ICOT), 2018 Publication	1%
3	coek.info Internet Source	1%
4	jurnal.stmikroyal.ac.id Internet Source	1%
5	www.coursehero.com Internet Source	1%
6	ejournal.gunadarma.ac.id Internet Source	1%
7	desaingambarrumahminimalis.com Internet Source	1%
8	ejnteti.jteti.ugm.ac.id Internet Source	

1 %

9

publikasi.dinus.ac.id

Internet Source

1 %

10

Wijoyo Santoso, Iskandar Simorangkir.
"PENGENDALIAN MONETER DALAM SISTEM
NILAI TUKAR YANG FLEKSIBEL (Konsiderasi
kemungkinan penerapan inflation targeting di
Indonesia)", Buletin Ekonomi Moneter dan
Perbankan, 2003

Publication

1 %

11

journal.itera.ac.id

Internet Source

1 %

12

pji.ub.ac.id

Internet Source

1 %

Exclude quotes On

Exclude matches Off

Exclude bibliography On

42537 MAXIMILIANUS RAYMOND

GRADEMARK REPORT

FINAL GRADE

/0

GENERAL COMMENTS

Instructor

PAGE 1

PAGE 2

PAGE 3
