

Penerapan Metode *Multiple Linear Regression* Untuk Prediksi Harga Sewa Kamar Kost

Muhamad Reza Fahlepi^{#1}, Andreas Widjaja^{*2}

Jurusan SI Teknik Informatika, Universitas Kristen Maranatha

Jl. Surya Sumantri No. 65 Bandung

¹Remcdohl@gmail.com

²andreas.widjaja@it.maranatha.edu

Abstract — . The increasing needs of temporary housing is one of the reasons for the increasing number of house or special building that provide *kost*, a temporary housing term in Indonesia. One of the issue that resurface is how to determine or knowing the fee set by the landlord. This study will propose linear regression method to solve the problem. Linear regression method that used are Multiple linear regression with the help of Scikit-learn. Determining factors to the rent fees that are chosen in this study are: A/C availability, internet availability, on-room toilet availability, gender- limited status, kitchen availability, hot water availability and the size of the room itself. Analysis method that are done for the model prove that model has the correlation coefficient of 0.70 and are better compared to several others testing done with RSME result of 334,578.63.

Keywords— machine learning, linear regression, multiple linear regression, scikit -learn.

I. PENDAHULUAN

Tidak sedikit pemilik lahan atau rumah membuka usaha untuk menyediakan tempat tinggal sementara yang disebut rumah kost. Semakin meningkatnya kebutuhan akan tempat tinggal merupakan salah satu alasan bertambahnya jumlah rumah atau bangunan khusus yang menawarkan jasa rumah kost [1]. Jasa kost melibatkan sejumlah pembayaran yang dibebankan untuk penyewa kost pada periode tertentu, yang biasanya dihitung perbulan [1].Salah satu permasalahan yang dihadapi adalah bagaimana mengetahui biaya yang akan dikeluarkan oleh calon penyewa kost dengan melihat fasilitas yang diinginkan oleh calon penyewa kost itu sendiri dan patokan harga yang sesuai bagi penyedia jasa sewa kost.

Metode yang diusulkan dalam penelitian ini adalah dengan menggunakan perhitungan regresi linear berganda. Metode tersebut dipilih untuk mengetahui nilai variabel terikat dengan menghitung pengaruh dari beberapa variable bebas. Sistem yang dikembangkan diharapkan dapat membantu memprediksi kisaran harga yang akan ditetapkan oleh penyedia jasa sewa dan membantu calon penyewa kost dalam mengetahui kisaran biaya yang dibutuhkan untuk menyewa rumah kost berdasarkan fasilitas kost.

II. TINJAUAN PUSTAKA

Kata “Regresi” dikemukakan oleh seorang antropolog dan ahli meteorologi terkenal dari Inggris Sir Francis Galton (1822-1911). Galton [2] melakukan eksperimen terhadap biji tanaman dengan berbagai macam ukuran tetapi dalam satu spesies yang sama. Galton mengemukakan regresi dengan melihat hasil eksperimen yang menjelaskan bahwa biji keturunan tidak cenderung menyerupai biji induknya dalam hal ukuran, namun lebih kecil dari pada induknya, jika induknya lebih besar dan biji keturunan lebih besar dari pada induknya, jika induknya lebih kecil.

Regresi linier merupakan salah satu teknik data mining untuk memprediksi suatu nilai di masa yang akan datang berdasarkan pada hubungan linier dari variabel yang lainnya [3]. Regresi linier akan

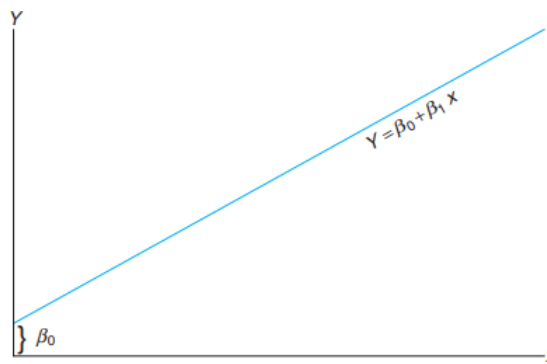
menghasilkan model hubungan antara variabel terikat dengan variabel bebas. Terdapat dua macam regresi linier dilihat dari jumlah variabel bebas (independent variable) yaitu regresi linier sederhana dan regresi linier berganda [4].

A. Regresi Linier Sederhana

Regresi linier sederhana adalah model hubungan antara satu variabel bebas dengan satu variabel terikat [5]. Persamaan Regresi linier sederhana dapat dilihat dengan persamaan 1 sebagai berikut :

$$Y = \beta_0 + \beta_1 x, \quad (1)$$

Y merupakan nilai yang diprediksi atau disebut variabel terikat (*dependent variable*) dan x merupakan nilai suatu data atau disebut variabel bebas (*independent variable*). β_0 adalah nilai rata-rata variabel Y jika variabel x bernilai 0 (intercept) dan β_1 adalah nilai koefisien untuk variabel bebas (*slope*).



Gambar 1. Hubungan Linier Intercept dan slope.

B. Regresi Linier Berganda

Regresi linier berganda adalah model hubungan antara lebih dari satu variabel bebas dengan satu variabel terikat [4]. Persamaan Regresi Berganda sederhana dapat dilihat dengan persamaan 2 sebagai berikut :

$$y = \beta_0 + \beta_1 x_1 + \dots + \beta_{p-1} x_{p-1} \quad (2)$$

Y_i merupakan variabel terikat (*dependent variable*) untuk pengamatan ke- i , $i=1,2,\dots,n$. $\beta_1, \beta_0, \dots, \beta_{p-1}$ merupakan parameter dan x_1, \dots, x_{p-1} merupakan variabel bebas (*independent variable*). Pada notasi matriks persamaan (2) dapat ditulis menjadi persamaan (3) sebagai berikut :

$$\hat{Y} = X \beta \quad (3)$$

$n \times 1 \quad n \times p \quad p \times 1$

Dengan :

$$\mathbf{Y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} \quad \boldsymbol{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix} \quad \mathbf{X} = \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{bmatrix} \quad (4)$$

dapat ditulis dengan persamaan matriks :

$$\mathbf{X}^T \mathbf{X} \hat{\boldsymbol{\beta}} = \mathbf{X}^T \mathbf{Y} \quad (5)$$

dimana jika $\mathbf{X}^T \mathbf{X}$ tidak bernilai 0 atau nonsingular, maka solusi persamaan normal (normal equations) yang dipakai untuk regresi linier berganda dari persamaan (5):

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{Y} \quad (6)$$

C. Scikit-learn

Salah satu modul Python penyedia berbagai macam algoritma machine learning yaitu Scikit-learn. Scikit-learn tersedia dalam bentuk library python. *Scikit-learn* menggunakan task-oriented interface yang konsisten sehingga mudah dalam melakukan perbandingan antar metode [6].

D. Flask

Flask adalah Web Application Framework untuk mempermudah dalam pengembangan atau pembangunan website. Flask ditulis dalam bahasa pemrograman Python dan tersedia dalam bentuk Library dan dapat di-install melalui *Package Manager* Python.

III. ANALISIS DAN RANCANGAN

Berbagai macam metode prediksi digunakan untuk pemecahan masalah dengan menganalisa data yang telah dikumpulkan. Secara umum metode machine learning yang digunakan untuk prediksi terbagi menjadi dua, yaitu time series dan klasifikasi. Prediksi Time Series mengacu proses dimana nilai-nilai yang didapat untuk masa yang akan datang berasal dari informasi yang diperoleh dari data masa lalu dan titik data saat ini [7]. Sedangkan model klasifikasi disajikan dengan serangkaian contoh yang telah diklasifikasikan dan dari contoh-contoh tersebut, model klasifikasi belajar untuk menetapkan label kelas untuk pengamatan baru. Contoh penerapan menggunakan prediksi *time series* seperti [8]. Contoh penerapan menggunakan prediksi klasifikasi seperti [9] yang melakukan klasifikasi untuk memprediksi performa mahasiswa.

Sistem yang dikembangkan ditujukan untuk calon penyewa kamar kost dan untuk pemilik kost dalam membantu mengetahui kisaran harga sewa kamar kost berdasarkan fasilitas yang dipilih. Untuk membantu dalam permasalahan memprediksi harga sewa kamar kost, tugas akhir ini akan membahas tentang pembuatan sistem prediksi harga kost berbasis web. Dari berbagai macam metode prediksi, sistem yang dikembangkan termasuk kedalam prediksi klasifikasi, tidak seperti [7].

Regresi linier berganda akan digunakan untuk perhitungan prediksi. Model akan dibangun menggunakan fungsi dari scikit-learn terhadap data yang disimpan dalam file csv.

A. Penunjang Penelitian

Untuk memperlancar dan membantu dalam proses pembuatan penelitian tugas akhir ini, maka diperlukan bahan dan peralatan yang meliputi :

a. Bahan

Bahan yang dibutuhkan dan digunakan dalam penelitian tugas akhir ini adalah data rumah kost yang berada disekitar kampus Universitas Kristen Maranatha yang beralamat di Jl. Surya Sumantri No.65, Sukawarna, Sukajadi, Kota Bandung, Jawa Barat.

b. Peralatan

Kebutuhan perangkat keras (hardware) dan kebutuhan perangkat lunak (software) diperlukan untuk kebutuhan penelitian tugas Akhir ini. Kebutuhan perangkat keras dan perangkat lunak .

Perangkat keras yang digunakan penulis:

- a. Laptop ASUS X550I Processor AMD Quad Core FX-9830P CPU @ 3.7GHz
- b. RAM 8GB
- c. Mouse

Perangkat lunak yang digunakan penulis:

- a. Sistem Operasi Windows 10 Creator 64-bit
- b. Microsoft Office Word Professional 2016

Software ini digunakan untuk proses penulisan hasil penelitian.

- c. Microsoft Office Excel Professional 2016
- Software ini digunakan untuk penulisan dataset.

- d. Anaconda Navigator

Software Anaconda Navigator merupakan package manager yang juga menyediakan library dan tools untuk python.

- e. Jupyter Notebook

Jupyter Notebook digunakan untuk proses pengujian dan visualisasi data yang digunakan dalam penelitian.

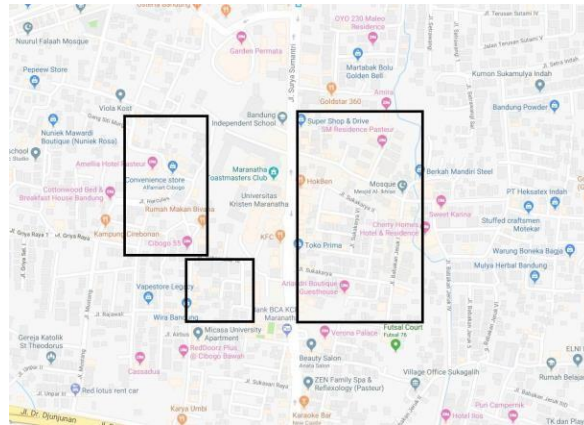
- f. Visual Studio Code

Visual Studio Code merupakan IDE yang digunakan penulis untuk editor kode.

B. Prosedur Pengumpulan Data

Data yang digunakan dalam penelitian ini merupakan data yang diperoleh menurut sifatnya, yaitu data kualitatif dan data kuantitatif.

Data kualitatif merupakan data yang diperoleh bukan dalam bentuk bilangan, seperti ada atau tidak ketersediaan fasilitas AC, internet, laundry, kamar mandi dalam, kost khusus, fasilitas dapur, air panas dan luas kamar. Data kuantitatif merupakan data yang diperoleh berupa bilangan, yaitu harga sewa.



Gambar 2. Lokasi Pengumpulan Data

Pengumpulan data diperoleh dengan wawancara secara langsung dari sumber yang menjadi objek penelitian. Sumber yang menjadi objek penelitian merupakan penjaga rumah kost atau pemilik rumah kost tersebut. Data pengujian diambil dari daerah sekitar Universitas Kristen Maranatha yang pemetaannya dapat dilihat pada gambar 2 yang ditandai dengan kotak berwarna hitam.

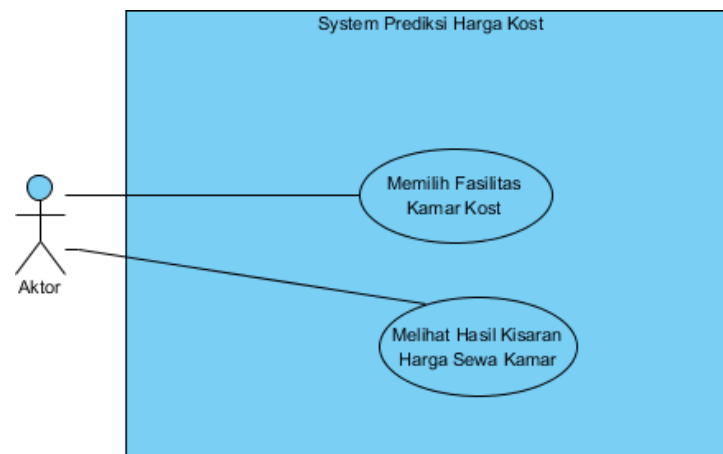
TABEL 1.
CONTOH MODEL DATA

AC (X_1)	Internet (X_2)	Laundry (X_3)	kamar mandi dalam (X_4)	Kost Khusus (X_5)	Fasilitas dapur (X_6)	Air Panas (X_7)	Luas Kamar (X_8)	Harga sewa (Y)
tidak	ya	tidak	ya	tidak	ya	Ya	kecil	1450000
tidak	ya	tidak	ya	tidak	ya	ya	sedang	1850000
tidak	ya	tidak	ya	tidak	ya	ya	besar	2100000
tidak	ya	tidak	tidak	tidak	ya	tidak	kecil	800000
tidak	ya	tidak	ya	tidak	ya	tidak	kecil	900000
tidak	ya	tidak	ya	ya	ya	tidak	kecil	560000
tidak	tidak	tidak	tidak	tidak	ya	tidak	kecil	750000
tidak	tidak	tidak	tidak	tidak	ya	tidak	besar	850000
ya	tidak	tidak	ya	tidak	ya	ya	sedang	1500000
ya	tidak	tidak	ya	tidak	ya	ya	besar	2000000
tidak	ya	tidak	tidak	ya	ya	tidak	kecil	500000
ya	ya	ya	ya	tidak	ya	ya	sedang	2000000
ya	ya	ya	ya	tidak	ya	ya	besar	3000000

Contoh model data yang dikumpulkan merupakan harga sewa kamar kost dan data fasilitas kost yang berupa tersedia atau tidak fasilitas AC, internet, kamar mandi dalam, kost khusus, fasilitas dapur dan air panas. Luas kamar yang di peroleh merupakan wawancara hasil perkiraan ukuran kamar kost oleh pemilik kost yang dikelompokkan menjadi kecil, sedang dan besar, dengan asumsi jika luas kamar dibawah atau sama dengan 3x3 maka kamar termasuk ukuran kecil, jika ukuran kamar lebih dari 3x3 atau kurang dari sama dengan 3x4 maka ukuran kamar termasuk ukuran sedang, dan jika ukuran

kamar lebih dari 3x4 maka ukuran kamar termasuk ukuran besar. Tabel contoh model data dapat dilihat pada tabel 1.

C. Use Case



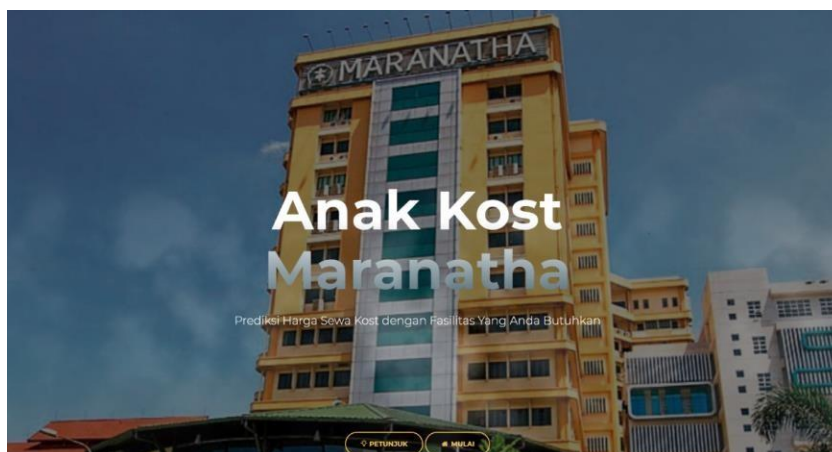
Gambar 3 Use case diagram dari Sistem Prediksi Harga Sewa Kamar Kost.

Berikut merupakan penjelasan Use case diagram pada gambar 3, calon penyewa kost dan penyedia jasa kost dianggap satu aktor pengguna yang sama. Aktor akan memilih kriteria fasilitas kost yang sesuai dan yang diinginkan. Model use case dapat dilihat pada gambar 3 .

IV. IMPLEMENTASI

A. Implementasi Tampilan Desain Antar Muka

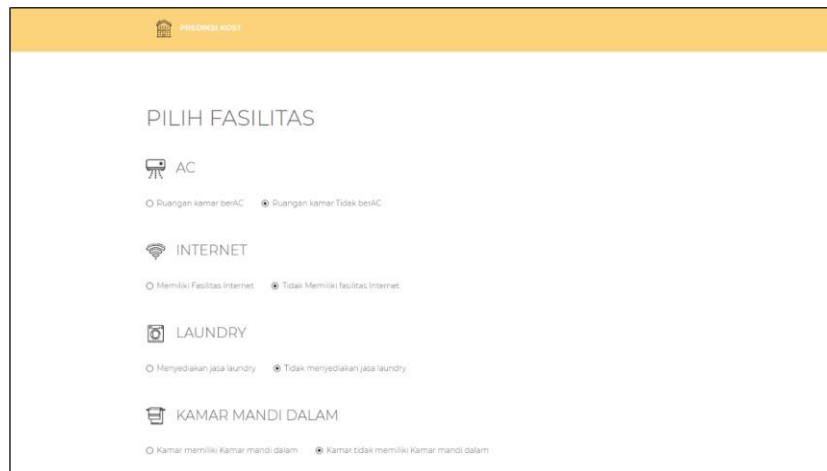
Gambar 4 merupakan tampilan awal aplikasi web berupa home dan penjelasan tentang teknik perhitungan yang digunakan dalam memprediksi harga sewa kamar kost. Terdapat 2 pilihan yang dapat dipilih user yaitu, petunjuk yang berupa penjelasan tentang aplikasi prediksi harga kost dan mulai untuk berpindah ke halaman selanjutnya pada web.



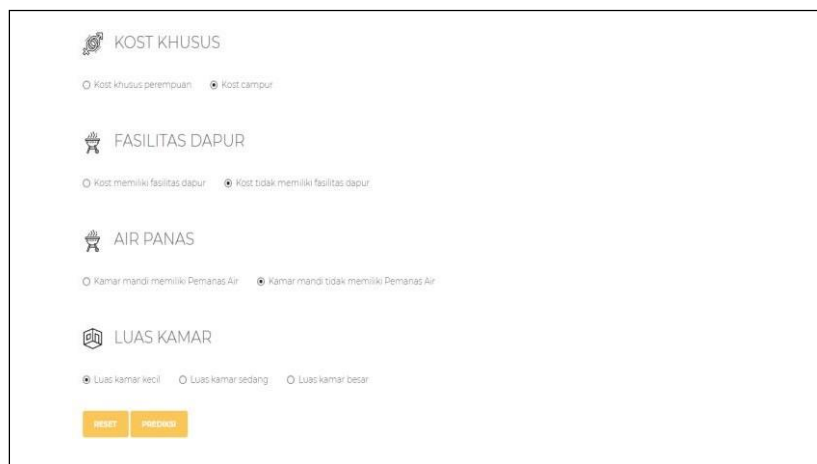
Gambar 4 Implementasi tampilan Home.

User dapat memilih fasilitas kamar kost yang akan diprediksi nilainya menggunakan radio button. Pilihan fasilitas kamar pada form di halaman web gambar 5 dan gambar 6 berupa fasilitas AC, internet,

laundry, kamar mandi dalam, kost khusus, fasilitas dapur, air panas dan luas kamar. Terdapat dua pilihan *button* yang dapat dipilih user yaitu, pilihan reset digunakan untuk mereset pilihan fasilitas kamar kost yang telah dipilih user menjadi default dan pilihan prediksi untuk memprediksi harga sewa kost sesuai dengan pilihan fasilitas kost yang di pilih oleh user.



Gambar 5. Implementasi tampilan pilih fasilitas kamar kost.



Gambar 6. Implementasi tampilan pilih fasilitas kamar kost.

Gambar 7 merupakan tampilan fasilitas yang dipilih oleh user dan hasil dari perhitungan menggunakan multiple linear regression ditampilkan pada halaman tersebut.



Gambar 7. Implementasi tampilan hasil prediksi.

B. Data Preprocessing

Data yang dikumpulkan akan dilist kedalam tabel. File data yang digunakan sebagai data training untuk perhitungan prediksi berupa file dalam bentuk format CSV. File csv akan disimpan ke dalam folder bernama static yang diberi nama datatest.csv. nilai data akan diubah sesuai seperti contoh data pada tabel 3.1, jika data bernilai ya, maka akan diubah menjadi nilai 1, jika nilai tidak akan diubah menjadi nilai 0, jika data bernilai kecil maka diubah menjadi nilai 0, jika data bernilai sedang maka diubah menjadi nilai 1, jika data bernilai besar maka diubah menjadi nilai 2 seperti pada tabel 2.

TABEL 2.

CONTOH HASIL DATA YANG TELAH DIUBAH.

AC (X_1)	Internet (X_2)	laundry (X_3)	kamar mandi dalam (X_4)	Kost Khusus (X_5)	Fasilitas dapur (X_6)	Air Panas (X_7)	Luas Kamar (X_8)	Harga sewa (Y)
0	1	0	1	0	1	1	0	1450000
0	1	0	1	0	1	1	1	1850000
0	1	0	1	0	1	1	2	2100000
0	1	0	0	0	1	0	0	800000
0	1	0	1	0	1	0	0	900000
0	1	0	1	1	1	0	0	560000
0	0	0	0	0	1	0	0	750000
0	0	0	0	0	1	0	2	850000
1	0	0	1	0	1	1	1	1500000
1	0	0	1	0	1	1	2	2000000
0	1	0	0	1	1	0	0	500000
1	1	1	1	0	1	1	1	2000000
1	1	1	1	0	1	1	2	3000000

V. PENGUJIAN

Pengujian model *multiple linear regression* membahas mengenai data yang digunakan dalam pengujian yang dilakukan menggunakan Jupyter. Pengujian yang dibahas dalam pengujian ini adalah membandingkan pengaruh lebih dari satu variabel bebas (x) terhadap variabel terikat dengan berdasarkan korelasi hubungan antara setiap variabel bebas (x) dan variabel terikat (y). Pengujian dengan mengukur tingkat keakuratan pada model *multiple linear regression* dengan menggunakan Jupyter juga akan di bahas pada subbab ini .

A. Skema Pengujian Model Multiple Linear regression

Untuk menguji data secara keseluruhan, metode perhitungan akurasi tersendiri akan diterapkan. Metode yang diterapkan akan menghitung nilai total jumlah prediksi yang benar pada data menggunakan multiple linear regression. Persamaan dapat dilihat seperti berikut :

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \beta_5 X_5 + \beta_6 X_6 + \beta_7 X_7 + \beta_8 X_8$$

Untuk menentukan koefisien-koefisien regresi ($\beta_0, \beta_1, \beta_2, \beta_3, \beta_4, \beta_5, \beta_6, \beta_7, \beta_8$) dapat diketahui dengan menggunakan kode berikut;

```
y_train = data['HARGA_SEWA']
x_train = data.drop(['HARGA_SEWA'], axis=1)
model = LinearRegression().fit(x_train,y_train)

model.intercept_      #untuk nilai  $\beta_0$ 
model.coef_           #untuk nilai  $\beta_1, \beta_2, \beta_3, \beta_4, \beta_5, \beta_6, \beta_7, \beta_8$ 
```

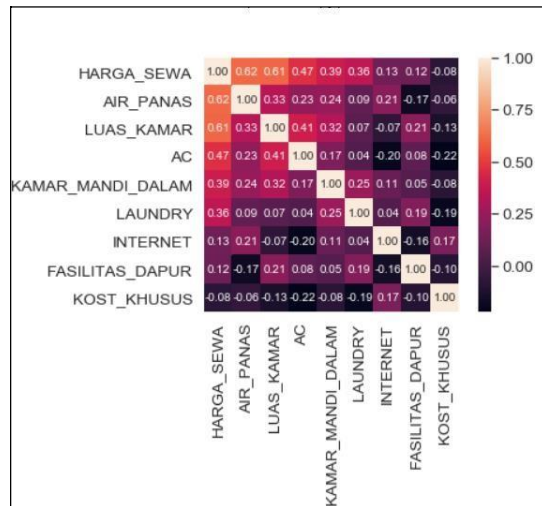
Dari kode diatas, maka diperolehlah nilai koefisien- koefisien linier sebagai berikut :

$$\begin{aligned} \beta_0 &= 317730.49102680665 & \beta_5 &= 91540.10463399 \\ \beta_1 &= 409977.29082988 & \beta_6 &= 146704.43069977 \\ \beta_2 &= 104249.30951046 & \beta_7 &= 420413.2581289 \\ \beta_3 &= 272084.32580051 & \beta_8 &= 236517.81310676 \\ \beta_4 &= 73673.77803368 & & \end{aligned}$$

Koefisien yang diperoleh maka dibentuklah model persamaan multiple linier regression :

$$\begin{aligned} Y &= 317730.49102680665 + 409977.29082988 X_1 \\ &+ 104249.30951046 X_2 + 272084.32580051 X_3 \\ &+ 73673.77803368 X_4 + 91540.10463399 X_5 \\ &+ 146704.43069977 X_6 + 420413.2581289 X_7 \\ &+ 236517.81310676 X_8 \end{aligned}$$

Dari persamaan diatas, jika X_1 hingga X_n bernilai nol, maka Y bernilai 317730.49102680665. Hasil korelasi dari data yang telah dimodelkan yaitu 0.7042373557064.



Gambar 8 Korelasi Hubungan antar variabel dengan heatmap.

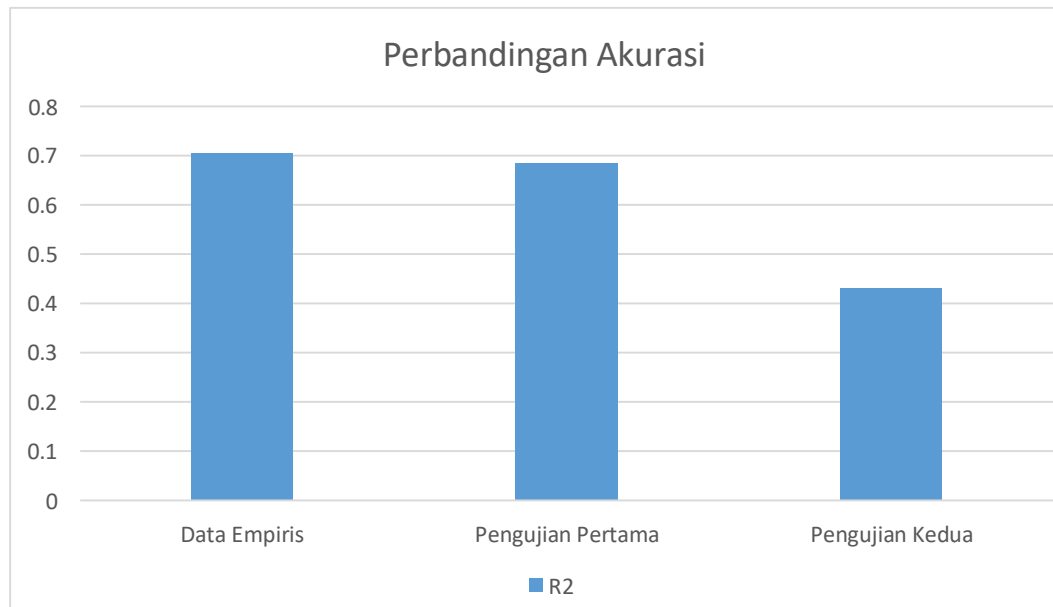
Gambar 8 Merupakan skema gambar hubungan antar variabel. Jika angka pada tiap kotak mendekati 1 maka hubungan antar variabel sangat kuat. Dilihat dari gambar 5.2 dapat diperoleh informasi sebagai berikut :

1. 'AIR_PANAS' dan 'LUAS_KAMAR' memiliki nilai 0.62 dan 0.61 terhadap variabel 'HARGA_SEWA', Ini menunjukkan bahwa harga sewa rumah kost lebih berpengaruh dengan besar kecilnya luas kamar dan ada tidaknya fasilitas air panas.
2. 'FASILITAS_DAPUR' dan 'INTERNET' memiliki nilai 0.12 dan 0.13 terhadap variabel 'HARGA_SEWA', ini menunjukkan bahwa ada tidaknya fasilitas dapur dan internet tidak terlalu berdampak terhadap harga sewa kost.
3. 'KOST_KHUSUS' memiliki nilai -0.08 terhadap variabel 'HARGA_SEWA', nilai negatif disini menunjukkan bahwa hubungan variabel kost khusus berbanding terbalik dengan harga sewa, artinya jika rumah kost termasuk kost khusus maka harga sewa bisa lebih rendah.
4. 'AIR_PANAS' dan 'LUAS_KAMAR' memiliki nilai 0.33, serta 'AC' dan 'LUAS_KAMAR' memiliki nilai 0.41, ini menunjukkan bahwa hubungan kedua variabel tersebut memiliki keterkaitan, setiap rumah kost mewah rata-rata memiliki ukuran kamar yang besar, memiliki fasilitas air panas dan AC.

B. Pengujian Dampak Korelasi Antar Variabel Terhadap Akurasi Model

Pengujian pertama akan dilakukan dengan memotong tiga variabel yang memiliki nilai korelasi yang terkecil berdasarkan variabel 'HARGA_SEWA' yaitu 'INTERNET', 'FASILITAS_DAPUR' dan 'KOST_KHUSUS'. Hasil akurasi dari data yang telah dipotong yaitu 0.6843382654356439. pengujian kedua, data pengujian kedua akan dilakukan dengan memotong 2 variabel yang memiliki korelasi terbesar berdasarkan variabel 'HARGA_SEWA' yaitu 'AIR_PANAS' dan 'LUAS_KAMAR'. Hasil akurasi dari data yang telah dipotong yaitu 0.4301899022667727.

Gambar grafik perbandingan dapat dilihat pada gambar 9. Pada data empiris nilai R^2 memiliki nilai 0.7042373557064.



Gambar 9. Visualisasi nilai R square..

C. Pengujian Root Mean Squared Error

metode Root Mean Squared Error mengukur keakuratan model dari data yang di prediksi dengan data yang benar , dengan hasil data yang telah dibagi menjadi 30% data test dan 70% menggunakan library sklearn dimana index data dirandom terlebih dahulu sebagai berikut:

TABEL 3.
 PERHITUNGAN RMSE

Y_test	Hasil	Y_test - hasil	Y_test – hasil
950000	1095305	-145305	21113487889.61638
1450000	1100505	349495	122146522295.16702
800000	750416	49584	2458547954.971706
900000	1100505	-200505	40202388542.09218
500000	524326	-24326	591736885.150232
500000	524326	-24326	591736885.150232
500000	733688	-233688	54609972593.686745
1700000	1556471	143529	20600572367.156105
1500000	1469456	30544	932908300.7968733
3000000	1985989	1014011	1028219235768.3816
1300000	1650820	-350820	123074415850.45236
2000000	1667400	332600	110622458808.54951
800000	1073093	-273093	74580004029.27838
560000	1309867	-749867	562301204155.7146
1500000	1324897	175103	30660896684.75213
1600000	1990225	-390225	152275757472.75583
700000	783898	-83898	7038894942.430935
800000	783898	16102	259270462.20977014
610000	437311	172689	29821459707.45871
1200000	755899	444101	197225538682.07977
1100000	1324897	-224897	50578871148.84919
1800000	1761596	38404	1474835920.2179308
1300000	1534260	-234260	54877556844.06465
1400000	1419246	-19246	370412261.3934089
Jumlah			2686628686452.376

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

Keterangan :

Y_i = data sebenarnya (Y_{test})

\hat{Y}_i = data hasil estimasi (Hasil)

n = jumlah data

\hat{Y}_i didapat dari persamaan regresi :

$$\begin{aligned} Y = & 317730.49102680665 + 409977.29082988 X_1 \\ & + 104249.30951046 X_2 + 272084.32580051 X_3 \\ & + 73673.77803368 X_4 + 91540.10463399 X_5 \\ & + 146704.43069977 X_6 + 420413.2581289 X_7 \\ & + 236517.81310676 X_8 \end{aligned}$$

$$\begin{aligned} & \sqrt{\frac{2686628686452.376}{24}} \\ & = 334578.6334115131 \end{aligned}$$

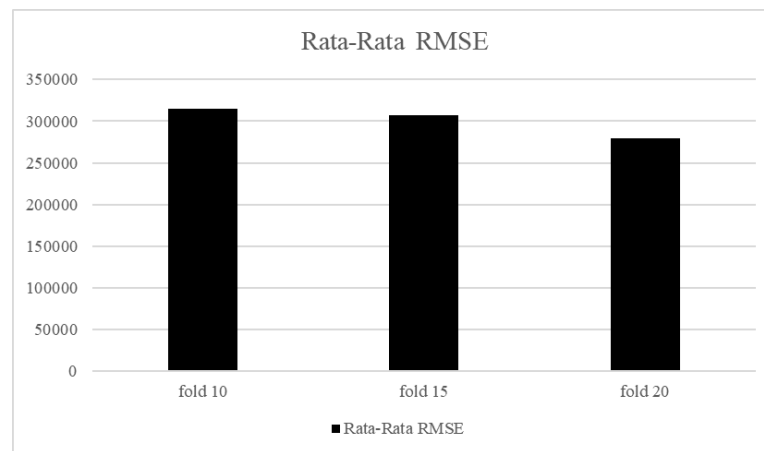
Hasil dari pengujian data menggunakan *Root Mean Squared Error* yaitu 334578.6334115131.

D. Pengujian dengan *k-Fold Cross Validation*

Pada pengujian berikutnya, diambil dengan tiga tahap pengujian menggunakan cross validation dengan melihat skor dari RMSE (Root Mean Squared Error) tiap tahapan. Tahap pertama diambil sebanyak 10 kali , untuk tahap kedua diambil data sebanyak 15 kali, tahap ketiga diambil data sebanyak 20 kali..

K-fold	Rata-rata RMSE
10	314627
15	306951
20	278859

berikut adalah gambar visualisasi rata-rata RMSE tiap pengujian yang dilakukan seperti gambar 10.



Gambar 10. Visualisasi rata-rata RMSE pada tiap pengujian.

VI. KESIMPULAN

Sistem yang dikembangkan dapat membantu calon penghuni kost atau pemilik dalam mengetahui kisaran harga sewa kamar kost dan mematok harga sewa kamar kos. Pengujian dengan data terkontrol membuktikan tingkat akurasi rata-rata sistem adalah 70% menggunakan metode *multiple linear regression*. Dari persamaan *multiple linear regression* yang telah didapat, jika X_1 hingga X_n bernilai nol, artinya jika kamar kost tidak memiliki fasilitas apapun dan luas kamar kecil maka kisaran prediksi harga bernilai 317730.49102680665. Berdasarkan hasil evaluasi, model data lebih tinggi tingkat akurasi dibandingkan dengan hasil perpotongan variabel data yang dilakukan pada pengujian pertama dan kedua. Dalam pengujian dampak korelasi pada hubungan antar variabel sangat menentukan biaya harga sewa.

Nilai RMSE yang didapat pada model data yang dipakai bernilai 334578. Pengujian cross validation berdasarkan RMSE memiliki hasil dengan selisih nilai yang tidak terlalu jauh.

DAFTAR PUSTAKA

- [1] A. Triansah, D. Cahyadi dan I. F. Astuti, "Membangun Aplikasi Web Dan Mobile Android Untuk Media Pencarian Kost Menggunakan Phonegap Dan Google Maps API," Jurnal Informatika Mulawarman, vol. 10, no. 1, p. 58, 2015..
- [2] F. Galton, "Regression toward Mediocrity in Hereditary Stature," The Journal of the Anthropological Institute of Great Britain and Ireland, vol. 15, pp. pp. 246-263, 1886.
- [3] M. K. Jiawei Han, Data Mining: Concepts and Techniques, 2nd, Waltham: Morgan Kaufmann, 2006..
- [4] J. A. Rice, Mathematical Statistics and Data Analysis Third Edition, Belmont: Cengage Learning, 2006.
- [5] R. E. Walpole, R. H. Myers, S. L. Myers dan K. Ye, Probability & Statistics for Engineers & Scientists Ninth Edition, Boston: Pearson Education, 1993.
- [6] F. Pedregosa, G. Varouquaux, A. Gramfort, V. Michael, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cornapeau, M. Brucher dan M. P.A, "Scikit-learn: Machine Learning in Python," Journal of Machine Learning Research, vol. 12, pp. 2825-2830, 2011.

- [7] E. Kayacan, B. Ulutas dan O. Kaynak, "Grey system theory-based models in time series prediction," *Expert System with Applications*, vol. 37, no. 2, pp. 1784-1789, 2010.
- [8] J. D. Salas, C.-h. Chung dan A. Cancelliere, "Correlations and Crossing Rates of Periodic-Stochastic Hydrologic Processes," *Journal of Hydrologic Processes*, vol. 10, no. 4, pp. 278 - 287, 2005.
- [9] K. Bunkar, R. Bunkar, U. K. Singh dan B. Pandya, "Data Mining: Prediction for Performance Improvement of Graduate Students using Classification," dalam Ninth Internal Conference on Wireless and Optical Communications Networks (WOCN), Indore, 2012.