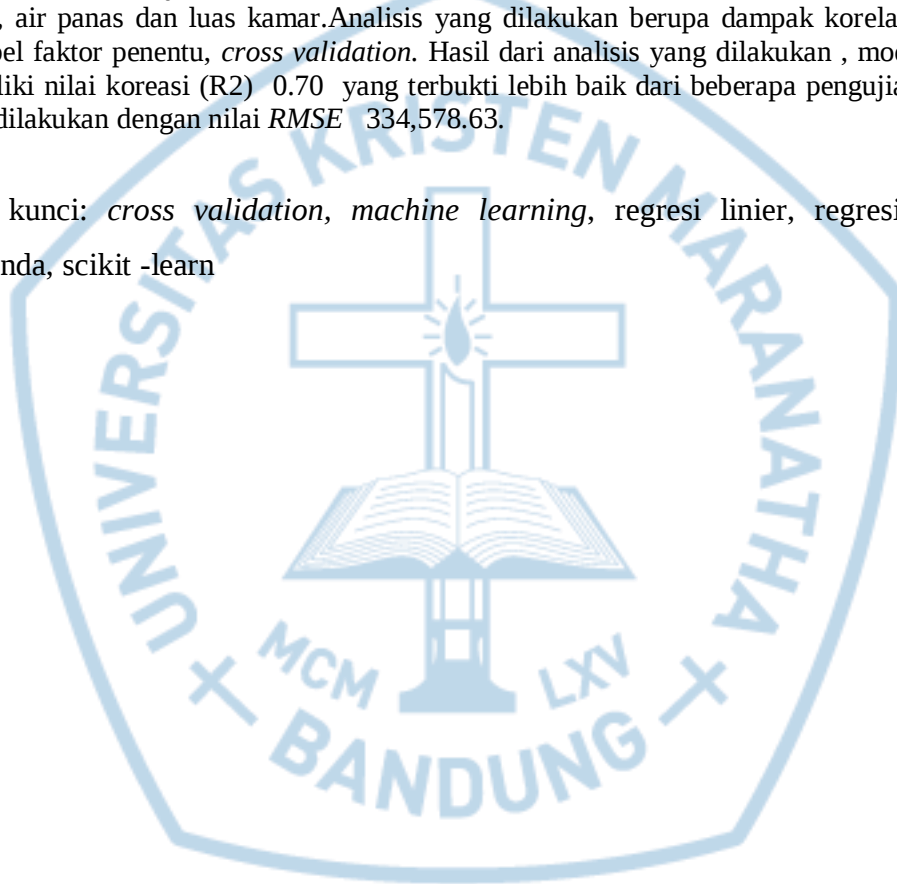


ABSTRAK

Regresi linier merupakan salah satu teknik *machine learning* untuk memprediksi suatu nilai di masa yang akan datang berdasarkan pada hubungan linier dari variabel yang lainnya. Salah satu penerapan metode regresi linier yaitu membangun sistem untuk memprediksi kisaran harga sewa kamar kost. Dengan aplikasi sistem prediksi harga sewa kamar kost bertujuan untuk membantu calon penyewa kost dan pemilik kost dalam mengetahui kisaran harga atau mematok harga sewa kamar kost berdasarkan fasilitas yang dipilih. Data yang diambil merupakan data empiris yang didapatkan secara langsung dari pemilik atau penjaga kost. Regresi linier yang digunakan adalah metode regresi linier berganda dengan menggunakan bantuan dari scikit-learn. faktor penentu dalam memprediksi harga kost yaitu tersedia atau tidaknya fasilitas AC, internet, kamar mandi dalam, kost khusus, fasilitas dapur, air panas dan luas kamar. Analisis yang dilakukan berupa dampak korelasi antar variabel faktor penentu, *cross validation*. Hasil dari analisis yang dilakukan, model data memiliki nilai korelasi (R^2) 0.70 yang terbukti lebih baik dari beberapa pengujian yang telah dilakukan dengan nilai *RMSE* 334,578.63.

Kata kunci: *cross validation*, *machine learning*, regresi linier, regresi linier berganda, scikit -learn



ABSTRACT

Linear regression is one of the machine learning technique that could be used to predict a value based on the linear relation from the other variables. One of the application for linear regression is to build a system for predicting a rent fees for kost (flat) room. Kost rent fees prediction system are envisioned to help a prospect renter and owner to know range and determine the rent fees based on facility provided. Data that will be used on this study is empirical data that are gathered by asking the owner or the caretaker. 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 (R^2) of 0.70 and are better compared to several others testing done with RSME result of 334,578.63.

Keywords: cross validation, machine learning, linear regression, multiple linear regression, scikit -learn



DAFTAR ISI

LEMBAR PENGESAHAN.....	i
PERNYATAAN ORISINALITAS LAPORAN PENELITIAN	ii
PERNYATAAN PUBLIKASI LAPORAN PENELITIAN.....	iii
PRAKATA	iv
ABSTRAK	v
ABSTRACT	vi
DAFTAR ISI	vii
DAFTAR GAMBAR.....	x
DAFTAR TABEL	xi
BAB 1 PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	2
1.3 Tujuan Pembahasan.....	2
1.4 Ruang Lingkup.....	2
1.5 Sumber Data	2
1.6 Sistematika Penyajian.....	3
BAB 2 KAJIAN TEORI	4
2.1 Regresi Linier.....	4
2.1.1 Regresi Linier Sederhana.....	4
2.1.2 Regresi Linier Berganda	5
2.2 Scikit-learn.....	7
2.3 Flask	7
BAB 3 ANALISIS DAN RANCANGAN SISTEM.....	8
3.1 Sistem yang dikembangkan	8

3.2 Penunjang Penelitian	9
3.2.1 Bahan	9
3.2.2 Peralatan.....	9
3.3 Prosedur Pengumpulan Data.....	10
3.4 Teknik Analisis Data.....	11
3.5 Interaksi Sistem Dengan Pengguna.....	12
3.5.1 <i>Use Case</i>	12
3.5.2 <i>Activity Diagram</i>	12
3.6 Desain Tampilan Antar Muka.....	13
BAB 4 IMPLEMENTASI.....	15
4.1 Implementasi Tampilan Desain Antar Muka.....	15
4.2 Instalasi library Scikit-Learn	17
4.3 Implementasi Program	17
4.3.1 Flask.....	17
4.3.2 Source Code Flask.....	18
4.4 <i>Data Preprocessing</i>	20
4.5 <i>Linear regression</i> pada Scikit-Learn.....	23
BAB 5 Pengujian	24
5.1 Pengujian Model <i>Multiple Linear regression</i>	24
5.1.1 Data Pengujian Model <i>Multiple Linear regression</i>	24
5.1.2 Skema Pengujian Model <i>Multiple Linear regression</i>	25
5.1.3 Pengujian Dampak Korelasi Antar Variabel Terhadap Akurasi Model.....	29
5.1.4 Pengujian <i>Root Mean Squared Error</i>	30
5.2 Pengujian dengan <i>k-Fold Cross Validation</i>	32
BAB 6 KESIMPULAN.....	35
6.1 Kesimpulan	35

6.2 Saran.....	35
DAFTAR PUSTAKA	36
LAMPIRAN A PENGUMPULAN DATA.....	A-1



DAFTAR GAMBAR

Gambar 2.1 Hubungan Linier β_0 ; <i>Intercept</i> ; β_1 ; <i>slope</i> . [8].....	5
Gambar 3.1 Daerah Pengumpulan Data.....	10
Gambar 3.2 <i>Use case diagram</i> sistem prediksi harga kost.....	12
Gambar 3.3 <i>Activity Diagram</i> pilih fasilitas kamar.	13
Gambar 3.4 Desain Antarmuka Tampilan <i>Home</i>	13
Gambar 3.5 Desain Antarmuka Pilih Fasilitas Kamar Kost.....	14
Gambar 3.6 Desain Antarmuka Tampilan Hasil Prediksi.	14
Gambar 4.1 Implementasi tampilan <i>Home</i>	15
Gambar 4.2 Implementasi tampilan pilih fasilitas kamar kost.....	16
Gambar 4.3 Implementasi tampilan pilih fasilitas kamar kost.....	16
Gambar 4.4 Implementasi tampilan hasil prediksi.	17
Gambar 4.5 Flask Running.....	17
Gambar 5.1 sebaran data pada kolom HARGA_KOST	27
Gambar 5.2 Korelasi Hubungan antar variabel dengan heatmap.	28
Gambar 5.3 Visualisasi nilai R2.	30
Gambar 5.4 <i>cross validation 10 fold</i>	32
Gambar 5.5 <i>cross validation 15 fold</i>	33
Gambar 5.6 <i>cross validation 20 fold</i>	33
Gambar 5.7 Visualisasi rata-rata RMSE pada tiap pengujian.....	34

DAFTAR TABEL

Tabel 3.1 Contoh model data.....	11
Tabel 4.1 Contoh hasil Data yang telah diubah.....	21
Tabel 4.2 Model data untuk <i>multiple linier regression</i>	22
Tabel 5.1 Deskripsi kolom pada data yang digunakan	24
Tabel 5.2 Contoh baris data yang digunakan	25
Tabel 5.3 Perhitungan RMSE.....	31

