

IMPLEMENTASI MULTIPLE LINEAR REGRESSION PADA MACHINE LEARNING UNTUK PREDIKSI HARGA KOST DI KOTA MALANG

Ridwan Maulana Samhadi¹, Bambang Minto Basuki², Sugiono³

^{1,2,3}Universitas Islam Malang

rdwanf49@gmail.com, bambang.minto@unisma.ac.id, gionounisma@unisma.ac.id

Abstract

With the rapid population, especially in the city of Malang, it is possible that land owners will use it to open a business providing temporary housing (boarding houses). One of the problems faced by prospective boarding house tenants is the price/cost, how to find out the costs that will be incurred for boarding tenants based on the desired facilities. Here, researchers use artificial intelligence technology or machine learning to predict the price of boarding houses in Malang, so that they can be used as a reference for economic considerations to choose the right boarding house. Boarding is a service that offers a room or place to live with a certain amount of payment for each period. The purpose of this study is to determine the price of boarding houses and the factors that influence them, and to determine the level of accuracy in boarding prices using the multiple linear regression method. In this study using quantitative methods with analytical methods, namely multiple linear regression. The programming used is Python as machine learning and Nuxtjs as the frontend as well as application testing with the Black box testing method. The result of this study is a web app-based system that can display predictive results regarding the desired facility, which can provide convenience in knowing the range of boarding house rental prices. The test results on the system show the percentage of accuracy is 79% and the result of MAPE measurement is 12%.

Keywords— Machine Learning, Multiple Linear Regression, Scikit Learn, Website.

Abstraksi

Pesatnya penduduk khususnya di kota Malang, tidak sedikit kemungkinan pemilik lahan akan digunakan untuk membuka usaha penyedia tempat tinggal sementara(kost). Salah satu permasalahan yang dihadapi calon penyewa kost yaitu harga/biaya, bagaimana mengetahui biaya yang akan dikeluarkan untuk penyewa kost berdasarkan fasilitas yang diinginkan. Disini peneliti memanfaatkan teknologi kecerdasan buatan atau *machine learning* untuk memprediksi harga sewa kost di kota Malang, agar bisa menjadi acuan pertimbangan ekonomi untuk memilih kost yang tepat. Kost adalah sebuah jasa yang menawarkan sebuah kamar atau tempat untuk ditinggali dengan sejumlah pembayaran untuk setiap periode tertentu. Adapun tujuan penelitian ini adalah untuk mengetahui harga sewa kost dan faktor yang mempengaruhi, dan Mengetahui tingkat akurasi pada harga kost menggunakan metode *multiple linear regression*. Dalam penelitian ini menggunakan metode kuantitatif dengan metode analisis yaitu *multiple linear regression*. Pemrograman yang digunakan adalah Python sebagai *machine learning* dan Nuxtjs sebagai *frontend* serta dengan pengujian aplikasi dengan metode *Black box testing*. Hasil dari penelitian ini adalah sebuah sistem berbasis web app yang dapat menampilkan hasil prediksi terkait fasilitas yang diinginkan, yang dapat memberikan kemudahan dalam mengetahui kisaran harga sewa kost. Hasil pengujian pada sistem menunjukkan persentase keakuratan sebesar 79% dan hasil pengukuran MAPE sebesar 12%.

Kata Kunci— Machine Learning, Multiple Linear Regression, Scikit Learn, Website.

I. PENDAHULUAN

Kota Malang merupakan kota tujuan wisata, dan juga dijuluki sebagai kota pelajar, dan termasuk kota terbesar kedua di Jawa Timur setelah Surabaya. karena banyaknya perguruan tinggi yang berada di kota Malang membuat para pendatang dari berbagai daerah berbondong-bondong untuk menimba ilmu. Permasalahan yang dihadapi sekarang yaitu semakin meningkatnya pertumbuhan penduduk di kota Malang yang dikarenakan tidak Cuma orang yang datang ke Malang untuk sekedar berwisata saja, namun juga melangsungkan pendidikan dan menetap, yang mengakibatkan peningkatan pertumbuhan penduduk. Menurut Badan Statistik Kota Malang berdasarkan jumlah migrasi masuk dari tahun 2019-2021 terus mengalami peningkatan, untuk Kota Malang dari 22.224 jiwa yang migrasi pada tahun 2019, sempat mengalami penurunan pada tahun 2020 menjadi 18.505 jiwa dan kembali meningkat menjadi 22.790 pada tahun 2021 [1]. Berdasarkan jumlah penduduk menurut jenis kelamin dari tahun 2018-2020 terus mengalami peningkatan yang sangat drastis, untuk Kota Malang dari 427.078 jiwa meningkat menjadi 874.890 jiwa pada tahun 2020 [2].

Tidak sedikit kemungkinan pemilik lahan akan digunakan untuk membuka usaha penyedia tempat tinggal sementara, bisa disebut kamar kost. Semakin meningkatnya kebutuhan sewa kost, menjadi alasan para pemilik kost terus menambah kapasitas kamar dan fasilitas pendukung lainnya. Untuk penyewa kost di awal akan diberitahu sejumlah biaya yang akan dibebankan untuk periode tertentu, biasanya dihitung perbulan.

Salah satu permasalahan yang dihadapi calon penyewa kost yaitu harga/biaya, bagaimana mengetahui biaya yang akan dikeluarkan untuk penyewa kost berdasarkan fasilitas yang diinginkan. Pada dasarnya harga kost merupakan suatu masalah yang menjadi pertanyaan ketika akan memesan tempat kost, tidak luput

kemungkinan orang akan mempertanyakan berapa harga sewa kost tersebut. Luas kamar dan fasilitas lainnya adalah suatu bagian yang akan berpengaruh terhadap harga per unit kamar.

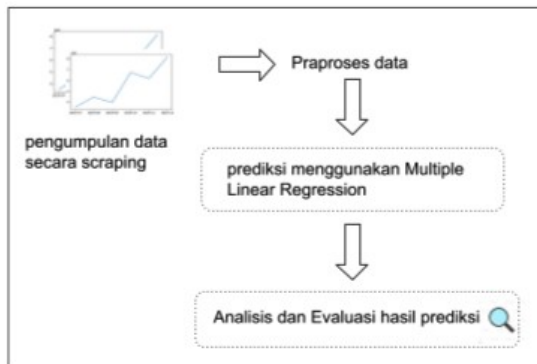
Dengan adanya teknologi *Artificial Intelligence* (kecerdasan buatan) dapat memudahkan para penggunanya untuk mencari solusi, secara garis besar untuk pengaplikasian kecerdasan buatan (AI) terbagi menjadi tujuh cabang diantaranya *machine learning*, *natural languages processing*, *expert system*, *vision*, *speech*, *planning and robotic*. Istilah *machine learning* pertama kali diperkenalkan oleh ilmuwan matematika seperti Adrien Marie Legendre, Thomas Bayes dan Andrey Markov pada tahun 1920-an dengan mengemukakan dasar-dasar *machine learning* dan konsepnya. Pada kasus kali ini menggunakan teknologi *machine learning*. *Machine learning* terdiri dari 4 kategori yaitu *supervised*, *unsupervised*, dan *deep learning*.

Oleh karena itu pada penelitian ini menggunakan *machine learning* untuk memproses dan menganalisis data untuk mengetahui tingkat akurasi pada suatu sistem. Penggunaan *machine learning* di era 4.0 saat ini sangat diperlukan untuk mempercepat dalam memproses suatu data yang sangat banyak serta dapat di implementasikan di berbagai *platform*. Hasil yang diperoleh dari penelitian ini adalah peningkatan akurasi sistem dari peneliti sebelumnya dan menerapkan pada aplikasi berbasis *web*.

II. METODE PENELITIAN

a. Analisis dan rancangan

Sistem yang akan dikembangkan ditujukan untuk calon penyewa kost dalam membantu mengetahui kisaran harga berdasarkan fasilitas yang dipilih. Adapun langkah-langkah yang akan dilakukan diawali dengan pengumpulan data, pra-proses data dan evaluasi hasil prediksi



Gambar 1. Alur metode penelitian

Metode yang digunakan dalam penelitian ini adalah *Multiple Linear Regression* pada *Machine Learning*. Konsep dasar metode MLR pada *Machine Learning* adalah mencari pengaruh antara variabel bebas dan variabel terikat dengan skala interval (Umi Narimawati, 2008). Metode MLR akan menghasilkan persentase dari hasil prediksi yang telah dilakukan, dengan cara mengalikan variabel bebas dengan nilai koefisien, terlebih dahulu data harus melalui proses normalisasi.

Langkah-langkah dalam metode MLR dari penelitian ini adalah sebagai berikut:

1. Menentukan data sebagai variabel bebas dan terikat
Pada tahap ini dilakukan pengumpulan data yang dibutuhkan. Data tersebut akan dikelompokkan masing-masing variabel. Data yang diperlukan untuk variabel bebas yaitu panjang, lebar, kamar mandi dalam, meja, kursi, dapur, wifi, ac, parkir motor, parkir mobil, lokasi, dan data variabel terikat adalah harga sewa/bulan.
2. Preprocessing
pada tahapan ini awal dari proses pengolahan data sesuai dengan variabel yang ditentukan. pada data ini data mentah yang berupa tipe data boolean akan diubah menjadi data numerik atau integer, sehingga bisa diproses pada tahapan selanjutnya. Berikut tahapan preprocessing sebagai berikut:

- a. Mengubah tipe data
mengubah tipe data bertujuan untuk mengkonversi tipe data

agar bisa di baca oleh sistem machine learning.

b. Normalisasi

Normalisasi bertujuan untuk membuat variabel dengan rentang nilai yang sama.

c. Train test split

Train test split bertujuan untuk melatih dan uji data yang telah di bagi 2 data latih dan data uji dengan perbandingan 80:20.

3. Modelling

Modelling ini bertujuan untuk mengetahui nilai koefisien dan intercept/error pada variabel.

4. Prediksi

Pada tahapan ini model multiple linear regression akan diuji dengan hasil pada train test split.

5. Evaluasi

Evaluasi ini bertujuan untuk mengetahui tingkat akurasi model dan kelayakan sistem.

b. Web Scraping

Web scraping adalah suatu proses pengambilan sebuah dokumen semi terstruktur dari internet, umumnya berupa halaman web dalam bahasa markup atau HTML, dan menganalisis untuk diambil data tertentu dari halaman tersebut untuk digunakan bagi kepentingan lain[9].



Gambar 2. Ilustrasi cara kerja web scraping

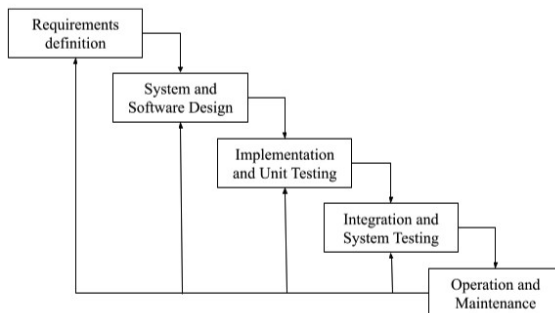
Cara kerja dari *web scraping* sebagai berikut:

1. *Create scraping template*: membuat template untuk mengambil informasi yang diapit dengan tag HTML.
2. *Explore site navigation*: menjelajah halaman pada situs yang akan diambil informasinya.

3. **Automate navigation and extraction:** untuk mengambil informasi secara otomatis dari website yang ditentukan.
4. **Extracted data and package history:** informasi yang telah didapat akan disimpan dalam tabel atau database.

c. Metode perancangan aplikasi

Pada penelitian ini, metode yang digunakan untuk perancangan aplikasi adalah *waterfall*. Sistem *waterfall* memiliki ciri khusus dalam pengerjaan dari suatu sistem yaitu dikerjakan secara berurutan.



Gambar 3. Model Waterfall

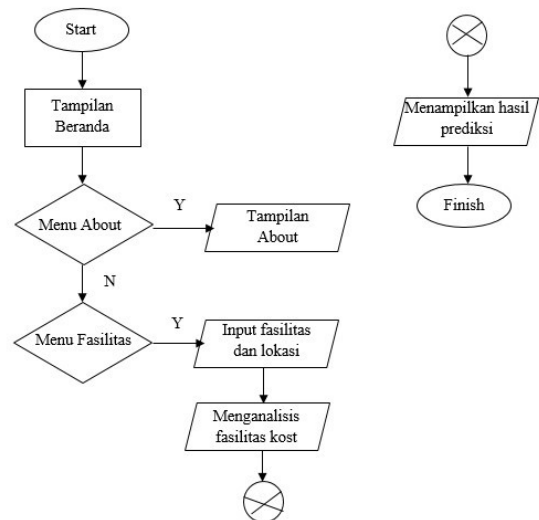
1. **Requirement definition:** proses pencarian kebutuhan untuk software, misalnya pengumpulan data informasi yang diperoleh kemudian dianalisa, sehingga didapat informasi lengkap mengenai spesifikasi kebutuhan software.
2. **System and software design:** pada tahap ini bertujuan untuk memberi gambaran tentang aplikasi yang akan dikembangkan dan bagaimana cara kerjanya.
3. **Implementation and unit testing:** perancangan software diimplementasikan pada serangkaian program atau unit program. Pengujian dilakukan bahwa program memenuhi spesifikasi.
4. **Integration and system testing:** menguji program keseluruhan untuk memastikan apakah software sesuai dengan desain dan kebutuhan.
5. **Operation and maintenance:** perawatan pada software pada waktu yang cukup panjang dan

berkala, juga perbaikan sistem dari sebelumnya serta meningkatkan sistem.

III. HASIL DAN PEMBAHASAN

Bagian ini membahas pengujian dan analisis sistem yang telah dirancang dari sistem yang telah dibuat

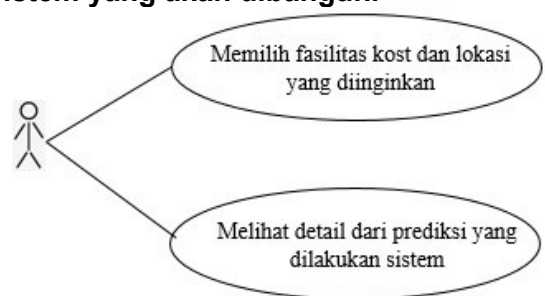
1. Flowchart Website



Gambar 4. Flowchart

2. Use Case Diagram

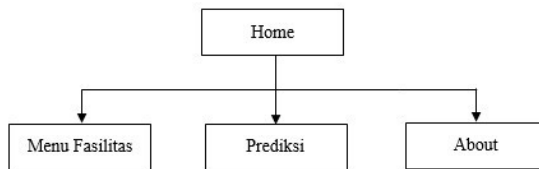
Use Case Diagram merupakan gambaran skenario pengguna suatu sistem informasi yang akan dibangun. Berikut use case diagram yang dari sistem yang akan dibangun.



Gambar 5. Use Case Diagram

3. Struktur Navigasi

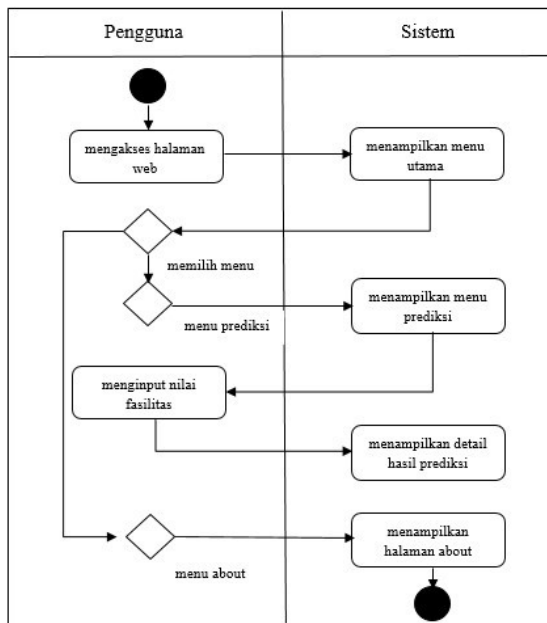
Pada sistem prediksi harga kost memiliki struktur navigasi diantaranya sebagai menu utama yaitu beranda yang memiliki beberapa navigasi sebagai penghubung antar halaman.



Gambar 6. Struktur Navigasi

4. Activity Diagram

Activity diagram merupakan suatu representatif dari keseluruhan tahapan alur kerja pada sistem.

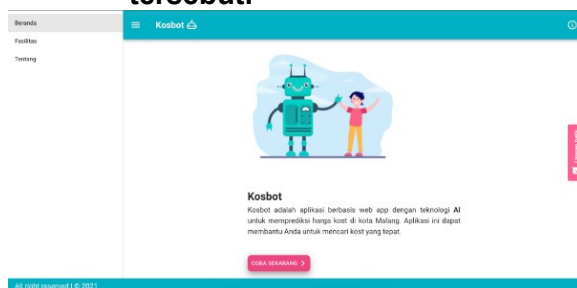


Gambar 7. Activity Diagram

5. Implementasi interface

a. Antarmuka halaman home

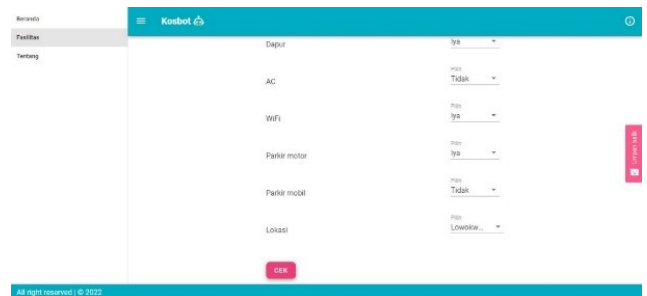
Antarmuka *Home* ini akan menampilkan halaman utama dari web Prediksi Harga Kost. Pada halaman utama ini menjelaskan tujuan, cara menggunakan serta teknologi yang digunakan pada web tersebut.



Gambar 8. Tampilan home

b. Antarmuka halaman menu fasilitas

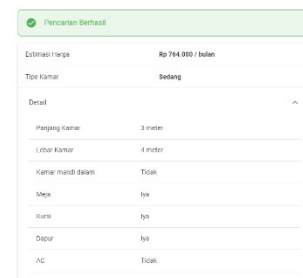
Pada halaman ini merupakan suatu komponen utama dalam pengembangan aplikasi berbasis web, halaman ini akan tampil ketika user mengklik tombol “coba sekarang” atau dari panel navigasi dengan menu “fasilitas”.



Gambar 9. Tampilan menu fasilitas

c. Antarmuka halaman detail

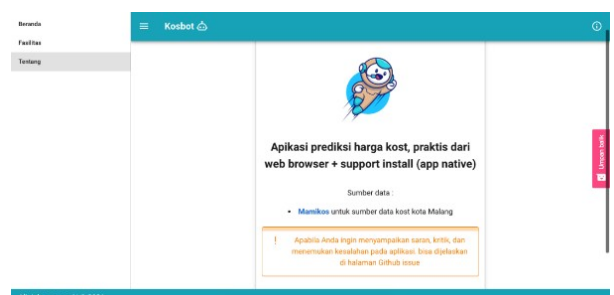
Halaman ini menampilkan rincian hasil prediksi dan fasilitas yang dipilih.



Gambar 10. Tampilan halaman detail

d. Antarmuka halaman about

Halaman ini menampilkan informasi seputar aplikasi, sumber data yang digunakan selama penelitian, dan kontak pengembang aplikasi.



Gambar 11. Tampilan menu *about*

6. Pengujian Sistem

Pada penelitian ini menggunakan pengujian black box yang dilakukan setiap proses dan kemungkinan terjadi kesalahan untuk tiap proses[10]. Adapun hasil dari pengujian pada sistem ini sebagai berikut:

a. Pengujian halaman *home*

Hasil Uji			
Data Masukan	Yang Diharapkan	Pengamatan	Kesimpulan
Sistem diakses	Halaman home menampilkan suatu informasi dan panduan mengenai aplikasi.	Antarmuka halaman home menampilkan informasi dan panduan mengenai aplikasi.	[√]Diterima []Ditolak
	Tombol "coba sekarang" mengarahkan pengguna ke halaman menu fasilitas.	Tombol "coba sekarang" berhasil mengarahkan pengguna ke halaman menu fasilitas.	[√]Diterima []Ditolak

b. Pengujian halaman menu fasilitas

Hasil Uji			
Data Masukan	Yang Diharapkan	Pengamatan	Kesimpulan
Sistem diakses	Tampilan halaman menu fasilitas berisi form input, dan tombol "Cari".	Antarmuka menu fasilitas menampilkan form input dan tombol "Cari".	[√]Diterima []Ditolak
	Tampilan halaman detail prediksi menampilkan estimasi harga, tipe kamar serta peta daerah kos yang telah dipilih.	Antarmuka halaman detail prediksi menampilkan estimasi harga, tipe kamar serta peta daerah kos yang telah dipilih.	[√]Diterima []Ditolak

c. Pengujian halaman *about*

Hasil Uji			
Data Masukan	Yang Diharapkan	Pengamatan	Kesimpulan
Sistem diakses	Tampilan menu <i>about</i> berisi informasi pengembang aplikasi.	Antarmuka halaman <i>about</i> menampilkan informasi pengembang aplikasi.	[√]Diterima []Ditolak

7. Perhitungan Evaluasi Kinerja Model Machine Learning

Untuk mengukur tingkat akurasi dari sistem machine learning diperlukan perhitungan $r2_score$. Berikut rumus $r2_score$:

$$r2 = 1 - (SS_{res} / SS_{mean})$$

$$SS_{res} = \frac{1}{n} \sum_{i=1}^n \hat{y}_i - y_i$$

$$SS_{mean} = \frac{1}{n} \sum_{i=1}^n \bar{y} - y_i$$

Dimana:

y = nilai asli

$y^{\wedge}$ = nilai prediksi

Dari 199 data kost di dapat tingkat akurasi prediksi sebesar 79%.

8. Perhitungan Error

Untuk mengukur error pada model dalam memprediksi dapat menggunakan MAPE (Mean Absolute Percentage Error). MAPE adalah persentase kesalahan rata-rata secara mutlak. MAPE dapat memberikan informasi seberapa besar kesalahan prediksi dibandingkan dengan nilai sebenarnya. Berikut rumus MAPE:

$$MAPE = \frac{1}{n} \sum_{i=1}^n \left| \frac{A_i - F_i}{A_i} \right| \times 100 \%$$

Dimana:

n = ukuran sample

A_i = nilai data aktual

F_i = nilai data peramalan

Dari 30 pengguna di dapat rata-rata persentase error sebesar 12%.

IV. KESIMPULAN

Berdasarkan hasil penelitian, dapat disimpulkan :

1. Pengguna menginput fasilitas yang dipilih yang nantinya sistem akan menerima inputan dan akan di proses dengan model machine learning yang selanjutnya akan menampilkan pada pengguna.
2. Dalam membangun sistem prediksi harga kost di kota Malang, tahap yang harus dilalui yaitu pengumpulan data, proses data dengan *machine learning*, lalu dilanjutkan dengan desain *interface*, lalu pembuatan website dengan menggunakan framework JavaScript yaitu Nuxt JS dan menggunakan layanan hosting dari firebase. dilanjutkan implementasi model *machine learning* pada aplikasi website yang telah dibuat.
3. Dari penelitian yang telah dilakukan fasilitas yang paling berpengaruh terhadap harga sewa yakni luas kamar, fasilitas kamar mandi dalam, AC, dan lokasi yang mengakibatkan hasil prediksi harga semakin tinggi dan yang tidak mempengaruhi harga yakni fasilitas meja, kursi, dapur, wifi, parkir sepeda motor, dan parkir mobil. Dan berdasarkan hasil pengujian terhadap 30 pengguna di dapat hasil pengukuran MAPE (Mean Absolute Percentage Error) didapat sebesar 12% yang menunjukkan bahwa sistem tersebut dalam interpretasi baik dan tingkat akurasi pada sistem machine learning dengan metode *multiple linear regression* dari 199 data didapat 79%.

DAFTAR PUSTAKA

- [1] Badan Pusat Statistik, "Badan Pusat Statistik." pp. 335–58, 2017, doi: 10.1055/s-2008-1040325.
- [2] B. P. Statistik and K. Malang, "Badan pusat statistik kota malang." 2018.
- [3] A. Izzah and R. Widyastuti, "Prediksi Harga Saham Menggunakan Improved Multiple Linear Regression untuk Pencegahan Data Outlier," *Kinet. Game Technol. Inf. Syst. Comput.*
- [4] C. Rozikin and A. Solichin, "Implementasi Algoritma Genetika dan Regresi Linier Berganda Untuk Prediksi Persediaan Bahan Makanan Pada Restoran Cepat Saji Implementation Of Genetic Algorithm and Multi-Linear Regression For Predicting Food Supplies At Fast Food Restaurants," *Semin. Nas. Multidisiplin Ilmu 2017*, no. April, pp. 10–17, 2017.
- [5] N. K. Arkarina, A. W. Widodo, and M. T. Furqon, "Implementasi Regresi Linier Berganda Untuk Prediksi Jumlah Peminat Mata Kuliah Pilihan," *J. Pengemb. Teknol. Inf. Dan Ilmu Komun.*, vol. 3, no. 11, pp. 10462–10467, 2019.
- [6] H. J. Chandra, "Prediksi skor pertandingan sepak bola liga inggris menggunakan multiple linear regression," 2018.
- [7] R. Andika, "PREDIKSI JUMLAH PRODUKSI di PT ASRINDO INDY RAYA MENGGUNAKAN METODE MULTIPLE LINEAR REGRESSION . SEMARANG OKTOBER 2019 FINAL PROJECT PREDICTION OF THE NUMBER OF PRODUCTION IN PT ASRINDO INDY RAYA USING MULTIPLE LINEAR REGRESSION METHOD . SEMARANG," 2019.
- [8] J. a Yani and K. Banjarbaru, "Aplikasi Data Mining Menggunakan Multiple Linear Regression Untuk Pengenalan Pola Curah Hujan," *Kumpul. J. Ilmu Komput.*, vol. 02, no. 01, pp. 34–44, 2015.
- [9] M. Turland, *Php|architect's Guide to Web Scraping*. 2010.
- [10] H. Bhasin, E. Khanna, and S. Sudha, "Black Box Testing based on Requirement Analysis and Design Specifications," *Int. J. Comput. Appl.*, vol. 87, no. 18, pp. 36–40, 2014, doi: 10.5120/15311-4024.