

Terbit online pada laman : <http://teknosi.fti.unand.ac.id/>

Jurnal Nasional Teknologi dan Sistem Informasi

| ISSN (Print) 2460-3465 | ISSN (Online) 2476-8812 |



Artikel Penelitian

Implementasi *SEM-Multiple Linear Regression* dalam Prediksi Jumlah Pendaftar Mahasiswa Baru di Perguruan Tinggi XYZ

Amelia Rahmadhani^{a*}, Dadang Syarif Sihabudin Sahid^a, Yohana Dewi Lulu Widyasari^a

^a Magister Terapan Teknik Komputer, Politeknik Caltex Riau, Pekanbaru, 28265, Indonesia

INFORMASI ARTIKEL

Sejarah Artikel:

Diterima Redaksi: 23 Juli 2023

Revisi Akhir: 29 Agustus 2023

Diterbitkan Online: 31 Agustus 2023

KATA KUNCI

Prediksi Mahasiswa Baru,
Multiple Linear Regression,
Structural Equation Modeling

KORESPONDENSI

E-mail: amelia.rahmadhani@alumni.pcr.ac.id*

A B S T R A C T

Bagi perguruan tinggi swasta (PTS), tidak menutup kemungkinan semakin banyak mahasiswa baru yang diterima, maka PTS tersebut akan terus eksis. Sebaliknya, jika PTS gagal menambah atau bahkan menurunkan jumlah mahasiswa baru setiap tahunnya, hal itu bisa berubah dengan tidak mampu beroperasi lagi bagi PTS dikarenakan pendapatan mereka satu-satunya hanya dari biaya kuliah mahasiswa. Tujuan penelitian ini diantaranya penentuan faktor-faktor yang mendukung prediksi pendaftaran mahasiswa di perguruan tinggi berdasarkan data sebelumnya, mengimplementasikan Multiple Linear Regression terhadap pendaftaran mahasiswa di perguruan tinggi, dan menganalisis tingkat akurasi hasil prediksi pendaftaran mahasiswa di perguruan tinggi. Penelitian ini menggunakan algoritma Multiple Linear Regression. Sebelum melakukan tahap prediksi, terlebih dahulu menentukan faktor-faktor yang mempengaruhi jumlah penerimaan mahasiswa baru menggunakan Structural Equation Modeling (SEM) dengan faktor promosi, biaya Pendidikan, tingkat kelulusan, informasi pendaftaran, jenis kelamin dan nilai akreditasi. Berdasarkan hasil SEM didapat faktor promosi, biaya Pendidikan, tingkat kelulusan, informasi pendaftaran, dan nilai akreditasi, dapat dilanjutkan ke tahap berikutnya karena faktor tersebut berpengaruh signifikan terhadap mahasiswa baru, sedangkan hasil prediksi menggunakan Multiple Linear Regression didapat bahwa nilai prediksi untuk tahun berikutnya adalah 486 orang calon mahasiswa baru, dengan hasil perhitungan MSE adalah 2657,79 dan MAE adalah 42,29, dimana semakin kecil hasil nilai MSE dan MAE yang diperoleh maka kesalahan pada sistem juga semakin sedikit serta R² adalah 0,9280 (92,80%) menandakan bahwa pengaruh semua struktur eksogen pada struktur endogen kuat.

1. PENDAHULUAN

Program studi di Indonesia diakreditasi oleh Badan Akreditasi Nasional Perguruan Tinggi atau BAN PT, serta Lembaga Akreditasi Mandiri atau LAM [1]. Menurut BAN PT, mutu dinilai dengan sembilan kriteria, dengan yang terpenting adalah mahasiswa. Menangkap kualitas program pendidikan yang diinginkan memerlukan analisis data pendidikan yang karakteristik dari proses pendidikan yang berkontribusi pada pencapaian mahasiswa. Salah satu aspek terpenting dalam penyelenggaraan kelembagaan adalah mahasiswa. Hal ini menunjukkan kualitas lembaga, yang dapat diukur dalam hal tingkat keberhasilan atau kegagalan yang dirasakan seseorang,

serta tingkat keberhasilan atau kegagalan yang dirasakan dalam kaitannya dengan orang yang mereka layani.

Perguruan tinggi merupakan kelanjutan dari jenjang pendidikan yang lebih tinggi setelah lulus dari sekolah menengah atas. Perguruan tinggi di Indonesia dapat berbentuk akademi dan sekolah kejuruan (vokasi). Tujuan pendidikan tinggi adalah untuk mempersiapkan para mahasiswa menjadi anggota masyarakat yang berpengetahuan luas dan yang dapat mengembangkan dan menyebarkan ilmu yang diperoleh. Selain penggunaan sistem informasi yang baik, efisien, dan efektif, kemajuan suatu perguruan tinggi juga ditentukan oleh banyaknya mahasiswa baru yang masuk ke perguruan tinggi tersebut, semakin banyak mahasiswa baru, semakin baik perguruan tinggi.

Ketika memilih institusi pendidikan tinggi, mahasiswa berusaha untuk mengumpulkan informasi terkait mengenai perguruan tinggi tempat mereka ingin melanjutkan studi. Informasi yang diperlukan berkaitan dengan pendidikan tinggi, biaya kuliah, perpustakaan, fasilitas, kegiatan kemahasiswaan, laboratorium, akreditasi, dan program studi yang tersedia di perguruan tinggi tersebut [2], [3]. Informasi ini merupakan salah satu faktor yang mempengaruhi calon mahasiswa dalam memilih program studi. Selain itu, yang mempengaruhi minat calon mahasiswa terdapat faktor-faktor pilihan perguruan tingginya, seperti teman, orang tua, kerabat, kesempatan kerja, beasiswa, dll.

Semua faktor ini juga dapat digunakan untuk memprediksi masa depan mahasiswa di perguruan tinggi. Faktor dan data tersebut di atas dapat dianalisis untuk mengetahui sejauh mana minat calon mahasiswa dalam memilih perguruan tinggi. Minat untuk melanjutkan pendidikan merupakan kecenderungan yang mengarahkan calon mahasiswa untuk melanjutkan studi ke perguruan tinggi dengan senang hati setelah tamat SMA. Crow mengemukakan bahwa, "Minat berhubungan dengan gaya gerak yang mendorong seseorang untuk menghadapi atau berurusan dengan orang, benda, kegiatan, pengalaman yang dirangsang oleh kegiatan itu sendiri", artinya, calon mahasiswa memiliki dorongan dan keinginan yang kuat untuk melanjutkan studi guna mencapai cita-citanya [4].

Bagi Perguruan Tinggi Swasta (PTS), tidak menutup kemungkinan semakin banyak mahasiswa baru yang diterima, maka PTS tersebut akan terus eksis. Sebaliknya, jika perguruan tinggi swasta (PTS) gagal menambah atau bahkan menurun jumlah mahasiswa baru setiap tahunnya, hal itu bisa berubah dengan tidak mampu beroperasi lagi bagi Perguruan Tinggi Swasta (PTS) dikarenakan pendapatan mereka satu-satunya hanya dari biaya kuliah mahasiswa. Memprediksi jumlah mahasiswa baru sangat diperlukan dikarenakan kenaikan dan penurunan jumlah mahasiswa baru juga berdampak terhadap penerimaan perguruan tinggi, dimana setiap tahunnya perguruan tinggi menyiapkan program kerja serta rencana kegiatan anggaran tahunan (RKAT) untuk satu tahun kedepan. Banyaknya mahasiswa baru yang diterima setiap tahunnya mempengaruhi proses perkuliahan di perguruan tinggi tersebut, tentunya perguruan tinggi harus membuat perencanaan perkuliahan dengan sangat baik, diantaranya menyiapkan fasilitas, sarana prasarana, dan tenaga pengajar. Sedangkan jika terjadi penurunan jumlah mahasiswa juga berdampak pada perguruan tinggi yaitu berkurang, minim, dan kesulitan dalam membiayai kegiatan operasional sementara perguruan tinggi mempunyai kewajiban pembayaran tetap setiap bulannya yang harus dikeluarkan seperti gaji, tunjangan, biaya listrik, perawatan elektronik, dan lain-lain. Ada beberapa faktor penyebab naik turunnya jumlah mahasiswa baru di sebuah perguruan tinggi, beberapa faktor tersebut yaitu biaya, ketidakstabilan biaya kuliah menjadi salah satu penyebab naik turunnya mahasiswa baru, akreditasi prodi, memiliki akreditasi yang sangat baik dapat membuat banyak mahasiswa berminat untuk masuk ke prodi tersebut (Kesuma et al., 2022). Jumlah pendaftar, calon mahasiswa diharuskan untuk melakukan pendaftaran terlebih dahulu sebelum melanjutkan ke tahap registrasi, hal ini dapat mempengaruhi jumlah mahasiswa baru yang masuk [6]. Kondisi pandemi merupakan salah satu yang

membawa dampak pada sulitnya pemenuhan biaya hidup sehari-hari termasuk di dalamnya yaitu biaya Pendidikan [2], [7].

Selain mempersiapkan fasilitas dan kualitas tenaga pendidik, perguruan tinggi membutuhkan strategi agar jumlah mahasiswa dapat bertambah setiap tahunnya. Salah satu metode penyusunan strategi yang dapat digunakan adalah metode peramalan. Metode peramalan dapat memprediksi apakah jumlah mahasiswa di perguruan tinggi setiap tahunnya mengalami kenaikan atau penurunan, hal ini dapat dijadikan pertimbangan atau perbaikan di masa yang akan datang [8], [9].

Dalam supervised learning terdapat beberapa metode yang dapat digunakan untuk prediksi atau peramalan, diantaranya *logistic regression*, *decision tree*, *random forest*, *k-nearest neighbors*, *support vector machine*, *neural network*, *naive bayes*, *linear regression*, *k-means clustering*, *hierarchical clustering*, *association rules* [10]. Dari sekian banyak metode yang dapat digunakan dalam prediksi dan peramalan tentunya setiap metode memiliki kelebihan dan kekurangan, bahwa metode yang digunakan harus dapat memprediksi dengan tingkat kesalahan yang rendah untuk dapat memprediksi jumlah pendaftar [11].

Penelitian terkait dengan prediksi jumlah mahasiswa baru sudah dilakukan oleh beberapa peneliti, seperti: Nainggolan dkk. melakukan prediksi terhadap jumlah mahasiswa baru dengan menggunakan metode regresi linear dengan variabel bebas yaitu promosi, jumlah lulusan, dan fasilitas perguruan tinggi. Hasil perancangan dan penelitian dengan menggunakan data mining digambarkan dengan nilai numerik yang menunjukkan perkiraan jumlah calon mahasiswa baru untuk tahun ajaran yang akan datang, akan tetapi pada penelitian ini tidak melakukan uji validitas terhadap variabel penelitian sehingga akurasi terhadap penggunaan variabel diragukan [12]. Selanjutnya, Kesuma dkk. melakukan penelitian yang berhubungan dengan prediksi jumlah mahasiswa baru dengan analisis regresi linear berganda, ada tiga variabel dalam penelitian ini yaitu pendaftar, biaya, dan mahasiswa baru serta pengujian menggunakan aplikasi rapidminer, dilihat dari hasil penelitian yang dilakukan terhadap hasil regresi linear berganda dan menggunakan aplikasi rapidminer menghasilkan prediksi yang sama, namun keterbatasan pada penelitian ini terdapat pada penggunaan 10 orang sampel dalam melakukan uji validitas terhadap kuesioner, seharusnya sampel yang digunakan minimal 30 orang [13], [14]. Pada penelitian ini, untuk memprediksi jumlah mahasiswa baru berdasarkan dari pengembangan variabel yang diteliti sebelumnya oleh peneliti-peneliti sebelumnya yaitu variabel biaya promosi, biaya kuliah, tingkat kelulusan, dan ditambahkan dengan variabel lain seperti informasi pendaftaran, jenis kelamin dan nilai akreditasi.

Dari permasalahan diatas, maka diperlukan analisis dengan menerapkan konsep *machine learning* dalam melakukan prediksi mahasiswa baru dengan menggunakan Algoritma Regresi Linear Berganda (*Multiple Linear Regression Algorithm*). *Machine learning* dapat memprediksi dengan akurasi tinggi, sehingga setelah internet machine learning akan menjadi teknologi yang terpenting [10]. Definisi *Machine learning* sebagai penerapan algoritma matematika dan komputer melalui pembelajaran yang berasal dari data dan menghasilkan prediksi masa depan [15]. Analisis dengan menggunakan variabel bebas berganda disebut

analisis regresi berganda. Beberapa teknik regresi linear digunakan untuk mengetahui apakah terdapat pengaruh yang signifikan terhadap variabel terikat (Y) antara dua atau lebih variabel bebas (X) [16].

Keuntungan dari metode regresi linear berganda adalah generalisasi dan isolasi pola data tertentu, memperoleh informasi meskipun tidak ada kepastian, dan kemampuan untuk melakukan perhitungan secara paralel untuk mempersingkat proses, sehingga cocok untuk peramalan skala besar [17]. Hasil keluaran dari prediksi menggunakan regresi linear berganda ini berupa angka sehingga tepat untuk memprediksi jumlah mahasiswa baru yang berbentuk angka.

Jumlah mahasiswa baru ini sebenarnya perlu diteliti, sehingga kita dapat mencari regresi dari faktor-faktor yang mempengaruhi jumlah penerimaan mahasiswa baru. Selain itu, juga dimungkinkan untuk mendapatkan data agregat perkiraan, yang berguna untuk melihat faktor mana yang memiliki pengaruh paling besar [18]. Dalam statistik, regresi adalah salah satu alat paling populer dalam ilmu sosial dan eksakta. Tidak ada salahnya menggunakan regresi untuk menguji angka jumlah mahasiswa baru, namun memiliki beberapa kelemahan, antara lain model regresi yang mensyaratkan variabel bebas tidak menunjukkan multikolinearitas, data harus berdistribusi normal, merupakan bagian yang sangat sulit menafsirkan, dan itu mengarah pada interpretasi yang tidak diberikan sesuai dengan kondisi [18]. Beberapa kelebihan analisis regresi yang tetap menggunakan metode ini antara lain: kemudahan penggunaan, menentukan kekuatan prediktor dan memprediksi trend masa depan [19]. Selain itu, cukup sulit untuk menentukan variabel independen dan variabel dependen di antara mereka sendiri dalam hal kausalitas, karena modelnya mungkin tidak cukup baik karena kesalahan pemilihan variabel yang digunakan untuk analisis, sehingga faktor-faktor tersebut diukur dan diamati kemudian [19].

Structural Equation Modeling (SEM) merupakan metode yang digunakan dalam penelitian ini untuk mengetahui faktor-faktor yang mempengaruhi penerimaan mahasiswa baru. Ada beberapa alasan utama penggunaan model persamaan struktural (SEM), antara lain: Pertama, kemampuannya mengakomodasi model yang relatif kompleks yang sulit dipecahkan dengan menggunakan metode analisis regresi linear. Kedua, SEM memiliki kemampuan untuk mengevaluasi hubungan antar

variabel yang memiliki banyak koneksi. Ketiga, kesalahan dalam setiap pengamatan tidak diabaikan melainkan dianalisis untuk memastikan model persamaan struktural (SEM) cukup tepat dalam menganalisis kuesioner yang berisi pengamatan tersebut. Langkah keempat melibatkan modifikasi model yang lancar untuk meningkatkan signifikansi statistiknya, sehingga menjadikannya lebih kuat untuk tujuan analitis. Elemen kelima, SEM memiliki kemampuan untuk menganalisis hubungan secara simultan [20], [21]. Dengan menerapkan *machine learning* menggunakan *Multiple Linear Regression* diharapkan dapat membantu dalam memprediksi jumlah mahasiswa baru pada tahun ajaran berikutnya.

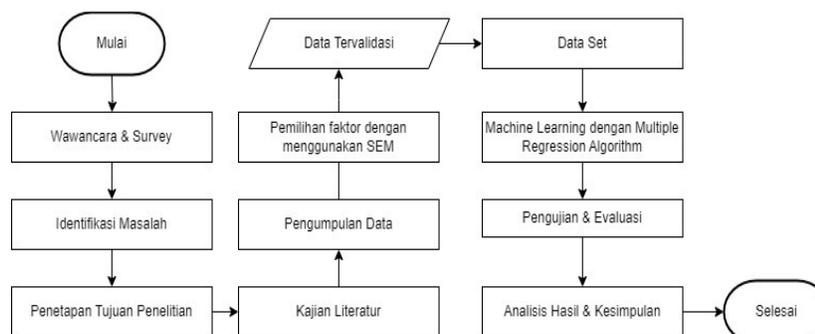
2. METODE

2.1. Bagan Alir Penelitian

Sistem prediksi penerimaan mahasiswa baru yang akan dibangun memerlukan beberapa persyaratan agar dapat menjadi sasaran perguruan tinggi. Persyaratan yang ditempatkan pada sistem peramalan terkait dengan algoritma yang digunakan yaitu regresi linear. Pemilihan algoritma ini dibuat berdasarkan tingkat akurasi dari penelitian sejenis sebelumnya.

Setelah memahami algoritma yang digunakan, tahap pengumpulan data yang diperlukan untuk penelitian berikut. Data yang digunakan disediakan oleh Perguruan Tinggi XYZ untuk memajukan penelitian ke tahap selanjutnya, yaitu tahap pembersihan data. Pada tahap ini pembersihan data dilakukan dengan cara mengisi blanko data dengan data yang telah ditentukan. Jika data yang digunakan tidak mengandung data kosong, maka data tersebut dianggap cocok untuk digunakan pada fase pelatihan dan algoritma yang digunakan pada fase pelatihan dapat memberikan hasil prediksi yang cukup akurat. Setelah pelatihan data, keakuratan hasil prediksi diperiksa kembali menggunakan perhitungan *Mean Absolute Error (MAE)*, *Mean Squared Error (RMSE)*, *Mean Absolute Percentage Error (MAPE)*, dan *R-Squared (R2)*.

Dalam melakukan prediksi tingkat penerimaan mahasiswa baru perguruan tinggi XYZ terdapat beberapa tahap yang harus dilakukan yaitu sebagai berikut yang telah digambarkan pada gambar 1.



Gambar 1. Bagan Alir Implementasi Algoritma terhadap Model

2.1.1. Pengumpulan Data

Informasi untuk penelitian ini diambil dari materi excel yang disiapkan oleh Perguruan Tinggi XYZ untuk tahun 2016-2022. Data yang digunakan terdiri dari biaya promosi, biaya kuliah, tingkat kelulusan, Informasi pendaftaran, jenis kelamin, nilai akreditasi dan jumlah mahasiswa baru.

2.1.2. Pembersihan Datad

Sebelum melakukan proses pelatihan data, dilakukan pengolahan data yang bertujuan untuk meningkatkan akurasi proses pelatihan data yang dilakukan. Pada saat melakukan proses preprocessing, tujuan metode SEM bertujuan untuk menemukan minimal subset dari variabel yang akan digunakan pada proses learning [23]. Minimal subset dari variabel yang dimaksudkan adalah memilih beberapa variabel yang ada, sehingga tidak semua variabel dalam dataset digunakan.

Faktor-faktor yang diperkirakan mempengaruhi jumlah mahasiswa baru adalah promosi (X₁), biaya kuliah (X₂), tingkat kelulusan (X₃), informasi pendaftaran (X₄), jenis kelamin (X₅), dan nilai akreditasi (X₆). Berikut dapat dilihat faktor-faktornya beserta indikator pengukurannya yang akan diukur dengan menggunakan SEM:

Tabel. 1 Faktor-faktor dxan Indikator

No	Faktor-faktor	Indikator
1	Promosi	Media yang digunakan
		Adanya bagian khusus
		Keterlibatan semua
2	Biaya Kuliah	Biaya SPP
		Biaya Buku
		Beasiswa
		Biaya Praktik
3	Tingkat Kelulusan	Cita-cita
		Prospek kerja lulusan
		Masa Studi
		Keberhasilan Alumni
4	Informasi Pendaftaran	Penyebaran Brosur
		Iklan disurat kabar
		Iklan dimedia sosial
		Promosi Langsung
5	Jenis Kelamin	Tidak membedakan
		Kemampuan yang sama
6	Nilai Akreditasi	Kualitas Dosen (SDM)
		Pelayanan Perguruan tinggi
		Popularitas Perguruan

Menurut Hair dkk sebaiknya ukuran sampel harus 100 atau lebih besar [24]. Penelitian ini dilaksanakan pada perguruan tinggi yang berada di Kota Pekanbaru dengan jumlah populasi 1.340 orang mahasiswa. Supaya menghasilkan sampel yang lebih akurat, diperlukan rumus-rumus untuk penentuan besarnya jumlah sampel, antara lain dengan rumus penentuan sampel, dari slovin 5%:

$$n = \frac{N}{1 + N(e)^2} \tag{1}$$

$$n = \frac{1340}{1 + 1340(5\%)^2}$$

$$n = 309$$

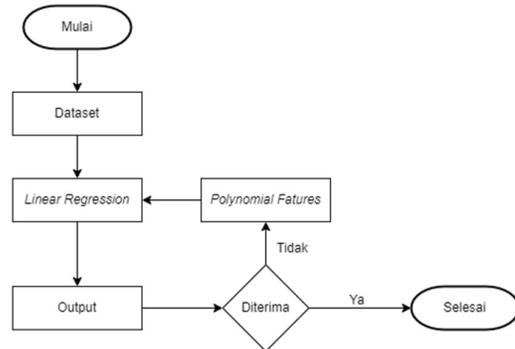
Keterangan :
n= Jumlah sampel

N= Jumlah populasi
e²= Tingkat kesalahan (5%)

Jadi jumlah sampel dalam penelitian ini adalah 309 orang responden.

2.1.3. Pemodelan dan Pelatihan data

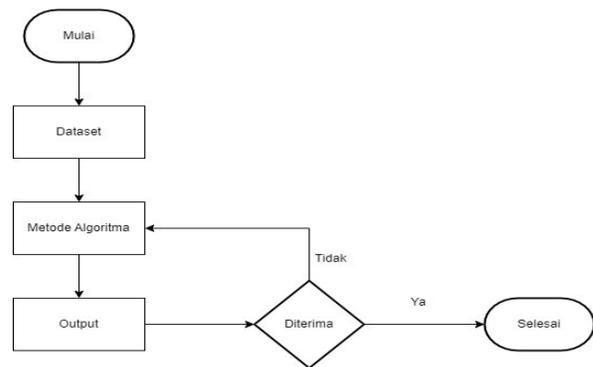
Langkah-langkah pemodelan dan pelatihan data adalah inti dari proses pembelajaran mesin. Pada penelitian ini, proses pelatihan dilakukan dengan menggunakan beberapa kombinasi parameter proses percobaan. Kumpulan data yang telah dibuat sebelumnya dengan variabel target dan variabel karakteristik yang dipilih. Pada proses pelatihan data algoritma regresi linear yang menggunakan parameter Normalize sebagai true, titik potong yang sesuai benar, dan derajat fitur polinomial adalah 3, tujuan penggunaan parameter fitur polinomial adalah proses keluaran maksimum yang memungkinkan untuk prediksi presisi tinggi. Pemilihan penggunaan fitur polinomial dalam proses pelatihan dikarenakan fitur polinomial tersebut kompatibel dengan algoritma perhitungan data dari dataset yang digunakan dalam penelitian. Gambar 2 adalah diagram alir untuk melatih data model.



Gambar 2. Diagram Alir Pemodelan dan Pelatihan Data

2.1.4. Implementasi

Setelah proses pelatihan data dilakukan dengan model algoritma regresi linear dan fitur polinomial, bentuk data pelatihan mengambil bentuk baru karena derajat fitur polinomial. Dengan demikian, diharapkan tingkat akurasi hasil ramalan akan meningkat dari bentuk yang baru. Gambar 3 adalah diagram alir model untuk pelatihan dataset.



Gambar 3. Diagram Alir Model Latih

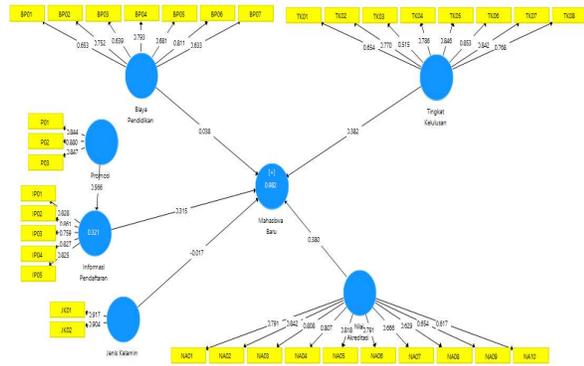
3. HASIL

3.1. Hasil Uji Validitas

Pengujian validitas dalam SEM-PLS dihitung melalui validitas konvergen dan validitas diskriminan.

1) Uji Validitas Konvergen

Sebuah indikator dinyatakan valid secara konvergen apabila nilai *loading* faktor $\geq 0,7$ dan nilai *average variances extracted (AVE)* $\geq 0,5$. Untuk hasil uji validitas konvergen dalam penelitian ini dapat dilihat pada gambar 4.8 dan tabel 4.6.



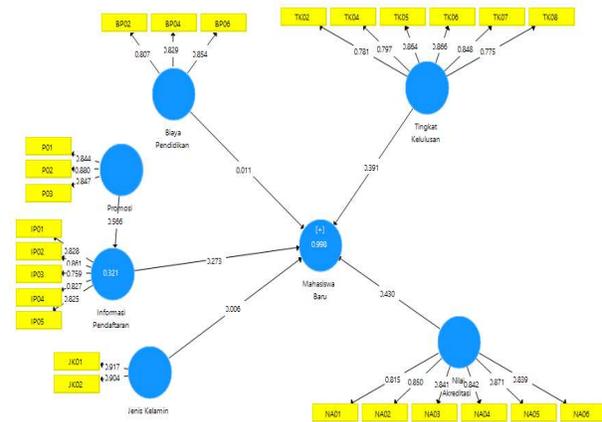
Gambar 4. Gambar Hasil Outer Loading (1)

Tabel 2. Uji Validitas Konvergen (1)

	Promos	Biaya	Tingk	Inform	Jenis	Nilai	Ma
P01	0,838						
P02	0,884						
P03	0,849						
BP01		0,661					
BP02		0,753					
BP03		0,631					
BP04		0,795					
BP05		0,687					
BP06		0,810					
BP07		0,623					
TK01			0,668				
TK02			0,765				
TK03			0,531				
TK04			0,780				
TK05			0,842				
TK06			0,850				
TK07			0,841				
TK08			0,765				
IP01				0,828			
IP02				0,857			
IP03				0,762			
IP04				0,829			
IP05				0,826			
JK01					0,914		
JK02					0,907		
NA01						0,785	
NA02						0,838	
NA03						0,801	
NA04						0,799	
NA05						0,807	
NA06						0,781	
NA07						0,680	
NA08						0,646	
NA09						0,673	
NA10						0,637	
P01							0,5
P02							0,5
P03							0,5
BP01							0,5
BP02							0,6
BP03							0,5
BP04							0,6
BP05							0,5
BP06							0,6

BP07	0,4
TK01	0,6
TK02	0,6
TK03	0,4
TK04	0,6
TK05	0,7
TK06	0,7
TK07	0,7
TK08	0,7
IP01	0,7
IP02	0,7
IP03	0,6
IP04	0,7
IP05	0,7
JK01	0,6
JK02	0,6
NA01	0,7
NA02	0,7
NA03	0,7
NA04	0,7
NA05	0,7
NA06	0,7
NA07	0,5
NA08	0,5
NA09	0,5
NA10	0,5

Berdasarkan tabel 2 dapat disimpulkan bahwa item pertanyaan yang digunakan pada faktor dalam penelitian ini ada yang valid secara konvergen dan ada juga yang tidak valid, dimana item di bawah 0,7, akan dihapus dan dilakukan pengulangan pengujian kepada item yang nilainya di atas 0,7. Untuk lebih jelasnya dapat dilihat pada gambar 4.8 dan tabel 4.7



Gambar 5. Gambar Hasil Outer Loading (2)

Berdasarkan tabel 5 dapat disimpulkan bahwa item pertanyaan yang digunakan pada variabel dalam penelitian ini valid secara konvergen, karena seluruh item pertanyaan memiliki nilai loading diatas 0,5 dan setiap variabel memiliki nilai *average variances extracted (AVE)* diatas 0,5 terlihat pada tabel 6.

Tabel 3. Uji Validitas Konvergen (2)

	Prom	Biaya	Ting	Inform	Jenis	Nilai	Mahasi
P01	0,844						
P02	0,880						
P03	0,847						
BP02		0,809					
BP04		0,828					
BP06		0,853					
TK02			0,774				
TK04			0,791				
TK05			0,866				
TK06			0,868				
TK07			0,852				
TK08			0,778				
IP01				0,828			

IP02	0,861
IP03	0,758
IP04	0,828
IP05	0,826
JK01	0,915
JK02	0,906
NA01	0,815
NA02	0,850
NA03	0,841
NA04	0,842
NA05	0,871
NA06	0,839
IP01	0,739
IP02	0,772
IP04	0,762
IP05	0,753
NA01	0,778
NA02	0,814
NA03	0,773
NA04	0,779
NA05	0,783
NA06	0,769
TK05	0,728
TK06	0,732
TK07	0,719
TK08	0,735

2) Uji Validitas Deskriminan

Dicriminant validity merupakan cara dalam menilai seberapa berbeda suatu konstruk dengan konstruk lainnya, yang dapat diketahui dengan cara membandingkan antara nilai AVE dari kedua konstruk dengan nilai kuadrat korelasi antara dua konstruk yang diuji. Validitas diskriminan dapat dilihat dari nilai cross loading dengan konstruknya, setiap indikator yang ada dalam suatu konstruk mempunyai perbedaan dengan indikator di

Tabel 5. Uji Validitas Deskriminan

	Promosi	Biaya Pendidikan	Tingkat Kelulusan	Informasi Pendaftaran	Jenis Kelamin	Nilai Akreditasi
P01	0,844	0,375	0,374	0,500	0,366	0,438
P02	0,880	0,442	0,373	0,459	0,355	0,412
P03	0,847	0,497	0,412	0,495	0,390	0,416
BP02	0,407	0,809	0,476	0,506	0,326	0,579
BP04	0,399	0,828	0,548	0,553	0,383	0,515
BP06	0,464	0,853	0,576	0,609	0,362	0,548
TK02	0,355	0,524	0,774	0,571	0,382	0,562
TK04	0,345	0,450	0,791	0,593	0,437	0,575
TK05	0,379	0,546	0,866	0,605	0,364	0,545
TK06	0,405	0,559	0,868	0,587	0,380	0,557
TK07	0,400	0,563	0,852	0,586	0,457	0,527
TK08	0,342	0,530	0,778	0,617	0,451	0,606
IP01	0,471	0,629	0,640	0,828	0,438	0,571
IP02	0,570	0,543	0,605	0,861	0,499	0,642
IP03	0,363	0,448	0,586	0,758	0,386	0,573
IP04	0,447	0,562	0,579	0,828	0,505	0,658
IP05	0,455	0,564	0,556	0,826	0,500	0,659
JK01	0,379	0,418	0,484	0,537	0,915	0,543
JK02	0,411	0,365	0,426	0,500	0,906	0,523
NA01	0,428	0,523	0,614	0,610	0,496	0,815
NA02	0,444	0,553	0,609	0,668	0,506	0,850
NA03	0,396	0,583	0,581	0,614	0,493	0,841
NA04	0,406	0,529	0,552	0,667	0,555	0,842
NA05	0,402	0,581	0,554	0,631	0,465	0,871
NA06	0,416	0,563	0,540	0,638	0,447	0,839

Berdasarkan tabel 6 dapat dijelaskan bahwa variabel dalam penelitian ini telah reliabel, dengan nilai *composite reliability* diatas 0,70 dan nilai *cronbach's alpha* diatas 0,70.

konstruk yang lainnya yang dapat ditunjukkan dengan nilai loading yang lebih tinggi dari konstruknya sendiri.

Tabel 4. Nilai AVE

Faktor	AVE
Promosi	0,735
Biaya Pendidikan	0,689
Tingkat Kelulusan	0,677
Informasi Pendaftaran	0,674
Jenis Kelamin	0,829
Nilai Akreditasi	0,711
Mahasiswa Baru	0,565

Berdasarkan hasil nilai cross loading antar indikator dengan konstruk pada table dibawah ini dapat disimpulkan bahwa masing-masing indikator yang ada dalam suatu konstruk memiliki perbedaan dengan indikator di konstruk yang lain yang ditunjukkan dengan skor loading yang lebih tinggi di konstruknya sendiri, maka dapat dikatakan bahwa instrumen yang digunakan dalam penelitian ini telah memenuhi kriteria validitas diskriminan.

3.2. Hasil Uji Reliabilitas

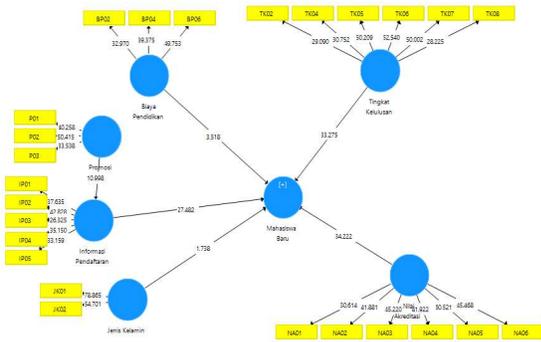
Pengujian reliabilitas menggunakan ukuran reliabilitas konstruk (*composite reliability*) dan/atau *cronbach's alpha*. Kriteria pengujian yaitu apabila nilai *composite reliability* $\geq 0,70$ dan/atau nilai *cronbach's alpha* $\geq 0,60$ maka dapat dinyatakan konstruk telah reliabel.

Tabel 6. Uji Reliabilitas

Faktor	composite reliability	cronbach's alpha
Promosi	0,893	0,820
Biaya Pendidikan	0,869	0,774
Tingkat Kelulusan	0,926	0,904
Informasi Pendaftaran	0,912	0,879
Jenis Kelamin	0,907	0,795
Nilai Akreditasi	0,937	0,919
Mahasiswa Baru	0,954	0,949

3.3. Hasil Uji Hipotesis

Uji hipotesis dilakukan dengan teknik bootstrapping. Data yang digunakan untuk bootstrapping adalah data yang sudah dilakukan tahapan Measurement. Uji hipotesis termasuk ke dalam Structural Model dan menunjukkan hubungan yang telah dihipotesiskan dengan praktik simulasi. Uji bootstrapping ini juga bertujuan untuk mengetahui arah hubungan dan signifikansi hubungan setiap variabel latennya. Berikut gambar struktur model SEM setelah dilakukan bootstrapping:



Gambar 8. Gambar Hasil Bootstrapping

Uji hipotesis dilakukan dengan melakukan perbandingan t-tatistic atau t-hitung yang sudah ditentukan. t-hitung yang dihasilkan dalam uji botstrapping harus lebih besar dari t-tabel one tail yaitu 1.65 untuk standar eror sebanyak 5% atau p value di bawah 0.05 [25].

Tabel 7. Hasil Pengujian Hipotesis

	Original Sample (O)	Sample Mean (M)	Standard Deviation (STDEV)	T Statistics (O-STDEV)	P Values	Keputusa
Promosi -> Informasi Pendaftaran	0,566	0,569	0,052	10,913	0,000	Mendukur
Biaya Pendidikan -> Mahasiswa Baru	0,011	0,011	0,003	3,517	0,000	Mendukur
Tingkat Kelulusan -> Mahasiswa Baru	0,391	0,391	0,012	33,756	0,000	Mendukur
Informasi Pendaftaran -> Mahasiswa Baru	0,273	0,273	0,010	27,751	0,000	Mendukur
Jenis Kelamin -> Mahasiswa Baru	0,006	0,006	0,003	1,755	0,079	Tidak Mendukur
Nilai Akreditasi -> Mahasiswa Baru	0,430	0,430	0,013	34,201	0,000	Mendukur

Dari tabel 7 dapat disimpulkan bahwa:

1. Promosi berpengaruh terhadap Informasi Pendaftaran dengan nilai 0,566 (jika promosi meningkat satu satuan unit maka informasi pendaftaran dapat meningkat sebesar 56,6%) Nilai signifikan dilihat dari nilai P Value sebesar 0,000 dimana nilainya lebih kecil dari 0,05.

2. Biaya Pendidikan berpengaruh positif signifikan terhadap mahasiswa baru dengan nilai 0,111 (jika Biaya Pendidikan meningkat satu satuan unit maka mahasiswa baru dapat meningkat sebesar 11,1%) Nilai signifikan dilihat dari nilai P Value sebesar 0,000 dimana nilainya lebih kecil dari 0,05.
3. Tingkat Kelulusan berpengaruh positif signifikan terhadap mahasiswa baru dengan nilai 0,391 (jika Tingkat Kelulusan meningkat satu satuan unit maka mahasiswa baru dapat meningkat sebesar 39,1%) Nilai signifikan dilihat dari nilai P Value sebesar 0,000 dimana nilainya lebih kecil dari 0,05.
4. Informasi Pendaftaran berpengaruh positif signifikan terhadap mahasiswa baru dengan nilai 0,273 (jika Informasi Pendaftaran meningkat satu satuan unit maka mahasiswa baru dapat meningkat sebesar 27,3%) Nilai signifikan dilihat dari nilai P Value sebesar 0,000 dimana nilainya lebih kecil dari 0,05.
5. Jenis Kelamin tidak berpengaruh positif signifikan terhadap mahasiswa baru dengan nilai 0,006 (jika Jenis Kelamin meningkat satu satuan unit maka mahasiswa baru dapat meningkat sebesar 0,6%) Nilai signifikan dilihat dari nilai P Value sebesar 0,079 dimana nilainya lebih besar dari 0,05.
6. Nilai Akreditasi berpengaruh positif signifikan terhadap mahasiswa baru dengan nilai 0,430 (jika Nilai Akreditasi meningkat satu satuan unit maka mahasiswa baru dapat meningkat sebesar 43%) Nilai signifikan dilihat dari nilai P Value sebesar 0,000 dimana nilainya lebih kecil dari 0,05.

Tabel 8 adalah nilai R Square dari hasil pengolahan data.

Tabel 8. Nilai R Square

	R Square	R Square Adjusted
Mahasiswa Baru	0,998	0,998

Nilai R Square pengaruh secara bersama-sama atau simultan faktor Promosi, Biaya Pendidikan, Tingkat Kelulusan, Informasi Pendaftaran, Jenis Kelamin, Nilai Akreditasi Terhadap Mahasiswa Baru adalah sebesar 0,998 dengan nilai adjusted r square 0,998. Maka, dapat dijelaskan bahwa semua konstruk exogen secara serentak mempengaruhi endogen sebesar 0,998 atau 99,8%. Oleh karena Adjusted R Square di atas dari 75% maka pengaruh semua konstruk eksogen terhadap endogen termasuk kuat.

4. PEMBAHASAN

4.1. Hasil Multiple Linier Regression dengan Google Colab

4.1.1. Importing Library

Proses awal dalam melakukan perhitungan analisis regresi berganda dengan menggunakan bahasa pemrograman python adalah menentukan library yang digunakan dalam proses analisis dan memindahkan data ke drive. Mengimpor library yang digunakan pada notebook dengan coding sebagai berikut:

```
# Import library yang dibutuhkan
# Data cleaning and manipulation
import numpy as np
import pandas as pd

# data visualization
%matplotlib inline
import matplotlib.pyplot as plt

import seaborn as sns
color = sns.color_palette()
sns.set_style('darkgrid')

import warnings

# machine learning
warnings.filterwarnings('ignore')
from sklearn import linear_model
from sklearn.model_selection import train_test_split
from sklearn.metrics import r2_score, mean_absolute_error, mean_squared_error
```

Gambar 9. Gambar Import library

4.1.2. Reading Data Set

Dataset yang digunakan pada penelitian ini adalah dataset jumlah mahasiswa baru. Proses ini diperlukan karena penggunaan google colab harus terkoneksi dengan internet dan data yang ada harus disimpan di drive google. File Data New.csv yang disimpan dikomputer lokal diunggah di drive dengan nama file yang sama. Untuk membaca dan menampilkan informasi dan data pada file csv bernama Data New, sebagai berikut:

```
# Mendefinisikan link dataset
df = pd.read_csv('Data New.csv')

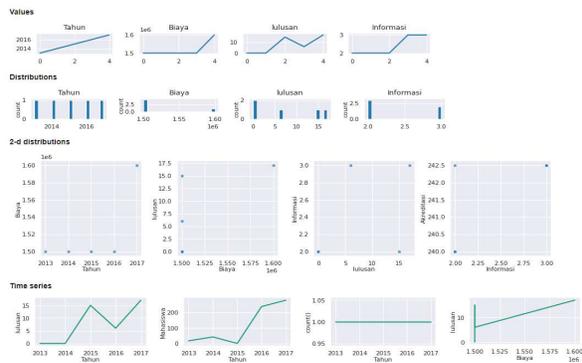
# Mengeprint kolom dataset
print(df.columns)

# Menampilkan 5 data teratas
df.head()
```

Index(['Tahun', 'Promosi', 'Biaya', 'lulusan', 'Informasi', 'Akreditasi', 'Mahasiswa'], dtype='object')

	Tahun	Promosi	Biaya	lulusan	Informasi	Akreditasi	Mahasiswa
0	2013	1	1500000	0	2	240.0	17
1	2014	1	1500000	0	2	240.0	42
2	2015	1	1500000	15	2	242.5	0
3	2016	1	1500000	6	3	242.5	238
4	2017	1	1600000	17	3	242.5	279

Gambar 10. Menampilkan 5 data teratas



Gambar 11. Distribusi data

Selanjutnya dengan kode berikutnya untuk menampilkan dimensi dataset dengan menginput:

```
# Menampilkan dimensi dataset
df.shape
```

(10, 7)

Gambar 12. Dimensi dataset

4.1.3. Checking Null Value

Langkah selanjutnya adalah menampilkan informasi detail tentang dataframe

```
# Menampilkan informasi detail tentang dataframe
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 10 entries, 0 to 9
Data columns (total 7 columns):
#   Column      Non-Null Count  Dtype
---  ---
0   Tahun       10 non-null     int64
1   Promosi     10 non-null     int64
2   Biaya       10 non-null     int64
3   Kelulusan   10 non-null     int64
4   Informasi   10 non-null     int64
5   Akreditasi   10 non-null     int64
6   Mahasiswa   10 non-null     int64
dtypes: int64(7)
memory usage: 688.0 bytes
```

Gambar 13. Informasi Detail tentang Data

Selanjutnya dengan kode berikutnya untuk memeriksa jumlah data yang missing value pada dataset dengan menginput:

```
# Memeriksa jumlah missing value pada dataset
df.isnull().sum()
```

```
Tahun      0
Promosi    0
Biaya      0
Kelulusan  0
Informasi  0
Akreditasi  0
Mahasiswa  0
dtype: int64
```

Gambar 14. Hasil Pemeriksaan Jumlah Data Missing Value

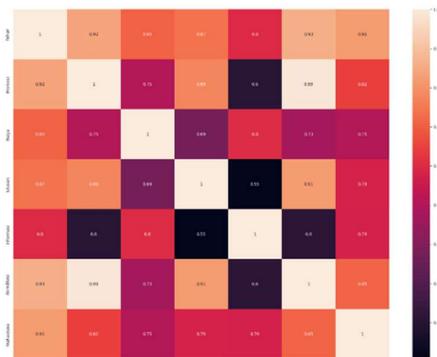
4.1.4. Data Exploration

Pada bagian ini melakukan eksplorasi data dari kumpulan data dengan cara Heatmap dan Plotting Pairwise serta melihat deskripsi dataset.

```
# Menampilkan heatmap dari dataset
cormap = df.corr()
plt.figure(figsize=(20,15))
sns.heatmap(cormap, annot=True)
plt.show()
```

Gambar 15. Heatmap data set

Hasil proses running skrip di atas ditampilkan pada gambar 16.



Gambar 16. Hasil Heatmap

Matrixplot merupakan plot yang berbentuk matrix umumnya digunakan untuk melihat korelasi antar variabel. Untuk menggunakan matrixplot() pertama kita harus cari korelasi antar atribut data dengan fungsi corr() dari hasil korelasi kita dapat membuat matrixplot untuk melihat data lebih dalam misalnya korelasi bisa dilihat dari perbedaan atau kedalaman warna.

Disini kita gunakan argumen annot=True untuk menampilkan korelasi antar atribut. Jika nilai korelasi mendekati 1 maka hubungan antar atribut semakin tinggi. Langkah selanjutnya adalah Describing Dataset sehingga output menampilkan informasi berikut untuk setiap kolom atau fitur diantaranya: count - Jumlah nilai yang tidak kosong, mean - Nilai rata-rata (rata-rata), std - Standar deviasi, min - nilai minimum, 25% - Persentil 25%*, 50% - Persentil 50%*, 75% - Persentil 75%*, maks - nilai maksimum.

```
# Menampilkan deskriptif statistik data
df.describe(percentiles=[])
```

	Tahun	Promosi	Biaya	lulusan	Informasi	Akreditasi	Mahasiswa
count	10.000000	10.000000	1.000000e+01	10.000000	10.000000	10.000000	10.000000
mean	2017.500000	1.800000	1.560000e+06	50.400000	2.700000	256.550000	250.700000
std	3.02765	0.918937	5.163978e+04	56.633716	0.483046	17.949698	202.529038
min	2013.000000	1.000000	1.500000e+06	0.000000	2.000000	240.000000	0.000000
50%	2017.500000	1.500000	1.600000e+06	32.500000	3.000000	250.250000	258.500000
max	2022.000000	3.000000	1.600000e+06	175.000000	3.000000	282.000000	612.000000

Gambar 17. Hasil Deskripsi Data

4.1.5. Data Processing

Data preprocessing adalah proses yang mengubah data mentah ke dalam bentuk yang lebih mudah dipahami. Data yang sesuai dikumpulkan, dibersihkan, dan diatur sesuai dengan kebutuhan bisnis. Sehingga data yang diolah nantinya sudah tidak mengandung missing value, outliers, duplikasi, dst.

```
# Drop id column
data = df.drop(['Tahun'], axis = 1)
target = data['Mahasiswa']
data.head()
```

	Promosi	Biaya	Kelulusan	Informasi	Akreditasi	Mahasiswa
0	1	4	0	2	2	17
1	1	4	0	2	2	42
2	1	4	15	2	2	0
3	1	4	6	3	2	238
4	1	4	17	3	2	279

Gambar 18. Hasil Pemrosesan Data

Selanjutnya dilakukan Splitting dataset dengan membagi menjadi 80% train and 20% test.

```
# Mensplit data menjadi training dan testing dengan porsi 80:20
X_train, X_test, y_train, y_test = \
    train_test_split(data, target, random_state=42, train_size=0.8, shuffle=True)
print ("train size={}, test_size={}, total_size={}".format(
    X_train.shape[0], X_test.shape[0], data.shape[0]))
)
```

train size=8, test_size=2, total_size=10

Gambar 19. Splitting Dataset

4.1.6. Models

Pada langkah ini mengembangkan model regresi linear.

```
from sklearn import linear_model

x=df[['Promosi','Biaya','lulusan','Informasi','Akreditasi']].values.reshape(-1,5) #reshape sesuai dengan jumlah variable X
x

array([[1.000e+00, 1.500e+06, 0.000e+00, 2.000e+00, 2.400e+02],
       [1.000e+00, 1.500e+06, 0.000e+00, 2.000e+00, 2.400e+02],
       [1.000e+00, 1.500e+06, 1.500e+01, 2.000e+00, 2.425e+02],
       [1.000e+00, 1.500e+06, 6.000e+00, 3.000e+00, 2.425e+02],
       [1.000e+00, 1.600e+06, 1.700e+01, 3.000e+00, 2.425e+02],
       [2.000e+00, 1.600e+06, 4.800e+01, 3.000e+00, 2.580e+02],
       [2.000e+00, 1.600e+06, 6.500e+01, 3.000e+00, 2.580e+02],
       [3.000e+00, 1.600e+06, 7.800e+01, 3.000e+00, 2.780e+02],
       [3.000e+00, 1.600e+06, 1.750e+02, 3.000e+00, 2.820e+02],
       [3.000e+00, 1.800e+06, 1.800e+02, 3.000e+00, 2.820e+02]])
```

Gambar 20. Gambar data Variabel X

```
y=df['Mahasiswa'].values.reshape(-1,1)
y

array([[ 17],
       [ 42],
       [  0],
       [238],
       [279],
       [314],
       [167],
       [352],
       [486],
       [612]])
```

Gambar 21. Data Variabel Y

```
a = lm.intercept_
a

array([-11179.86637684])

b = lm.coef_
b

array([[ -7.00032795e+02,  5.72498789e-04, -1.02810790e+00,
         1.44924702e+02,  4.46620434e+01]])
```

Gambar 22. Koefisien beta

Sehingga persamaan multiple linear regression $y = -1.1179 - 7.0003X1 + 5.7249X2 - 1.0281X3 + 1.4492X4 + 4.4662X5$.

4.1.7. Akurasi Penelitian dan Multiple regression dengan Statemodels

```
y_prediksi = model.predict(x)
```

Gambar 23. Prediksi

MSE adalah metode evaluasi lain yang digunakan dalam data science. MSE menghitung rata-rata dari selisih kuadrat antara nilai prediksi dan nilai aktual. Dengan kata lain, MSE menghitung berapa rata-rata kesalahan kuadrat dalam prediksi. Semakin kecil nilai MSE, semakin baik kualitas model tersebut.

```
MSE = mean_squared_error(y, y_prediksi)
MSE

2657.7995420489533
```

Gambar 24. Nilai MSE

Hasil perhitungan MSE adalah 2657,799 dimana semakin kecil hasil nilai MSE yang diperoleh maka kesalahan pada sistem juga semakin sedikit. MAE adalah salah satu metode evaluasi yang umum digunakan dalam data science. MAE menghitung rata-rata dari selisih absolut antara nilai prediksi dan nilai aktual. Dengan kata lain, MAE menghitung berapa rata-rata kesalahan absolut dalam prediksi. Semakin kecil nilai MAE, semakin baik kualitas model tersebut.

```
MAE = mean_absolute_error(y, y_prediksi)
MAE
42.29426685874732
```

Gambar 25. Nilai MAE

Hasil perhitungan MAE adalah 42,294 dimana semakin kecil hasil nilai MAE yang diperoleh maka kesalahan pada system juga semakin sedikit.

```
R2 = r2_score(y, y_prediksi)
R2
0.928004539413744
```

Gambar 26. Nilai R2

R² adalah sebuah nilai yang menyatakan seberapa sesuai hasil prediksi model mendekati data yang sebenarnya. Semakin besar R², maka hasil prediksi semakin dekat dengan data yang sebenarnya. Semakin besar R² maka model semakin bagus.

4.1.8. Multiple Regression dengan statsmodels.api

```
import statsmodels.api as sm

x, y = np.array(x), np.array(y)

x = sm.add_constant(x)
x
array([[1.000e+00, 1.000e+00, 1.500e+06, 0.000e+00, 2.000e+00, 2.400e+02],
       [1.000e+00, 1.000e+00, 1.500e+06, 0.000e+00, 2.000e+00, 2.400e+02],
       [1.000e+00, 1.000e+00, 1.500e+06, 1.500e+01, 2.000e+00, 2.425e+02],
       [1.000e+00, 1.000e+00, 1.500e+06, 6.000e+00, 3.000e+00, 2.425e+02],
       [1.000e+00, 1.000e+00, 1.600e+06, 1.700e+01, 3.000e+00, 2.425e+02],
       [1.000e+00, 2.000e+00, 1.600e+06, 4.800e+01, 3.000e+00, 2.580e+02],
       [1.000e+00, 2.000e+00, 1.600e+06, 6.500e+01, 3.000e+00, 2.580e+02],
       [1.000e+00, 3.000e+00, 1.600e+06, 7.000e+01, 3.000e+00, 2.780e+02],
       [1.000e+00, 3.000e+00, 1.600e+06, 1.750e+02, 3.000e+00, 2.820e+02],
       [1.000e+00, 3.000e+00, 1.600e+06, 1.080e+02, 3.000e+00, 2.820e+02]])
```

Gambar 27. Array Variabel X

```
model = sm.OLS(y, x).fit()
print(model.summary())
```

Gambar 28. Model OLS

Hasil proses running skrip di atas ditampilkan pada gambar 4.30:

```
=====
OLS Regression Results
=====
Dep. Variable:          y      R-squared:                0.928
Model:                OLS      Adj. R-squared:           0.838
Method:               Least Squares   F-statistic:             10.31
Date:                 Tue, 25 Jul 2023   Prob (F-statistic):       0.0211
Time:                 07:15:42         Log-likelihood:          -53.616
No. Observations:     10         AIC:                     119.2
Df Residuals:         4          BIC:                     121.0
Df Model:              5
Covariance Type:      nonrobust
=====
```

Gambar 29. Hasil OLS Regresi

Hasil estimasi dari model linier menghasilkan sebesar 0.928, artinya bahwa 92,80% variasi variabel dependen (mahasiswa baru) dapat dijelaskan oleh beberapa variasi independen (promosi, biaya pendidikan, tingkat kelulusan, informasi pendaftaran dan nilai akreditasi), sedangkan sisanya sebesar 7,20 % dijelaskan oleh variabel lain diluar model.

4.1.9. Prediction Jumlah Mahasiswa Baru

```
predictedMahasiswa = pd.DataFrame(model.predict(X_test), columns=['Predicted Mahasiswa'])
actualMahasiswa = pd.DataFrame(y_test)
actualMahasiswa = actualMahasiswa.reset_index(drop=True)
df_actual_vs_predicted = pd.concat([actualMahasiswa,predictedMahasiswa],axis =1)
df_actual_vs_predicted.T
```

	0	1
0	486.000000	42.000000
Predicted Mahasiswa	486.584789	-12.411163

Gambar 30. Hasil Prediksi

Dari hasil prediksi dapat kita lihat bahwa penerimaan mahasiswa baru pada tahun berikutnya adalah 485,5 atau dibulatkan menjadi 486 Orang Mahasiswa. Meskipun hasil prediksi untuk jumlah mahasiswa baru pada tahun berikutnya adalah 486. Perguruan Tinggi harus memantau perkembangan seiring waktu dan memastikan bahwa strategi penerimaan tetap relevan dengan perubahan lingkungan dan kebutuhan mahasiswa.

Strategi kampus terhadap penerimaan mahasiswa baru bisa mencakup beberapa aspek yang telah Anda sebutkan. Berikut adalah beberapa strategi yang dapat diimplementasikan oleh kampus terkait faktor-faktor yang Anda sebutkan:

1. Promosi:

- Meningkatkan visibilitas dan citra kampus melalui kampanye pemasaran yang tepat sasaran. Ini dapat mencakup pemasaran digital, iklan di media sosial, brosur, atau kolaborasi dengan pihak-pihak eksternal untuk memperluas jangkauan promosi.
- Melakukan kunjungan ke sekolah-sekolah menengah atas dan pameran pendidikan untuk memperkenalkan kampus kepada calon mahasiswa dan memberikan informasi mengenai program studi yang ditawarkan.

2. Biaya pendidikan:

- Menyediakan program pembiayaan yang fleksibel dan penjelasan yang transparan mengenai biaya pendidikan dan potensi penghematan melalui program bantuan.
- Menawarkan beasiswa dan bantuan keuangan bagi calon mahasiswa yang berprestasi atau memiliki keterbatasan finansial untuk memastikan aksesibilitas pendidikan.
- Menawarkan beasiswa pekerja yang diperuntukkan bagi pegawai/karyawan perusahaan.

3. Tingkat kelulusan:

- Memperkuat program pendukung akademik untuk membantu mahasiswa mengatasi tantangan akademik dan meningkatkan tingkat kelulusan.
- Menyediakan mentor atau penasihat akademik yang membantu mahasiswa dalam merencanakan program studi dan pencapaian tujuan akademik mereka.

4. Informasi pendaftaran:

- Memperbarui dan menyederhanakan proses pendaftaran agar lebih mudah diakses dan dipahami oleh calon mahasiswa.
- Memberikan informasi yang jelas dan komprehensif mengenai persyaratan, tenggat waktu, dan dokumen yang diperlukan untuk pendaftaran.

5. Jenis kelamin:

- Menerapkan kebijakan penerimaan yang adil dan tanpa diskriminasi berdasarkan jenis kelamin.
- Memastikan lingkungan kampus yang inklusif dan mendukung bagi semua mahasiswa, tanpa memandang jenis kelamin.

6. Nilai akreditasi:

- Menjaga dan meningkatkan standar akademik dan kualitas program studi untuk mendukung nilai akreditasi kampus.
- Melakukan evaluasi internal secara berkala untuk mengidentifikasi area yang perlu ditingkatkan guna memenuhi persyaratan akreditasi.

Dalam merancang strategi penerimaan mahasiswa baru, kampus harus berfokus pada nilai-nilai akademik, etika, dan keadilan. Dengan memahami dan mengatasi berbagai faktor yang mempengaruhi penerimaan mahasiswa, kampus dapat memastikan pengalaman belajar yang positif dan sukses bagi para mahasiswa yang mendaftar.

Penelitian ini merupakan studi literatur yang merujuk beberapa penelitian terdahulu di antaranya penelitian yang dilakukan oleh Nainggolan dkk. (2021) melakukan prediksi terhadap jumlah mahasiswa baru dengan menggunakan metode regresi linear dengan variabel bebas yaitu promosi, jumlah lulusan, dan fasilitas perguruan tinggi, data yang digunakan adalah 8 tahun terakhir, dimana hasil dari penelitian dan perancangan data mining ini yaitu berupa angka yang menunjukkan estimasi jumlah mahasiswa baru di tahun berikutnya, akan tetapi pada penelitian ini tidak melakukan uji validitas terhadap variabel penelitian sehingga akurasi terhadap penggunaan variabel diragukan [12].

Selanjutnya, Kesuma dkk. (2022), melakukan penelitian yang berhubungan dengan prediksi jumlah mahasiswa baru dengan analisis regresi linear berganda, adapun variabel yang digunakan terdiri dari tiga variabel yaitu biaya, pendaftar, dan mahasiswa baru serta pengujian menggunakan aplikasi rapidminer, dimana data yang digunakan adalah 10 tahun terakhir, dilihat dari hasil penelitian yang dilakukan terhadap hasil regresi linear berganda dan menggunakan aplikasi rapidminer menghasilkan prediksi yang sama, namun keterbatasan pada penelitian ini terdapat pada penggunaan 10 orang sampel dalam melakukan uji validitas terhadap kuesioner, seharusnya sampel yang digunakan minimal 30 orang [14], [13].

Yobioktabera dan Wibowo (2021), penelitian ini berfokus pada memprediksi akseptabilitas calon mahasiswa kedokteran ditinjau dari faktor-faktor yang mempengaruhinya dengan jumlah data yang valid berjumlah 146 data. Data mining dengan algoritma K Nearest Neighbors (KNN) digunakan untuk dapat memprediksi masa depan siswa. Dari percobaan yang dilakukan diperoleh RMSE terbaik dengan 70 data latih yaitu 0,218 / - 0,000. Dari penelitian ini dapat disimpulkan bahwa k-NN dapat digunakan untuk memprediksi penerimaan mahasiswa kedokteran di masa mendatang dengan akurasi terbaik sebesar 76,1%, walaupun hasilnya belum optimal. Secara umum, hasil tes menunjukkan hasil sedang dan tidak optimal. Namun, faktor penentu penerimaan siswa di masa depan dalam penelitian ini menunjukkan nilai yang kurang konsisten, sehingga faktor yang diselidiki harus diperiksa untuk mencapai hasil yang lebih baik lagi [26].

Lusiana dkk. (2022), algoritma yang digunakan dalam penelitian ini adalah Regresi Linear Berganda dengan menggunakan data 6 tahun terakhir. Sekolah merupakan tempat yang cocok untuk penerapan metode ini, oleh karena itu penelitian ini dilakukan di Sekolah Yayasan Madrasah Ibtidaiyah Sinaksak. Dalam penelitian ini, peneliti menggunakan metode data mining asosiasi

regresi linear berganda, dimana penerapan algoritma linear regresi berganda dapat diterapkan dalam menghitung prediksi jumlah pendaftar dengan sangat baik [27].

Riandari dkk. (2022), elemen terpenting dari perguruan tinggi adalah mahasiswa, oleh karena itu setiap perguruan tinggi harus terus meningkatkan pelayanan di masa yang akan datang, salah satunya dengan menggunakan *decision support*. Peramalan jumlah calon mahasiswa di perguruan tinggi dilakukan dengan memanfaatkan pendekatan data mining dan regresi linear berganda dengan data yang digunakan 13 tahun terakhir. Penggunaan 2 variabel bebas yaitu biaya administrasi (X1), nilai akreditasi (X2), dan jumlah mahasiswa yang terdaftar setiap tahun sebagai variabel terikat (Y). Pendekatan yang diusulkan memiliki struktur yang sederhana dan memberikan hasil prediksi awal yang dapat diterima. Selain itu, dapat digunakan untuk memprediksi jumlah calon mahasiswa yang akan mendaftar di masa yang akan datang. Oleh karena itu pihak Perguruan tinggi dapat membuat perencanaan yang lebih baik. Bagi peneliti selanjutnya disarankan untuk membuat perbandingan metode prediksi, misalnya *Neural Network*, *Logistic Regression* dan sebagainya untuk mengetahui metode mana yang lebih tepat untuk memprediksi. Akan tetapi akan tetapi pada penelitian ini tidak melakukan uji validitas terhadap variabel penelitian sehingga akurasi terhadap penggunaan variabel diragukan [27].

Selanjutnya Santoso dan Yulia (2018) memprediksi kinerja mahasiswa di pada mata kuliah yang diterapkan. Berdasarkan hasil pengujian, model regresi berganda memiliki kinerja yang lebih baik daripada regresi linear sederhana. Apalagi dengan bertambahnya jumlah model regresi linear, RMSE cenderung menurun secara bertahap [28].

Kemudian penelitian Khan (2021), kami menerapkan teknik penambangan data dan algoritma pembelajaran mesin menggunakan perangkat lunak R, yang digunakan untuk memprediksi, di sini kami menerapkan model regresi untuk menguji beberapa faktor pada kumpulan data yang kami asumsikan mempengaruhi kinerja mahasiswa. Model dibangun di atas dataset yang ada yang berisi banyak faktor dan nilai akhir. Faktor-faktor yang diuji adalah perhatian terhadap pendidikan tinggi, ketidakhadiran, waktu belajar, tingkat pendidikan orang tua, pekerjaan orang tua, dan jumlah kegagalan di masa lalu. Hasil penelitian menunjukkan bahwa hanya waktu belajar dan absen yang dapat mempengaruhi kinerja mahasiswa. Namun pada penelitian ini data yang digunakan adalah data yang sudah lama, data tidak disajikan secara detail serta terlalu sedikit data yang digunakan [29].

Selanjutnya Darman dkk. (2019), analitik pembelajaran adalah pengukuran, pengumpulan, analisis, dan pelaporan data tentang peserta didik dan konteksnya, untuk tujuan memahami dan mengoptimalkan pembelajaran dan lingkungan tempat terjadinya dengan menggunakan model regresi linear berganda. Dengan data 10 orang murid dapat dilihat bahwa yang diprediksi dan yang diamati nilai ujian memiliki perbedaan yang relatif kecil yang menunjukkan bahwa model regresi layak untuk digunakan, akan tetapi pada penelitian ini sangat sedikit data yang digunakan sehingga inkonsisten dalam memilih sampel yang tepat [30].

Tabel 9. Persamaan penelitian terdahulu dengan penelitian yang peneliti lakukan

Kesamaan dengan penelitian Terdahulu	Penelitian Terdahulu
Promosi	Nainggolan dkk. (2021)
Biaya Kuliah	Kesuma dkk. (2022)
Nilai Akreditasi	Riandari dkk. (2022)
Jenis Kelamin	Lusiana dkk. (2022)
Metode Multiple Regression	Nainggolan dkk. (2021), Kesuma dkk. (2022), Santoso dan Yulia (2018), Lusiana dkk. (2022), Riandari dkk. (2022), Darman dkk. (2019)
<i>Structural Equation Model</i>	Sahid (2020), Tabita dan Halim (2014)

Kemudian Sahid (2020), perilaku belajar saat berinteraksi dengan *e-learning* menjadi variabel utama dalam penentuan multikriteria atribut. Semua variabel yang diilustrasikan dalam model struktur menunjukkan hubungan yang positif, sehingga semua hubungan sebab akibat yang diajukan dalam hipotesis diterima. Hasil ini juga memberikan indikasi bahwa intensitas belajar sebagai representasi perilaku belajar dapat dijadikan acuan untuk mengidentifikasi perilaku pengguna selama berinteraksi dengan *e-learning*. Penelitian ini menggunakan SEM dalam menentukan komponen-komponen yang terkait dengan model penerimaan teknologi (TAM) [31].

Selanjutnya Tabita dan Halim (2014), penelitian ini bertujuan untuk mengetahui faktor-faktor yang mempengaruhi siswa SMA di Surabaya dalam memilih Universitas. Pengumpulan data dilakukan dengan menggunakan kuesioner kepada 283 siswa SMA se-Surabaya, sebagai sampel yang dipilih secara acak. Analisis menggunakan *Structural Equation Modeling* (SEM) dengan metode *Partial Least Square* (PLS). Pengujian dilakukan dengan menggunakan *factor loading*, composite reliability, dan koefisien jalur. Hasil penelitian menunjukkan bahwa proses merupakan faktor yang paling mempengaruhi siswa SMA dalam memilih Universitas. Yang paling berpengaruh bagi seorang siswa SMA Negeri dalam memilih Perguruan Tinggi adalah Perguruan Tinggi Akademik dan faktor yang paling berpengaruh bagi siswa SMA Swasta adalah prosesnya [32].

Adapun keterbaharuan dalam penelitian ini berkaitan dengan penelitian yang dijelaskan sebelumnya, dimana penelitian sebelumnya memiliki keterbatasan pada data, metode, dan hasil penelitian yang dilakukan seperti penelitian Nainggolan dkk. (2021), Kesuma dkk. (2022), Yobioktabera dan Wibowo (2021), Riandari dkk. (2022), Khan (2021).

Secara lebih jelasnya kesamaan penelitian terdahulu dengan penelitian yang peneliti lakukan dapat terlihat pada tabel 9. Selanjutnya ada beberapa penelitian sebelumnya memperlihatkan bahwa metode *multiple regression* dapat digunakan dengan sangat baik seperti penelitian yang dilakukan oleh Santoso dan Yulia (2018), Lusiana dkk. (2022), Darman dkk. (2019). Beberapa kelebihan analisis regresi yang tetap menggunakan metode ini antara lain: kemudahan penggunaan, menentukan kekuatan prediktor dan memprediksi trend masa depan,

sedangkan kelemahan analisis regresi yang paling menonjol adalah hasil prediksi analisis regresi merupakan nilai estimasi, jadi masih ada kemungkinan data sebenarnya tidak cocok [19]. Selain itu, cukup sulit untuk menentukan variabel independen dan variabel dependen di antara mereka sendiri dalam hal kausalitas, karena modelnya mungkin tidak cukup baik karena kesalahan pemilihan variabel yang digunakan untuk analisis, sehingga faktor-faktor tersebut diukur dan diamati kemudian [19].

Dalam upaya untuk mengatasi keterbatasan dan meningkatkan penerapan analisis regresi berganda, penelitian ini memanfaatkan *Structural Equation Modeling* (SEM) untuk memilih faktor-faktor yang memengaruhi penerimaan mahasiswa baru sebelum melakukan proses prediksi. Penelitian ini merinci identifikasi faktor-faktor yang signifikan dengan menggunakan SEM, yang meliputi faktor-faktor seperti promosi, biaya kuliah, tingkat kelulusan, informasi pendaftaran, dan nilai akreditasi. Selanjutnya, sebagai pendekatan lanjutan, penelitian ini menerapkan analisis *multiple regression* dan prediksi dengan menggunakan platform *Google Colab*. Langkah ini bertujuan untuk lebih memahami hubungan dan dampak relatif dari faktor-faktor yang telah diidentifikasi sebelumnya terhadap penerimaan mahasiswa baru. Melalui kombinasi SEM, analisis *multiple regression*, dan prediksi menggunakan *Google Colab*, penelitian ini mengarahkan fokus pada pemahaman mendalam tentang peran faktor promosi, biaya kuliah, tingkat kelulusan, informasi pendaftaran, dan nilai akreditasi dalam konteks penerimaan mahasiswa baru.

5. KESIMPULAN

Berdasarkan hasil penelitian dan pengujian yang telah dilakukan, maka dapat diambil beberapa kesimpulan bahwa faktor-faktor yang mendukung prediksi pendaftaran mahasiswa di perguruan tinggi adalah faktor promosi, biaya pendidikan, tingkat kelulusan, informasi pendaftaran, jenis kelamin, dan nilai akreditasi. Faktor-faktor penerimaan mahasiswa baru berdasarkan hasil *Structural Equation Modeling* adalah promosi, biaya pendidikan, tingkat kelulusan, informasi pendaftaran, dan nilai akreditasi. Berdasarkan hasil penelitian dengan menggunakan *Multiple Linear Regression Algorithm* terhadap pendaftaran mahasiswa di perguruan tinggi dengan persamaan sebagai berikut regression $y = -1.1179 - 7.0003X_1 + 5.7249X_2 - 1.0281X_3 + 1.4492X_4 + 4.4662X_5$. Hasil prediksi jumlah mahasiswa baru pada tahun berikutnya adalah 485,5 atau dibulatkan menjadi 486 Orang Mahasiswa dengan hasil perhitungan MSE adalah 2657,79 dan MAE adalah 42.29, serta R² adalah 0.9280 (92,80%).

DAFTAR PUSTAKA

- [1] M. A. Badan Akreditasi Nasional Perguruan Tinggi, "Peraturan Badan Akreditasi Nasional Perguruan Tinggi Nomor 9 Tahun 2020 Tentang Kebijakan Pengalihan Akreditasi Program Studi Dari Badan Akreditasi Nasional Perguruan Tinggi Ke Lembaga Akreditasi Mandiri," *BAN-PT*, pp. 4–6, 2020.
- [2] W. Sudarwati and D. E. Tikwalau, "Faktor-Faktor Yang Mempengaruhi Siswa-Siswa SMU/SMK Terhadap Keputusan Pemilihan Perguruan Tinggi Swasta," *JISI: Jurnal Integrasi Sistem Industri, 1(1)*, vol. 2014, pp. 68–81, 2014.

- [3] A. D. Suharti, "Peran Perpustakaan Perguruan Tinggi Dalam Mendukung Akreditasi Program Studi," *Buletin Perpustakaan Universitas Islam Indonesia*, vol. 2, no. 2, pp. 47–62, 2019.
- [4] Djaali, *Psikologi Pendidikan*. Jakarta: Bumi Aksara. 2012.
- [5] H. Di Kesuma, D. Apriadi, H. Juliansa, and E. Etriyanti, "Implementasi Data Mining Prediksi Mahasiswa Baru Menggunakan Algoritma Regresi Linear Berganda," *Jurnal Ilmiah Binary STMIK Bina Nusantara Jaya Lubuklinggau*, vol. 4, no. 2, pp. 62–66, 2022.
- [6] A. A. Azhar and A. T. Sikumbang, "Kecenderungan Peminatan Mahasiswa Program Studi Komunikasi Islam Pascasarjana Universitas Islam Negeri Sumatera Utara (UINSU) Medan Tahun 2010-2016," Universitas Islam Negeri (UIN) Sumatera Utara, 2018.
- [7] M. W. P. Putra and K. S. Kasmiarno, "Pengaruh Covid-19 Terhadap Kehidupan Masyarakat Indonesia: Sektor Pendidikan, Ekonomi Dan Spiritual Keagamaan," *POROS ONIM: Jurnal Sosial Keagamaan*, vol. 1, no. 2, pp. 144–159, 2020.
- [8] S. Makridakis, R. J. Hyndman, and F. Petropoulos, "Forecasting in social settings: The state of the art," *Int J Forecast*, vol. 36, no. 1, pp. 15–28, 2020, doi: [10.1016/j.ijforecast.2019.05.011](https://doi.org/10.1016/j.ijforecast.2019.05.011).
- [9] A. Purba, "Perancangan Aplikasi Peramalan Jumlah Calon Mahasiswa Baru yang mendaftar menggunakan Metode Single Exponential Smothing (Studi Kasus: Fakultas Agama Islam UISU)," *JURIKOM (Jurnal Riset Komputer)*, 2(6), vol. 2, no. 6, pp. 8–12, 2015.
- [10] P. Santoso, H. Abijono, and N. L. Anggreini, "Algoritma Supervised Learning Dan Unsupervised Learning Dalam Pengolahan Data," *Unira Malang*, vol. 4, no. 2, 2021.
- [11] D. Winarso, "Perbandingan Metode Regresi Linier Dan Weighted Moving Average Dalam Meramalkan Jumlah Mahasiswa Pada Periode Tertentu," *Prosiding Celscitech*, vol. 2, p. tech_70-tech_74, 2017.
- [12] R. Z. Nainggolan, K. Ibutama, and M. G. Suryanata, "Implementasi Data Mining Dengan Metode Regresi Linier Berganda Dalam Estimasi Mahasiswa Baru Pada Sekolah Tinggi Agama Islam Raudhatul Akmal BatangKuis," *Jurnal Cyber Tech*, vol. 1, no. 1, 2021.
- [13] H. Di Kesuma, D. Apriadi, H. Juliansa, E. Etriyanti, K. Palembang, and P. S. Selatan, "Implementasi Data Mining Prediksi Mahasiswa Baru Menggunakan Algoritma Regresi Linear Berganda Implementation of Data Mining Predictions for New Students Using Multiple Linear Regression Algorithm," *Jurnal Ilmiah Binary STMIK Bina Nusantara Jaya Lubuklinggau*, 4(2), 62–66, vol. 0, no. 02, pp. 62–66, 2022, doi: [10.52303/jb.v4i2.74](https://doi.org/10.52303/jb.v4i2.74).
- [14] M. Singarimbun and E. Shofian, *Metode Penelitian Survei*. Jakarta: LP3ES. 1995.
- [15] A. Roihan, P. A. Sunarya, and A. S. Rafika, "Pemanfaatan Machine Learning dalam Berbagai Bidang: Review paper," vol. 5, no. April, pp. 75–82, 2020.
- [16] L. Lela, Ruslan, and I. Yahya, "Analisis Faktor-Faktor Yang Mempengaruhi Skor Kebahagiaan Remaja (Studi Kasus Siswa SMAN 1 Wangi-Wangi)," no. 2005, pp. 611–620, 2019.
- [17] I. L. L. Gaol, S. Sinurat, and E. R. Siagian, "Implementasi Data Mining Dengan Metode Regresi Linear Berganda Untuk Memprediksi Data Persediaan Buku Pada PT. Yudhistira Ghalia Indonesia Area Sumatera Utara," *KOMIK (Konferensi Nasional Teknologi Informasi dan Komputer)*, vol. 3, no. 1, Nov. 2019, doi: [10.30865/komik.v3i1.1579](https://doi.org/10.30865/komik.v3i1.1579).
- [18] R. Efendi, "Model Prediksi Fertilitas Menggunakan Regresi, Rough Sets dan Rough-Regresi," in *Seminar Nasional Teknologi Informasi Komunikasi dan Industri*, 2018, pp. 666–672.
- [19] S. E. Saputra, "Analisis Lossetime Optimal Alat Gali Muat Excavator 1250 dan Alat Angkut HD 785 pada Kegiatan Pemindahan Tanah Penutup Menggunakan Regresi Linier Berganda pada Pit MTBU Penambangan PT. Pama Persada Nusantara," *repository.sttind.ac.id*, 2022.
- [20] J. Hair, W. Black, R. Anderson, and B. Babin, "Multivariate data analysis (8, ilustrasi ed.)," *Cengage Learning EMEA*, vol. 27, no. 6, pp. 1951–1980, 2018.
- [21] J. F. Hair, C. M. Ringle, and M. Sarstedt, "Partial Least Squares Structural Equation Modeling: Rigorous Applications, Better Results and Higher Acceptance," *Long Range Planning*, vol. 46, no. 1–2, Elsevier Ltd, pp. 1–12, 2013. doi: [10.1016/j.lrp.2013.01.001](https://doi.org/10.1016/j.lrp.2013.01.001).
- [22] J. F. Hair, J. J. Risher, M. Sarstedt, and C. M. Ringle, "When to use and how to report the results of PLS-SEM," 2018.
- [23] R. C. Prihandari, "Data Mining: Konsep Dan Aplikasi Menggunakan Rapidminer (Series: Supervised Learning Dan Unsupervised Learning)," INSTUT RISET DAN PUBLIKASI INDONESIA (IRPI), 2022.
- [24] J. F. Hair, J. J. Risher, M. Sarstedt, and C. M. Ringle, "When to use and how to report the results of PLS-SEM," *European Business Review*, vol. 31, no. 1. Emerald Group Publishing Ltd., pp. 2–24, Jan. 14, 2019. doi: [10.1108/EBR-11-2018-0203](https://doi.org/10.1108/EBR-11-2018-0203).
- [25] J. Hair, G. Hult, and C. Ringle, *A Primer on Partial Least Squares Structural Equation Modeling (PLS-SEM)*. 2nd Edition. Sage Publications, Thousand Oaks., 2017.
- [26] A. Yobioktabera and A. W. Wibowo, "Penerapan Data Mining Untuk Memprediksi Penerimaan Calon Mahasiswa Baru Fakultas Kedokteran," pp. 16–19, 2021.
- [27] F. Oktavia Lusiana et al., "Journal of Artificial Intelligence and Engineering Applications Analysis Of Multiple Regression Data Mining Methods On The Prediction Of Ibtidaiyah School Registration," 2022.
- [28] L. W. Santoso and Yulia, "Predicting Student Performance Using Data Mining," vol. X, no. X, pp. 1–4, 2018.
- [29] S. Khan, "Study Factors for Student Performance Applying Data Mining Regression Model Approach," vol. 21, no. 2, pp. 188–192, 2021.
- [30] H. Darman, S. Musa, R. Ramasamy, and R. Rajeswari, "Predicting Students' Final Grade in Mathematics Module using Multiple Linear Regression," *Int. J. Recent Technol. Eng*, 7(5), 331–335, 2019.
- [31] D. S. S. Sahid, "Learner Behavior in e-Learning as a Multicriteria Attribute based on Perspective of Flow Experience," *International Journal of Advanced Computer Science and Applications*, 11(12), vol. 11, no. 12, pp. 277–284, 2020.
- [32] A. Tabita and S. Halim, "Faktor-Faktor yang Mempengaruhi Siswa SMA dalam Memilih Perguruan Tinggi," *Prosiding Seminar Nasional Teknik Industri UK Petra 2014*, 2014.