

# Pemodelan Estimasi Kelulusan Mahasiswa Berbasis Data Akademik Melalui Regresi Linier Berganda

Rita Hariningrum<sup>1</sup>, Candra Yogatama<sup>2</sup>, Sukarno Budi Utomo<sup>3</sup>

<sup>1,2</sup>Teknik Kelistrikan Kapal, Universitas Ivet,

Jl. Pawiyatan Luhur IV No.17 Bendan Duwur Kec. Gajahmungkur Semarang, 50235,  
Jawa Tengah, Indonesia

<sup>3</sup>Teknik Elektro & Teknik Industri Fakultas Teknologi Industri Universitas Islam Sultan Agung. Jl.  
Raya Kaligawe KM.4 Semarang. Jawa Tengah, Indonesia

*E-mail: ritahariningrum@gmail.com<sup>1</sup>, cyogatama@gmail.com<sup>2</sup>, sukarno@unissula.ac.id<sup>3</sup>*

**Abstract** - The utilization of academic data of students has been leveraged to generate valuable information that supports the management in monitoring the study duration of students in the information systems study program. Through the application of multiple linear regression methods, a Y equation model has been developed to estimate the graduation time of students, supplemented by the creation of a prototype. The analysis conducted using nine independent variables—namely SKS1, SKS2, SKS3, SKS4, IPS1, IPS2, IPS3, IPS4, and the number of courses repeated—for students from the 2008 to 2012 cohorts resulted in the formula:  $Y = 13.49 + 0.099X1 - 0.068X2 + 0.025X3 - 0.059X4 - 0.585X5 - 0.443X6 - 0.155X7 - 0.368X8 - 0.082X9$ . This formula indicates the model's error level based on Mean Squared Error (MSE) and Root Mean Squared Error (RMSE) at 0.1166 and 0.3415, respectively. The prototype developed using PHP, with Sublime Text as the text editor and XAMPP as the local server, demonstrates significant potential in supporting the monitoring process of student study durations if fully integrated and supported by management.

**Keywords:** Data mining, multiple linear regression, prediction, monitoring, study duration

**Intisari** - Penggunaan data akademis mahasiswa telah dimanfaatkan untuk menghasilkan informasi berharga yang mendukung pengelolaan dan pemantauan durasi pendidikan mahasiswa di program studi sistem informasi. Melalui aplikasi metode regresi linier berganda, telah dikembangkan sebuah model persamaan Y yang berfungsi untuk memperkirakan waktu kelulusan mahasiswa, yang juga didukung oleh pembuatan prototipe.

Analisis yang dilaksanakan menggunakan sembilan variabel independen—yaitu SKS1, SKS2, SKS3, SKS4, IPS1, IPS2, IPS3, IPS4, dan jumlah mata kuliah yang diulang—untuk mahasiswa angkatan 2019 hingga 2023 menghasilkan formula:  $Y = 13.49 + 0.099X1 - 0.068X2 + 0.025X3 - 0.059X4 - 0.585X5 - 0.443X6 - 0.155X7 - 0.368X8 - 0.082X9$ . Formula ini menunjukkan tingkat kesalahan model berdasarkan Mean Squared Error (MSE) dan Root Mean Squared Error (RMSE) sebesar 0.1166 dan 0.3415, berturut-turut. Prototipe yang dikembangkan menggunakan PHP, dengan Sublime Text sebagai editor teks dan XAMPP sebagai server lokal, memperlihatkan potensi besar dalam mendukung proses pemantauan durasi studi mahasiswa jika terintegrasi dan didukung sepenuhnya oleh manajemen.

**Kata kunci:** penambahan data, regresi linier berganda, prediksi, pemantauan, durasi pendidikan

## I. PENDAHULUAN

Dalam era digital saat ini, evolusi penyimpanan data telah mencapai titik di mana kumpulan data menjadi aset penting yang dapat diolah menjadi informasi berharga. Analisis data ini menghasilkan wawasan yang membantu dalam pengambilan keputusan strategis baik untuk keperluan saat ini maupun masa depan. Khususnya, data akademik mahasiswa dari program studi sistem informasi menawarkan potensi yang signifikan untuk diubah menjadi informasi yang dapat dimanfaatkan secara strategis. Sebagai contoh, data ini bisa digunakan untuk

memonitor tingkat kelulusan dalam program studi sistem informasi, yang mana memiliki dampak langsung terhadap nilai Akreditasi dari program tersebut. Perbandingan antara jumlah mahasiswa yang mendaftar dengan yang lulus dalam satu angkatan memberikan gambaran mengenai efektivitas dan kualitas pendidikan yang disediakan.

Berbagai studi telah dilakukan untuk meramalkan tingkat kelulusan mahasiswa dengan menggunakan metode dan atribut yang berbeda-beda. Sebuah studi menggunakan empat atribut—jalur pendaftaran, asal sekolah, asal kota, dan indeks prestasi—untuk memprediksi waktu kelulusan mahasiswa, menghasilkan akurasi sebesar 82,32%, yang menunjukkan klasifikasi yang baik dengan menggunakan metode Naive Bayes Classifier [1]. Studi lainnya memanfaatkan metode Jaccard Coefficient untuk memperkirakan waktu kelulusan, mencapai tingkat akurasi sebesar 80,65% [2]. Selain itu, gabungan metode algoritma bayesian network dengan k-nearest neighbors berhasil digunakan untuk estimasi durasi studi mahasiswa, sementara analisis terhadap prediksi kelulusan berdasarkan indeks prestasi selama dua semester pertama, nilai ujian nasional, jurusan sekolah, lulusan sekolah, dan jalur masuk perguruan tinggi juga telah dilakukan.

Pembuatan model sistem pengambilan keputusan yang berbasis algoritma data mining untuk mengestimasi durasi masa studi mahasiswa menunjukkan potensi lain dari pemanfaatan data akademik. Dengan menggunakan regresi linier berganda, model ini mampu memperkirakan waktu kelulusan, memberikan panduan bagi mahasiswa tentang perkiraan semester kelulusan. Kajian-kajian ini tidak hanya memperkuat nilai data dalam dunia akademik tetapi juga membuka jalan bagi pengembangan metode baru dalam meningkatkan kualitas pendidikan dan pengelolaan institusi pendidikan tinggi [3].

Dalam konteks ini, data akademik mahasiswa dari program studi sistem informasi dapat dianggap sebagai sumber informasi yang kaya, yang, jika dianalisis dengan benar, dapat mengungkapkan wawasan berharga untuk peningkatan dan inovasi. Salah satu aplikasi khusus dari analisis data ini adalah dalam pengembangan sistem pemantauan tingkat kelulusan, yang tidak hanya berfungsi sebagai alat pengukur efektivitas program pembelajaran tetapi juga sebagai indikator kualitas pendidikan yang disajikan oleh suatu program studi. Pengakuan ini terutama penting dalam konteks akreditasi, di mana tingkat kelulusan mahasiswa menjadi salah satu faktor penentu dalam menilai kualitas program studi sistem informasi.

Untuk mengilustrasikan potensi analisis data ini, berbagai penelitian telah dijalankan dengan pendekatan yang beragam. Sebagai contoh, penelitian yang memfokuskan pada prediksi waktu kelulusan mahasiswa dengan memanfaatkan atribut spesifik seperti jalur pendaftaran dan asal sekolah telah memberikan hasil yang menjanjikan dengan tingkat akurasi yang tinggi. Pendekatan ini tidak hanya memberikan nilai tambah dalam hal pemahaman tentang dinamika kelulusan tetapi juga menawarkan panduan praktis bagi pengambilan keputusan strategis dalam pengelolaan program pendidikan tinggi.

## II. SIGNIFIKANSI STUDI

Proses ekstraksi informasi dari data melibatkan penerapan berbagai teknik yang berasal dari bidang statistik, matematika, pembelajaran mesin (machine learning), dan kecerdasan buatan (artificial intelligence) [4]. Teknik-teknik ini memungkinkan kita untuk mengeksplorasi dan menemukan wawasan berharga dari kumpulan data yang besar, sebuah proses yang dikenal sebagai penambangan data (data mining). Dalam konteks analisis data akademik, metode Estimasi Regresi Linier menjadi alat yang efektif untuk mengidentifikasi pola dan mengembangkan formula yang bisa digunakan untuk memprediksi hasil masa depan, seperti durasi masa studi mahasiswa.

Sebelum data dapat dianalisis lebih lanjut, langkah awal yang penting adalah pra-pemrosesan data, yang melibatkan pengurutan dan pembersihan data dari informasi yang tidak relevan atau mengganggu, sehingga data yang tersisa lebih fokus dan relevan untuk analisis. Langkah ini krusial dalam memastikan keakuratan analisis yang akan dilakukan.

Selanjutnya, penentuan data uji yang akan digunakan untuk menguji model regresi linier dihasilkan dari kumpulan data pelatihan sangat penting. Biasanya, data dibagi dengan proporsi 70% untuk pelatihan dan 30% untuk pengujian, memungkinkan model yang dibangun untuk diuji efektivitasnya dalam kondisi yang mirip dengan aplikasi nyata. Variabel prediktor (X) dalam konteks ini merujuk pada variabel yang mempengaruhi variabel lain (variabel independen), sedangkan variabel kriteria (Y) adalah variabel yang dipengaruhi (variabel dependen). Dengan mengidentifikasi hubungan antara variabel prediktor dan kriteria, kita dapat mengembangkan persamaan regresi linier yang dapat digunakan untuk memperkirakan tingkat kelulusan mahasiswa di program studi Sistem Informasi.

Melalui proses ini, model yang dihasilkan tidak hanya memberikan wawasan tentang faktor-faktor yang mempengaruhi durasi studi mahasiswa tetapi juga membantu dalam pengembangan strategi yang lebih efektif untuk meningkatkan kinerja akademik dan efisiensi program studi.

$$Y = b_0 + b_1X_1 + b_2X_2 + b_3X_3 + b_4X_4 + b_5X_5 + b_6X_6 + b_7X_7 + b_8X_8 + b_9X_9 + b_{10}X_{10} \quad (1)$$

Dimana:

Y = variable terikat (Dependen)

X = Variable tidak terikat (Independen)

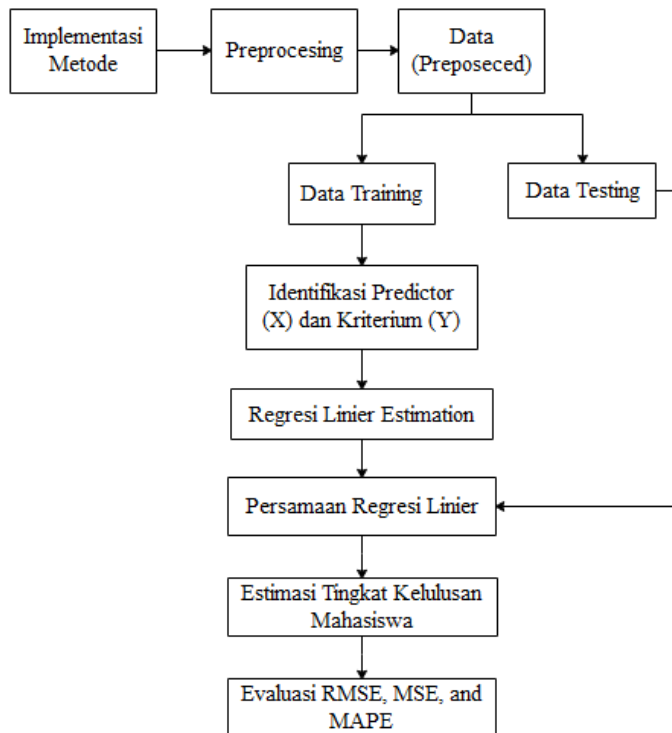
$b_0$  = Konstanta

$b_1 \dots b_n$  = Koefisien arah regresi (kemiringan), besaran respon yang ditimbulkan

Menggunakan persamaan yang telah dibahas sebelumnya, kita dapat melakukan perhitungan estimasi untuk tingkat kelulusan mahasiswa di Program Studi Informatika menggunakan pendekatan regresi linier berganda. Proses validasi model ini melibatkan penggunaan Mean Square Error (MSE) dan Root Mean Squared Error (RMSE) sebagai metrik untuk mengevaluasi performa prediksi model [6].

Kualitas dari suatu prediksi dianggap tinggi jika nilai MSE yang diperoleh rendah, terutama bila dibandingkan dengan nilai yang dihasilkan oleh metode prediksi lain. Nilai MSE yang lebih rendah menandakan bahwa tingkat kesalahan antara nilai yang diprediksi oleh model dengan nilai sebenarnya (observasi) lebih kecil, yang berarti akurasi estimasi model tersebut lebih tinggi [7]. Demikian pula, nilai RMSE yang rendah mengindikasikan bahwa perbedaan antara hasil prediksi model dengan data aktual cukup kecil, menjadikan model tersebut dapat diandalkan dalam melakukan estimasi.

Evaluasi ini penting karena menyediakan wawasan tentang seberapa dekat prediksi yang dihasilkan oleh model dengan kondisi nyata, memungkinkan para peneliti dan praktisi untuk menilai efektivitas model regresi linier berganda dalam konteks estimasi tingkat kelulusan mahasiswa. Dengan demikian, penggunaan metrik evaluasi ini memainkan peran krusial dalam proses validasi model prediktif dan merupakan langkah penting dalam memastikan bahwa model yang dikembangkan dapat diandalkan dan akurat dalam penerapannya [8].



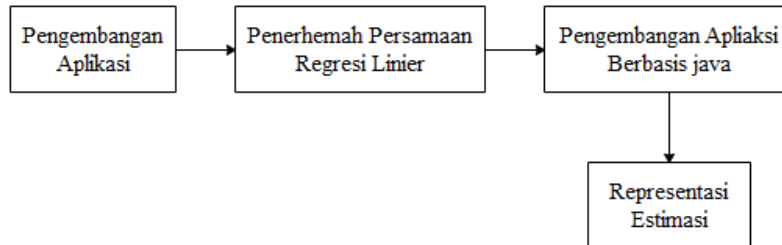
Gambar 1. Alur implementasi metode

Unified Modeling Language (UML) merupakan sebuah bahasa pemrograman visual yang digunakan untuk memodelkan dan mendefinisikan sistem melalui penggunaan diagram serta teks penjelas. UML berperan vital dalam tahapan perencanaan pengembangan prototype, memungkinkan para pengembang untuk merancang dan menyusun kerangka kerja sistem sebelum memasuki fase pengembangan yang lebih kompleks. Fungsinya sebagai alat perencanaan dan komunikasi antar tim pengembang menjadikan UML sebuah komponen penting dalam proses pembuatan perangkat lunak [10]. Metode data mining telah terbukti menjadi pendekatan yang lebih cepat dan efisien bila dibandingkan dengan K-Means dalam hal memberikan rekomendasi [11]. Algoritma ini mengidentifikasi itemset yang sering muncul bersama-sama dengan menghitung nilai support dan confidence untuk setiap kombinasi produk. Nilai support menunjukkan seberapa sering itemset muncul dalam kumpulan data, sementara confidence menunjukkan seberapa sering aturan asosiasi terbukti benar [12].

Prototype dalam konteks sistem informasi merujuk pada versi awal dari sistem yang dapat berfungsi, namun masih dalam tahap awal dan memerlukan pengembangan lebih lanjut [13]. Tujuan utama dari prototype adalah untuk memberikan gambaran awal tentang bagaimana sistem akan beroperasi, memfasilitasi evaluasi awal terhadap desain sistem, dan mendapatkan umpan balik yang bisa digunakan untuk perbaikan. Ini merupakan langkah penting dalam proses desain sistem informasi, memberikan kesempatan untuk menyesuaikan dan memperbaiki fitur-fitur sistem sebelum pengembangan lebih lanjut [14]. Kombinasi algoritma Bayesian Network dan K-Nearest Neighbors menunjukkan bahwa integrasi kedua algoritma ini efektif dalam memperkirakan durasi pendidikan mahasiswa, memberikan akurasi yang lebih baik dalam prediksi masa studi dibandingkan dengan penggunaan masing-masing algoritma secara terpisah [15].

Dalam pengembangan prototype tersebut, bahasa pemrograman Hypertext Preprocessor (PHP) sering kali digunakan bersamaan dengan sistem manajemen basis data MySQL. Kombinasi PHP dan MySQL menyediakan platform yang kuat untuk pengembangan aplikasi web, memungkinkan pembuatan sistem yang interaktif, dinamis, dan mampu mengelola data dengan efisien. PHP, sebagai bahasa pemrograman server-side, berperan dalam pemrosesan

logika aplikasi, sedangkan MySQL bertindak sebagai basis data backend, tempat penyimpanan data sistem. Penggunaan kedua teknologi ini dalam pembuatan prototype memungkinkan pengembangan sistem yang dapat diuji dan dievaluasi dalam lingkungan yang mirip dengan operasi nyata.



Gambar 2. Alur pengembangan implementasi metode

### III. HASIL DAN PEMBAHASAN

Bagian Hasil dan Pembahasan memuat hasil-hasil dari penelitian serta pembahasan menyeluruh dari masing-masing hasil yang didapatkan dari penelitian yang dibahas. Berikan pembahasan yang mendetail tentang hasil-hasil yang didapatkan hingga dapat menjawab permasalahan yang disebutkan di bagian Pendahuluan.

#### A. Multiple Linear Regression Analysis

Langkah awal dalam analisis regresi linier berganda adalah pra-pengolahan data, yang mencakup penghapusan atribut, penghapusan rekaman, dan pembersihan data. Dataset awal terdiri dari 1.753 rekaman mahasiswa dari Program Studi Sistem Informasi di Universitas Dian Nuswantoro, yang meliputi angkatan tahun 2019 hingga 2023, diperoleh dari Pusat Data dan Informasi Universitas. Setelah melalui tahap pra-pengolahan, dataset tersebut disempurnakan menjadi 1.000 rekaman. Kemudian, dataset dibagi menjadi dua bagian: set pelatihan yang terdiri dari 700 rekaman dan set pengujian yang terdiri dari 300 rekaman. Atribut-atribut berikut ini dipilih untuk digunakan dalam penelitian ini:

TABEL I  
PENGUNAAN ATRIBUT

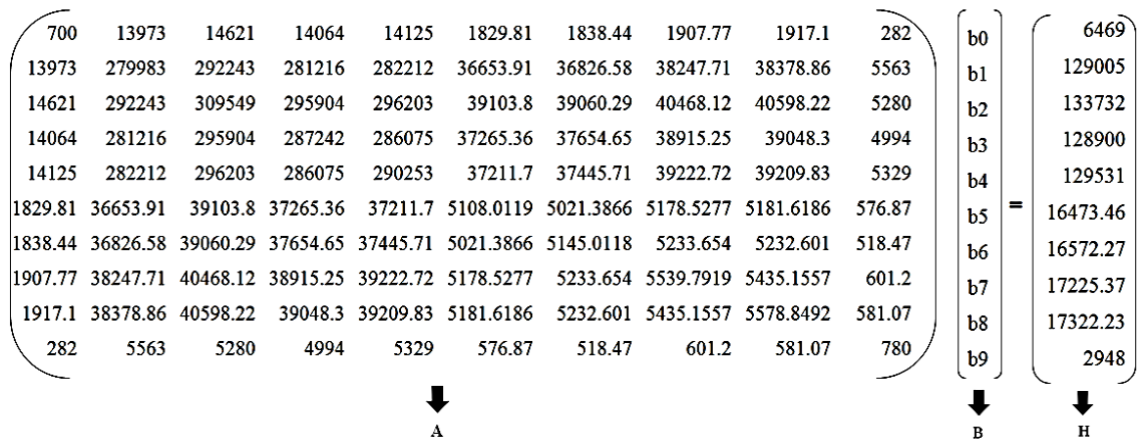
Atribut	Notasi	Atribut	Notasi
Estimasi lulus di semester ke	Y	IPS 2	X <sub>6</sub>
SKS 1	X <sub>1</sub>	IPS 3	X <sub>7</sub>
SKS 2	X <sub>2</sub>	IPS 4	X <sub>8</sub>
SKS 3	X <sub>3</sub>	Jumlah mata kuliah mengulang	X <sub>9</sub>
SKS 4	X <sub>4</sub>	Konstanta	A
IPS 1	X <sub>5</sub>	Koefisien regresi	b <sub>1</sub> ...b <sub>n</sub>

Bentuk persamaan secara umum pada regresi linier berganda sehingga hasil persamaan untuk menghitung estimasi kelulusan mahasiswa sebagai berikut:

$$Y = b_0 + b_1 (\text{SKS 1}) + b_2 (\text{SKS 2}) + b_3 (\text{SKS 3}) + b_4 (\text{SKS 4}) + b_5 (\text{IPS 1}) + b_6 (\text{IPS 2}) + b_7 (\text{IPS 3}) + b_8 (\text{IPS 4}) + b_9 (\text{Makul Mengulang}).$$

Mengembangkan persamaan regresi linier berganda bertujuan untuk memperkirakan waktu kelulusan mahasiswa, melibatkan penggunaan sebuah formula yang terdiri dari sepuluh variabel. Proses ini dimulai dengan penentuan variabel-variabel yang relevan dan kemudian mengintegrasikannya ke dalam sebuah model matematika. Model ini dirancang untuk memprediksi dengan seakurat mungkin, berdasarkan data yang tersedia, kapan seorang

mahasiswa diharapkan menyelesaikan studinya. Melalui pendekatan regresi linier berganda, persamaan yang dihasilkan akan menggabungkan variabel-variabel tersebut dalam bentuk yang memungkinkan penghitungan estimasi kelulusan mahasiswa dengan lebih efisien dan efektif. Persamaan ini akan menjadi alat penting dalam analisis data akademik, memberikan wawasan yang berharga untuk pengambilan keputusan dalam konteks akademis dan administratif.



Gambar 3. Perkalian matriks dengan koefisien regresi

$$Ab = H$$

$$B = A^{-1} H$$

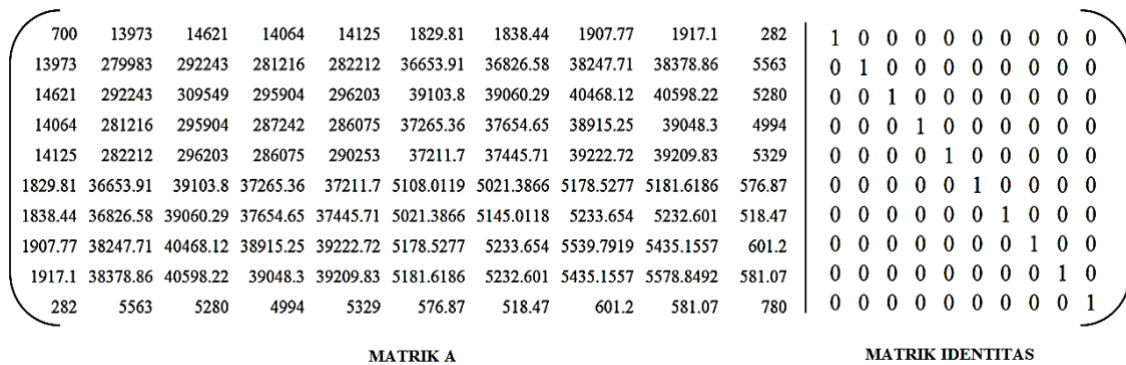
Dimana:

- A = Matriks (diketahui)
- H = Vektor kolom (diketahui)
- b = Vektor kolom (tidak diketahui)
- A<sup>-1</sup> = Kebalikan (invers) dari matriks A

- 1  $700b_0 + 13973b_1 + 14621b_2 + 14064b_3 + 14125b_4 + 1829.81b_5 + 1838.44b_6 + 1907.77b_7 + 1917.1b_8 + 282b_9 = \mathbf{6469}$
- 2  $13973b_0 + 279983b_1 + 292243b_2 + 281216b_3 + 282212b_4 + 36653.91b_5 + 36826.58b_6 + 38247.71b_7 + 38378.86b_8 + 5563b_9 = \mathbf{129005}$
- 3  $14621b_0 + 292243b_1 + 309549b_2 + 295904b_3 + 296203b_4 + 39103.8b_5 + 39060.29b_6 + 40468.12b_7 + 40598.22b_8 + 5280b_9 = \mathbf{133732}$
- 4  $14064b_0 + 281216b_1 + 295904b_2 + 287242b_3 + 286075b_4 + 37265.36b_5 + 37654.65b_6 + 38915.25b_7 + 39048.3b_8 + 4994b_9 = \mathbf{128900}$
- 5  $14125b_0 + 282212b_1 + 296203b_2 + 286075b_3 + 290253b_4 + 37211.7b_5 + 37445.71b_6 + 39222.72b_7 + 39209.83b_8 + 5329b_9 = \mathbf{129531}$
- 6  $1829.81b_0 + 36653.91b_1 + 39103.8b_2 + 37265.36b_3 + 37211.7b_4 + 5108.0119b_5 + 5021.3866b_6 + 5178.5277b_7 + 5181.6186b_8 + 576.87b_9 = \mathbf{16344,77}$
- 7  $1838.44b_0 + 36826.58b_1 + 39060.29b_2 + 37654.65b_3 + 37445.71b_4 + 5021.3866b_5 + 5145.0118b_6 + 5233.654b_7 + 5232.601b_8 + 518.47b_9 = \mathbf{16275,83}$
- 8  $1907.77b_0 + 38247.71b_1 + 40468.12b_2 + 38915.25b_3 + 39222.72b_4 + 5178.5277b_5 + 5233.654b_6 + 5539.7919b_7 + 5435.1557b_8 + 601.2b_9 = \mathbf{16867,31}$
- 9  $1917.1b_0 + 38378.86b_1 + 40598.22b_2 + 39048.3b_3 + 39209.83b_4 + 5181.6186b_5 + 5232.601b_6 + 5435.1557b_7 + 5435.1557b_8 + 581.07b_9 = \mathbf{17166,23}$
- 10  $282b_0 + 5563b_1 + 5280b_2 + 4994b_3 + 5329b_4 + 576.87b_5 + 518.47b_6 + 601.2b_7 + 581.07b_8 + 780b_9 = \mathbf{2948}$

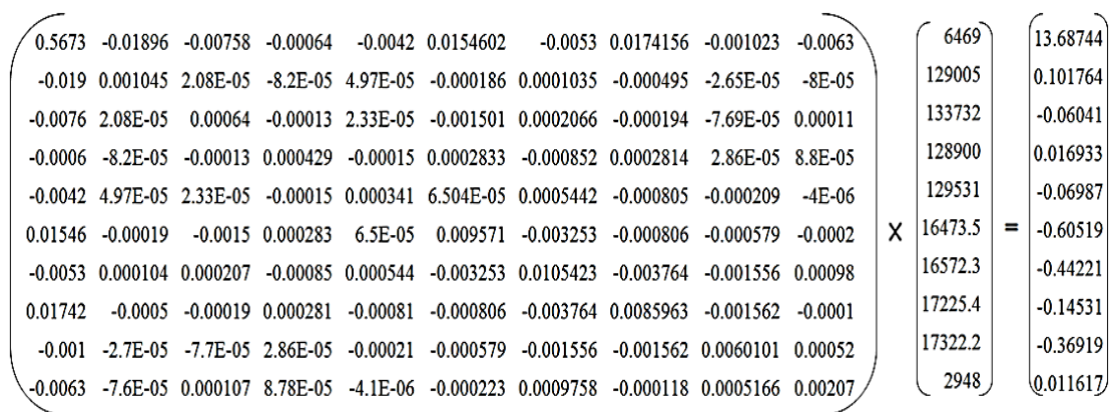
Dalam proses pencarian matriks invers dari A, langkah yang diambil melibatkan operasi baris elementer yang diterapkan pada matriks A bersamaan dengan matriks identitas. Melalui

serangkaian operasi ini, tujuannya adalah untuk mengubah matriks A menjadi matriks identitas. Setelah matriks A berhasil diubah menjadi matriks identitas melalui operasi baris elementer, posisi matriks identitas yang semula berada di sisi kanan akan bertransformasi menjadi matriks invers dari A. Proses ini merupakan teknik standar dalam aljabar linier untuk menghitung matriks invers, memanfaatkan hubungan antara matriks A dan matriks identitas untuk menghasilkan inversnya.



Gambar 4. Perkalian matrik A dengan matrik identitas

Setelah memperoleh matriks invers, langkah selanjutnya adalah mengalikannya dengan serangkaian sumbu-sumbu, termasuk  $\Sigma Y$ ,  $\Sigma X1Y$ ,  $\Sigma X2Y$ ,  $\Sigma X3Y$ ,  $\Sigma X4Y$ ,  $\Sigma X5Y$ ,  $\Sigma X6Y$ ,  $\Sigma X7Y$ ,  $\Sigma X8Y$ , dan  $\Sigma X9Y$ . Proses perkalian ini bertujuan untuk mendapatkan nilai-nilai yang akan digunakan dalam persamaan regresi linier berganda untuk Y. Dengan demikian, hasil dari perkalian ini akan menyediakan koefisien-koefisien yang diperlukan untuk membentuk persamaan akhir Y. Langkah ini merupakan bagian penting dalam proses pembuatan model regresi, dimana nilai-nilai ini memainkan peran kunci dalam menentukan hubungan antara variabel independen dan variabel dependen.



Gambar 5. Perkalian matrik inver dengan H (vektor kolom)

Persamaan (Y) regresi linier berganda untuk menentukan estimasi kelulusan Mahasiswa Sistem Informasi dapat menggunakan persamaan berikut ini:  
 $Y = 13.49 + 0.099 X1 + (-0.068) X2 + 0.025 X3 + (-0.059) X4 + (-0.585) X5 + (-0.443) X6 + (-0.155) X7 + (-0.368) X8 + (-0.082) X9$

**B. Evaluasi Model dan Proses Validasi**

Dalam proses evaluasi dan validasi model yang dikembangkan, metrik yang dipilih untuk menguji kinerjanya adalah Mean Square Error (MSE) dan Root Mean Square Error (RMSE). Untuk melakukan perhitungan ini, Microsoft Excel digunakan dengan memasukkan data uji

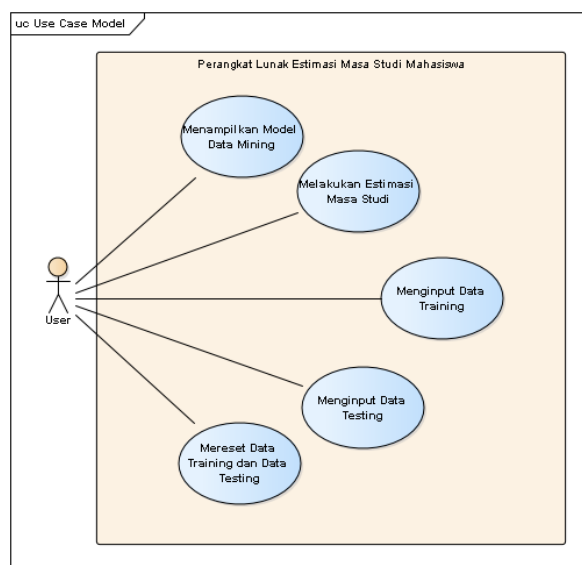


sejumlah 300 data. Berdasarkan persamaan yang telah ditetapkan sebelumnya, didapatkan hasil nilai MSE sebesar 0.1168 dan RMSE sebesar 0.3418. Referensi dalam literatur menunjukkan bahwa nilai RMSE yang lebih tinggi menandakan bahwa model memiliki tingkat ketidakakuratan yang lebih besar dalam melakukan estimasi atau prediksi. Oleh karena itu, dalam konteks evaluasi model regresi, nilai RMSE yang rendah dianggap menunjukkan kinerja model yang lebih akurat dan andal.

### C. Desain prototipe

Perancangan use case merupakan sebuah diagram aktifitas aktor dalam prototype untuk mengetahui kegiatan aktor dalam menjalankan prototype. Gambar yang disajikan adalah sebuah diagram Use Case untuk sebuah prototype yang berkaitan dengan estimasi masa studi mahasiswa. Diagram ini menggambarkan interaksi antara pengguna (User) dan sistem dengan mengidentifikasi fungsi-fungsi atau 'use cases' yang dapat dijalankan oleh pengguna dalam sistem tersebut. Dari diagram dapat dilihat ada lima use cases utama:

- Menampilkan Model Data Mining: Fungsi ini menunjukkan bahwa pengguna dapat melihat atau mempresentasikan model data mining yang telah dibuat.
- Melakukan Estimasi Masa Studi: Fungsi ini memungkinkan pengguna untuk melakukan estimasi atau perhitungan terkait dengan durasi masa studi mahasiswa, mungkin berdasarkan model data mining yang telah disajikan.
- Menginput Data Training: Ini menunjukkan bahwa pengguna memiliki kemampuan untuk memasukkan atau meng-upload data yang akan digunakan sebagai data latih (training) dalam pembuatan atau peningkatan model prediksi.
- Menginput Data Testing: Serupa dengan penginputan data training, namun data ini digunakan untuk menguji dan memvalidasi model yang telah dibuat, untuk menentukan seberapa akurat model tersebut dalam melakukan prediksi.
- Mereset Data Training dan Data Testing: Use case ini menyediakan fungsi bagi pengguna untuk mengatur ulang atau menghapus data training dan testing yang ada dalam sistem, mungkin untuk memulai proses pembuatan model dari awal atau untuk menghapus data lama dan memperbarui dengan data yang baru.



Gambar 6. Use Case diagram prototype

Desain antarmuka ini mencerminkan sistem yang interaktif dan user-friendly yang memfasilitasi pengguna dalam menghitung estimasi masa studi dengan memasukkan data yang



relevan ke dalam model regresi linier berganda yang telah dikembangkan. Gambar yang ditampilkan merupakan desain antarmuka untuk menu formulir estimasi pada aplikasi berbasis web yang menggunakan Regresi Linier Berganda. Antarmuka ini dibagi menjadi beberapa bagian:

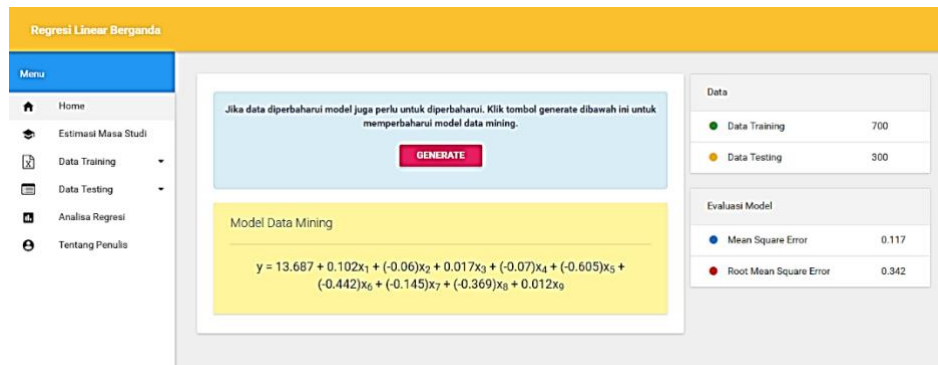
- Menu Navigasi: Pada sisi kiri, terdapat menu navigasi vertikal yang memberikan akses pengguna ke berbagai fitur aplikasi seperti Home, Estimasi Masa Studi, Data Training, Data Testing, Analisa Regresi, dan Testing Hipotesis.
- Form Estimasi Masa Studi: Bagian tengah menampilkan formulir yang digunakan untuk memasukkan data yang diperlukan dalam melakukan estimasi masa studi mahasiswa. Form ini terdiri dari berbagai bidang untuk memasukkan jumlah SKS yang diambil per semester (SKS Semester 1, SKS Semester 2, dst.) dan Indeks Prestasi (IPS) per semester.
- Jumlah Mata Kuliah Mengulang: Ada juga bidang khusus untuk memasukkan jumlah mata kuliah yang diulang oleh mahasiswa, yang mungkin digunakan sebagai salah satu variabel dalam model regresi.
- Tombol Aksi: Di bagian bawah formulir, terdapat tombol "Submit" untuk mengirimkan data yang diinputkan ke sistem dan tombol "Reset" untuk membersihkan formulir dari data yang telah dimasukkan.
- Upload File dengan Format CSV: Di sisi kanan atas, terdapat fitur yang memungkinkan pengguna untuk meng-upload file data dalam format CSV, yang kemudian akan digunakan oleh sistem untuk proses estimasi. Ini menyediakan cara alternatif bagi pengguna untuk memasukkan data secara massal daripada memasukkan secara manual satu per satu.
- Tombol Aksi untuk Upload: Terdapat tombol "Choose File" untuk memilih file yang ingin di-upload dari komputer pengguna dan tombol "Submit" khusus untuk memproses file yang telah dipilih, bersamaan dengan tombol "Reset" untuk membatalkan pilihan file.

Gambar 7. Desain tampilan menu form estimasi

Keseluruhan tampilan antarmuka ini dirancang untuk memberikan pengguna kemudahan dalam mengakses dan memahami model data mining yang telah dibuat, serta untuk memudahkan proses evaluasi kinerja model berdasarkan metrik yang telah ditetapkan. Gambar tersebut menampilkan bagian dari antarmuka aplikasi web yang berkaitan dengan model data mining menggunakan regresi linier berganda. Dari elemen yang dapat dilihat, berikut adalah penjelasan dari berbagai komponen:

- Menu Navigasi: Sama seperti gambar sebelumnya, terdapat menu navigasi di sisi kiri yang menyediakan akses ke berbagai fitur utama aplikasi seperti Home, Estimasi Masa Studi, Data Training, Data Testing, Analisa Regresi, dan Testing Hipotesis.
- Seksi Model Data Mining: Di bagian tengah, terdapat sebuah panel yang menampilkan model persamaan regresi linier berganda. Persamaan ini disajikan dalam bentuk matematik yang memberikan nilai konstanta dan koefisien untuk setiap variabel independen ( $X_1$  hingga  $X_9$ ), yang digunakan untuk melakukan estimasi atau prediksi.
- Tombol "Generate": Di bawah deskripsi, ada tombol "Generate" yang kemungkinan memungkinkan pengguna untuk menjalankan fungsi atau algoritma data mining yang akan menghasilkan model regresi linier berdasarkan data yang ada atau yang baru diinputkan.

- d. Kotak Informasi: Di sisi kanan terdapat kotak yang memberikan informasi ringkas mengenai jumlah data yang digunakan untuk training dan testing, serta nilai evaluasi model yang meliputi Mean Square Error (MSE) dan Root Mean Square Error (RMSE), yang merupakan indikator kinerja dari model yang dihasilkan.



Gambar 8. Model persamaan data mining

#### IV. KESIMPULAN

Kesimpulan penelitian ini mengindikasikan bahwa aplikasi dari metode data mining menggunakan regresi linier berganda telah berhasil membentuk sebuah model persamaan regresi (Y) untuk memperkirakan waktu kelulusan mahasiswa di program studi Sistem Informasi. Model ini memasukkan sembilan variabel independen yang didefinisikan sebagai berikut: X1 sampai X4 mewakili jumlah SKS yang diambil pada semester pertama hingga keempat, X5 sampai X8 merepresentasikan Indeks Prestasi Semester (IPS) dari semester pertama hingga keempat, dan X9 melambangkan jumlah mata kuliah yang diulang. Penggunaan sampel data yang luas, mencakup 1000 entri dari kohort tahun 2019 hingga 2023 dari Departemen Sistem Informasi, memungkinkan pembentukan persamaan regresi yang robust, yakni:

$$Y = 13.49 + 0.099X_1 - 0.068X_2 + 0.025X_3 - 0.059X_4 - 0.585X_5 - 0.443X_6 - 0.155X_7 - 0.368X_8 - 0.082X_9.$$

Model ini diharapkan dapat dijadikan alat yang efektif dalam membantu pengambilan keputusan akademik dan strategi pendukung mahasiswa.

#### REFERENCES

- [1] Vira, "Pemodelan data mining untuk prediksi Kelulusan mahasiswa dengan metode Naive bayes classifier," Tugas Akhir Sistem Informasi Universitas Dian Nuswantoro. Semarang. 2015
- [2] K. Hafidh, "Memprediksi Masa Studi Mahasiswa Menggunakan Metode Jaccard Coefficient (Studi Kasus: Mahasiswa Program Studi Teknik Informatika Jurusan Teknik Elektro Fakultas Teknik Universitas Tanjungpura)," Kalimantan Tengah. 2015
- [3] Windarti M, "Prediksi Masa Studi Mahasiswa Menggunakan Kombinasi Algoritma Bayesian Network Dan K- Nearest Neighbors," Tesis Universitas Adma Jaya Yogyakarta, Program Studi Magister Teknik Informatika. Yogyakarta. 2016
- [4] Tedyyana, Agus, et al. "Enhance Telecommunication Security Through the Integration of Support Vector Machines." *International Journal of Advanced Computer Science & Applications* 15.3 (2024).

- [5] Pramitarini Y. Dkk. Analisa Rekam Medis Untuk Menentukan Status Gizi Anak Balita Menggunakan Naive Bayes Classifier. Prosiding Seminar Nasional Manajemen Teknologi XVII. Program Studi MMT-ITS, Surabaya 2 Februari 2013
- [6] Ali Fikri, "Penerapan Data Mining Untuk Mengetahui Tingkat Kekuatan Beton Yang Dihasilkan Dengan Metode Estimasi Menggunakan Linear Regression," [eprints.dinus.ac.id/12789/1/jurnal12969.pdf](http://eprints.dinus.ac.id/12789/1/jurnal12969.pdf). Semarang. 2013
- [7] A. A. Ghofur dan U. D. Widiarti, "Sistem Peramalan Untuk Pengadaan Material Unit Injection di PT. XYZ," *Jurnal Ilmiah Komputer dan Informatika (KOMPUTA)*, vol. III, no. 2, pp. 13-18, 2013.
- [8] James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). *An Introduction to Statistical Learning* (p. 68). New York: Springer.
- [9] Shalahuddin, M. dan Rosa A.S. *Rekayasa Perangkat Lunak Terstruktur dan Berorientasi Objek*. Bandung: Informatika. 2013
- [10] Anisya, "Aplikasi Sistem Database Rumah Sakit Terpusat Pada Rumah SAKIT Umum (RSU) 'Aisyiyah Padang Dengan Menerapkan Open Source (PHP - MySQL)," *Jurnal Momentum*, vol. 15, no. 2, pp. 1-10, 2013.
- [11] I. K. Juni Arta, G. Indrawan, and G. R. Dantes, "Data Mining Rekomendasi Calon Mahasiswa Berprestasi di STMIK Denpasar Menggunakan Metode Technique for Others Reference by Similarity to Ideal Solution," *Jurnal Ilmu Komputer Indonesia (JIKI)*, vol. 4, no. 1, pp. [page numbers], Feb. 2019, ISSN (Print): 2615-2703, ISSN (Online): 2615-2711.
- [12] P. N. Harahap and S. Sulindawaty, "Implementasi Data Mining Dalam Memprediksi Transaksi Penjualan Menggunakan Algoritma Apriori (Studi Kasus PT. Arma Anugerah Abadi Cabang Sei Rampah)," *MATICS: Jurnal Ilmu Komputer dan Teknologi Informasi*, vol. 11, no. 2, pp. 46-50, Dec. 2019, doi: 10.18860/mat.v11i2.7821.
- [13] I. Sungkawa dan R. T. Megasari, "Penerapan Ukuran Ketetapan Nilai Ramalan Data Deret Waktu Dalam Seleksi Model Peramalan Volume Penjualan PT. Satria Mandiri Citra Mulia," *ComTech*, vol. 2, no. 2, hal. 636-645, 2011.
- [14] A. A. Ghofur dan U. D. Widiarti, "Sistem Peramalan Untuk Pengadaan Material Unit Injection di PT. XYZ," *Jurnal Ilmiah Komputer dan Informatika (KOMPUTA)*, vol. 3, no. 2, hal. 13-18, 2013.
- [15] Windarti M., "Prediksi Masa Studi Mahasiswa Menggunakan Kombinasi Algoritma Bayesian Network Dan K-Nearest Neighbors," Tesis, Program Studi Magister Teknik Informatika, Universitas Atma Jaya Yogyakarta, Yogyakarta, 2016.

#### UCAPAN TERIMA KASIH

Ucapan terimakasih saya berikan kepada Univet yang telah memberikan dukungan dalam penelitian ini. Khususnya pada Jurusan Kemaritiman yang telah memfasilitasi penelitian ini.