

Prediksi Semester Tugas Akhir Mahasiswa Berdasarkan Transkrip Nilai Menggunakan *Linear Regression, Kernel Ridge Regression dan Decision Tree Regression*

Eki Ahmad Zaki Hamidi^{1*)}, Edi Mulyana²⁾, Dilla Restu Agusthiani³⁾, Aldi Fahruci Muharam⁴⁾

^{1,2,3)}Jurusan Teknik Elektro

UIN Sunan Gunung Djati Bandung

Jalan AH. Nasution 104 Bandung 40614

⁴⁾Jurusan Teknik Informatika, UIN Sunan Gunung Djati Bandung

^{*)}Korespondensi : ekiahmadzaki@uinsgd.ac.id

Abstrak

Penelitian ini bertujuan untuk memprediksi semester penyelesaian tugas akhir mahasiswa berdasarkan data transkrip nilai dengan menggunakan tiga algoritma regresi: Linear Regression, Kernel Ridge Regression, dan Decision Tree Regression. Kinerja setiap model dievaluasi menggunakan metrik Mean Squared Error (MSE) dan Mean Absolute Error (MAE). Hasil eksperimen menunjukkan bahwa Kernel Ridge Regression memiliki performa terbaik dengan nilai MSE sebesar 2,271 dan MAE sebesar 1,251. Sebagai perbandingan, Linear Regression menghasilkan MSE sebesar 5,137 dan MAE sebesar 1,859, sedangkan Decision Tree Regression menghasilkan MSE sebesar 4,1 dan MAE sebesar 1,2. Temuan ini mengindikasikan bahwa Kernel Ridge Regression merupakan metode yang paling efektif untuk memprediksi semester penyelesaian berdasarkan transkrip nilai, memberikan hasil yang lebih akurat dan andal. Penelitian ini berkontribusi pada bidang akademik dengan menunjukkan potensi model pembelajaran mesin dalam memprediksi perkembangan akademik mahasiswa dan mendukung pengambilan keputusan yang lebih baik dalam pengelolaan akademik. Dengan demikian, hasil penelitian ini dapat dijadikan dasar bagi institusi pendidikan untuk mengembangkan strategi akademik yang lebih adaptif dan berbasis data guna meningkatkan efisiensi serta efektivitas pendidikan tinggi secara keseluruhan.

Kata kunci : Prediksi Semester, Tugas Akhir Mahasiswa, *Linear Regression, Kernel Ridge Regression, Decision Tree Regression*, Transkrip Nilai.

Abstract

This study aims to predict the semester of thesis completion for students based on transcript data using three regression algorithms: Linear Regression, Kernel Ridge Regression, and Decision Tree Regression. The performance of each model is evaluated using the Mean Squared Error (MSE) and Mean Absolute Error (MAE) metrics. Experimental results show that Kernel Ridge Regression performs the best, with an MSE of 2.271 and an MAE of 1.251. In comparison, Linear Regression yields an MSE of 5.137 and an MAE of 1.859, while Decision Tree Regression produces an MSE of 4.1 and an MAE of 1.2. These findings indicate that Kernel Ridge Regression is the most effective method for predicting the completion semester based on transcript data, providing more accurate and reliable results. This research contributes to the academic field by demonstrating the potential of machine learning models in predicting students' academic progress and supporting better decision-making in academic management. Thus, the findings of this study can serve as a foundation for educational institutions to develop more adaptive and data-driven academic strategies to enhance the efficiency and effectiveness of higher education.

Keywords : Semester Prediction, Student Thesis, *Linear Regression, Kernel Ridge Regression, Decision Tree Regression, Academic Transcripts.*

Info Makalah:

Dikirim : 01-20-2025;
Revisi 1 : 02-03-2025;
Revisi 2 : mm-dd-yy;
Diterima : 02-05-2025.

Penulis Korespondensi:

Telp : +62-852-2412-6792
e-mail : ekiahmadzaki@uinsgd.ac.id

I. PENDAHULUAN

Penyelesaian tugas akhir mahasiswa menjadi salah satu indikator penting dalam menilai keberhasilan akademik dan kualitas pendidikan di perguruan tinggi. Tantangan utama dalam dunia pendidikan tinggi adalah memahami dan memprediksi faktor-faktor yang memengaruhi waktu penyelesaian tugas akhir. Prediksi yang akurat mengenai semester penyelesaian tugas akhir dapat membantu pengelola program studi untuk mengidentifikasi mahasiswa yang berpotensi mengalami keterlambatan. Dengan demikian, intervensi yang tepat waktu dapat dilakukan untuk mendukung mahasiswa menyelesaikan studi mereka sesuai jadwal. Salah satu pendekatan yang menjanjikan dalam memprediksi semester penyelesaian tugas akhir adalah penerapan pembelajaran mesin (*machine learning*) berbasis data transkrip nilai[1].

Transkrip nilai mahasiswa berisi informasi penting mengenai kinerja akademik sepanjang masa studi, seperti nilai mata kuliah, indeks prestasi, dan jumlah kredit yang telah diselesaikan. Data ini memiliki potensi besar untuk digunakan dalam prediksi akademik karena mencerminkan pola dan tren kinerja mahasiswa. Namun, sifat data transkrip yang sering kali kompleks dan non-linear memerlukan metode analisis yang tepat untuk menghasilkan prediksi yang akurat. Oleh karena itu, penelitian ini memanfaatkan tiga algoritma regresi, yaitu *Linear Regression*, *Kernel Ridge Regression*, dan *Decision Tree Regression*, untuk memprediksi semester penyelesaian tugas akhir berdasarkan data transkrip nilai mahasiswa [2].

Ketiga algoritma ini dipilih karena mewakili pendekatan yang berbeda dalam regresi. *Linear Regression* dipilih dalam penelitian ini karena sifatnya yang sederhana dan kemampuannya dalam memodelkan hubungan linear antara variabel independen (nilai transkrip) dan variabel dependen (semester tugas akhir). Model ini sering digunakan sebagai baseline untuk membandingkan performa model lainnya. *Kernel Ridge Regression*, sebagai pengembangan dari *Linear Regression*, menawarkan fleksibilitas tambahan dengan menggunakan fungsi kernel untuk menangkap pola data yang lebih kompleks dan non-linear. Sementara itu, *Decision Tree Regression* dipilih karena kemampuannya menangani data non-linear dengan baik dan menghasilkan model yang dapat diinterpretasikan dengan mudah. Setiap algoritma memiliki kelebihan dan kekurangan yang dapat diukur melalui berbagai metrik evaluasi [3][4].

Dalam penelitian ini, performa ketiga model dievaluasi menggunakan metrik *Mean Squared Error* (MSE) dan *Mean Absolute Error* (MAE). MSE digunakan untuk mengukur tingkat kesalahan prediksi secara keseluruhan, sedangkan MAE memberikan gambaran rata-rata kesalahan absolut dalam prediksi. Dengan membandingkan hasil ketiga model ini, penelitian bertujuan untuk menentukan algoritma yang paling sesuai untuk memprediksi semester tugas akhir mahasiswa berdasarkan data transkrip. Selain itu, evaluasi terhadap efisiensi dan kecepatan komputasi juga dilakukan untuk memastikan implementasi model dapat diterapkan secara praktis dalam pengelolaan akademik[5][6].

Makalah ini diharapkan dapat memberikan kontribusi signifikan dalam bidang prediksi akademik, terutama dalam mendukung pengelolaan pendidikan di perguruan tinggi. Dengan menggunakan pendekatan berbasis pembelajaran mesin, hasil penelitian ini dapat membantu institusi pendidikan dalam mengidentifikasi mahasiswa yang berisiko mengalami keterlambatan tugas akhir dan memberikan solusi yang lebih terarah. Selain itu, hasil penelitian ini juga diharapkan dapat menjadi referensi bagi penelitian lanjutan dalam mengembangkan model prediksi yang lebih akurat dan efisien untuk berbagai kebutuhan akademik.

II. STUDI LITERATUR

Penelitian yang dilakukan oleh Muhammad Naufal Faruqhy, dkk, tentang "*Prediksi Prestasi Nilai Akademik Mahasiswa Menggunakan Metode Multiple Linear Regression*," bertujuan memprediksi prestasi akademik mahasiswa berdasarkan jalur masuk perguruan tinggi, sehingga mahasiswa dapat memperoleh gambaran mengenai Indeks Prestasi Semester (IPS) pada semester berikutnya. Variabel independen yang digunakan dalam penelitian ini adalah jalur masuk perguruan tinggi dan IPS

sebelumnya, dengan sampel penelitian berasal dari mahasiswa Fakultas Teknik Universitas Bengkulu. Pengembangan sistem menggunakan metode *waterfall*, sementara pengujian simpangan error dilakukan menggunakan *Mean Error* (ME), *Mean Absolute Deviation* (MAD), *Mean Squared Error* (MSE), *Root Mean Squared Error* (RMSE), dan *Mean Absolute Percent Error* (MAPE). Pengujian fungsional sistem dengan metode *Black Box* menunjukkan keberhasilan 100% berdasarkan skenario pengujian yang dirancang. Hasil eksperimen menunjukkan tingkat error sebesar ME 0,12, MAD 0,15, MSE 1,53, RMSE 1,24, dan MAPE 4,05%, yang mengindikasikan bahwa sistem ini memiliki tingkat akurasi yang baik dalam memprediksi prestasi akademik mahasiswa[7].

Penelitian lain yang dilakukan oleh Anita Qoiriah dan tim membahas prediksi nilai akhir mahasiswa pada mata kuliah Pemrograman Dasar menggunakan metode regresi. Variabel terikat dalam penelitian ini adalah nilai akhir mahasiswa, sedangkan variabel bebas meliputi satu nilai tugas teori, dua nilai tugas praktikum, dan satu nilai UTS. Data yang digunakan terdiri dari hasil pembelajaran sebanyak 105 mahasiswa, yang kemudian dibagi secara acak menjadi dua kelompok: 75% sebagai data pelatihan (*training*) dan 25% sebagai data pengujian (*testing*). Dari empat model yang dikembangkan, model ketiga menunjukkan performa terbaik dengan nilai RMSE sebesar 1,949. Model ini menggunakan kombinasi prediktor berupa nilai tugas teori, dua nilai praktikum, dan nilai UTS[8].

Penelitian yang dilakukan oleh Althof Thabibi dan tim membandingkan performa model *Multiple Linear Regression* (MLR) dan *Decision Tree Regression* (DTR) dalam memprediksi harga saham Telkom, Indosat, dan XL. Penelitian ini memanfaatkan machine learning berbasis regresi untuk memprediksi data secara akurat. Pengujian dilakukan menggunakan tiga skema pembagian data, yaitu 80:20, 70:30, dan 60:40, dengan evaluasi *error* pada setiap model regresi untuk menentukan model terbaik dalam mengolah dataset. Hasil penelitian menunjukkan perbedaan performa yang signifikan antara MLR dan DTR. Model MLR menghasilkan performa terbaik pada metrik MAE dengan skema 70:30, yaitu sebesar 0.008, sedangkan pada metrik RMSE dan MAPE, performa terbaik diperoleh menggunakan skema 60:40 dengan nilai masing-masing 2.591 dan 0.852. Sementara itu, model DTR menunjukkan hasil terbaik menggunakan skema 80:20, dengan nilai MAE sebesar 0.013, RMSE sebesar 4.014, dan MAPE sebesar 1.287. Kesimpulan dari penelitian ini menunjukkan bahwa model MLR lebih unggul untuk digunakan pada dataset saham perusahaan telekomunikasi selama lima tahun terakhir, karena memiliki nilai *error* yang lebih kecil dibandingkan dengan model DTR, meskipun hasilnya bervariasi pada skema pembagian data yang berbeda[9].

Penelitian yang dilakukan oleh William Jaya Mangala Putra dan tim menganalisis algoritma *regresi linear* dan *Decision Tree* dalam memprediksi penjualan produk. Penelitian ini bertujuan untuk memprediksi hasil penjualan di *Lookma Boutique* guna mengevaluasi efektivitas penjualan berdasarkan popularitas produk, yang diukur melalui total produk yang dibayar dan pesanan yang dibuat. Perbandingan dilakukan antara algoritma Regresi Linear dan Decision Tree, yang keduanya dikenal unggul dalam prediksi. Hasil penelitian menunjukkan bahwa evaluasi menggunakan metrik *Mean Absolute Error* (MAE) menghasilkan nilai sebesar 0.84, *Mean Squared Error* (MSE) sebesar 1.79, dan *R² Score* sebesar 1.00, yang mengindikasikan kinerja prediksi yang sangat baik[10].

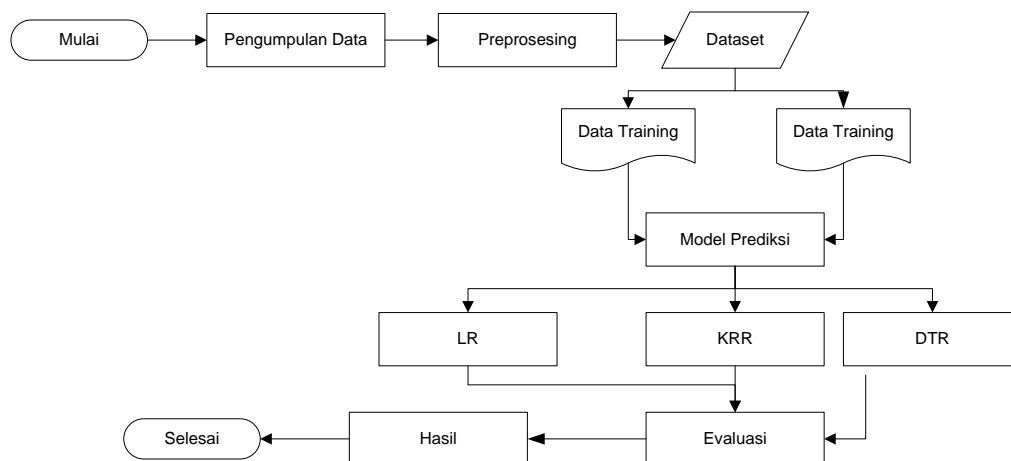
Penelitian yang ada sebelumnya, seperti yang dilakukan oleh Muhammad Naufal Faruqhy dan tim, berfokus pada prediksi prestasi akademik mahasiswa berdasarkan jalur masuk perguruan tinggi dan Indeks Prestasi Semester (IPS) sebelumnya dengan menggunakan metode *Multiple Linear Regression*. Namun, penelitian ini hanya mempertimbangkan variabel terbatas dan tidak mengeksplorasi faktor akademik yang lebih holistik, seperti data transkrip nilai lengkap yang mencakup seluruh riwayat akademik mahasiswa. Hal ini memberikan celah dalam literatur untuk memperluas cakupan variabel yang digunakan dalam prediksi perkembangan akademik, khususnya untuk prediksi penyelesaian tugas akhir mahasiswa yang melibatkan lebih dari sekedar prestasi masa lalu, tetapi juga dinamika kinerja akademik secara keseluruhan.

Selain itu, penelitian yang dilakukan oleh Anita Qoiriah dan tim hanya memfokuskan pada prediksi nilai akhir mahasiswa dalam mata kuliah tertentu, dengan variabel terbatas pada nilai tugas teori, praktikum, dan UTS, tanpa mempertimbangkan aspek longitudinal atau performa keseluruhan mahasiswa dalam menyelesaikan studi mereka. Penelitian ini tidak menyentuh bagaimana seluruh transkrip nilai dapat digunakan untuk memprediksi sesuatu yang lebih kompleks, seperti waktu penyelesaian tugas akhir. Penelitian yang diusulkan bertujuan untuk mengisi gap ini dengan memfokuskan pada prediksi semester penyelesaian tugas akhir mahasiswa berdasarkan data transkrip nilai lengkap, serta membandingkan kinerja tiga algoritma regresi—*Linear Regression*, *Kernel Ridge Regression*, dan *Decision Tree Regression*—dalam menghasilkan prediksi yang lebih akurat.

III. METODE

A. Diagram Alir Penelitian

Penelitian ini mengusulkan model prediksi menggunakan algoritma *Linear Regression*, *Kernel Ridge Regression* dan *Decision Tree Regression*, yang dievaluasi berdasarkan nilai *Mean Squared Error* (MSE) dan *Mean Absolute Error* (MAE) untuk mencapai tingkat akurasi yang lebih tinggi. Gambaran arsitektur sistem yang merepresentasikan penelitian ini disajikan pada Gambar 1.



Gambar 1 Diagram Alir Penelitian

Gambar 1 memperlihatkan diagram alir metode penelitian berjudul "Prediksi Semester Ujian Tugas Akhir Mahasiswa Berdasarkan Transkrip Nilai Menggunakan *Linear Regression*, *Kernel Ridge Regression*, dan *Decision Tree Regression*." Langkah yang dilakukan dalam penelitian ini adalah merumuskan tujuan serta menetapkan hipotesis bahwa ketiga metode regresi tersebut—*Linear Regression*, *Kernel Ridge Regression*, dan *Decision Tree Regression*—dapat digunakan untuk memprediksi semester ujian tugas akhir mahasiswa berdasarkan data transkrip nilai mahasiswa.

B. Preprocessing

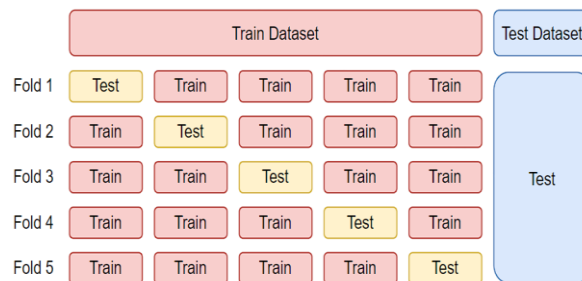
Sebelum pengujian dilakukan, data mentah disaring dengan menghapus faktor-faktor yang tidak relevan untuk algoritma *Linear Regression*, *Kernel Ridge Regression*, dan *Decision Tree Regression*, seperti NIM, Nama, Angkatan, Tanggal Masuk, dan Tanggal Sidang. Tabel 1 berikut menampilkan hasil proses seleksi data yang relevan, di mana status kelulusan diubah menjadi label dalam perangkat lunak RapidMiner untuk digunakan sebagai dasar dalam membangun model prediksi dengan *Linear Regression*, *Kernel Ridge Regression*, dan *Decision Tree Regression*.

Tabel 1 Data Nilai Mahasiswa

No	Smt	IPK	Yud	El	...	TA	SKS
0	9	3,64	2	12	...	24	469
1	11	3,32	2	6	...	24	422
2	9	3,36	2	12	...	24	434
3	9	3,45	2	9	...	24	439
4	9	3,46	2	12	...	24	445

Proses *preprocessing* pada Tabel 1 melibatkan pembulatan nilai dalam kolom "semester" dan mengubahnya menjadi tipe data integer. Selain itu, dilakukan pemetaan nilai yudisium dari bentuk teks ke bentuk numerik berdasarkan aturan yang telah ditentukan, yaitu "sangat memuaskan" dipetakan menjadi 2, "pujian" menjadi 3, dan "memuaskan" menjadi 1. Hasil pemetaan ini diterapkan pada kolom yudisium. Sementara itu, nilai Tugas Akhir dihitung berdasarkan nilai mahasiswa (A, B, atau C) yang kemudian dikalikan dengan bobot 6 SKS. Data yang telah diproses kemudian dipisahkan menjadi fitur (x) dan target (y), di mana kolom "semester" digunakan sebagai target, sedangkan kolom lainnya digunakan sebagai fitur.

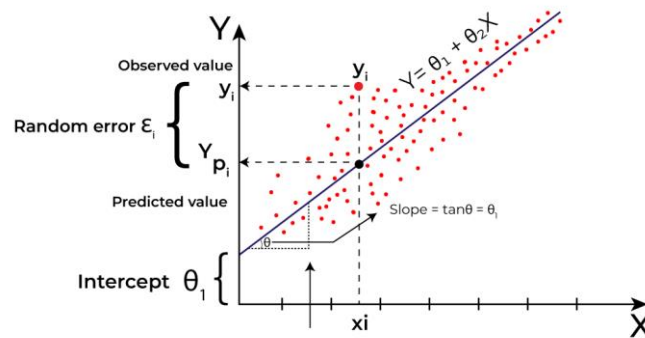
Dataset selanjutnya dibagi menjadi dua bagian, yaitu 80% untuk data pelatihan (train) dan 20% untuk data pengujian (*test*). Data pengujian diolah lebih lanjut menggunakan metode *K-Fold Cross Validation* (KCV), di mana dataset dibagi menjadi K subset yang digunakan secara bergantian untuk melatih model dan mengevaluasi performanya secara iteratif. Dalam penelitian ini, dataset dibagi menjadi lima subset untuk pelatihan (*train*) dan pengujian (*test*), seperti yang ditunjukkan pada Gambar 2.



Gambar 2 Dataset 5-Fold Cross Validation

C. Linear Regression

Regresi linier adalah metode statistik yang digunakan untuk memodelkan hubungan antara variabel dependen dan satu atau lebih variabel independen. Teknik ini memberikan wawasan yang mendalam dalam analisis dan prediksi data. Model regresi ini menghasilkan koefisien yang secara jelas menunjukkan pengaruh masing-masing variabel independen terhadap variabel dependen. Garis Regresi Terbaik menggambarkan hubungan linear antara variabel-variabel tersebut, di mana kemiringan garis mencerminkan besarnya perubahan pada variabel dependen untuk setiap perubahan satu unit pada variabel independen[11][12].



Gambar 3 *Linear Regression*

D. Kernel Ridge Regression

Kernel Ridge Regression (KRR) adalah metode yang efektif dalam *scikit-learn* untuk menyelesaikan masalah regresi, terutama saat menghadapi hubungan non-linear antara fitur dan variabel target. Teknik ini memungkinkan pemodelan hubungan non-linear yang kompleks, sehingga menjadi alat yang sangat berguna dalam analisis data. *Ridge Regression* (RR) sendiri adalah metode regresi sederhana namun kuat untuk prediksi, dan ketika kernel diterapkan pada RR (menjadi KRR), teknik ini mampu memetakan data input dalam ruang *non-linear* yang telah ditransformasikan, menjadikannya sangat cocok untuk analisis data deret waktu [13][14].

E. Decision Tree Regression

Decision Tree Regression merupakan algoritma yang banyak digunakan dalam pembelajaran mesin untuk tugas pemodelan prediktif. Ini merupakan alat yang ampuh yang dapat menangani masalah klasifikasi dan regresi, sehingga serbaguna untuk berbagai aplikasi. Namun, seperti algoritma lainnya, *Decision Tree Regression* memiliki kelebihan dan kekurangan. *Decision Tree Regression* adalah algoritma serbaguna yang dapat digunakan secara strategis dalam berbagai skenario tergantung pada sifat data dan persyaratan spesifik masalah yang dihadapi [15][16][17].

F. Metrik Evaluasi

Dalam penelitian ini, kinerja model regresi dievaluasi menggunakan dua metrik utama, yaitu *Mean Squared Error* (MSE) dan *Mean Absolute Error* (MAE). MSE mengukur rata-rata kuadrat selisih antara nilai aktual dan nilai prediksi, di mana nilai yang lebih kecil menunjukkan bahwa model memiliki akurasi yang lebih baik dalam melakukan estimasi. Sementara itu, MAE menghitung rata-rata selisih absolut antara nilai aktual dan prediksi, memberikan gambaran yang lebih intuitif mengenai tingkat kesalahan dalam satuan semester. Kedua metrik ini digunakan untuk menilai sejauh mana model dapat memprediksi semester penyelesaian tugas akhir mahasiswa dengan tingkat kesalahan yang minimal. Model dengan nilai MSE dan MAE yang lebih kecil dianggap lebih akurat dalam memprediksi semester penyelesaian tugas akhir mahasiswa. Perbandingan antara kedua metrik ini membantu dalam memahami sejauh mana tingkat kesalahan yang dihasilkan oleh masing-masing model serta karakteristik kesalahan prediksi yang terjadi. Selain itu, evaluasi menggunakan kedua metrik ini juga memberikan wawasan mengenai efektivitas pendekatan regresi yang digunakan dalam penelitian ini [18].

IV. HASIL DAN DISKUSI

A. Machine Learning Model

Model *Machine Learning* ini dikembangkan, dilatih, dan dievaluasi menggunakan *TensorFlow* untuk implementasi *Linear Regression*, *Kernel Ridge Regression*, dan *Decision Tree Regression*. Model didefinisikan dengan menggunakan Sequential API, seperti yang ditunjukkan pada Gambar 4, yang terdiri dari enam lapisan Dense. Lima lapisan pertama memiliki masing-masing 512 unit dengan fungsi aktivasi ReLU, sedangkan lapisan terakhir memiliki satu unit output. Model dikompilasi menggunakan optimizer Adam dengan *learning rate* sebesar 1×10^{-3} , menggunakan fungsi kerugian *Mean Squared*

Prediksi Semester Tugas Akhir Mahasiswa Berdasarkan Transkrip Nilai Menggunakan Linear Regression, Kernel Ridge Regression dan Decision Tree Regression
(Eki Ahmad Zaki Hamidi, Edi Mulyana, Dilla Restu Agusthiani, Aldi Fahrudi Muharam : Halaman 53 - 62)

Error (MSE) dan metrik evaluasi *Mean Absolute Error* (MAE). Proses pelatihan dilakukan pada data pelatihan selama 1000 epoch untuk setiap *fold* dalam validasi.

Layer (type)	Output Shape	Param #
dense_6 (Dense)	(None, 512)	32,768
dense_7 (Dense)	(None, 512)	262,656
dense_8 (Dense)	(None, 512)	262,656
dense_9 (Dense)	(None, 512)	262,656
dense_10 (Dense)	(None, 512)	262,656
dense_11 (Dense)	(None, 1)	513

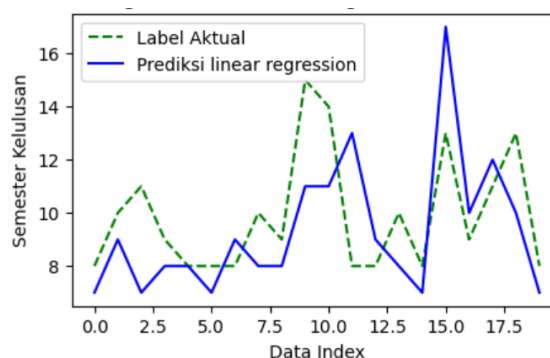
Total params: 3,251,717 (12.40 MB)
Trainable params: 1,083,905 (4.13 MB)
Non-trainable params: 0 (0.00 B)
Optimizer params: 2,167,812 (8.27 MB)

Gambar 4 Arsitektur model menggunakan Sequential API dengan beberapa lapisan Dense.

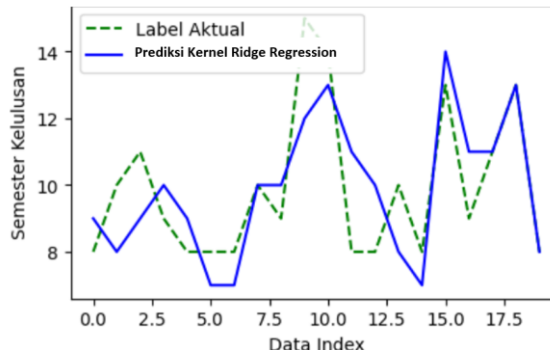
Evaluasi model dalam penelitian berjudul "Prediksi Semester Ujian Tugas Akhir Mahasiswa Berdasarkan Transkrip Nilai Menggunakan *Linear Regression*, *Kernel Ridge Regression*, dan *Decision Tree Regression*" merupakan tahap yang sangat penting. Tahap ini bertujuan untuk menilai seberapa baik model mampu memprediksi semester ujian tugas akhir berdasarkan data transkrip nilai yang telah melalui tahap *preprocessing*. Proses evaluasi dilakukan menggunakan dua metrik utama, yaitu *Mean Squared Error* (MSE) dan *Mean Absolute Error* (MAE), yang memberikan wawasan komprehensif terhadap kinerja model. *Mean Squared Error* (MSE) adalah metrik evaluasi yang menghitung rata-rata kuadrat selisih antara nilai prediksi dan nilai aktual. Dengan memberikan bobot lebih besar pada kesalahan besar, MSE sangat efektif dalam mengidentifikasi prediksi yang meleset jauh dari nilai sebenarnya. Nilai MSE yang lebih rendah menunjukkan bahwa model mampu menghasilkan prediksi yang lebih akurat secara konsisten, dengan sedikit kesalahan signifikan.

Sementara itu, *Mean Absolute Error* (MAE) adalah metrik yang menghitung rata-rata selisih absolut antara nilai prediksi dan nilai sebenarnya. Berbeda dengan MSE, MAE memberikan gambaran yang lebih langsung tentang rata-rata tingkat kesalahan prediksi tanpa memperbesar pengaruh dari kesalahan ekstrem. Hal ini menjadikan MAE sebagai metrik yang intuitif untuk memahami kesalahan prediksi dalam satuan yang sama dengan variabel target. Kombinasi antara MSE dan MAE memberikan pendekatan evaluasi yang menyeluruh. MSE membantu mendeteksi kesalahan besar yang mungkin terjadi, sementara MAE memberikan gambaran rata-rata kesalahan secara umum. Dengan menggunakan kedua metrik ini, penelitian dapat mengevaluasi kinerja model secara mendalam, memastikan bahwa model yang dihasilkan tidak hanya akurat tetapi juga dapat diandalkan dalam memprediksi semester ujian tugas akhir mahasiswa.

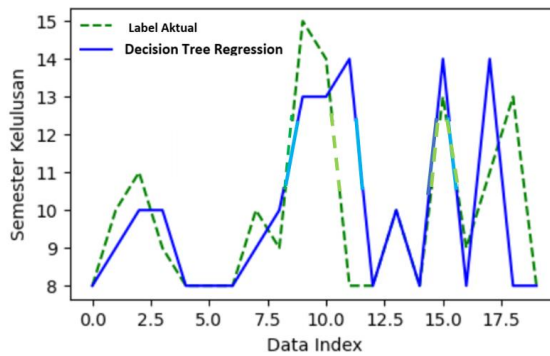
Dalam penelitian ini, perbandingan prediksi dengan label aktual digunakan untuk mengukur efektivitas kedua model tersebut dalam memprediksi semester ujian tugas akhir mahasiswa. Seperti pada gambar 5,6 dan 7 dibawah ini.



Gambar 5 Perbandingan Prediksi dan Label Data Aktual (*Linear Regression*)



Gambar 6 Perbandingan Prediksi dan Label Data Aktual (*Kernel Ridge Regression*)



Gambar 7 Perbandingan Prediksi dan Label Data Aktual (*Decision Tree Regression*)

Gambar 5,6, dan 7, menunjukkan perbandingan antara hasil prediksi model dengan label data aktual untuk tiga algoritma regresi yang digunakan dalam penelitian, yaitu *Linear Regression*, *Kernel Ridge Regression*, dan *Decision Tree Regression*. Perbandingan kinerja model, evaluasi dilakukan untuk mengukur efektivitas masing-masing algoritma regresi dalam memprediksi semester penyelesaian tugas akhir mahasiswa. Analisis dilakukan berdasarkan dua metrik utama, yaitu *Mean Squared Error* (MSE) dan *Mean Absolute Error* (MAE), yang digunakan untuk menilai tingkat kesalahan prediksi dari setiap model. Tabel 2 menyajikan ringkasan hasil evaluasi dari *Linear Regression*, *Kernel Ridge Regression*, dan *Decision Tree Regression*, termasuk nilai MSE, MAE, serta interpretasi terhadap performa masing-masing model.

Tabel 2 Perbandingan Kinerja Model dalam Prediksi Semester Penyelesaian Tugas Akhir

Model	Mean Squared Error (MSE)	Mean Absolute Error (MAE)	Kesimpulan
<i>Linear Regression</i>	5.5	1.9	Model dapat mengikuti tren data aktual, tetapi terdapat perbedaan signifikan di beberapa titik data, menunjukkan tingkat kesalahan yang relatif tinggi.
<i>Kernel Ridge Regression</i>	2.35	1.25	Model memiliki performa terbaik, mampu menangkap pola non-linear dengan lebih baik dan menghasilkan tingkat kesalahan yang lebih kecil.
<i>Decision Tree Regression</i>	4.1	1.2	Model cukup baik dalam menangkap pola data, namun fluktuasi tajam di beberapa titik masih menyebabkan

			tingkat kesalahan yang lebih tinggi dibandingkan Kernel Ridge Regression.
--	--	--	---

Berdasarkan hasil yang ditampilkan dalam Tabel 2, dapat disimpulkan bahwa *Kernel Ridge Regression* memiliki performa terbaik dibandingkan dua model lainnya, dengan nilai MSE sebesar 2.35 dan MAE sebesar 1.25. Model ini mampu menangkap pola data secara lebih akurat, termasuk hubungan non-linear yang kompleks, sehingga menghasilkan tingkat kesalahan yang lebih rendah.

Sementara itu, *Linear Regression* menunjukkan performa yang paling rendah dengan MSE sebesar 5.5 dan MAE sebesar 1.9, yang mengindikasikan bahwa model ini kurang mampu menangkap pola yang lebih kompleks dalam data transkrip nilai mahasiswa. *Decision Tree Regression* memiliki performa yang lebih baik dibandingkan *Linear Regression* dengan MSE sebesar 4.1 dan MAE sebesar 1.2, namun masih kurang optimal dibandingkan *Kernel Ridge Regression*. Dengan demikian, hasil ini menunjukkan bahwa penggunaan *Kernel Ridge Regression* lebih direkomendasikan dalam memprediksi semester penyelesaian tugas akhir mahasiswa karena memberikan estimasi yang lebih akurat dan stabil.

V. KESIMPULAN

Penelitian ini menunjukkan bahwa *Kernel Ridge Regression* adalah algoritma terbaik untuk memprediksi semester kelulusan mahasiswa berdasarkan data transkrip nilai. Dibandingkan dengan *Linear Regression* dan *Decision Tree Regression*, *Kernel Ridge Regression* memiliki performa yang lebih baik, dengan MSE 2.35 dan MAE 1.25, yang menunjukkan prediksi yang lebih akurat dan kesalahan yang lebih kecil. *Linear Regression*, meskipun mampu mengikuti tren data, memiliki kesalahan yang lebih tinggi (MSE 5.5, MAE 1.9), sehingga kurang efektif dalam memberikan prediksi yang akurat. Di sisi lain, *Decision Tree Regression*, meskipun lebih baik dari *Linear Regression* (MSE 4.1, MAE 1.2), tetap kalah dalam hal akurasi dibandingkan *Kernel Ridge Regression*.

Hasil penelitian ini menunjukkan bahwa untuk tugas prediksi semester kelulusan mahasiswa, *Kernel Ridge Regression* lebih dapat diandalkan dalam menangani hubungan non-linear antar data, memberikan kesalahan prediksi yang lebih kecil, dan lebih efektif dalam menganalisis pola yang kompleks. Penelitian selanjutnya dapat mempertimbangkan faktor-faktor non-akademik dan pendekatan feature engineering yang lebih mendalam untuk meningkatkan akurasi prediksi. Dengan pengembangan lebih lanjut, diharapkan model ini dapat diimplementasikan dalam sistem akademik universitas untuk membantu dosen pembimbing dan pihak akademik dalam memberikan rekomendasi yang lebih akurat terkait progres studi mahasiswa.

UCAPAN TERIMA KASIH

Terimakasih kepada LPPM UIN Sunan Gunung Djati Bandung yang sudah memfasilitasi penelitian ini.

DAFTAR PUSTAKA

- [1] T. Anagnostopoulos, D. Papakyriakopoulos, Y. Psaromiligkos, and S. Retalis, "Exploiting LSTM Neural Network Algorithm Potentiality for Early Identification of Delayed Graduation in Higher Education," *WSEAS Trans. Inf. Sci. Appl.*, vol. 21, pp. 524–532, 2024, doi: 10.37394/23209.2024.21.48.
- [2] P. Kalyane, J. Damania, H. Patil, M. Wardule, and P. Shahane, "Student's Performance Prediction Using Decision Tree Regressor," *Commun. Comput. Inf. Sci.*, vol. 2092 CCIS, pp. 286–302, 2024, doi: 10.1007/978-3-031-64070-4_18.
- [3] L. Delnoij *et al.*, "Predicting Completion: The Road to Informed Study Decisions in Higher Online Education," *Front. Educ.*, vol. 6, no. July, pp. 1–17, 2021, doi: 10.3389/educ.2021.668922.
- [4] R. Bakri, N. P. Astuti, and A. S. Ahmar, "Machine Learning Algorithms with Parameter Tuning to Predict Students' Graduation-on-time: A Case Study in Higher Education," *J. Appl. Sci. Eng. Technol. Educ.*, vol. 4, no. 2, pp. 259–265, 2022, doi: 10.35877/454ri.asci1581.
- [5] A. B. Hassanat *et al.*, "A Novel Outlier-Robust Accuracy Measure for Machine Learning Regression Using

- a Non-Convex Distance Metric,” *Mathematics*, vol. 12, no. 22, pp. 1–20, 2024, doi: 10.3390/math12223623.
- [6] S. M. Robeson and C. J. Willmott, “Decomposition of the mean absolute error (MAE) into systematic and unsystematic components,” *PLoS One*, vol. 18, no. 2 February, pp. 1–8, 2023, doi: 10.1371/journal.pone.0279774.
- [7] M. N. Faruqhy, D. Andreswari, and J. P. Sari, “Prediksi Prestasi Nilai Akademik Mahasiswa Berdasarkan Jalur Masuk Perguruan Tinggi Menggunakan Metode Multiple Linear Regression (Studi Kasus: Fakultas Teknik Universitas Bengkulu),” *Rekursif J. Inform.*, vol. 9, no. 2, pp. 172–183, 2021, doi: 10.33369/rekursif.v9i2.17108.
- [8] A. Qoiriah and Y. Yamasari, “Prediksi Nilai Akhir Mahasiswa Dengan Metode Regresi (Studi Kasus Mata Kuliah Pemrograman Dasar),” *J. Inf. Eng. Educ. Technol.*, vol. 5, no. 1, pp. 40–43, 2021, doi: 10.26740/jieet.v5n1.p40-43.
- [9] A. Thabibi and R. Supriyanto, “Perbandingan Model Multiple Linear Regression Dan Decision Tree Regression (Studi Kasus: Prediksi Harga Saham Telkom, Indosat, Dan Xl),” *J. Ilm. Teknol. dan Rekayasa*, vol. 28, no. 1, pp. 78–92, 2023, doi: 10.35760/tr.2023.v28i1.6081.
- [10] W. J. M. Putra, “Analisa Algoritma Regresi Linear dan Decision Tree Dalam Prediksi Penjualan Produk (Studi Kasus : Lookma Boutique),” *Skripsi*, p. 31, 2022, [Online]. Available: <https://lib.mercubuana.ac.id/>
- [11] N. Roustaei, “Application and interpretation of linear-regression analysis,” *Med. Hypothesis, Discov. Innov. Ophthalmol.*, vol. 13, no. 3, pp. 151–159, 2024, doi: 10.51329/mehdiophthal1506.
- [12] D. Mustofani, H. Hariyani, A. Afif, D. I. Oktaviasari, and B. Y. Ariadhita, “Analisis Data Hubungan Antar Variabel Pada Pengetahuan Swamedikasi,” *Unisda J. Math. Comput. Sci.*, vol. 10, no. 1, pp. 12–17, 2024, doi: 10.52166/ujmc.v10i1.6701.
- [13] A. A. Masrur Ahmed, E. Sharma, S. Janifer Jabin Jui, R. C. Deo, T. Nguyen-Huy, and M. Ali, “Kernel Ridge Regression Hybrid Method for Wheat Yield Prediction with Satellite-Derived Predictors,” *Remote Sens.*, vol. 14, no. 5, 2022, doi: 10.3390/rs14051136.
- [14] Putri R.A, Winahju W.S, and Mashuri Muhammad, “Penerapan Metode Ridge Regression dan Support Vector Regression (SVR) untuk Prediksi Indeks Batubara di PT XYZ,” *J. Sains Dan Seni Its*, vol. 9, no. 1, pp. 64–71, 2020.
- [15] B. B. Acharya and G. D. Shaileshbhai, “Comparative Analysis of Machine Learning Algorithms : KNN , SVM , Decision Tree and Logistic,” no. November, 2024.
- [16] I. Septian *et al.*, “Decision Tree Regression untuk Prediksi Prevalensi Stunting di Provinsi Nusa Tenggara Timur,” vol. 10, no. 2, pp. 413–427, 2024.
- [17] I. Azure, “Predictive modeling for industrial productivity: Evaluating linear regression and decision tree regressor approaches,” *J. Applied Math*, vol. 2, no. 4, p. 1435, 2024, doi: 10.59400/jam.v2i4.1435.
- [18] F. Rachmawati, J. Jaenudin, N. B. Ginting, and P. Laksono, “Machine Learning for the Model Prediction of Final Semester Assessment (FSA) using the Multiple Linear Regression Method,” *J. Tek. Inform.*, vol. 17, no. 1, pp. 1–9, 2024, doi: 10.15408/jti.v17i1.28652.