

Penerapan Regresi terhadap Faktor-faktor yang Mempengaruhi Peningkatan Skor PISA dalam Sistem Pendidikan Indonesia

Nadya Khairani
Fakultas Informatika
Telkom University
Bandung, Indonesia

nadyakhairani@student.telkomuniversity.ac.id

Nakhwa Azizah
Fakultas Informatika
Telkom University
Bandung, Indonesia

nakhwaazizah@student.telkomuniversity.ac.id

Najma Zahira Layalia Marsyafathanya
Fakultas Informatika
Telkom University
Bandung, Indonesia

najmazlm@student.telkomuniversity.ac.id

Abstrak—Pendidikan adalah dasar yang penting untuk pembangunan sosial, ekonomi, dan kultural suatu negara. Laporan ini membahas tantangan yang dihadapi Indonesia dalam meningkatkan kualitas pendidikan guna mencapai standar global, dengan fokus pada hasil Program for International Student Assessment (PISA). Berdasarkan data dari Kaggle, analisis regresi linier digunakan untuk memprediksi skor PISA dan mengidentifikasi faktor-faktor yang secara signifikan memengaruhi kualitas pendidikan. Hasil analisis menunjukkan bahwa GDP, tingkat literasi, angka kematian bayi, kepemilikan ponsel, dan populasi adalah faktor-faktor yang secara signifikan dapat memengaruhi skor PISA. Model prediksi menunjukkan nilai Mean Squared Error (MSE) sebesar 0.0704 dan Root Mean Squared Error (RMSE) sebesar 0.2653, menandakan ketepatan yang cukup baik dalam melakukan prediksi. Berdasarkan analisis ini, direkomendasikan untuk meningkatkan kompetensi guru, kurikulum, dan fasilitas pendidikan; mengatasi ketimpangan pendidikan antar daerah; memanfaatkan teknologi digital; serta meningkatkan program dan layanan kesehatan. Strategi ini diharapkan dapat meningkatkan skor PISA Indonesia dan memperkuat daya saing global di masa mendatang.

Kata kunci—PISA, pendidikan, Indonesia, regresi

I. LATAR BELAKANG

Pendidikan adalah landasan penting untuk perkembangan suatu negara secara sosial, ekonomi, dan kultural. Sebagai tulang punggung pembangunan, sistem pendidikan yang berkualitas bukan hanya menciptakan individu yang terampil secara akademis, tetapi juga membentuk warga negara yang terampil, kritis, dan berdaya saing. Namun, banyak negara, termasuk Indonesia, masih menghadapi tantangan dalam menyediakan pendidikan yang merata dan berkualitas. Pentingnya pendidikan dalam skala global semakin ditekankan oleh hasil studi internasional seperti Program for International Student Assessment (PISA), yang memberikan gambaran komprehensif tentang kualitas sistem pendidikan di berbagai negara [1].

Keikutsertaan Indonesia dalam PISA memungkinkan pemantauan kualitas Pendidikan dari waktu ke waktu dan membandingkannya dengan negara lain. Pada tahun 2018, Indonesia sebagai salah satu negara yang berpartisipasi dalam PISA, menemukan dirinya berada pada peringkat yang masih cukup rendah dibandingkan dengan negara-negara lain. Kemudian pada tahun 2022 Indonesia mengalami peningkatan dengan nilai skor literasi naik 5 sampai 6 posisi dibanding PISA 2018 [2]. Meskipun begitu, Indonesia masih harus terus meningkatkan kualitas pendidikan dan masih terdapat tantangan seperti hasil pada studi internasional TIMSS yang menunjukkan rendahnya

kemampuan literasi matematika siswa indonesia [3]. Maka dari itu, perubahan signifikan sangat diperlukan untuk meningkatkan kualitas pendidikan, khususnya agar Indonesia dapat bersaing lebih baik pada tahun 2025.

Salah satu masalah yang ada dalam sistem pendidikan Indonesia adalah pendekatan pembelajaran yang terlalu fokus pada pemberian soal dan jawaban, yang mengakibatkan siswa lebih didorong untuk menghafal atau mengingat teknik dan strategi cepat dalam menyelesaikan soal. Jika hal tersebut terus berlangsung bisa mengurangi pola pikir kritis pada siswa dalam menyelesaikan suatu masalah apapun dan juga berpengaruh pada hasil skor PISA yang didapatkan oleh Indonesia. Selain itu faktor kurangnya ekonomi di Indonesia juga berpengaruh terhadap kemajuan pendidikan. Pada faktanya, masih banyak pelajar yang orang tuanya hidup dibawah garis kemiskinan sehingga tidak dapat membayar untuk mengikuti bimbingan belajar [4]. Masih banyak lagi aspek lain yang mempengaruhi kualitas pendidikan di suatu negara, maka dari itu perlu adanya perhatian khusus dari pemerintah agar dapat memantau dan menganalisis aspek mana saja yang berpengaruh dalam pengurangan atau peningkatan skor PISA, hal tersebut dapat menggunakan salah satu metode analisis regresi dengan melibatkan beberapa variabel yang dimiliki dan membandingkannya dengan negara lain, sehingga pemerintah Indonesia dapat mudah menyoroti pada aspek-aspek yang perlu ditingkatkan di tahun tersebut dan tahun-tahun berikutnya.

II. STUDI LITERATUR

A. PISA Score

Program for International Student Assessment (PISA) diluncurkan oleh OECD pada tahun 1997, dan pertama kali diadakan pada tahun 2000 dengan

melibatkan lebih dari 80 negara untuk menguji kemampuan siswa yang berusia 15 tahun dalam membaca, matematika, dan sains setiap tiga tahun sekali [7].

B. Regresi Linear

Analisis regresi merupakan sebuah metode statistik yang digunakan untuk meneliti hubungan antara dua atau lebih variabel dalam suatu penelitian [5]. Fungsi utama metode regresi adalah untuk memprediksi dampak variabel yang menjadi prediktor terhadap variabel yang menjadi kriteria, atau untuk menguji apakah terdapat hubungan fungsional antara variabel independen (X) dan variabel dependen (Y) [6].

III. DATA DAN METODE PENELITIAN

A. Dataset

Pada proyek ini digunakan dua jenis dataset, diantaranya adalah *countries.csv* dan *pisa20162018.csv*. Kedua data tersebut diambil dari situs Kaggle. Data *countries.csv* merupakan data yang berisi *GDP*, *literacy rate*, *death rate*, *birth rate*, *population*, *region*, dan lain-lain. Dataset *pisa20162018.csv* berisi skor PISA, nama negara, peringkat negara, tahun pengujian, serta nama mata pelajaran.

B. Tahapan Pengerjaan

Dalam tahapan pengerjaan dapat menggambarkan bagaimana analisis dilakukan. Tahapan-tahapan tersebut antara lain:

1. Pengambilan Data

Pengambilan kedua data dilakukan pada situs Kaggle. Kemudian dilakukan join pada kedua dataset tersebut untuk menambah pengetahuan atau fitur pada Machine Learning agar model dapat berlatih dengan dataset yang disediakan. Setelah dilakukan join, terdapat total 1167 baris dan 24 kolom pada dataframe.

2. Pembersihan Data

Tahapan pembersihan data melibatkan beberapa langkah yaitu, mengubah nama kolom, menghilangkan spasi tambahan pada nama negara dan wilayah, mengubah format data dari string dengan koma menjadi float, dan mengisi nilai null pada kolom tertentu dengan rata-rata kolom.

3. Penggabungan Data

Penggabungan data ini dilakukan karena proyek ini memakai dua dataset yang berada pada file yang berbeda sehingga penggabungan data dilakukan untuk menyatukan kedua file tersebut agar dapat dilakukan analisis lebih lanjut.

4. Prediksi Skor PISA Siklus Ke-6

Untuk memprediksi skor PISA siklus ke-6, dilakukan model regresi linier menggunakan data yang telah dibersihkan dan digabungkan. Dari dataset, dipilih fitur populasi, GDP, dan tingkat literasi sebagai variabel independen. Selanjutnya melakukan standarisasi fitur menggunakan *StandardScaler* agar memiliki skala yang sama. Kemudian Regresi Linier digunakan untuk memprediksi skor PISA, dan kinerja model dievaluasi dengan *Mean Squared Error* (MSE) dan *Root Mean Squared Error* (RMSE). Hasil prediksi setiap negara dibandingkan dengan skor asli untuk menilai akurasi model.

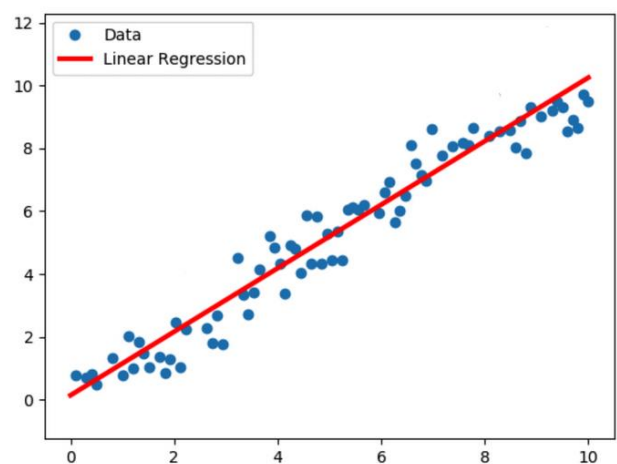
5. Prediksi Faktor yang Mempengaruhi Skor PISA

Untuk memprediksi faktor yang mempengaruhi skor PISA, dilakukan seleksi fitur menggunakan metode *SelectKBest* dengan fungsi *f_regression* untuk mengidentifikasi faktor yang paling signifikan. Analisis regresi linier dilakukan dengan fitur terpilih untuk memprediksi skor PISA dan kinerja model dievaluasi dengan menampilkan koefisien regresi.

Faktor-faktor signifikan dengan $p\text{-value} < 0.05$ diidentifikasi untuk memahami dampaknya terhadap skor PISA dan memberikan masukan bagi kebijakan pendidikan.

C. Regresi Linear

Metode Regresi Linear memungkinkan untuk menguji hipotesis apakah terdapat pengaruh antara satu variabel dengan variabel lainnya. Hasil dari analisis regresi diekspresikan dalam bentuk persamaan regresi, yang memungkinkan untuk memahami dan memprediksi hubungan antar variabel dengan lebih baik [5], definisi sederhananya adalah teknik dalam statistika untuk memprediksi hubungan antar variabel. Bentuk dari regresi juga dapat didefinisikan dengan sebuah grafik yang bernama scatter plot seperti gambar berikut:



Gambar 1. Regresi Linear

IV. HASIL DAN PEMBAHASAN

A. Hasil

Berikut ini akan dibahas mengenai hasil dari analisis data menggunakan metode regresi linear untuk memprediksi skor PISA dan faktor-faktor yang mempengaruhinya.

1. Evaluasi Model Prediksi Skor PISA

Setelah dilakukan pembersihan dan penggabungan data, serta pembuatan model

regresi linear, model diuji untuk memprediksi skor PISA pada siklus ke-6. Hasil evaluasi model menunjukkan bahwa *Mean Square Error* (MSE) bernilai 0.0704 dan *Root Mean Square Error* (RMSE) bernilai 0.2653. Nilai MSE dan RMSE ini menandakan bahwa model yang dibangun memiliki ketepatan yang cukup baik dalam memprediksi skor PISA.

Table 1. Hasil Prediksi Skor PISA Siklus Ke-6

Negara	Prediksi Skor PISA Siklus 6	Skor Asli PISA
Iceland	0.500138	0.114844
Portugal	0.373364	0.492681
United Kingdom	0.718057	0.492681
Austria	0.553996	0.592112

2. Faktor yang Mempengaruhi Skor PISA

Berdasarkan hasil analisis regresi, terdapat beberapa faktor signifikan yang mempengaruhi skor PISA, yaitu:

- **GDP (Gross Domestic Product):** GDP menunjukkan kemampuan ekonomi suatu negara. Negara dengan nilai GDP yang tinggi cenderung memiliki sumber daya yang lebih baik untuk mendukung pendidikan
- **Tingkat Literasi:** Tingkat literasi menggambarkan persentase penduduk yang mampu membaca dan menulis. Literasi yang tinggi memberikan hubungan yang positif dengan skor PISA yang lebih baik

- **Infant Mortality Rate (Angka Kematian Bayi):** Angka kematian bayi yang rendah mencerminkan sistem kesehatan yang baik dan berkontribusi pada kesejahteraan dan kemampuan belajar anak-anak.
- **Kepemilikan Ponsel:** Tingginya kepemilikan ponsel menunjukkan akses yang besar terhadap informasi dan teknologi digital, sehingga mendukung proses pembelajaran
- **Populasi:** Populasi suatu negara juga dapat mempengaruhi PISA, meskipun hubungan ini bisa kompleks dan dipengaruhi oleh berbagai faktor lainnya.

Table 2. Hasil Prediksi Faktor Berpengaruh

Fitur	P-value	Koefisien Regresi
infant_mortality	2.554770e-144	-0.150017
phones_per_1000	9.446353e-130	0.268638
GDP	7.994708e-101	0.080637
literacy	2.478612e-87	0.170365
pop_density	2.035530e-13	0.424787

B. Pembahasan

1. Hasil Model

Hasil model regresi linear dengan nilai MSE 0.0704 dan RMSE 0.2653 menunjukkan bahwa model prediksi skor PISA siklus ke-6 memiliki tingkat kesalahan

yang rendah, yang menandakan prediksi yang cukup akurat.

Untuk prediksi faktor-faktor yang memengaruhi, ketika hasil $p\text{-value} < 0,05$ maka faktor atau variabel tersebut berpengaruh secara signifikan. Koefisien regresi merupakan banyaknya skor yang berkurang maupun yang bertambah ketika terdapat penambahan satu unit pada setiap faktor. Misalkan, `pop_density` memiliki koefisien 0.424787. Hal ini berarti setiap peningkatan satu unit dalam kepadatan populasi terkait dengan peningkatan skor PISA sebesar 0.424787, dengan asumsi variabel lainnya tetap konstan. Faktor-faktor yang mempengaruhi skor PISA, seperti GDP, tingkat literasi, infant mortality rate, kepemilikan ponsel, dan populasi, semuanya memberikan kontribusi yang signifikan terhadap kualitas pendidikan.

2. Strategi Peningkatan Skor PISA

Berdasarkan hasil analisis, terdapat beberapa strategi yang bisa diterapkan untuk meningkatkan skor PISA di Indonesia:

- **Peningkatan Kualitas Pendidikan:** Meningkatkan kualitas pendidikan dengan meningkatkan kompetensi guru, kurikulum yang relevan, dan fasilitas pendidikan yang memadai.
- **Pemerataan Pendidikan:** Mengatasi ketimpangan pendidikan antar daerah dengan memberikan dukungan khusus pada daerah yang tertinggal.
- **Pemanfaatan Teknologi Digital:** Meningkatkan akses dan pemanfaatan teknologi digital dalam proses

pembelajaran untuk mendukung siswa dalam mengembangkan keterampilan.

- **Peningkatan Program dan Layanan Kesehatan:** Meningkatkan layanan kesehatan, terutama bagi anak-anak dan ibu hamil, agar dapat mendukung perkembangan anak yang sehat dan siap belajar.

V. KESIMPULAN

Analisis menggunakan data dari situs Kaggle yang mencakup variabel ekonomi dan pendidikan seperti GDP, tingkat literasi, dan skor PISA, serta metode regresi linier untuk menganalisis hubungan antara seluruh variabel dan memprediksi skor PISA. Model prediksi menunjukkan nilai Mean Squared Error (MSE) sebesar 0.0704 dan Root Mean Squared Error (RMSE) sebesar 0.2653, menunjukkan ketepatan model yang cukup baik. Faktor-faktor signifikan yang mempengaruhi skor PISA meliputi GDP, tingkat literasi, angka kematian bayi, kepemilikan ponsel, dan populasi.

Dalam meningkatkan skor PISA, direkomendasikan untuk meningkatkan kualitas pendidikan melalui peningkatan kompetensi guru, kurikulum yang relevan, dan fasilitas pendidikan yang memadai; pemerataan pendidikan dengan memberikan dukungan khusus pada daerah tertinggal; pemanfaatan teknologi digital dalam proses pembelajaran; serta peningkatan program dan layanan kesehatan untuk mendukung perkembangan anak yang sehat dan siap belajar. Dengan fokus pada faktor-faktor signifikan ini, diharapkan Indonesia dapat mencapai skor PISA yang lebih baik dan meningkatkan daya saing global di masa depan.

REFERENSI

- [1] E. Ismawati, H. Hersulastuti, I. P. Amertawengrum, and K. A. Anindita, "Portrait of Education in Indonesia: Learning from PISA Results 2015 to Present," *International Journal of Learning, Teaching and Educational Research*, vol. 22, no. 1, pp. 321–340, Jan. 2023, doi: <https://doi.org/10.26803/ijlter.22.1.18>.
- [2] Kemdikbud, "Peringkat Indonesia pada PISA 2022 Naik 5-6 Posisi Dibanding 2018," *Kementerian Pendidikan, Kebudayaan, Riset, dan Teknologi*, Dec. 05, 2023. <https://www.kemdikbud.go.id/main/blog/2023/12/peringkat-indonesia-pada-pisa-2022-naik-56-posisi-dibanding-2018>
- [3] Eivers, C. (2010). PISA Untuk Siswa Indonesia - Jurnal Universitas Ngudi Waluyo. Retrieved December 16, 2023, from <https://jurnal.unw.ac.id/index.php/janacitta/article/download/13/182>
- [4] H. L. T. B. Ziyad, "PISA Dan Sistem Pendidikan Indonesia. Tahukah Kamu Apa Itu PISA ?," *Thariq Bin Ziyad*, Oct. 25, 2022. <https://thariq.sch.id/pisa-dan-sistem-pendidikan-indonesia-tahukah-kamu-apa-itu-pisa/> (accessed Mar. 12, 2024).
- [5] R. Hans, "Teknik Analisis Data : Pengertian dan Manfaat Analisis Regresi," *dqlab.id*, Jan. 06, 2022. <https://dqlab.id/teknik-analisis-data-pengertian-dan-manfaat-analisis-regresi>
- [6] YQ, "ANALISIS REGRESI." Available: <https://staffnew.uny.ac.id/upload/198401312014042002/pendidikan/ANALISIS%20REGRESI-YQ.pdf>
- [7] Kemendikbud, "Programme for International Student Assessment (PISA)," 12 June 2024. [Online]. Available: <https://bskap.kemdikbud.go.id/pisa>.