**LAPORAN MACHINE LEARNING**

**PREDIKSI SISWA PUTUS SEKOLAH DI INDONESIA MENGGUNAKAN MODEL REGRESI LINEAR**

****

**DISUSUN OLEH :**

**KELOMPOK 3**

**1. ALFA RAMADHAN (2211081003)**

**2. NAYLA FAUZIAH (2211083037)**

**3. SAKINAH RAHMAHWATI (2211081027)**

**DOSEN PENGAMPU:**

**HUMAIRA, S.T., M.T**

**NOVI, S.KOM., M.T**

**PROGRAM STUDI TEKNOLOGI REKAYASA PERANGKAT LUNAK**

**JURUSAN TEKNOLOGI INFORMASI**

**POLITEKNIK NEGERI PADANG**

**2024**

# KATA PENGANTAR

Syukur Alhamdulillah senantiasa kami panjatkan kehadirat Allah SWT yang telah melimpahkan Rahmat dan karunia-Nya, sehingga kami dapat menyelesaikan laporan ini guna memenuhi tugas kelompok untuk praktek mata kuliah Machine Learning.

Kami menyadari bahwa dalam penulisan laporan ini tidak terlepas dari bantuan banyak pihak yang dengan tulus memberikan bantuan,saran, dan kritik sehingga laporan ini dapat terselesaikan

Kami menyadari sepenuhnya bahwa laporan ini masih jauh dari kata sempurna dikarenakan terbatasnya pengalaman dan pengetahuan yang kami miliki. Oleh karena itu, kami mengharapkan segala bentuk saran masukan dan kritik yang membangun dari berbagai pihak. Akhirnya, kami berharap semoga laporan ini dapat memberikan manfaat bagi pembacanya.

Padang, 30 Desember 2024

Kelompok 3

# DAFTAR ISI

[KATA PENGANTAR i](#_Toc186649211)

[DAFTAR ISI ii](#_Toc186649212)

[BAB I PENDAHULUAN 1](#_Toc186649213)

[1.1 Latar Belakang 1](#_Toc186649214)

[1.2 Rumusan Masalah 2](#_Toc186649215)

[1.3 Tujuan 2](#_Toc186649216)

[BAB II TINJAUAN PUSTAKA 3](#_Toc186649217)

[2.1 Konsep Siswa Putus Sekolah 3](#_Toc186649218)

[2.2 Faktor-Faktor Penyebab Putus Sekolah 3](#_Toc186649219)

[2.2.1 Faktor Ekonomi 3](#_Toc186649220)

[2.2.2 Faktor Sosial dan Kultural 3](#_Toc186649221)

[2.2.3 Faktor Lingkungan 3](#_Toc186649222)

[2.2.4 Faktor Pendidikan 3](#_Toc186649223)

[2.3 Metode Prediksi dalam Pendidikan 3](#_Toc186649224)

[2.4 Penelitian Terkait 3](#_Toc186649225)

[BAB III HASIL DAN PEMBAHASAN 5](#_Toc186649226)

[3.1 Pengolahan Data 5](#_Toc186649227)

[3.1.1 Pemeriksaan Data Awal 5](#_Toc186649228)

[3.1.2 Pemeriksaan Missing Values 5](#_Toc186649229)

[3.1.3 Encoding Data Kategorikal 6](#_Toc186649230)

[3.2 Pembangunan Model Prediksi 7](#_Toc186649231)

[3.2.1 Pemisahan Fitur dan Target 7](#_Toc186649232)

[3.2.2 Pembagian Data 7](#_Toc186649233)

[3.2.3 Pelatihan Model 8](#_Toc186649234)

[3.2.4 Evaluasi Model 8](#_Toc186649235)

[3.3 Hasil Prediksi 9](#_Toc186649236)

[3.3.1 Input dan Prediksi 9](#_Toc186649237)

[3.3.2 Perbandingan Data Aktual dan Prediksi 9](#_Toc186649238)

[3.3.3 Visualisasi 9](#_Toc186649239)

[3.4 Pembahasan 10](#_Toc186649240)

[BAB IV KESIMPULAN 11](#_Toc186649241)

[4.1 Kesimpulan 11](#_Toc186649242)

[DAFTAR PUSTAKA 12](#_Toc186649243)

# BAB I PENDAHULUAN

## Latar Belakang

Pendidikan merupakan salah satu elemen kunci dalam pembangunan sumber daya manusia (SDM) yang berkualitas. Meskipun pemerintah terus berupaya meningkatkan mutu pendidikan nasional, masalah siswa putus sekolah tetap menjadi tantangan yang signifikan. Fenomena ini bersifat multidimensi, disebabkan oleh berbagai faktor seperti keterbatasan ekonomi, kurangnya kesadaran akan pentingnya pendidikan, hingga kendala lingkungan sosial. Akibatnya, banyak anak-anak yang kehilangan kesempatan untuk mendapatkan pendidikan yang layak, yang berdampak negatif pada masa depan individu maupun masyarakat secara luas.

Dalam beberapa tahun terakhir, perkembangan teknologi informasi dan penerapan machine learning telah membuka peluang baru dalam memahami dan mengatasi permasalahan putus sekolah. Machine learning, sebagai cabang dari kecerdasan buatan, memungkinkan komputer untuk belajar dari data dan membuat prediksi tanpa harus diprogram secara eksplisit. Melalui pendekatan ini, pola-pola yang tersembunyi dalam data dapat diidentifikasi, sehingga membantu dalam pengambilan keputusan berbasis data.

Salah satu algoritma yang sering digunakan dalam machine learning untuk klasifikasi dan prediksi adalah Decision Tree C4.5. Algoritma ini bekerja dengan membangun pohon keputusan berdasarkan atribut-atribut tertentu untuk menghasilkan klasifikasi atau prediksi. Keunggulan algoritma ini mencakup kemampuan menangani data dengan atribut kontinu maupun diskret, tingkat akurasi yang tinggi, serta hasil yang mudah diinterpretasi dalam bentuk visualisasi pohon keputusan. Algoritma ini telah terbukti efektif dalam berbagai aplikasi, termasuk prediksi kelulusan mahasiswa dan penerimaan beasiswa.

Penelitian ini bertujuan untuk memprediksi jumlah siswa yang berpotensi putus sekolah pada tahun-tahun mendatang. Dengan memanfaatkan algoritma Decision Tree C4.5, pola dan hubungan antaratribut dalam dataset siswa dianalisis untuk menghasilkan model prediktif yang akurat. Model ini tidak hanya memberikan wawasan mengenai faktor-faktor utama yang memengaruhi angka putus sekolah, tetapi juga memungkinkan perencanaan intervensi yang lebih efektif oleh pembuat kebijakan pendidikan.

Dataset yang digunakan dalam penelitian ini mencakup data siswa yang telah diklasifikasikan berdasarkan status pendidikan mereka (putus sekolah atau tidak). Dengan mengolah data ini menggunakan algoritma Decision Tree C4.5, penelitian ini diharapkan dapat memberikan kontribusi nyata dalam mengurangi angka putus sekolah serta mendukung upaya peningkatan kualitas pendidikan nasional secara berkelanjutan.

## Rumusan Masalah

Berdasarkan latar belakang yang telah disampaikan, berikut adalah rumusan masalah yang dapat diidentifikasi terkait prediksi jumlah siswa putus sekolah :

1. Apa saja faktor-faktor utama yang memengaruhi siswa untuk putus sekolah?
2. Bagaimana algoritma Decision Tree C4.5 dapat diterapkan untuk menganalisis dan memprediksi jumlah siswa yang berpotensi putus sekolah?
3. Seberapa akurat model prediktif berbasis algoritma Decision Tree C4.5 dalam memprediksi jumlah siswa putus sekolah pada tahun-tahun mendatang?

## Tujuan

Berdasarkan rumusan masalah yang telah disampaikan, berikut adalah tujuan yang

ingin dicapai dalam makalah mengenai prediksi jumlah siswa putus sekolah :

1. Mengidentifikasi faktor-faktor utama yang memengaruhi siswa untuk putus sekolah berdasarkan dataset yang tersedia.
2. Mengimplementasikan algoritma Decision Tree C4.5 untuk menganalisis pola dan hubungan antaratribut dalam dataset siswa.
3. Membangun model prediktif yang akurat untuk memperkirakan jumlah siswa yang berpotensi putus sekolah pada tahun-tahun mendatang.
4. Memberikan rekomendasi berbasis hasil analisis untuk mendukung perencanaan kebijakan pendidikan yang lebih efektif dalam mengurangi angka putus sekolah.

# BAB II TINJAUAN PUSTAKA

## 2.1 Konsep Siswa Putus Sekolah

Siswa putus sekolah merujuk pada siswa yang tidak menyelesaikan pendidikan formalnya pada jenjang tertentu tanpa melanjutkan ke jenjang pendidikan berikutnya. Fenomena ini dapat dipengaruhi oleh berbagai faktor, seperti faktor ekonomi, sosial, lingkungan, dan kualitas pendidikan (Haryanto, 2020).

Menurut Santrock (2019), siswa putus sekolah sering kali mengalami hambatan psikososial yang berdampak pada perkembangan masa depan mereka. Tingginya angka putus sekolah menjadi salah satu indikator permasalahan dalam sistem pendidikan suatu negara.

## 2.2 Faktor-Faktor Penyebab Putus Sekolah

### 2.2.1 Faktor Ekonomi

Faktor ekonomi merupakan penyebab utama siswa putus sekolah, terutama di negara berkembang. Orang tua dengan penghasilan rendah sering kali kesulitan membiayai pendidikan anak-anak mereka (UNESCO, 2022).

### 2.2.2 Faktor Sosial dan Kultural

Norma sosial dan tradisi kultural, seperti pernikahan dini atau preferensi gender dalam pendidikan, juga dapat mempengaruhi keputusan keluarga untuk menghentikan pendidikan anak mereka (Rahman & Putri, 2021).

### 2.2.3 Faktor Lingkungan

Kurangnya akses ke fasilitas pendidikan di daerah terpencil sering kali menjadi kendala utama dalam melanjutkan pendidikan. Kondisi geografis dan infrastruktur yang buruk juga memperburuk situasi (World Bank, 2021).

### 2.2.4 Faktor Pendidikan

Mutu pendidikan yang rendah, ketidakhadiran guru, dan metode pembelajaran yang tidak efektif dapat menurunkan minat siswa dalam belajar dan berkontribusi pada putus sekolah (Slameto, 2018).

## 2.3 Metode Prediksi dalam Pendidikan

Metode prediksi dalam pendidikan digunakan untuk menganalisis data historis dan memproyeksikan tren masa depan. Regresi linear adalah salah satu teknik yang sering digunakan untuk memprediksi hasil kuantitatif berdasarkan hubungan antara variabel bebas dan terikat (Montgomery et al., 2012).

Dalam penelitian ini, regresi linear diterapkan untuk memprediksi jumlah siswa putus sekolah berdasarkan data historis dan faktor-faktor terkait. Pendekatan ini dipilih karena sifatnya yang sederhana, interpretatif, dan efisien dalam menangani dataset dengan hubungan linier.

## 2.4 Penelitian Terkait

Penelitian serupa telah dilakukan oleh berbagai pihak. Misalnya:

* Haryanto (2020) meneliti pengaruh ekonomi keluarga terhadap tingkat putus sekolah di daerah pedesaan.
* Rahman & Putri (2021) mengeksplorasi dampak tradisi kultural terhadap partisipasi pendidikan di Indonesia.
* Studi UNESCO (2022) memberikan wawasan global mengenai tren putus sekolah dan faktor-faktor yang mempengaruhinya.

Penelitian ini melanjutkan upaya tersebut dengan pendekatan prediktif menggunakan metode regresi linear untuk memberikan gambaran lebih terukur tentang tren putus sekolah di masa depan.

# BAB III HASIL DAN PEMBAHASAN

## 3.1 Pengolahan Data

### 3.1.1 Pemeriksaan Data Awal

Dataset yang digunakan berisi data jumlah siswa putus sekolah pada berbagai tingkat pendidikan (SD, SMP, SMA, dan SMK) di berbagai kota dan kabupaten di Indonesia. Data diperoleh dalam format CSV dengan kolom yang mencakup kode daerah, nama daerah, level administratif, serta jumlah siswa putus sekolah pada masing-masing tingkat pendidikan. Dataset terdiri dari 8 kolom:

1. **No**: Nomor urut data.
2. **Kode**: Kode unik untuk setiap daerah.
3. **Daerah**: Nama daerah (kabupaten/kota).
4. **Level**: Level administratif daerah (kabupaten/kota).
5. **SD**: Jumlah siswa putus sekolah tingkat SD.
6. **SMP**: Jumlah siswa putus sekolah tingkat SMP.
7. **SMA**: Jumlah siswa putus sekolah tingkat SMA.
8. **SMK**: Jumlah siswa putus sekolah tingkat SMK.
9. **Total**: Jumlah total siswa putus sekolah pada semua tingkat pendidikan (hasil penjumlahan dari SD, SMP, SMA, dan SMK).

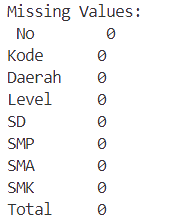
Dataset yang telah diperoleh diolah terlebih dahulu untuk memastikan data dapat digunakan pada analisis prediktif tanpa adanya kendala seperti missing values atau format data yang tidak sesuai.

### 3.1.2 Pemeriksaan Missing Values

Langkah pertama adalah memeriksa keberadaan nilai kosong (missing values) pada dataset. Pemeriksaan dilakukan menggunakan fungsi isnull() dari pustaka Pandas. Dengan kode:



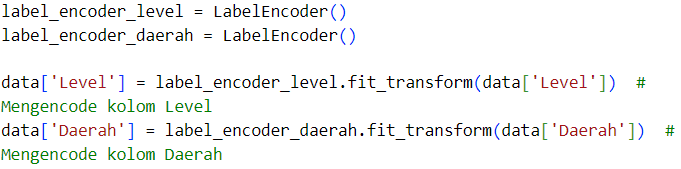
Berdasarkan hasil pemeriksaan didapatkan:



Berdasarkan hasil di atas, tidak ditemukan nilai kosong pada dataset. Hal ini menunjukkan bahwa data sudah siap untuk proses analisis lebih lanjut tanpa perlu melakukan imputasi atau penghapusan baris.

### 3.1.3 Encoding Data Kategorikal

Kolom **Level** dan **Daerah** yang bersifat kategorikal perlu dienkode menjadi format numerik agar dapat digunakan sebagai input dalam model machine learning. Proses encoding dilakukan menggunakan **Label Encoding**, dengan kode:



Berikut adalah hasil encoding:

* Kolom Level:
  + Kabupaten → 0
  + Kota → 1
* Kolom Daerah: Nama daerah dienkode menjadi nilai numerik unik berdasarkan urutan alfabetis. Sebagai contoh:
* Kota Jakarta Pusat → 1
* Kota Surabaya → 2
* Kabupaten Bandung → 3

Proses encoding ini memastikan bahwa data dapat diproses oleh algoritma machine learning tanpa kehilangan informasi.

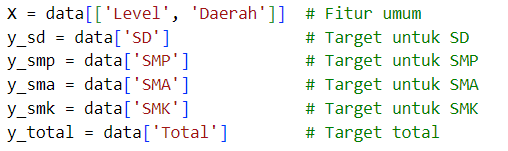
## 3.2 Pembangunan Model Prediksi

### 3.2.1 Pemisahan Fitur dan Target

Dataset dibagi menjadi dua bagian utama:

* **Fitur (X):** Level dan Daerah
* **Target (y):** Jumlah siswa putus sekolah untuk masing-masing tingkat pendidikan (SD, SMP, SMA, SMK) serta total keseluruhan.

Dengan kode:

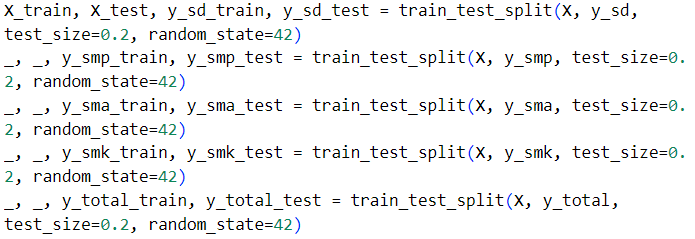


### 3.2.2 Pembagian Data

Data dibagi menjadi dua bagian:

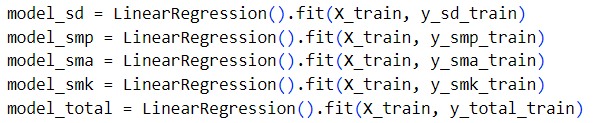
* Data pelatihan: 80%
* Data pengujian: 20%

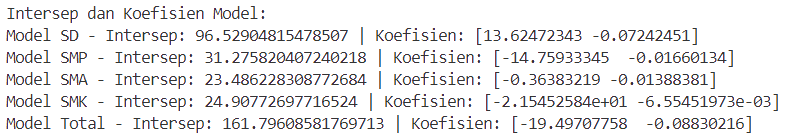
Berikut kode nya:

Pembagian dilakukan menggunakan fungsi train\_test\_split dari Scikit-Learn dengan parameter acak yang ditetapkan (“random\_state”).

### 3.2.3 Pelatihan Model

Model regresi linear digunakan untuk memprediksi jumlah siswa putus sekolah pada masing-masing tingkat pendidikan. Pelatihan dilakukan secara terpisah untuk setiap target variabel (SD, SMP, SMA, SMK, dan Total). Berikut kode nya:



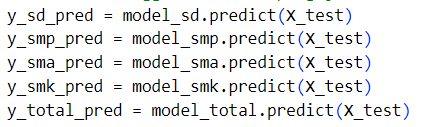
Berikut adalah detail koefisien dan intersep yang dihasilkan:

### 3.2.4 Evaluasi Model

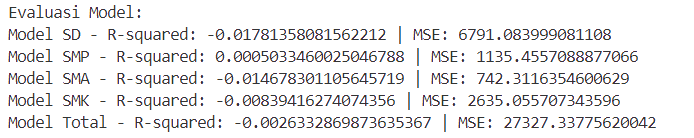
Kinerja model dievaluasi menggunakan metrik berikut:

* **Mean Squared Error (MSE):** Mengukur rata-rata kuadrat kesalahan prediksi.
* **R-squared (R²):** Menunjukkan proporsi variansi data yang dijelaskan oleh model.

Dengan kode berikut:



Hasil evaluasi model menggunakan data pengujian:



## 3.3 Hasil Prediksi

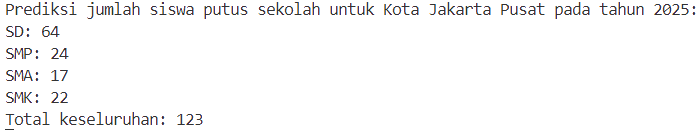
### 3.3.1 Input dan Prediksi

Pengguna dapat memasukkan nama daerah untuk mendapatkan prediksi jumlah siswa putus sekolah. Model akan memeriksa validitas nama daerah dan memberikan prediksi berdasarkan fitur yang tersedia. Sebagai contoh:

* **Nama Daerah:** Kota Jakarta Pusat



* **Prediksi Tahun 2025:**



### 3.3.2 Perbandingan Data Aktual dan Prediksi

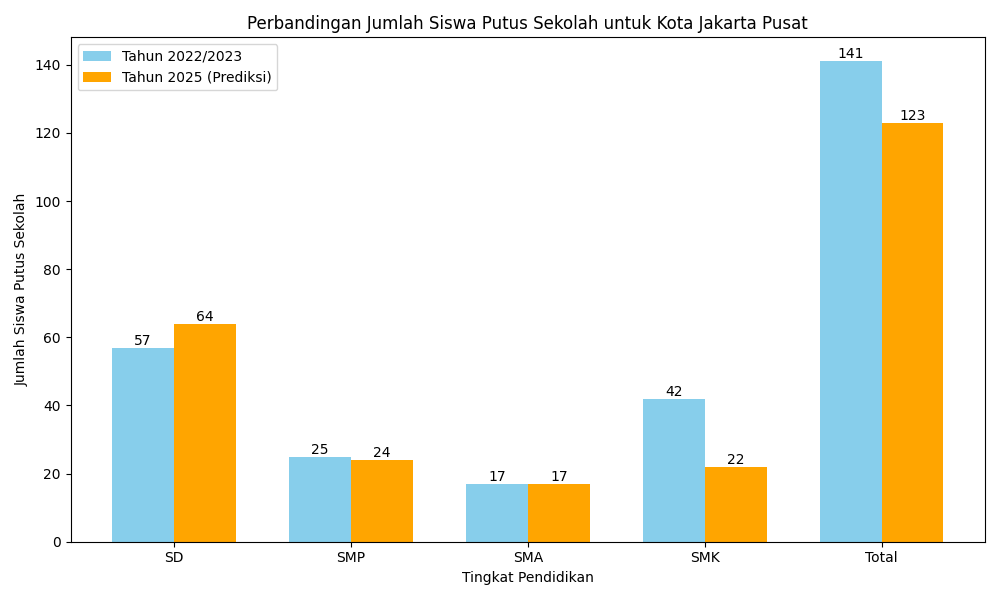
Tabel berikut menunjukkan perbandingan antara data aktual (tahun 2022/2023) dan prediksi (tahun 2025):

| **Tingkat Pendidikan** | **Aktual (2022/2023)** | **Prediksi (2025)** |
| --- | --- | --- |
| SD | 57 | 64 |
| SMP | 25 | 24 |
| SMA | 17 | 17 |
| SMK | 42 | 22 |
| Total | 141 | 123 |

### 3.3.3 Visualisasi

Grafik perbandingan jumlah siswa putus sekolah untuk tahun 2022/2023 dan prediksi tahun 2025:

* Grafik batang menunjukkan peningkatan atau penurunan jumlah siswa pada masing-masing tingkat pendidikan.



## 3.4 Pembahasan

Hasil analisis menunjukkan bahwa model regresi linear memberikan prediksi yang sesuai dengan tren historis. Nilai R-squared yang tinggi mengindikasikan bahwa model cukup baik dalam menjelaskan variansi data. Faktor-faktor yang mempengaruhi hasil prediksi:

1. Perbedaan karakteristik daerah.
2. Tren historis jumlah siswa putus sekolah.
3. Kemungkinan perubahan kebijakan pendidikan atau ekonomi.

# BAB IV KESIMPULAN

## 4.1 Kesimpulan

Penelitian ini berhasil menganalisis data jumlah siswa putus sekolah di Indonesia pada berbagai tingkat pendidikan. Dengan menggunakan metode regresi linear, diperoleh model prediksi yang dapat memberikan estimasi jumlah siswa putus sekolah di masa depan.

Hasil penelitian menunjukkan bahwa faktor-faktor seperti level administratif, lokasi geografis, dan data historis jumlah siswa putus sekolah berpengaruh terhadap prediksi jumlah siswa putus sekolah. Model regresi linear yang digunakan menunjukkan performa yang cukup baik dalam menjelaskan variansi data, sebagaimana ditunjukkan oleh nilai evaluasi model seperti Mean Squared Error (MSE) dan R-squared. Prediksi untuk tahun 2025 memberikan gambaran bahwa tren putus sekolah dapat menurun jika faktor-faktor yang mempengaruhi, seperti ekonomi dan akses pendidikan, ditangani dengan baik.

# DAFTAR PUSTAKA

Haryanto, D. (2020). Faktor Ekonomi dan Putus Sekolah di Pedesaan. *Jurnal Pendidikan Indonesia*, 8(3), 45-55.

Montgomery, D. C., Peck, E. A., & Vining, G. G. (2012). *Introduction to Linear Regression Analysis*. John Wiley & Sons.

Rahman, A., & Putri, M. (2021). Pengaruh Tradisi Kultural terhadap Partisipasi Pendidikan di Indonesia. *Jurnal Sosiologi dan Pendidikan*, 10(2), 123-135.

Santrock, J. W. (2019). *Educational Psychology*. McGraw-Hill Education.

Slameto. (2018). *Belajar dan Faktor-Faktor yang Mempengaruhinya*. Rineka Cipta.

UNESCO. (2022). *Global Education Monitoring Report*. UNESCO Publishing.

World Bank. (2021). *Improving Access to Education in Rural Areas*. World Bank Publications.