

# EDUCATION AND LIBRARY JOURNAL

## Analisis Prediktif Tren Pendidikan di Indonesia Menggunakan KNN: Studi Kasus Data Pendidikan 2021-2023

Mukhtada Billah Nasution<sup>1</sup>, Akhiyar Waladi<sup>2</sup>, Ulfa Khaira<sup>3</sup>, Pradita Eko Prasetyo Utomo<sup>4</sup>

<sup>1</sup> Sistem Informasi, Fakultas Sains dan Teknologi, Universitas Jambi, Indonesia

Corresponding author email : mukhtadanasution@gmail.com<sup>1</sup>

### ARTICLE INFO

#### Article history:

Received : 21 Oct 2024

Revised : 2 Dec 2024

Accepted : 31 Dec 2024

#### Keywords :

Indonesian education,  
education trends, k-nearest  
neighbor (knn) algorithm,  
education digitalization,  
education data analysis

### ABSTRACT

*This research focuses on the importance of education in improving the competitiveness of the younger generation in Indonesia, especially in facing the challenges of globalization and the digital revolution. Education trends in Indonesia during the 2021-2023 period have been dominated by two main factors, namely digitalization and equal access to education. A data-driven approach is used to predict education trends in 2024, using the K-Nearest Neighbor (KNN) algorithm to analyze data from the Central Statistics Agency (BPS) regarding the percentage of the population aged 25 years and over who have at least a high school education, categorized by gender. The result of this research will predict the trend of education in each region in 2024 whether it is decreasing, stable, or increasing. Through data collection and literature study, this research identifies relevant patterns and presents statistically-based predictions that can serve as a reference for stakeholders in the development of education in Indonesia. The results of this study are also expected to provide insights for policymakers in formulating effective strategies to address the education gap and promote inclusive digitalization..*



This work is licensed under a

## A. INTRODUCTION

Menurut Sukmayadi (2020), siap atau tidaknya seseorang di era digital bukanlah pilihan, melainkan konsekuensi progres. Menjadikan literasi berkualitas dan pendidikan tentang media dan digital sebagai bagian dari kurikulum merupakan solusi untuk mempersiapkan sistem pendidikan nasional yang mampu menghadapi tantangan era digital ke depannya. Tantangan sistem pendidikan tidak hanya menjadi tanggung jawab pemerintah saja, melainkan kolaborasi antara para sarjana, masyarakat, institusi pendidikan tinggi, dan keterlibatan publik dapat mengembangkan sistem pendidikan nasional yang konstruktif. Faisal (2019), dalam studinya menyebutkan bahwa terdapat kesetaraan dalam pendaftaran untuk sekolah dasar dan menengah pertama untuk semua kalangan, akan tetapi perbedaan untuk pendidikan menengah ke atas dan universitas. Partisipasi pada pendidikan menengah ke atas di wilayah perkotaan mencapai 88% sedangkan di wilayah pedesaan hanya 77%. Ini berarti lebih sedikit siswa yang melanjutkan ke sekolah menengah ke atas dibandingkan di perkotaan. Hal ini menjadi lebih parah pada perguruan tinggi, di mana tingkat partisipasi hanya mencapai 37% di wilayah perkotaan dan 17% di wilayah pedesaan, angka yang sangat kecil. Dalam studi yang dilakukan oleh Sukasni dan Efendy (2017), terdapat beberapa faktor yang menyebabkan kualitas pendidikan Indonesia yang tidak seberkualitas pendidikan negara maju, yakni kurikulum yang tidak relevan dengan tuntutan masyarakat, biayanya yang mahal, kontroversi dalam implementasi ujian nasional, dan banyak fasilitas pendidikan yang tidak layak.

Pramana dkk (2021), menyebutkan bahwa pendidikan yang bagus dan berkualitas merupakan pendidikan yang menghasilkan sumber daya manusia yang mampu bersaing di tingkat internasional. Ia juga menambahkan, pendidikan dianggap sebagai fondasi utama dalam pembentukan sumber daya manusia yang dapat memperkuat ketahanan bangsa dan negara. Namun, berkebalikan dengan hal tersebut, kualitas pendidikan di Indonesia masih tertinggal jika dibandingkan dengan neegara-negara maju. Pendidikan sangat penting bagi generasi muda di era globalisasi ini. Mutu Pendidikan seseorang sangat mempengaruhi kemampuan dan daya saing mereka baik itu di ranah lokal, nasional, maupun global. Pendidikan penting bagi generasi muda untuk bangkit dan meningkatkan potensi belajar guna mencapai mutu diri yang diinginkan. Pendidikan sendiri merupakan upaya untuk bangkit menuju pemerataan pendidikan untuk mencapai standarisasi pendidikan yang dapat bersaing di ranah global. Memiliki mutu pendidikan yang baik dan mampu bersaing di ranah global tentu saja menjadi nilai tambah suatu negara. Dengan pendidikan yang bermutu akan menghasilkan generasi-generasi yang kreatif dan inovatif. Pendidikan juga merupakan proses yang panjang demi mengubah pola pikir dan perilaku masyarakat ke arah yang lebih. Menurut Nugroho dkk (2021), salah satu upaya peningkatan mutu pendidikan adalah dengan mengimplementasikan konsep STEM (Science, Technology, Engineering, and Mathematics) ke dalam kurikulum, salah satunya kurikulum 2013. Dengan aplikasi STEM pada kurikulum dapat meningkatkan kemampuan berpikir kritis. Pendidikan merupakan berbagai latihan yang dilakukan untuk menghasilkan generasi yang berbudaya dan bermoral. Pendidikan yang berorientasi pada kualitas dan kompetensi global ini juga memegang peranan kunci dalam mengangkat citra dan daya saing bangsa di kancah internasional. Oleh sebab itu, pengembangan dan inovasi dalam dunia pendidikan Indonesia menjadi fokus utama bagi pemerintah dan berbagai pemangku kepentingan untuk mencapai standar mutu yang dapat bersaing secara global. Selain memberikan pengetahuan dasar, pendidikan juga diharapkan mampu menumbuhkan keterampilan berpikir kritis, kreativitas, serta kemampuan berinovasi. Dengan demikian, pendidikan tidak hanya menjadi alat untuk memperoleh pekerjaan tetapi juga sebagai sarana menciptakan masyarakat yang tangguh, adaptif, dan siap menghadapi berbagai tantangan.

Muttaqin (2018), dalam penelitiannya menemukan bahwa variasi ketimpangan pendidikan di tingkat kota dan kabupaten setelah desentralisasi. Dalam beberapa tahun terakhir, tren pendidikan di Indonesia telah mengalami berbagai perubahan yang signifikan sebagai upaya menjawab tuntutan tuntutan globalisasi dan revolusi digital. Sepanjang periode 2021 hingga 2023, dalam rentang waktu tersebut berbagai nilai yang bervariasi telah menjadi tren pendidikan di Indonesia. Tidak sedikit generasi muda yang berusia di bawah 25 tahun tidak memiliki akses pendidikan yang layak. Tahun 2021 sendiri adalah awal-mula digitalisasi pendidikan di Indonesia. Hal ini tidak lepas dari COVID-19 yang melanda Indonesia, sehingga tidak dapat melaksanakan kelas secara tatap muka, yang berujung pada pengadaan kelas secara daring atau bahkan tidak sama sekali. Teknologi yang dianggap sebagai suatu alternatif kini menjadi bagian yang tak terpisahkan dari pendidikan. Tren digitalisasi juga memungkinkan siswa di daerah terpencil untuk mengakses materi pendidikan yang sama dengan mereka yang berada di wilayah kota.

Selain digitalisasi, kurikulum pendidikan di Indonesia juga mulai berfokus pada pengembangan keterampilan praktis. Hal ini dapat dilihat dari peningkatan minat pada pendidikan vokasional dan pengembangan kurikulum yang mendukung keterampilan yang relevan dengan industri, seperti pemrograman, analisis data, dan keterampilan teknis lainnya. Perubahan ini mencerminkan kebutuhan akan lulusan yang siap bekerja dan memiliki keahlian yang dibutuhkan dalam dunia kerja modern. Pendidikan vokasional ini tidak hanya menciptakan tenaga kerja yang terampil tetapi juga mengurangi tingkat pengangguran dan meningkatkan

produktivitas ekonomi. Menurut Madhakomala dkk (2022), dalam studinya tentang alternatif permasalahan pendidikan di Indonesia, beberapa hal yang dapat dilakukan dengan menyelaraskan paradigma pendidikan, mendorong peran lembaga pendidikan tinggi, meningkatkan kualitas guru, dan menambah relevansi edukasi terhadap dunia kerja.

Mengelompokkan tren pendidikan dari 2021 hingga 2023 ini dapat membantu untuk memahami faktor-faktor kunci yang mendorong transformasi pendidikan di Indonesia. Tren digitalisasi menegaskan peran teknologi sebagai elemen utama dalam proses pembelajaran, sedangkan fokus pada keterampilan praktis menunjukkan penyesuaian pendidikan terhadap kebutuhan industri. Pemerataan akses pendidikan mencerminkan komitmen pemerintah dalam menciptakan kesetaraan pendidikan bagi seluruh rakyat Indonesia. Melalui klasifikasi ini, arah perkembangan pendidikan nasional dan faktor-faktor yang perlu diperhatikan untuk mempersiapkan generasi muda yang mampu bersaing di tingkat global dapat dilihat.

## **B. METHODOLOGY**

Penelitian ini menggunakan pendekatan kuantitatif untuk menghasilkan prediksi tren pendidikan dengan menggunakan algoritma K-Nearest Neighbor (KNN). Pendekatan kuantitatif dipilih karena penelitian ini berfokus pada analisis data statistik untuk mendapatkan hasil prediksi yang objektif dan akurat. Dengan algoritma KNN, penelitian ini mengidentifikasi pola dari data historis untuk memperkirakan perkembangan persentase pendidikan di masa depan, khususnya untuk tahun 2024. Tren data akan dihitung menggunakan rumus growth rate atau percentage change yang akan menghitung perubahan nilai antara dua waktu.

### **2.1. Algoritma K-Nearest Neighbor (KNN)**

K-Nearest Neighbor (KNN) adalah sebuah algoritma dalam machine learning yang termasuk dalam kategori supervised learning dan umum digunakan dalam tugas klasifikasi serta regresi (Gallego, Rico-Juan, Valero-Mas, 2022). Algoritma ini bekerja dengan prinsip sederhana yaitu untuk melakukan prediksi pada data baru, KNN akan mencari sejumlah K tetangga terdekat berdasarkan jarak tertentu, lalu memilih kelas yang paling umum atau rata-rata nilai pada data tetangga tersebut. Jarak antara data ditentukan menggunakan berbagai metrik, seperti jarak Euclidean, Manhattan, atau Minkowski yang bertujuan untuk mengukur kesamaan atau kedekatan data baru terhadap data yang sudah ada. Pada tugas klasifikasi, seperti memprediksi tren pendidikan, KNN menentukan kelas atau kelompok data baru berdasarkan mayoritas kelas dari K tetangga terdekatnya. Oleh (Uddin, Haque, Lu, Moni & Gide, 2022), KNN cocok digunakan pada dataset yang relatif kecil atau menengah, dan bekerja baik ketika data memiliki pola atau distribusi yang dapat didekati secara linear atau memiliki cluster yang jelas.

### **2.2. Pengumpulan Data**

Sumber data utama penelitian ini adalah data yang diambil dari Badan Pusat Statistik (BPS) yang memuat persentase penduduk berusia 25 tahun ke atas yang memiliki pendidikan minimal SMA, dikelompokkan berdasarkan gender. Data ini diambil untuk periode tiga tahun, yaitu dari tahun 2021 hingga 2023. Sumber ini dipilih karena data dari BPS memiliki validitas tinggi dan secara rutin dipublikasikan untuk keperluan analisis demografis dan sosial-ekonomi. Lalu selanjutnya dilakukan studi literatur yang dilakukan dengan membaca berbagai jurnal ilmiah terkait tren pendidikan, aplikasi machine learning dalam analisis pendidikan, serta efektivitas algoritma KNN dalam prediksi tren berbasis data. Studi literatur ini bertujuan untuk memberikan wawasan tambahan tentang metode yang akan diterapkan dan memberikan dukungan konseptual yang kuat bagi penelitian ini.

### **2.3. Alat dan Perangkat Lunak**

Seluruh analisis dalam penelitian ini dilakukan menggunakan perangkat lunak Python, yang menyediakan library seperti scikit-learn untuk implementasi algoritma KNN dan optimalisasi parameter. Library Pandas digunakan untuk memproses dan mengelola data, sementara Matplotlib dan Seaborn digunakan untuk visualisasi hasil prediksi. Library-library ini dipilih karena sifatnya yang open-source dan kemampuannya untuk menangani analisis data dalam jumlah besar dengan efisien. Penggunaan perangkat lunak ini memungkinkan penelitian untuk memperoleh hasil analisis yang cepat dan efisien, serta visualisasi data yang mudah dipahami sehingga dapat mendukung interpretasi hasil dengan lebih baik.

## C. RESULT AND DISCUSSION

### 3.1. Praproses Data

Dalam data yang digunakan terdapat 549 baris data dari berbagai kabupaten dari berbagai provinsi di Indonesia. Oleh karena itu data akan dikelompokkan berdasarkan gender dan jenis area, yakni provinsi dan kabupaten.

#### 3.1.1. Pengelompokan Data

*Tabel 1 Dataset Original*

No	Region	Male_2021	Male_2022	Male_2023	Female_2021	Female_2022	Female_2023
0	ACEH	46.93	47.72	48.61	42.05	42.91	44.57
1	Simeulue	52.05	50.98	52.16	40.00	40.43	42.67
2	Aceh Singkil	42.18	38.35	39.57	38.77	37.36	38.62
3	Aceh Selatan	38.20	37.46	39.08	32.60	29.16	34.83
4	Aceh Tenggara	54.36	55.48	55.44	43.27	46.46	47.54
...	...	...	...	...	...	...	...
544	Dogyai	13.54	17.49	17.10	6.94	8.76	11.05
545	Intan Jaya	19.66	21.65	14.29	3.76	6.18	2.32
546	Deiyai	20.16	15.54	18.52	2.27	4.51	3.23
547	Kota Jayapura	69.53	76.55	73.70	66.56	69.28	69.70
548	INDONESIA	41.30	42.06	42.62	34.87	36.95	37.60

Praproses data dimulai dari memuat *tools* yang akan digunakan yakni *library* sklearn dan pandas. *Library* ini sudah terkenal digunakan untuk proses data. Dari *library* sk-learn akan dimuat beberapa *tools* yang akan digunakan. Selanjutnya akan dilakukan muat dataset yang akan digunakan dilakukan prediksi tren di 2024. Setelah dilakukan muat data, selanjut akan dilakukan penyeseiaan pada kolom dan menghapus baris-baris yang bernilai *null* atau *NA*.

Dapat dilihat dari dataset yang dimuat terdapat kumpulan data yang berjumlah 549 data, di dalam data tersebut juga terdapat baris untuk data provinsi dan keseluruhan rata-rata yakni Indonesia. Kemudian dataset akan dikelompokkan berdasarkan kategori area dan gender.

Kemudian selanjutnya dilakukan proses ekstrak provinsi dari dataset mentah. Dapat dilihat pola-pola yang terdapat di dalam dataset adalah bahwa nama provinsi diketik dalam huruf besar, sehingga cara mengekstraknya dapat dilakukan dengan hanya cek apakah baris 'Region' memiliki huruf besar semua atau tidak. Hal ini dapat dicapai dengan *method* isupper() pada Python.

*Tabel 2 Data Laki-Laki*

	Region	2021	2022	2023
1	Simeulue	52.05	50.98	52.16
2	Aceh Singkil	42.18	38.35	39.57
3	Aceh Selatan	38.20	37.46	39.08
4	Aceh Tenggara	54.36	55.48	55.44
5	Aceh Timur	33.47	32.20	32.36
...	...	...	...	...

543	Puncak	8.81	8.53	4.66
544	Dogiyai	13.54	17.49	17.10
545	Intan Jaya	19.66	21.65	14.29
546	Deiyai	20.16	15.54	18.52
547	Kota Jayapura	69.53	76.55	73.70

Tabel 1 Data Perempuan

	Region	2021	2022	2023
1	Simeulue	40.00	40.43	42.67
2	Aceh Singkil	38.77	37.36	38.62
3	Aceh Selatan	32.60	29.16	34.83
4	Aceh Tenggara	43.27	46.46	47.54
5	Aceh Timur	28.85	29.14	31.23
...	...	...	...	...
543	Puncak	4.54	3.65	2.06
544	Dogiyai	6.94	8.76	11.05
545	Intan Jaya	3.76	6.18	2.32
546	Deiyai	2.27	4.51	3.23
547	Kota Jayapura	66.56	69.28	69.70

Tabel 4 Data Laki-Laki berdasarkan Provinsi

	Region	2021	2022	2023
0	ACEH	46.93	47.72	48.61
24	SUMATERA UTARA	51.38	54.36	55.75
58	SUMATERA BARAT	43.74	45.32	46.08
78	RIAU	46.39	45.76	45.75
91	JAMBI	39.74	40.53	41.37
...	...	...	...	...
474	SULAWESI BARAT	35.98	38.03	39.36
481	MALUKU	51.50	51.49	49.38
493	MALUKU UTARA	42.82	44.94	43.70
504	PAPUA BARAT	57.48	56.54	56.12
518	PAPUA	36.70	38.80	39.30

Tabel 5 Data Perempuan berdasarkan Provinsi

	Region	2021	2022	2023
0	ACEH	42.05	42.91	44.57
24	SUMATERA UTARA	47.50	49.82	51.68
58	SUMATERA BARAT	44.97	48.52	49.08
78	RIAU	43.04	43.24	43.27
91	JAMBI	33.73	35.28	35.93
...	...	...	...	...
474	SULAWESI BARAT	31.03	32.75	34.37
481	MALUKU	49.42	48.71	47.34
493	MALUKU UTARA	35.05	40.03	38.45
504	PAPUA BARAT	46.65	49.02	49.64
518	PAPUA	26.44	30.80	32.70

### 3.1.2. Menghitung tren data

Menghitung tren dari data didasarkan pada nilai dari kolom 2021 dan 2023. Metode penghitungan tren yang digunakan adalah *growth rate* atau *percentage change* antara dua waktu. Metode tersebut menghitung persentase perubahan nilai dari tahun 2021 ke tahun 2023.

$$Growth\ Rate\ (Tren) = \frac{Value_{2023} - Value_{2021}}{Value_{2021}} \times 100$$

Lalu selanjutnya akan dilakukan pengelompokan nilai dari hasil perhitungan tersebut. Untuk indikator trennya sebagai berikut:

Tabel 6 Indikator Pengelompokan Nilai

	<b>Ketentuan</b>	<b>Batas</b>
<b>-2</b>	Sangat Menurun	$\leq 20$
<b>-1</b>	Menurun	$> -20$ dan $< -2$
<b>0</b>	Stabil	$\geq -2$ dan $\leq 2$
<b>1</b>	Menaik	$>2$ dan $\leq 20$
<b>2</b>	Sangat Menaik	$> 20$

Kemudian selanjutnya akan indikator-indikator tersebut akan dipasangkan ke *dataframe*, lalu untuk melihat seberapa besar persentasi perubahan tiap tahun dapat dihitung dengan menggunakan metode perhitungan perubahan rata-rata.

$$Avg\ Change = \frac{Tahun_{akhir} - Tahun_{awal}}{Jumlah\ Tahun}$$

Tabel 7 Data Laki-Laki

	<b>Region</b>	<b>2021</b>	<b>2022</b>	<b>2023</b>	<b>Tren</b>	<b>Label</b>	<b>Avg Change</b>
<b>1</b>	Simeulue	52.05	50.98	52.16	0.211335	0	0.055
<b>2</b>	Aceh Singkil	42.18	38.35	39.57	-6.187767	-1	-1.305
<b>3</b>	Aceh Selatan	38.20	37.46	39.08	2.303665	1	0.440
<b>4</b>	Aceh Tenggara	54.36	55.48	55.44	1.986755	0	0.540
<b>5</b>	Aceh Timur	33.47	32.20	32.36	-3.316403	-1	-0.555

Tabel 8 Data Perempuan

	<b>Region</b>	<b>2021</b>	<b>2022</b>	<b>2023</b>	<b>Tren</b>	<b>Label</b>	<b>Avg Change</b>
<b>1</b>	Simeulue	40.00	40.43	42.67	6.675000	1	1.335
<b>2</b>	Aceh Singkil	38.77	37.36	38.62	-0.386897	0	-0.075
<b>3</b>	Aceh Selatan	32.60	29.16	34.83	6.840491	1	1.115
<b>4</b>	Aceh Tenggara	43.27	46.46	47.54	9.868269	1	2.135
<b>5</b>	Aceh Timur	28.85	29.14	31.23	8.249567	1	1.190

Tabel 9 Data Laki-Laki berdasarkan Provinsi

	<b>Region</b>	<b>2021</b>	<b>2022</b>	<b>2023</b>	<b>Tren</b>	<b>Label</b>	<b>Avg Change</b>
<b>0</b>	ACEH	46.93	47.72	48.61	3.579800	1	0.840
<b>24</b>	SUMATERA UTARA	51.38	54.36	55.75	8.505255	1	2.185
<b>58</b>	SUMATERA BARAT	43.74	45.32	46.08	5.349794	1	1.170
<b>78</b>	RIAU	46.39	45.76	45.75	-1.379608	0	-0.320
<b>91</b>	JAMBI	39.74	40.53	41.37	4.101661	1	0.815

Tabel 10 Data Perempuan berdasarkan Provinsi

	<b>Region</b>	<b>2021</b>	<b>2022</b>	<b>2023</b>	<b>Tren</b>	<b>Label</b>	<b>Avg Change</b>
<b>0</b>	ACEH	42.05	42.91	44.57	5.992866	1	1.260
<b>24</b>	SUMATERA UTARA	47.50	49.82	51.68	8.800000	1	2.090

58	SUMATERA BARAT	44.97	48.52	49.08	9.139426	1	2.055
78	RIAU	43.04	43.24	43.27	0.534387	0	0.115
91	JAMBI	33.73	35.28	35.93	6.522384	1	1.100
....							

### 3.1.3. Pembagian *Train Data* dan *Test Data*

Untuk bekerja dengan metode *K-Nearest Neighbor*, maka akan diperlukan untuk melakukan pembagian Data Latih dan Data Tes. Hal tersebut dapat dilakukan dengan mengambil X yaitu data tanpa kolom data yang ingin diprediksi dan Y yaitu kolom data yang akan diprediksi, dalam hal ini adalah data[semua kolom kecuali 'label'] dan data['label']. Untuk baris mana saja yang akan dipisah ke data latih dan data tes dapat dilakukan dengan menggunakan argumen `random_state` dan `test_size` pada `method train_test_split` dari `library sklearn.mode_selection`.

### 3.1.4. Normalisasi Data

Sebelum dilakukan proses data lebih lanjutnya, ada baiknya jika dilakukan normalisasi terlebih dahulu agar akurasi dari model nantinya lebih akurat. Untuk normalisasi dapat dilakukan dengan menggunakan metode Z-Score dan MinMax. Dalam penelitian ini akan digunakan metode Z-Score (*Standar Scaler*) untuk menormalisasi data.

Sejalan dengan proses normalisasi tersebut akan dipasangan metode prediksi yang digunakan, dalam hal ini KNN (*K-Nearest Neighbor*) dengan menggunakan Pipeline. Pipeline adalah *tools* yang praktis dan mempermudah perhitungan dan normalisasi data.

## 3.2. Pemrosesan Data

Setelah langkah-langkah praproses selesai, data dapat dilanjutkan untuk diproses dan dicari hasil akhirnya. Untuk langkah ini akan menggunakan GridSearch dan Cross-Validation untuk mencari parameter paling optimal yang digunakan untuk KNN, yaitu nilai K dan banyak Cross-Validation akan dipakai adalah sebanyak 5 kali. Sebelum mendefinisikan GridSearch tersebut akan dibuat terlebih dahulu objek untuk menampung parameter model.

```

pipeline = Pipeline([
    ('scaler', StandardScaler()),
    ('knn', KNeighborsClassifier())
])

param_grid = {
    'knn__n_neighbors': range(1, 50),
    'knn__weights': ['uniform', 'distance'],
    'knn__p': [1, 2]
}

gs_male = GridSearchCV(pipeline, param_grid, cv=5, scoring='accuracy')
gs_male.fit(X_train_male, y_train_male)

gs_female = GridSearchCV(pipeline, param_grid, cv=5, scoring='accuracy')
gs_female.fit(X_train_female, y_train_female)

```

Dikarenakan data yang dipakai telah dipisah berdasarkan gender, maka GridSearch juga akan dibedakan menjadi `gs_male` dan `gs_female`, untuk kategori yang dibedakan berdasarkan provinsi akan mengikuti GridSearch berdasarkan gendernya. Dengan model GridSearch tersebut dapat dilihat berapa persentase akurasi.

```

print(f"Best Male Model: {gs_male.best_params_}, Accuracy: {male_accuracy}")
print(f"Best Female Model: {gs_female.best_params_}, Accuracy: {female_accuracy}")
Best Male Model: {'knn__n_neighbors': 12, 'knn__p': 2, 'knn__weights': 'distance'}, Accuracy:
0.9290322580645162

```

Best Female Model: {'knn\_\_n\_neighbors': 9, 'knn\_\_p': 1, 'knn\_\_weights': 'uniform'}, Accuracy: 0.8580645161290322

Dapat dilihat bahwa nilai k yang digunakan paling optimal pada 12 untuk model Laki-Laki dan 9 untuk model Perempuan. Setelah GridSearch selesai diimplementasikan, maka bisa dilanjutkan untuk menentukan model paling cocok untuk prediksi tren data. Hal ini dapat dilakukan dengan menggunakan *method* predict() dari GridSearch.

```
male_model = gs_male.best_estimator_
female_model = gs_female.best_estimator_
male_predictions = male_model.predict(X_test_male)
female_predictions = female_model.predict(X_test_female)
male_accuracy = accuracy_score(y_test_male, male_predictions)
female_accuracy = accuracy_score(y_test_female, female_predictions)
```

Langkah-langkah yang diperlukan untuk prediksi tren sudah dipenuhi sehingga prediksi tren bisa dilakukan. Prediksi tren sendiri dapat dilakukan dengan mengunakan model yang telah dibuat, yakni male\_model dan female\_model, dipasangkan dengan *method* apply dari dataframee dan memberikan argumen yang cocok.

```
data_male['2024 Trend'] = data_male.apply(lambda row: male_model.predict([[row['2021'],
row['2022'], row['2023'], ((row['2023']-row['2021'])/row['2021']) * 100, ((row['2023']-
row['2021'])/2]])[0],axis=1)
```

```
data_male_prov['2024 Trend'] = data_male_prov.apply(lambda row:
male_model.predict([[row['2021'], row['2022'], row['2023'], ((row['2023']-row['2021'])/row['2021']) *
100, ((row['2023']-row['2021'])/2]])[0],axis=1)
```

```
data_female['2024 Trend'] = data_female.apply(lambda row: female_model.predict([[row['2021'],
row['2022'], row['2023'], ((row['2023']-row['2021'])/row['2021']) * 100, ((row['2023']-
row['2021'])/2]])[0],axis=1)
```

```
data_female_prov['2024 Trend'] = data_female_prov.apply(lambda row:
female_model.predict([[row['2021'], row['2022'], row['2023'], ((row['2023']-row['2021'])/row['2021'])
* 100, ((row['2023']-row['2021'])/2]])[0],axis=1)
```

### 3.3. Interpretasi Data

#### 3.3.1. Hasil Perhitungan Data

##### a) Data Laki Laki

Tabel 11 Hasil Perhitungan Data Laki-Laki

	Region	2021	2022	2023	Tren	Label	Avg Change	Trend 2024
1	Simeulue	52.05	50.98	52.16	0.211335	0	0.055	0
2	Aceh Singkil	42.18	38.35	39.57	-6.187767	-1	-1.305	-1
3	Aceh Selatan	38.20	37.46	39.08	2.303665	1	0.440	1
4	Aceh Tenggara	54.36	55.48	55.44	1.986755	0	0.540	0
5	Aceh Timur	33.47	32.20	32.36	-3.316403	-1	-0.555	-1
...	...	...	...	...	...	...	...	...
543	Puncak	8.81	8.53	4.66	-47.105562	-2	-2.075	-1
544	Dogiyai	13.54	17.49	17.10	26.292467	2	1.780	2
545	Intan Jaya	19.66	21.65	14.29	-27.314344	-2	-2.685	-2
546	Deiyai	20.16	15.54	18.52	-8.134921	-1	-0.820	-1
547	Kota Jayapura	69.53	76.55	73.70	5.997411	1	2.085	1

##### b) Data Laki Laki Berdasarkan Provinsi

Tabel 12 Hasil Perhitungan Data Laki-Laki berdasarkan Provinsi

	Region	2021	2022	2023	Tren	Label	Avg Change	Trend 2024
--	--------	------	------	------	------	-------	------------	------------

0	ACEH	46.93	47.72	48.61	3.579800	1	0.840	1
24	SUMATERA UTARA	51.38	54.36	55.75	8.505255	1	2.185	1
58	SUMATERA BARAT	43.74	45.32	46.08	5.349794	1	1.170	1
78	RIAU	46.39	45.76	45.75	-1.379608	0	-0.320	0
91	JAMBI	39.74	40.53	41.37	4.101661	1	0.815	1
...	...	...	...	...	...	...	...	...
474	SULAWESI BARAT	35.98	38.03	39.36	9.394108	1	1.690	1
481	MALUKU	51.50	51.49	49.38	-4.116505	-1	-1.060	-1
493	MALUKU UTARA	42.82	44.94	43.70	2.055114	1	0.440	1
504	PAPUA BARAT	57.48	56.54	56.12	-2.366040	-1	-0.680	0
518	PAPUA	36.70	38.80	39.30	7.084469	1	1.300	1

## c) Data Perempuan

Tabel 13 Hasil Perhitungan Data Perempuan

	Region	2021	2022	2023	Tren	Label	Avg Change	Trend 2024
1	Simeulue	40.00	40.43	42.67	6.675000	1	1.335	1
2	Aceh Singkil	38.77	37.36	38.62	-0.386897	0	-0.075	0
3	Aceh Selatan	32.60	29.16	34.83	6.840491	1	1.115	1
4	Aceh Tenggara	43.27	46.46	47.54	9.868269	1	2.135	1
5	Aceh Timur	28.85	29.14	31.23	8.249567	1	1.190	1
...	...	...	...	...	...	...	...	...
543	Puncak	4.54	3.65	2.06	-54.625551	-2	-1.240	1
544	Dogiyai	6.94	8.76	11.05	59.221902	2	2.055	2
545	Intan Jaya	3.76	6.18	2.32	-38.297872	-2	-0.720	1
546	Deiyai	2.27	4.51	3.23	42.290749	2	0.480	2
547	Kota Jayapura	66.56	69.28	69.70	4.717548	1	1.570	1

## d) Data Perempuan Berdasarkan Provinsi

Tabel 14 Hasil Perhitungan Data Perempuan berdasarkan Provinsi

	Region	2021	2022	2023	Tren	Label	Avg Change	Trend 2024
0	ACEH	42.05	42.91	44.57	5.992866	1	1.260	1
24	SUMATERA UTARA	47.50	49.82	51.68	8.800000	1	2.090	1
58	SUMATERA BARAT	44.97	48.52	49.08	9.139426	1	2.055	1
78	RIAU	43.04	43.24	43.27	0.534387	0	0.115	-1
91	JAMBI	33.73	35.28	35.93	6.522384	1	1.100	1
...	...	...	...	...	...	...	...	...
474	SULAWESI BARAT	31.03	32.75	34.37	10.763777	1	1.670	1
481	MALUKU	49.42	48.71	47.34	-4.208822	-1	-1.040	-1
493	MALUKU UTARA	35.05	40.03	38.45	9.700428	1	1.700	1
504	PAPUA BARAT	46.65	49.02	49.64	6.409432	1	1.495	1
518	PAPUA	26.44	30.80	32.70	23.676248	2	3.130	2

## 3.3.2. Hasil Analisis

Dari data yang telah dihasilkan, telah didapatkan tren-tren untuk tahun 2024. Berikut adalah rekap data berdasarkan indikator:

Tabel 15 Hasil Analisis Data

No	Indikator	M	F	MP	FP	Total M	Total F
0	Sangat Menurun	2	0	0	0	2	0
1	Menurun	129	93	4	3	133	96
2	Stabil	82	29	11	0	93	29
3	Menaik	260	290	19	30	279	320
4	Sangat Menaik	41	102	0	1	41	103
		<b>514</b>	<b>514</b>	<b>34</b>	<b>34</b>	<b>548</b>	<b>548</b>

Tren dikatakan memiliki hasil yang positif pada indikator jumlah Stabil, Menaik, dan Sangat Menaik lebih banyak dari Menurun dan Sangat Menurun. Untuk data laki-laki, jumlah indikator yang menunjukkan arah positif mencapai 75,5%, sedangkan arah negatif sebesar 25,5%. Secara spesifik, indikator Sangat Menurun menunjukkan angka sebesar 2, dan Menurun sebanyak 129, sehingga jumlah kategori negatif adalah 131. Sementara itu, indikator positif dengan kategori Stabil mencapai 82, Menaik sebanyak 260, dan Sangat Menaik sebanyak 41, menjadikan total nilai positif sebesar 383 dari total data 514. Dengan demikian, arah positif sebesar 75,5% ini menunjukkan adanya peningkatan minat atau pencapaian dalam pendidikan bagi gender laki-laki, yang memproyeksikan adanya pertumbuhan pendidikan yang cukup signifikan untuk tahun mendatang.

Untuk data perempuan, hasil menunjukkan tren positif yang lebih kuat dibandingkan dengan laki-laki. Tidak ada data pada kategori *Sangat Menurun*, dan kategori *Menurun* menunjukkan nilai sebesar 93, menjadikan jumlah kategori negatif sebesar 18%. Sementara itu, pada kategori positif, *Stabil* tercatat sebanyak 29, *Menaik* sebanyak 290, dan *Sangat Menaik* sebanyak 102, dengan total 421 dari total data 514, atau sebesar 82%. Ini mengindikasikan bahwa perempuan memiliki tren peningkatan dalam pendidikan yang lebih tinggi, yakni 7,4% lebih banyak dibandingkan dengan laki-laki. Dengan demikian, prediksi untuk tahun 2024 memperlihatkan bahwa perempuan akan memiliki potensi lebih besar dalam peningkatan pencapaian pendidikan dibandingkan laki-laki. Dibandingkan dengan data tren 2023 yang belum diprediksi:

Tabel 16 Data Tren 2023 Belum Prediksi

No	Indikator	M	F	MP	FP	Total M	Total F
0	Sangat Menurun	3	2	0	0	3	2
1	Menurun	129	84	5	1	134	85
2	Stabil	79	53	9	4	88	57
3	Menaik	261	273	20	28	281	301
4	Sangat Menaik	42	102	0	1	42	103
		<b>514</b>	<b>514</b>	<b>34</b>	<b>34</b>	<b>548</b>	<b>548</b>

Penelitian ini memberikan wawasan yang cukup signifikan mengenai tren pendidikan di Indonesia dengan melihat prediksi data untuk tahun 2024. Perbandingan dengan data tren tahun 2023 yang belum diprediksi memperlihatkan pola yang cukup konsisten. Secara umum, tren pendidikan menunjukkan perkembangan positif pada kedua gender, yang mana data perempuan menunjukkan peningkatan yang lebih signifikan. Pada data tahun 2023, kelompok perempuan yang stabil dan meningkat lebih tinggi pada kelompok yang menurun. Sementara pada data laki-laki, persentasi yang menurun lebih besar dibandingkan dengan perempuan, yang mengindikasikan adanya tantangan khusus bagi laki-laki di beberapa wilayah.

Hasil ini juga memberikan indikasi adanya perbaikan akses dan kualitas pendidikan untuk kedua gender, meskipun perempuan cenderung lebih menunjukkan kenaikan tren. Hal ini berpotensi disebabkan oleh upaya pemerintah dalam menggalakkan program pendidikan dan meningkatkan kualitas fasilitas belajar di berbagai wilayah. Hal ini sejalan dengan pendapat Fitri (2021), tentang rendahnya kualitas pendidikan Indonesia yang dapat diperbaiki dan diselesaikan dengan baik asalkan kerja sama berbagai pihak ada baik itu pemerintah, tenaga pendidik, peserta didik, maupun masyarakat. Dari hasil penelitian yang dilakukan terlihat bahwa tren pendidikan Indonesia pergi ke arah yang positif, ini menandakan bahwa kualitas pendidikan Indonesia telah menjadi lebih baik dari sebelumnya.

#### D. CONCLUSION

Kesimpulan penelitian ini menunjukkan bahwa tren pendidikan di Indonesia untuk tahun 2024 diprediksi akan mengalami peningkatan yang signifikan, dengan hasil yang lebih positif pada kategori stabil, menaik, dan sangat menaik. Berdasarkan hasil analisis, tren pendidikan pada data laki-laki menunjukkan arah positif sebesar 75,5% dan arah negatif sebesar 25,5%. Untuk data perempuan, tren menunjukkan arah positif sebesar 82%, yang mengungguli tren laki-laki dengan selisih 7,4%. Hal ini menunjukkan bahwa perempuan diprediksi akan memiliki tingkat pendidikan yang lebih tinggi dibandingkan laki-laki pada tahun mendatang. Prediksi ini didapatkan menggunakan algoritma K-Nearest Neighbor (KNN) yang menganalisis data historis dari tahun 2021 hingga 2023 untuk memperkirakan perkembangan pendidikan berdasarkan pola pertumbuhan persentase. Dengan pendekatan kuantitatif, hasil prediksi yang diperoleh diharapkan dapat menjadi panduan bagi pengambil kebijakan dalam memperkuat upaya peningkatan pendidikan di Indonesia, terutama dalam mendukung partisipasi pendidikan yang setara antar-gender di masa depan.

Analisis mendalam terhadap data menunjukkan bahwa, meskipun terdapat sejumlah daerah yang menunjukkan penurunan, tren umum mengindikasikan perbaikan dalam akses dan kualitas pendidikan, terutama di kalangan perempuan. Ini mencerminkan upaya yang lebih besar dalam mendorong partisipasi perempuan dalam pendidikan, sejalan dengan kebijakan pemerintah dan program-program pendidikan yang ditujukan untuk meningkatkan kesetaraan gender. Selanjutnya, tren kenaikan yang signifikan pada tingkat pendidikan perempuan menunjukkan perubahan positif dalam persepsi masyarakat terhadap pentingnya pendidikan bagi perempuan, yang diharapkan akan berdampak pada peningkatan kualitas sumber daya manusia di Indonesia. Prediksi tren pendidikan yang meningkat ini menandakan kualitas pendidikan Indonesia perlahan membaik dan memiliki masa depan yang cerah.

#### REFERENCES

- Faisal, & Martin, S. N. (2019). Science Education in Indonesia: Past, Present, and Future. *Asia-Pacific Science Education*, 1-29. doi:10.1186/s41029-019-0032-0
- Fitri, S. F. (2021). Problematika Kualitas Pendidikan di Indonesia. *Jurnal Pendidikan Tambusai*, 5(1), 1617-1620.
- Gallego, A. J., Rico-Juan, J. R., & Valero-Mas, J. J. (2022). Efficient k-nearest neighbor search based on clustering and adaptive k values. *Pattern Recognition*, 122(108356). doi:10.1016/j.patcog.2021.108356
- Muttaqin, T. (2018). Determinants of Unequal Access to and Quality of Education in Indonesia. *The Indonesian Journal of Development Planning*, 2(1), 1-23. doi:https://doi.org/10.36574/jpp.v2i1.27
- Nugroho, O. F., Permanasari, A., Firman, H., & Riandi. (2021). The Importance of Stem Based Education in Indonesia. *Pedagonal : Jurnal Ilmiah Pendidikan*, 5(2), 56-61. doi:0.33751/pedagonal.v5i2.3779
- Pramana, C., Chamidah, D., Suyatno, Renaldi, F., & Syaharuddin. (2021). Strategies to Improves Education Quality in Indonesia: A Review. *Turkish Online Journal of Qualitative Inquiry*, 12(3), 1977-1994.
- Rizkianti, P. A., Asbari, M., & Priambudi, N. P. (2024). Pendidikan Indonesia Masih Buruk? *Journal Of Informattion System And Management*, 3(2), 35-38.
- Statistik, B. P. (2023, Mei 6). *Persentase penduduk usia 25 tahun ke atas dengan pendidikan SMA ke atas menurut jenis kelamin (Persen), 2021-2023*. (Badan Pusat Statistik) Dipetik 10 24, 2023, dari <https://www.bps.go.id/id/statistics-table/2/MjE5OSMy/persentase-penduduk-usia-25-tahun-keatas-dengan-pendidikan-sma-ke-atas-menurut-jenis-kelamin--persen-.html>

- Sukasni, A., & Efendy, H. (2017). The Problematic of Education System in Indonesia and Reform Agenda. *International Journal of Education*, 9(3), 183-199. doi:doi:10.5296/ije.v9i3.11705
- Sukmayadi, V., & Yahya, A. H. (2020). Indonesia Education Landscape and the 21st Century Challenges. *Journal of Social Education Research*, 11(4), 219-234.
- Uddin, S., Haque, I., Lu, H., Moni, M. A., & Gide, E. (2022). Comparative performance analysis of K-nearest neighbour (KNN) algorithm and its different variants for disease prediction. *Scientif Reports*, 12(6256), 1-11. doi:0.1038/s41598-022-10358-x