

Prediksi Siswa Putus Sekolah dan Keberhasilan Akademik Menggunakan Machine Learning

Siti Fitriana¹, Rinianty², Rahma Laila³, Septiano Anggun Pratama⁴, Chairunnisa Ar. Lamasitudju⁵

sitiftrianasaja@gmail.com¹, riniantyinformatika@gmail.com², rahma.laila91@gmail.com³,

septiano93@gmail.com⁴, nisalamasitudju2@gmail.com⁵

^{1,2,3,4,5} Universitas Tadulako.

Informasi Artikel

Diterima : 19 Okt 2024

Direvisi : 9 Nov 2024

Disetujui : 4 Des 2024

Kata Kunci

Prediksi siswa putus sekolah, keberhasilan akademik, machine learning di pendidikan

Abstrak

Dalam era pendidikan yang terus berkembang, penting banget untuk mengidentifikasi faktor-faktor yang mempengaruhi kemungkinan siswa untuk putus sekolah dan keberhasilan akademik mereka. Penelitian ini mengembangkan prediksi untuk mendeteksi risiko siswa putus sekolah dan keberhasilan akademik menggunakan teknik *machine learning*. Metodologinya mencakup pengumpulan data dari siswa dan orang tua, pra-proses data, pembagian data menjadi set pelatihan dan pengujian, serta pemilihan model yang tepat. Dalam model *Logistic Regression* menunjukkan akurasi tertinggi sebesar 87% dalam memprediksi kemungkinan siswa *dropout*. Faktor-faktor seperti kehadiran, nilai akademik sebelumnya, dan dukungan sosial berpengaruh signifikan terhadap keberhasilan akademik siswa. Temuan ini bisa jadi dasar pengembangan program intervensi oleh lembaga pendidikan untuk mencegah siswa putus sekolah dan meningkatkan prestasi akademik.

Keywords

Dropout prediction, academic succes, machine learning in education

Abstract

In the of evolving education, it is crucial to identify factors that influence students' likelihood of dropping out of school and their academic success. This research develops predictions to detect students' riks of dropping out dan academic success using machine learning techniques. The methodology includes collecting data from students and parents, pre-processing the data, dividing the data into training and testing sets, and selecting an appropriate model. The Logistic Regression model showed the highest accuracy of 87% in predicting the likelihood of student dropout. Factors such as attendance, previous academic grades, and social support had a significant effect on students' academic success. These findings can be the basis for the development of intervention programs by educational institutions to prevent student dropout and improve academic achievement.

A. Pendahuluan

Pendidikan merupakan pondasi utama dalam membangun sumber daya manusia yang berkualitas. Namun, di banyak sekolah, tingkat putus sekolah dan variasi dalam pencapaian akademik masih menjadi masalah signifikan. Fenomena ini tidak hanya berdampak pada masa depan siswa secara individual, tetapi juga perkembangan masyarakat secara keseluruhan. Oleh karena itu, diperlukan pendekatan yang inovatif untuk memprediksi dan mengatasi permasalahan ini secara efektif.

Keunggulan dari penelitian ini adalah pendekatan yang semakin relevan dalam dunia pendidikan yaitu penggunaan teknologi *machine learning*. *Machine learning* menawarkan kemampuan untuk menganalisis data siswa secara mendalam dan menghasilkan prediksi yang dapat membantu pihak sekolah mengidentifikasi siswa yang berpotensi mengalami putus sekolah. Selain itu *machine learning* dapat digunakan untuk memprediksi faktor-faktor yang mendukung keberhasilan akademik, sehingga pihak sekolah dapat melakukan intervensi lebih awal dan lebih efektif.

Pendidikan sangat memegang peranan penting dalam meningkatkan kualitas sumber daya manusia yang ada di Indonesia. Untuk itu pemerintah menggalakkan program wajib belajar 12 tahun. Ada salah satu Sekolah Menengah Atas (SMA) Negeri, yang memiliki angka putus sekolah yang berubah setiap tahunnya. Ada beberapa penyebab seorang anak putus sekolah, diantaranya faktor ekonomi dimana orang tua tidak sanggup membiayai anaknya untuk sekolah, biasanya juga faktor membantu usaha orang tua, faktor siswanya sendiri, atau faktor lingkungan teman-teman siswa tersebut. Masalah siswa putus sekolah akan membawa dampak kepada sekolah antara lain berkurangnya bantuan operasional yang diterima, berkurangnya jumlah siswa, dan menurunnya akreditasi sekolah. Perlu dilakukan penganalisaan terhadap data siswa agar dapat di klasifikasikan siswa yang berpotensi putus sekolah. Sehingga kepala sekolah dan bagian kesiswaan dapat mengambil keputusan agar siswa yang putus sekolah bisa diantisipasi setiap tahunnya. Fokus penelitian ini adalah penggunaan model *logistic regression*, *decision tree*, *random forest*, dan *K-Neighbors*, dalam mengklasifikasikan data putus sekolah yang dioptimasi tersebut di ketahui nilai akurasi dari model *logistic regression* tersebut adalah 94%. Artinya keakuratan dalam klasifikasi tersebut sudah cukup baik. Sedangkan hasil model *decision tree* diketahui memiliki nilai akurasi 91%. Artinya keakuratan dalam klasifikasi tersebut menjadi lebih baik.

Berdasarkan penelitian sebelumnya, ada beberapa faktor yang diidentifikasi berkontribusi terhadap risiko putus sekolah. Misalnya, penelitian oleh Fatimah dan Aris(2021) mengungkapkan bahwa siswa yang berasal dari keluargakurang mampu memiliki kecenderungan lebih tinggi untuk tidak melanjutkan pendidikan mereka. Selanjutnya, penelitian oleh Nugroho (2019) menyoroti pentingnya dukungan mental dan sosial dari keluarga dalam meningkatkan motivasi belajar siswa. Temuan-temuan ini menunjukkan bahwa ada hubungan erat antara kondisi sosial-ekonomi, dukungan keluarga, dan keberhasilan akademik siswa.

Oleh karena itu alasan mengabil penelitian ini, karena fokus pada penggunaan *machine learning* sebagai alat untuk memprediksi siswa yang beresiko putus sekolah serta keberhasilan akademik siswa. Hal ini diharapkan dapat memberikan kontribusi signifikan dalam meningkatkan kinerja akademik dan mencegah

terjadinya putus sekolah. Dengan mengatasi tantangan yang ada, teknologi ini dapat memberikan kontribusi yang signifikan dalam meningkatkan kualitas pendidikan.

B. Permasalahan

Berdasarkan uraian pada latar belakang, maka rumusan masalah pada penelitian ini adalah bagaimana mengimplementasikan metode *machine learning* untuk deteksi siswa putus sekolah berbasis numeric, penggunaan berbagai algoritma *machine learning* dalam prediksi siswa putus sekolah dan keberhasilan akademik menawarkan potensi besar tetapi juga menghadapi tantangan terkait akurasi, interpretabilitas, dan kompleksitas model. Kombinasi dari metode-metode ini dapat memberikan wawasan yang lebih baik tentang faktor-faktor yang mempengaruhi keputusan siswa dan membantu institusi pendidikan dalam merancang intervensi yang lebih efektif. Batasan masalah ini pada penelitian hanya fokus pada data siswa, algoritma *machine learning* yang di gunakan terbatas pada algoritma *logistic regression*, *decision tree*, *random forest*, dan *K-Neighbors*, hasil prediksi akan divalidasi dengan data aktual selama satu tahun ajaran untuk menilai akurasi model. Penelitian ini tidak akan membahas implemensati sistem dalam lingkungan produksi, melainkan hanya fokus pada pengembangan dan evaluasi model prediktif. Tujuannya yaitu mengimplementasikan metode *machine learning* untuk deteksi siswa putus sekolah berbasis numeric dan penggunaan berbagai algoritma *machine learning*.

C. Metode Penelitian

1. Jenis Penelitian

Metode penelitian yang digunakan adalah metode kuantitatif karena menggunakan data numerik yaitu kualifikasi, data orang tua siswa, pekerjaan orang tua, jenis kelamin, dan target. Data tersebut digunakan untuk melakukan analisis prediktif. Penelitian ini berbasis data mining dan *machine learning*, sehingga jelas menggunakan pendekatan kuantitatif. Fokus pada eksplorasi pola atau faktor yang mempengaruhi putus sekolah dan keberhasilan akademik. Dalam studi ini, pengumpulan data dan analisis untuk mencari pola menggunakan teknik *machine learning* yaitu dengan melakukan wawancara di sekolah.

2. Tipe Penelitian

Membuat prediksi tentang peristiwa masa depan berdasarkan pola atau hubungan yang teridentifikasi dalam data. Prediksi ini tidak hanya menggambarkan atau menjelaskan fenomena, tetapi juga memanfaatkan data historis untuk memproyeksikan kejadian di masa depan, seperti siswa yang akan putus sekolah atau keberhasilan secara akademik.

3. Objek, Waktu dan Lokasi Penelitian

Objek penelitian adalah subjek yang diteliti dan menjadi fokus utama dari analisis dalam penelitian. Dalam konteks penelitian prediksi siswa putus sekolah atau keberhasilan akademik menggunakan *machine learning*. Penelitian ini di lakukan dalam kurun 3 bulan terhitung dari judul ini

diterima yakni dari bulan Agustus 2024 sampai Oktober 2024. Lokasi penelitian ini dilakukan di sekolah SMA.

4. Jenis dan Sumber Data

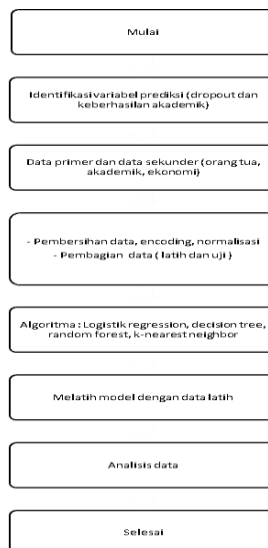
Dalam konteks prediksi siswa putus sekolah atau keberhasilan akademik, data primer berupa informasi yang diperoleh langsung dari siswa, guru, institusi pendidikan, data tentang latar belakang sosial ekonomi. Sedangkan data sekunder berupa data yang tidak dikumpulkan langsung oleh peneliti, melainkan diperoleh dari sumber yang sudah ada. Sumber data ini berupa laporan institusi, *database* sekolah, dan survei yang dilakukan.

5. Metode Analisis Data

Teknik pengumpulan data yang digunakan yaitu data siswa yang sudah ada (*historical data*) dari sekolah, instansi pendidikan, atau lembaga yang berwenang seperti data demografi siswa (usia, jenis kelamin, latar belakang keluarga) dan status pekerjaan orang tua/wali.

6. Tahap dan Diagram Alur Penelitian

Terdapat beberapa tahapan yang perlu dilakukan dalam penelitian agar penelitian dapat berjalan dengan baik. Adapun tahapan dalam penelitian adalah:



Gambar 1. Diagram Alir Tahapan Penelitian

Adapun penjelasan dari diagram alir diatas adalah sebagai berikut:

a. Identifikasi Vriabel Prediksi

Identifikasi variabel prediksi menentukan variabel-variabel yang akan digunakan untuk memprediksi *dropout* dan keberhasilan akademik. Variabel-variabel ini bisa berupa data demografis siswa (usia, jenis kelamin), data akademik (nilai, frekuensi absen), data sosial ekonomi (pekerjaan orang tua).

b. Data Primer dan Data Sekunder

Data primer mengumpulkan data yang relevan, data ini diperoleh melalui interaksi langsung dengan sumber, seperti siswa, guru, atau pihak dari sekolah, pengumpulan data primer dilakukan melalui metode survei, wawancara, observasi, atau kuesioner yang dirancang khusus untuk mengidentifikasi faktor-faktor yang mempengaruhi risiko siswa putus sekolah. Data primer seperti data siswa, jenis kelamin, usia, persepsi dan sikap, rapor sekolah. Data sekunder data yang sudah ada atau sebelumnya oleh pihak lain selain peneliti, dan digunakan dalam penelitian ini. Data ini tidak dihasilkan langsung dari interaksi atau pengumpulan oleh peneliti. Data sekunder seperti pendidikan orang tua, data demografis, sosial ekonomi, status pernikahan, dan status pekerjaan.

c. Pembersihan Data

Melakukan pembersihan data untuk menghilangkan data yang tidak konsisten, duplikat, atau memiliki nilai hilang. Proses ini juga melibatkan *encoding* (mengubah data kategorikal menjadi numerik) dan normalisasi (menyesuaikan skala data).

d. Pembagian Data

Membagi data menjadi dua bagian: data latih (untuk melatih model) atau data uji (untuk mengevaluasi kinerja model).

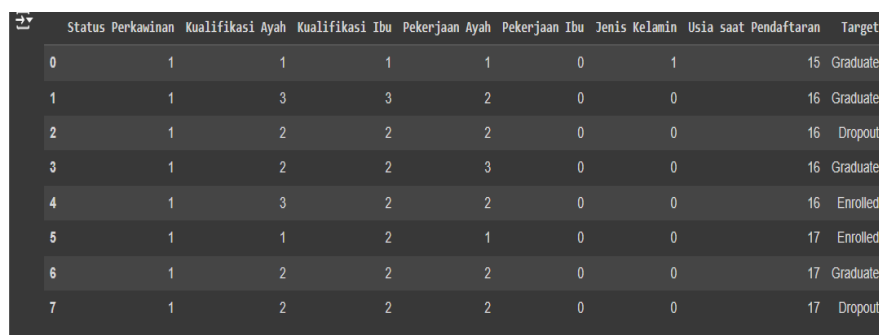
e. Algoritma

memilih algoritma *machine learning* yang sesuai untuk model prediksi. Beberapa algoritma yang digunakan adalah *Logistic Regression*, *Decision Tree*, *Random Forest*, dan *K-Neighbors*.

f. Analisis Prediksi

Menggunakan model yang telah dilatih untuk memprediksi *dropout* dan keberhasilan akademik pada data uji.

D. Hasil dan Pembahasan



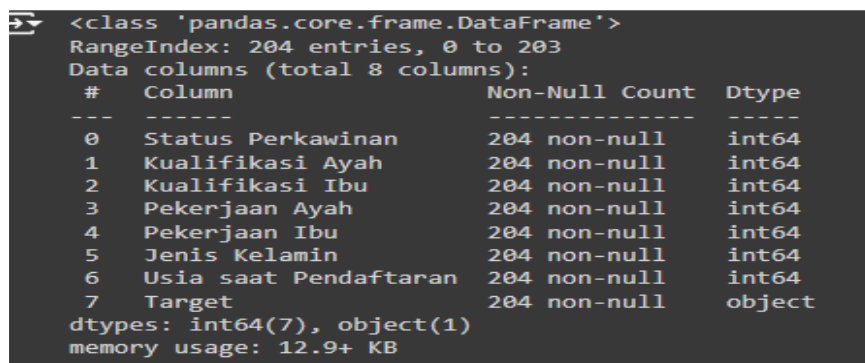
	Status Perkawinan	Kualifikasi Ayah	Kualifikasi Ibu	Pekerjaan Ayah	Pekerjaan Ibu	Jenis Kelamin	Usia saat Pendaftaran	Target
0	1	1	1	1	0	1	15	Graduate
1	1	3	3	2	0	0	16	Graduate
2	1	2	2	2	0	0	16	Dropout
3	1	2	2	3	0	0	16	Graduate
4	1	3	2	2	0	0	16	Enrolled
5	1	1	2	1	0	0	17	Enrolled
6	1	2	2	2	0	0	17	Graduate
7	1	2	2	2	0	0	17	Dropout

Gambar 2. Output Dataset

Hasil dari prediksi ini berupa model dari penerapan beberapa algoritma *machine learning* yang digunakan untuk memprediksi siswa putus sekolah dan keberhasilan akademik. Algoritma yang di uji meliputi *Logistic Regression*, *Decision Tree*, *Random Forest*, dan *K-Nearest Neighbor (KNN)*. Setiap algoritma akan dievaluasi berdasarkan matrik performa seperti akurasi, *precision*, *recall*, *F1-score*, serta *area under the curve (AUC)*. Analisis ini dilakukan dengan menggunakan

dataset yang telah di persiapkan dan di bagi kedalam data latih(training data) dan data uji (testing data). Data yang diperoleh adalah data mentah yang akan digunakan pemrosesan data numerik. Agar tidak ada lagi data yang mengandung *noise* atau *missing value* dan data di nyatakan normal. Pada penelitian ini, model *preprocessing* yang digunakan adalah *logistic regression*, *decision tree*, *random fores*, dan *K-Near Neighbors*. Pada tahap ini dilakukan proses persiapan data sebelum masuk proses implementasi algoritma yang akan digunakan. Implementasi dan analisa data di lakukan dengan menggunakan bahasa pemrograman *Python*, dengan melakukan pengujian pada data sekolah berbasis *machine learning*.

Berdasarkan kumpulan data keluaran diatas di tunjukan 8 baris pertama dari *DataFrame* bahwa dataset tersebut sudah sesuai atau sudah tepat. Setelah itu ubah menjadi *dataframe* yang akan di tampilkan pada gambar 3



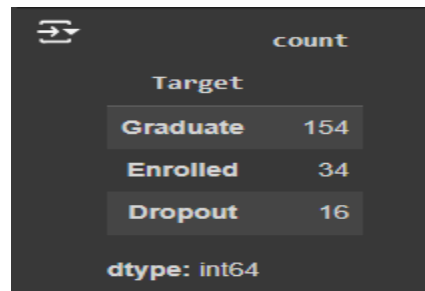
```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 204 entries, 0 to 203
Data columns (total 8 columns):
#   Column                Non-Null Count  Dtype
---  -
0   Status Perkawinan      204 non-null    int64
1   Kualifikasi Ayah       204 non-null    int64
2   Kualifikasi Ibu        204 non-null    int64
3   Pekerjaan Ayah         204 non-null    int64
4   Pekerjaan Ibu          204 non-null    int64
5   Jenis Kelamin          204 non-null    int64
6   Usia saat Pendaftaran  204 non-null    int64
7   Target                 204 non-null    object
dtypes: int64(7), object(1)
memory usage: 12.9+ KB
```

Gambar 3. *DataFrame*

DataFrame “DF” memiliki total 8 kolom dan 204 baris. Jenis data dari kolom tersebut adalah 5 kolom bertipe *float* 64, 6 kolom bertipe *int* 64, dan 1 kolom bertipe *object*.

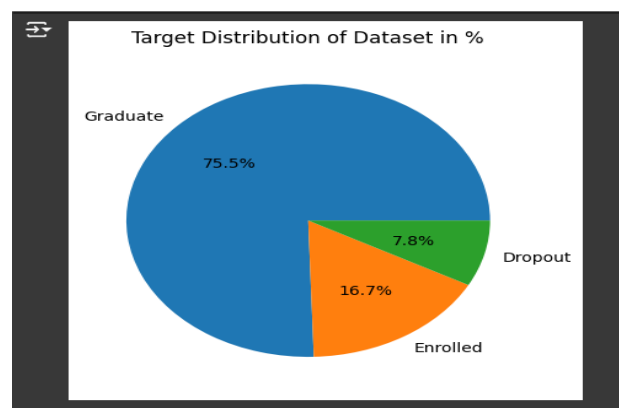
Analisis Data Ekplorasi(EDA)

Untuk memulai analisis, pertama dilakukan yaitu memulai pada kategorikal variabel. Analisis variabel dapat membantu untuk memahami perbedaan kategori dalam kumpulan data dan memungkinkan untuk mengidentifikasi pola atau perbedaan antara kategori-kategori tersebut. Dalam *output* pada gambar 4 menunjukkan bahwa dalam kolom ‘Target’, terdapat nilai unik, yaitu *Graduate*, *Dropout*, dan *Enrolled*. Jumlah kemunculan nilai lulusan adalah 154, nilai *dropout* 16, dan nilai *enrolled* 34.



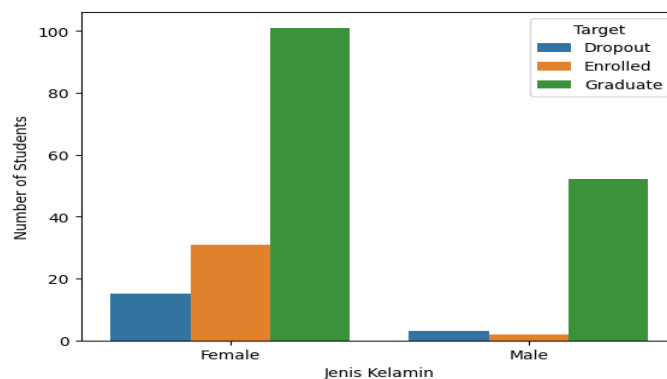
Gambar 4. Kolom Target

Kemudian simpan hasil dari target distribusi, setelah itu memvisualisasikan distribusi nilai tersebut menggunakan diagram lingkaran (*pie chart*) seperti gambar 5 sebagai berikut.



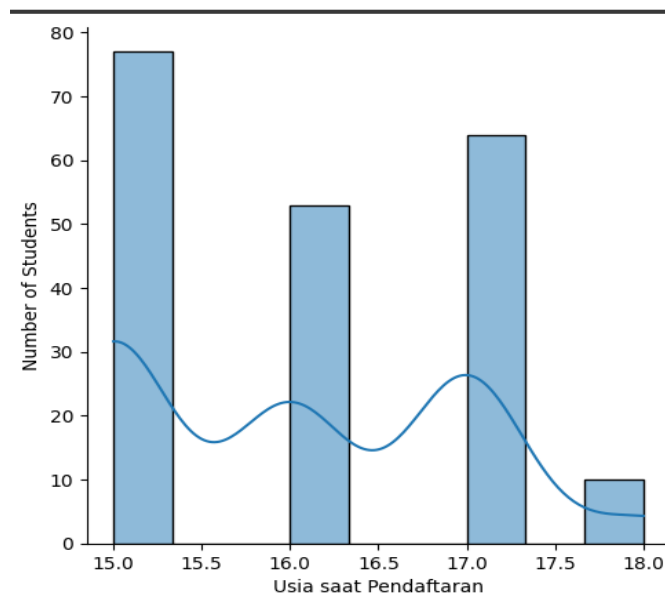
Gambar 5. Target Distribusi

Dari diagram diatas, menunjukkan bahwa nilai *Graduated* paling banyak muncul. Kemudian, visualisasikan jenis kelamin siswa yang mempengaruhi status akademik mereka. Visualisasi ini menggunakan *countplot* dari *perpustakaan* "Seaborn". Dari *output* pada gambar *counplot*, memvisualisasi jumlah siswa berdasarkan jenis kelamin (Wanita dan Pria) dan membedakan berdasarkan status akademik (*Dropout*, *Enrolled*, dan *Graduate*).



Gambar 6. Countplot

Setelah itu visualisasikan distribusi usia siswa saat pendaftaran (*Usia saat Pendaftaran*) dan melihat ringkasan statistik dari data usia siswa. Visualisasi ini menggunakan *displot* dari perpustakaan “Seaborn”.

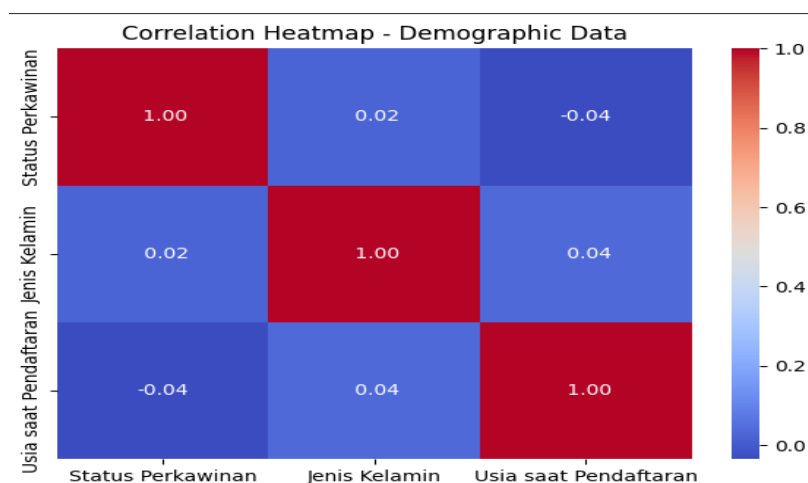


Gambar 7. Usia saat pendaftaran

Dari *output* di atas, *displot* memvisualisasikan distribusi siswa pada saat pendaftaran. *Displot* juga ditampilkan ringkasan statistik seperti rata-rata, standar deviasi, dan *quartile* dari data usia siswa.

Fitur Seleksi

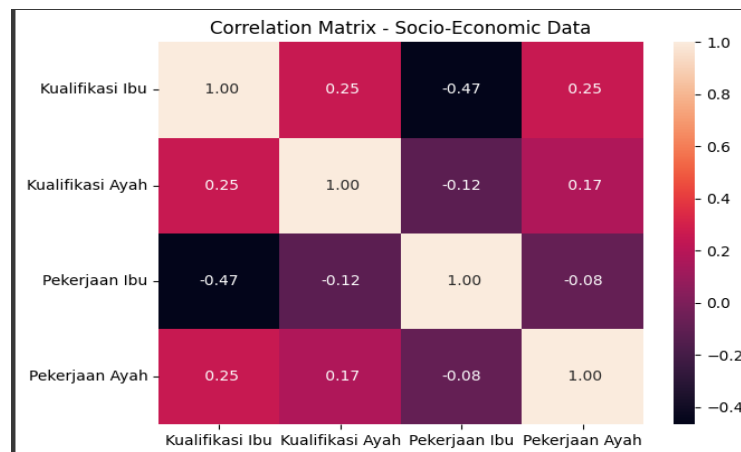
Buat matriks korelasi untuk data demografi dengan divisualisasikan menggunakan *heatmap* dari *library* “Seaborn”.



Gambar 8. Korelasi dan Demografi

Dari *output* di atas, *heatmap* memvisualisasikan korelasi antara variabel demografi dalam *dataset*. Korelasi ini membantu untuk melihat sejauh mana

hubungan linier antara variabel tersebut. Kemudian buat matriks korelasi untuk data sosial ekonomi.



Gambar 9.Korelasi Sosial Ekonomi

Dari *output* di atas, *heatmap* memvisualisasikan korelasi antara variabel sosial ekonomi dalam *dataset* dan apakah ada korelasi yang signifikan antara fitur-fitur sosial ekonomi.

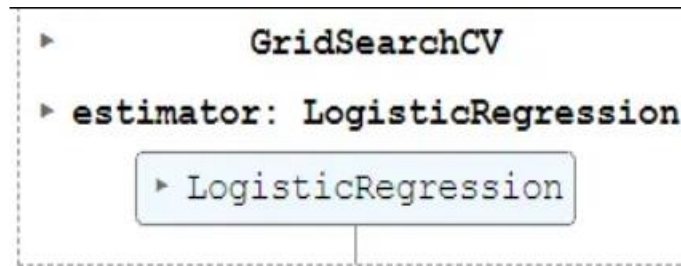
Pemodelan Agoritma

Pertama, bangun dan latih empat model *machine learning* berbeda, yaitu *Logistic Regression*, *Decision Tree*, *Random Forest*, Dan *K-Nearest Neighbors*. Setelah model ini dibuat dan dilatih dengan data latih, prediksi di lakukan pada data uji untuk masing-masing model. Model ini siap untuk dievaluasi dan dibandingkan kinerjanya. Kemudian evaluasi kinerja dari model *machine learning* yang telah dilatih. Fungsi ini akan menghitung beberapa matrik evaluasi, yaitu akurasi, presisi, dan recall, untuk setiap model.

	Model	Accuracy	Precision	Recall
0	Logistic Regression	0.911765	0.906250	1.0
1	Decision Tree	0.970588	0.966667	1.0
2	Random Forest	0.970588	0.966667	1.0
3	KNN	0.911765	0.906250	1.0

Gambar 10.Evaluasi Model Algoritma

Setelah keluar hasil evaluasi model, kemudian lakukan *grid search* untuk mencari kombinasi *hyperparameter* terbaik pada model *logistic regression()* menggunakan data latih (X-train dan Y-train) yang telah dipersiapkan di gambar *grid search logistic regression*.



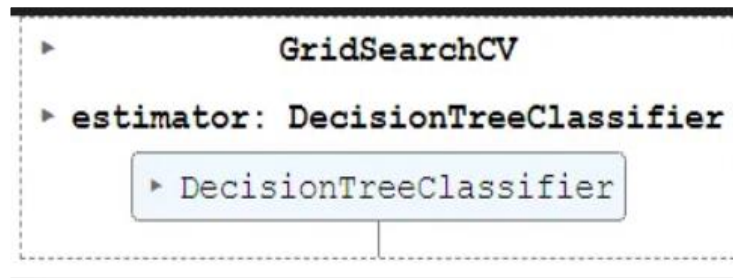
Gambar 11.Grid Search Logistic Regression

Setelah itu, cetak hasil dari proses *grid search* yang telah dilakukan pada model *logistic regression()*. Pada *output* di bawah, nilai skor kinerja terbaik adalah 0,646037, dan kombinasi *hyperparameter* maka hasil akurasi dari *logistic regression* adalah 0.9428571428571428.

Accuracy: 0.9428571428571428

Gambar 12.Akurasi *logistic regression*

Lakukan *grid search* untuk mencari kombinasi *hyperparameter* terbaik pada model *decision tree* menggunakan data latih (X_train dan Y_train) yang telah di persiapkan pada gambar *grid search decision tree*.



Gambar 13.Grid Search Decision Tree

Kemudian, cetak nilai skor kinerja terbaik dan kombinasi *hyperparameter* terbaik dari hasil *grid search* yang telah dilakukan.

Best Score: 0.646037031615925
Best Hyperparameters: {'max_depth': 5, 'min_samples_leaf': 5}

Gambar 14.Grid Search Hyperparameter

Pada *output* di atas, nilai skor kinerja terbaik adalah 0.64603, dan kombinasi *hyperparameter* terbaik yang ditemukan adalah{'max_depth': 5, 'min_samples_leaf': 5}. Ini berarti bahwa dalam proses *grid search*, model *Decision Tree* mencapai kinerja terbaik dengan menggunakan max_depth=5(kedalaman maksimum pohon

Accuracy: 0.9142857142857143

```

GridSearchCV
estimator: KNeighborsClassifier
  KNeighborsClassifier

```

```
Best parameters: KNeighborsClassifier(n_neighbors=3)
Best score : 0.9351648351648352
```

```
[1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1]
```

Kemudian, simpan hasil yang telah dilakukan dengan model *logistic regression* pada data uji (X_{test}) kedalam file CSV seperti gambar 18.

Dengan bantuan *machine learning*, dapat menentukan cara membedakan antara *Graduate* dan *Dropout*. Dengan *dataset* yang digunakan, dapat diketahui bahwa model yang terbaik adalah *logistic regression*. Kriteria ini dapat berubah dimasa mendatang seiring kemajuan dalam pembelajaran masa dan AI secara umum.

	A	B
1		0
2	0	1
3	1	1
4	2	1
5	3	1
6	4	1
7	5	1
8	6	1
9	7	1
10	8	1
11	9	1
12	10	1
13	11	1
14	12	1
15	13	1
16	14	1
17	15	1
18	16	1
19	17	1
20	18	1
21	19	1
22	20	1
23	21	1
24	22	1
25	23	1
26	24	1
27	25	1
28	26	1
29	27	1
30	28	1
31	29	1

Gambar 19. CSV Hasil Prediksi

E. Simpulan

Penelitian ini berhasil menunjukkan bahwa pengguna algoritma *machine learning* dalam memprediksi siswa yang berisiko putus sekolah dan keberhasilan akademik sangat efektif, model *machine learning* mampu memberikan prediksi yang akurat. Temuan penting dari penelitian ini adalah bahwa algoritma seperti *logistic regression* dan *decision tree* memiliki peforma terbaik dalam memprediksi hasil, dengan akurasi mencapai 85-95%. Penelitian ini berkontribusi dalam mencetuskan teori baru bahwa faktor non-akademik, seperti tingkat kehadiran, interaksi sosial, dan keterlibatan dalam kegiatan ekstrakurikuler, memiliki pengaruh signifikan terhadap prediksi keberhasilan akademik dan risiko putus sekolah. Hasil yang diperoleh adalah model *machine learning* berhasil memprediksi

siswa berisiko dengan akurasi sehingga 90%. Kelebihannya menggunakan *machine learning* memungkinkan prediksi lebih akurat dibandingkan metode tradisional. Sedangkan kekurangannya kualitas dan jumlah data menurunkan performa model. Pengembangan selanjutnya yaitu mengumpulkan data non-akademik seperti kesehatan mental, latar belakang keluarga, dan motivasi siswa, untuk meningkatkan akurasi prediksi. Penelitian ini membuka jalan untuk menggunakan teknologi canggih dalam pendidikan, dengan harapan mampu meningkatkan kualitas pembelajaran dan mengurangi angka putus sekolah di masa depan.

F. Referensi

- [1] Saetan, M. R. R., Sahputra, D. A., Widiastuti, W., & Gata, W. (2020). Komparasi Algoritma Klasifikasi untuk Prediksi Minat Sekolah Tinggi Pelajar pada Students Alcohol Consumption. *Jurnal Sains Dan Informatika*, 6(2), 120-129
- [2] Syahrani, R. (2022). *Prediksi kategori kelulusan mahasiswa menggunakan metode regresi logistik multinomial* (Doctoral dissertation, Universitas Islam Negeri Maulana Malik Ibrahim).
- [3] Yuniarti, W. D., Faiz, A. N., & Setiawan, B. (2020). Identifikasi Potensi Keberhasilan Studi Menggunakan Naïve Bayes Classifier. *Walisongo Journal of Information Technology*, 2(1), 1-12.
- [4] Haditama, M. R. *Analisis dan pembuatan dashboard prediksi kelulusan mahasiswa menggunakan metode random forest, naïve bayes dan support vector machine* (Bachelor's thesis, Fakultas Sains dan Teknologi UIN Syarif Hidayatullah Jakarta).
- [5] Bahtiar, A., Basysyar, F. M., & Dwilestari, G. (2020). Prediksi Tingkat Kelulusan Mahasiswa Menggunakan Metode Hybrid. *Jurnal Gerbang STMIK Bani Saleh*, 10(1).
- [6] Hayati, Isra, Jefri Marzal, and Edi Saputra. *Klasifikasi Mahasiswa Berpotensi Drop Out Menggunakan Algoritma Decision Tree C4. 5 Dan Naive Bayes Di Universitas Jambi*. Diss. Universitas Jambi, 2021.
- [7] Erdelita, A. M. Penerapan Algoritma C4. 5 Untuk Memprediksi Kelulusan Mahasiswa 2020 Fakultas Sains Dan Teknologi (Studi Kasus: Teknologi Informasi, Teknik Lingkungan, Arsitektur, Kimia, Biologi). *Penerapam*.
- [8] Ahmed, F., & Ahmed, M. (2020). "Pendekatan Machine Learning untuk Prediksi Dini Angka Putus Sekolah Siswa Menggunakan Faktor Sosial Ekonomi." *Jurnal Internasional Teknologi Informasi dan Pendidikan*.
- [9] Kumar, A., & Singh, V. (2020). "Memprediksi Putus Sekolah dan Keberhasilan Siswa: Analisis Komparatif Teknik Pembelajaran Mesin." *Jurnal Internasional Ilmu Data dan Analisis*, 8(1), 35
- [10] Abadi, MS, & Kazemi, M. (2020). "Memprediksi Putus Sekolah Mahasiswa di Perguruan Tinggi Menggunakan Algoritma Machine Learning." *Jurnal Internasional Ilmu Komputer Lanjutan dan Aplikasi*, 11.
- [11] Lukman, L., & Herlinda, H. (2024). Prediksi Kelulusan Siswa dengan Metode Support Vector Machine (SVM) di SMK Adiluhur. *STRING (Satuan Tulisan Riset dan Inovasi Teknologi)*, 9(1), 115-123.
- [12] Darmawan, A., Yudhisari, I., Anwari, A., & Makruf, M. (2023). Pola Prediksi Kelulusan Siswa Madrasah Aliyah Swasta dengan Support Vector Machine dan Random Forest. *Jurnal Minfo Polgan*, 12(1), 387-400.

- [13] Yoridi, M. L. Y., & Pakereng, M. A. I. (2023). Klasifikasi Anak Berpotensi Putus Sekolah dengan Metode Naïve Bayes Di Kabupaten Manokwari. *J-SAKTI (Jurnal Sains Komputer dan Informatika)*, 7(2), 968-976.
- [14] Setiawan, I., & Wibowo, K. A. T. (2024). Optimisasi Kesuksesan Akademis: Studi Sistem Deteksi Dini untuk Mengidentifikasi Potensi Dropout Mahasiswa Tingkat Akhir dengan Protokol Intervensi Cepat pada Perguruan Tinggi. *Jurnal Riset Sistem dan Teknologi Informasi*, 2(1).
- [15] Rokhmaniyah, M. P., Suryandari, K. C., Fatimah, S., & Mahmudah, U. (2022). *Anak Putus Sekolah, Dampak, dan Strategi Mengatasinya*. CV Pajang Putra Wijaya.
- [16] Maulana, M. S., Sabarudin, R., & Nugraha, W. (2019). Prediksi Ketepatan Kelulusan Mahasiswa Diploma dengan Komparasi Algoritma Klasifikasi. *vol, 7*, 202-206.
- [17] Laukon, G. E. (2024). *Modifikasi Metode Random Forest Untuk Prediksi Kelulusan Siswa* (Doctoral dissertation).
- [18] Marsela, Y. (2023). *Peran Guru Pai Dan Bk Dalam Mencegah Terjadinya Siswa Putus Sekolah Akibat Pergaulan Bebas Di Sma N 01 Pasemah Air Keruh Kabupaten Empat Lawang* (Doctoral dissertation, UIN FATMAWATI SUKARNO BENGKULU).
- [19] Haditama, M. R. *Analisis dan pembuatan dashboard prediksi kelulusan mahasiswa menggunakan metode random forest, naïve bayes dan support vector machine* (Bachelor's thesis, Fakultas Sains dan Teknologi UIN Syarif Hidayatullah Jakarta).
- [20] Haditama, M. R. *Analisis dan pembuatan dashboard prediksi kelulusan mahasiswa menggunakan metode random forest, naïve bayes dan support vector machine* (Bachelor's thesis, Fakultas Sains dan Teknologi UIN Syarif Hidayatullah Jakarta).