

## OPTIMASI KLASIFIKASI GAYA BELAJAR MAHASISWA INKLUSIF BERDASARKAN MODEL VAK DENGAN STRATIFIED SPLIT DAN MULTILAYER PERCEPTRON

Fandy Setyo Utomo<sup>1</sup>, Velizha Sandy Kusuma<sup>2</sup>, Wiga Maulana Baihaqi<sup>\*3</sup>, Pungkas Subarkah<sup>4</sup>

<sup>1,2,3</sup> Universitas Amikom Purwokerto, Purwokerto

Email: <sup>1</sup>fandy\_setyo\_utomo@amikompurwokerto.ac.id, <sup>2</sup>21sa1158@mhs.amikompurwokerto.ac.id,

<sup>3</sup>wiga@amikompurwokerto.ac.id, <sup>3</sup>subarkah@amikompurwokerto.ac.id

<sup>\*</sup>Penulis Korespondensi

(Naskah masuk: 10 Desember 2024, diterima untuk diterbitkan: 30 Oktober 2025)

### Abstrak

Identifikasi gaya belajar mahasiswa dengan mempertimbangkan fitur disabilitas memiliki peran penting dalam menciptakan pengalaman belajar yang inklusif dan personal. Namun, ketidakseimbangan data dalam kategori gaya belajar dan disabilitas menimbulkan tantangan yang signifikan bagi model klasifikasi. Penelitian ini bertujuan mengatasi tantangan tersebut dengan menerapkan teknik stratified split untuk menjaga keseimbangan distribusi kelas, khususnya pada variabel disabilitas dan gaya belajar. Algoritma Random Forest dan Multilayer Perceptron (MLP) digunakan untuk mengklasifikasikan gaya belajar mahasiswa berdasarkan model Visual, Auditory, dan Kinesthetic (VAK). Data yang digunakan berasal dari Open University Learning Analytics Dataset (OULAD), yang diproses melalui penggabungan data, pengkodean label, dan transformasi fitur untuk meningkatkan kinerja model. Evaluasi model dilakukan menggunakan metrik akurasi, presisi, recall, dan F1-score. Hasil penelitian menunjukkan bahwa model MLP mencapai kinerja sempurna dengan skor 100% pada semua metrik, sementara Random Forest menunjukkan performa sangat baik dengan skor 99%. Implementasi stratified split terbukti efektif dalam menjaga keseimbangan distribusi data, memastikan representasi yang memadai untuk semua kelas, termasuk mahasiswa dengan disabilitas. Penelitian ini memberikan kontribusi penting dalam mengembangkan model klasifikasi gaya belajar yang lebih akurat dan mendukung pendekatan pembelajaran yang lebih inklusif.

**Kata kunci:** gaya belajar, *stratified split*, *Random Forest*, MLP, disabilitas, *e-learning*

## OPTIMIZATION OF INCLUSIVE STUDENT LEARNING STYLE CLASSIFICATION BASED ON VAK MODEL WITH STRATIFIED SPLIT AND MULTILAYER PERCEPTRON

### Abstract

Identifying students' learning styles by considering disability features plays an important role in creating an inclusive and personalized learning experience. However, the imbalance of data in learning style and disability categories poses significant challenges for classification models. This research aims to overcome these challenges by applying a stratified split technique to maintain a balanced class distribution, especially in the disability and learning style variables. Random Forest and Multilayer Perceptron (MLP) algorithms are used to classify student learning styles based on the Visual, Auditory, and Kinesthetic (VAK) model. The data used comes from the Open University Learning Analytics Dataset (OULAD), which is processed through data merging, label coding, and feature transformation to improve model performance. Model evaluation was conducted using accuracy, precision, recall, and F1-score metrics. The results showed that the MLP model achieved perfect performance with a score of 100% on all metrics, while Random Forest showed excellent performance with a score of 99%. The implementation of stratified split proved effective in maintaining the balance of data distribution, ensuring adequate representation for all classes, including students with disabilities. This research makes an important contribution in developing more accurate learning style classification models and supporting more inclusive learning approaches.

**Keywords:** learning styles, *stratified split*, *Random Forest*, MLP, disability, *e-learning*

## 1. PENDAHULUAN

Proses pembelajaran *e-learning* telah menjadi fokus utama dalam pendidikan modern yang memungkinkan fleksibilitas dan aksesibilitas yang lebih baik bagi siswa. Salah satu model yang banyak digunakan dalam konteks ini adalah model Visual, Auditory, Kinesthetic (VAK). Model ini memberikan kerangka dasar untuk memahami bagaimana siswa menerima dan mengolah informasi secara berbeda berdasarkan preferensi mereka sehingga dapat mengoptimalkan pengalaman belajar siswa dengan memenuhi kebutuhan belajar mereka, termasuk untuk siswa dengan disabilitas (Dahlia et al., 2023)

Penelitian menunjukkan bahwa menyesuaikan metode pengajaran dengan gaya belajar VAK dapat meningkatkan efektivitas pembelajaran. Ketika materi disampaikan melalui saluran yang sesuai dengan preferensi siswa, pemahaman dan retensi informasi menjadi lebih baik (Akbar & Nasution, 2021). Pada penelitian sebelumnya ditemukan bahwa mayoritas siswa memiliki gaya belajar visual, yang menunjukkan pentingnya pendekatan yang sesuai dengan gaya belajar mereka untuk meningkatkan hasil belajar (Hidayati et al., 2021). Dalam konteks ini, identifikasi gaya belajar mahasiswa dengan mempertimbangkan status disabilitas mereka menjadi hal yang sangat penting, karena kebutuhan pendidikan siswa dengan disabilitas bisa berbeda dibandingkan dengan siswa tanpa disabilitas (Irmanda et al., n.d.)

Namun, tantangan muncul dalam klasifikasi gaya belajar, terutama ketika terdapat ketidakseimbangan data antara kategori gaya belajar atau dalam konteks disabilitas. Ketidakseimbangan kelas ini, seperti jumlah mahasiswa dengan gaya belajar visual yang lebih dominan dibandingkan dengan auditory atau kinesthetic, dapat mengurangi efektivitas model dalam mengklasifikasikan kategori yang kurang terwakili (Devy et al., 2022). Untuk mengatasi masalah ini, penelitian ini menerapkan teknik stratified split pada kolom disabilitas dan study type guna menjaga distribusi kelas yang seimbang antara kategori gaya belajar dan disabilitas. Pendekatan ini bertujuan untuk memberikan kesempatan yang adil bagi model dalam mempelajari setiap kategori, baik yang lebih dominan maupun yang kurang terwakili, dengan mempertimbangkan aspek disabilitas (Subramaniam et al., 2023)

Dalam penelitian ini, metode klasifikasi seperti Random Forest dan Multilayer Perceptron (MLP) digunakan untuk mengklasifikasikan gaya belajar mahasiswa berdasarkan model VAK, dengan perhatian khusus pada peran disabilitas sebagai salah satu fitur penting. Penggunaan Random Forest

memungkinkan identifikasi fitur yang signifikan dalam prediksi gaya belajar, sedangkan MLP mampu menangkap pola yang lebih kompleks dalam data (Susanti et al., 2023). Dengan pendekatan ini, penelitian diharapkan dapat menghasilkan model yang lebih akurat dan mendukung strategi pembelajaran yang lebih personal dan adaptif. Hasil klasifikasi yang akurat akan membantu pengembangan metode pengajaran yang lebih responsif dan inklusif, khususnya bagi siswa dengan disabilitas, sehingga dapat menciptakan lingkungan belajar yang optimal dan mendukung semua siswa. (Destriani et al., 2021).

## 2. LANDASAN PUSTAKA

### 2.1. Teori Gaya Belajar

Teori gaya belajar merupakan konsep penting dalam pendidikan yang berfokus pada cara individu menyerap, memproses, dan mengingat informasi. Dalam konteks *e-learning*, penting untuk mengembangkan sistem yang dapat menyesuaikan dengan gaya belajar siswa. Sebuah studi menunjukkan bahwa personalisasi dalam *e-learning* dapat membantu mengakomodasi karakteristik belajar siswa, sehingga meningkatkan motivasi dan hasil belajar (Nyoman Pardomuan et al., 2020). Penelitian lain juga menegaskan bahwa pengalaman belajar online yang disesuaikan dengan gaya belajar siswa, seperti visual dan auditori, dapat meningkatkan motivasi belajar secara signifikan (Syarifuddin et al., n.d.). Dengan memahami dan mengakomodasi gaya belajar siswa, pendidik dapat menciptakan pengalaman belajar yang lebih menarik dan produktif, yang pada gilirannya dapat meningkatkan hasil belajar siswa secara keseluruhan.

### 2.2. Model VAK

Gaya belajar VAK pertama kali diperkenalkan oleh Neil Fleming pada tahun 1987 sebagai kerangka untuk memahami perbedaan cara individu menyerap informasi. VAK mengklasifikasikan gaya belajar menjadi tiga kategori yaitu, Visual (melalui penglihatan), Aural (melalui pendengaran), dan Kinesthetic (melalui pengalaman fisik) (Childs-Kean et al., 2020). Penerapan strategi pembelajaran yang disesuaikan dengan gaya belajar VAK terbukti dapat meningkatkan keterlibatan dan motivasi siswa. Misalnya, penggunaan media visual bagi siswa dengan gaya belajar visual atau penerapan pembelajaran aktif untuk siswa kinestetik dapat membantu mereka lebih memahami materi pelajaran (Hassanzadeh et al., 2019). Penelitian menunjukkan

bahwa ketika strategi pengajaran disesuaikan dengan preferensi belajar siswa, hasil akademik mereka cenderung meningkat, terutama bagi mereka yang memiliki kebutuhan pendidikan khusus (Van Kessel et al., 2019). Hal ini sejalan dengan temuan Childs-Kean et al. yang menunjukkan bahwa penggunaan kerangka gaya belajar dalam pendidikan kesehatan dapat meningkatkan hasil belajar.

### 2.3. Stratified Split

Stratified Split adalah teknik pemisahan data yang mempertahankan proporsi kelas target dalam data pelatihan dan pengujian, sesuai distribusi aslinya. Pada penelitian ini, stratified split diterapkan untuk memastikan bahwa kelas dalam variabel target, khususnya yang terkait dengan disabilitas dan gaya belajar, tetap seimbang di setiap subset data. Hal ini sangat penting karena ketidakseimbangan kelas dapat menyebabkan bias dalam pelatihan model, di mana model cenderung belajar lebih baik pada kelas mayoritas dan kurang akurat dalam memprediksi kelas minoritas. Penelitian sebelumnya telah menekankan bahwa stratifikasi membantu dalam mencapai distribusi label yang konsisten baik dalam pelatihan maupun pemecahan uji yang sangat penting untuk evaluasi kinerja yang andal pada estimasi kinerja stabilitas dalam pencitraan kardiovaskular (Singh et al., 2021).

### 2.4. Random Forest

Algoritma Random Forest adalah metode ensemble yang menggabungkan banyak pohon keputusan untuk meningkatkan akurasi klasifikasi dan mengurangi risiko *overfitting*. Dalam penelitian ini, Random Forest digunakan untuk mendeteksi gaya belajar mahasiswa pada platform *e-learning* berdasarkan status disabilitas, dengan memanfaatkan data yang telah melalui proses stratified split untuk menangani ketidakseimbangan kelas. Setiap pohon dalam Random Forest dibangun menggunakan subset acak dari data pelatihan, yang memungkinkan model untuk menangkap pola berbeda di setiap subset (Kaur & Kumar, n.d.).

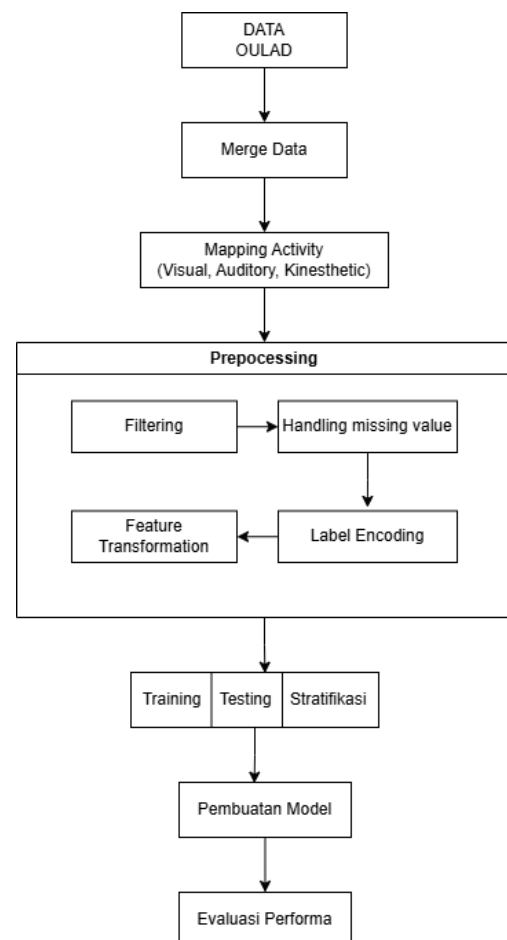
### 2.5. Multilayer Perception

Algoritma Multilayer Perceptron (MLP) adalah jenis jaringan saraf tiruan yang terdiri dari lapisan input, satu atau lebih lapisan tersembunyi, dan lapisan output. Dalam penelitian ini, MLP digunakan untuk mengklasifikasikan gaya belajar mahasiswa di platform *e-learning* berdasarkan status disabilitas. MLP bekerja dengan memproses fitur input secara berlapis-lapis melalui jaringan node

yang terhubung. Setiap node menggunakan fungsi aktivasi, seperti ReLU atau sigmoid, untuk memodelkan hubungan non-linear yang ada di antara fitur (Gomede et al., 2020).

## 3. METODE PENELITIAN

Dalam penelitian ini, klasifikasi gaya belajar mahasiswa berdasarkan model VAK (Visual, Auditory, Kinesthetic) dilakukan dengan menggunakan algoritma Random Forest dan Multilayer Perceptron (MLP). Alur penelitian yang menunjukkan tahapan pemrosesan dan pengklasifikasian gaya belajar ini dapat dilihat pada Gambar 1. Algoritma Random Forest dan MLP dipilih untuk membandingkan performa klasifikasi terhadap data yang telah diproses melalui berbagai tahap



Gambar 1. Metode Penelitian

### 3.1 Pengumpulan Data

Penelitian ini menggunakan data Open University Learning Analytics Dataset (OULAD), yang tersedia secara *open source*. Dataset ini mencakup informasi tentang mahasiswa, mata kuliah, serta interaksi mahasiswa dengan *Virtual Learning Environment* (VLE) dari *Open University* (OU) di Inggris.

### 3.2 Penggabungan Data

Proses penggabungan data melibatkan penggabungan dua set data atau lebih menjadi satu basis data terpadu untuk membentuk dataset yang komprehensif. Penggabungan ini tidak hanya menambah detail baru ke data yang sudah ada, tetapi juga memastikan kelengkapan dan akurasi informasi dengan menghapus duplikasi atau entri yang salah. Dengan demikian, dataset yang dihasilkan dapat digunakan untuk analisis lebih lanjut mengenai gaya belajar mahasiswa (Jurczyk et al., 2020).

### 3.3 Mapping Activity

Proses di mana variabel atau nilai dalam dataset dikaitkan dengan kategori atau kelas tertentu untuk tujuan klasifikasi atau analisis lebih lanjut. Pada penelitian ini, mapping dilakukan untuk mengklasifikasikan jenis aktivitas yang tercatat dalam kolom *activity\_type* ke dalam gaya belajar VAK (Visual, Auditory, Kinesthetic). Proses ini penting untuk menentukan bagaimana mahasiswa berinteraksi dengan platform pembelajaran berdasarkan gaya belajar mereka (Rashad Sayed et al., 2024).

### 3.4 Filtering Data

Filtering adalah proses selektif untuk memilih, menyaring, atau memodifikasi dataset guna memenuhi tujuan analisis atau pemodelan yang ditentukan. Proses ini melibatkan penggunaan kriteria atau metode tertentu untuk mengurangi redundansi dalam data, menghilangkan data yang tidak sesuai dengan fokus penelitian. Filtering membantu memperbaiki kualitas dataset dengan memastikan bahwa hanya data yang valid dan penting yang digunakan dalam analisis selanjutnya, sehingga meningkatkan efisiensi dan akurasi model yang dikembangkan.

### 3.5 Handling Missing Value

Missing values, atau nilai yang hilang, merujuk pada keadaan di mana suatu entitas data tidak memiliki nilai untuk atribut tertentu. Penanganan nilai hilang penting agar analisis data tidak terdistorsi (Desiani et al., 2021).

### 3.6 Label Encoding

Proses Pengkodean Label adalah langkah penting dalam pemrosesan data sebelum digunakan dalam model pembelajaran mesin. Label Encoding mengubah data kategorikal, yang awalnya berupa teks atau simbol, menjadi angka sehingga dapat diproses oleh algoritma machine learning yang umumnya hanya dapat memproses data numerik (Maulana Baihaqi et al., 2024).

### 3.7 Feature Transformation

Transformasi fitur merupakan langkah penting dalam preprocessing data, khususnya dalam penelitian yang melibatkan analisis gaya belajar mahasiswa berdasarkan status disabilitas. Dua metode transformasi yang umum digunakan adalah log transformasi dan power transformasi. Log transformasi digunakan untuk mengatasi distribusi data yang miring, seperti data yang terdistribusi skewed, dengan tujuan menstabilkan variansi dan membuat data lebih normal, cocok untuk analisis statistik. Sebaliknya, power transformasi lebih fleksibel, seperti Box-Cox dan Yeo-Johnson, yang dapat menangani nilai nol atau negatif. Kedua metode ini membantu menyiapkan data untuk model machine learning dalam penelitian klasifikasi gaya belajar mahasiswa berdasarkan disabilitas..

### 3.8 Stratified Split

Stratified split adalah teknik pemisahan data yang memastikan bahwa proporsi kelas target dalam data pelatihan dan data pengujian tetap seimbang, sesuai dengan distribusi aslinya. Teknik ini penting untuk mengatasi masalah ketidakseimbangan kelas, sehingga model dapat belajar secara efektif dari data yang representatif. Stratified split sering digunakan untuk memastikan bahwa setiap subset data mencerminkan distribusi yang ada pada dataset lengkap (Norinder, 2023).

### 3.9 Pembuatan Model

Pada tahap pembuatan model, algoritma Random Forest dan Multilayer Perceptron (MLP) digunakan untuk klasifikasi gaya belajar mahasiswa berdasarkan data yang telah diproses. Model-model ini dipilih karena kemampuannya dalam menangani data yang kompleks dan multidimensional, serta kemampuannya untuk memberikan prediksi yang akurat meskipun dengan data yang tidak seimbang (Sathe & Adamuthe, 2021a).

### 3.10 Evaluasi Performa

Setelah model dikembangkan menggunakan algoritma Random Forest dan Multilayer Perceptron (MLP), evaluasi performa dilakukan untuk mengukur akurasi dan efektivitas kedua model dalam mengklasifikasikan gaya belajar mahasiswa. Evaluasi ini melibatkan penggunaan metrik seperti akurasi, presisi, recall, dan F1-score pada data uji yang telah dipisahkan melalui stratified splitting (Özdemir et al., 2019).

## 4. HASIL DAN PEMBAHASAN

### 4.1 Pengumpulan Data

Penelitian ini menggunakan dataset Open University Learning Analytics Dataset (OULAD), yang merupakan data publik yang disediakan oleh Open University (OU) di Inggris. Dataset ini mencakup data dari berbagai mata kuliah yang diselenggarakan oleh OU selama periode 2013 hingga 2014, di mana setiap mata kuliah disebut sebagai module dan dapat ditawarkan beberapa kali dalam setahun. Data OULAD menawarkan informasi mendetail mengenai karakteristik dan aktivitas belajar mahasiswa, serta hasil akademik (Zhang et al., 2023).

### 4.2 Merging Data

Pada tahap ini, data dari beberapa tabel dalam dataset OULAD, yaitu vle, studentVLE, dan studentInfo, digabungkan untuk membentuk dataset terpadu. Penggabungan ini menghasilkan dataset komprehensif yang mencakup interaksi mahasiswa dengan aktivitas di VLE serta informasi demografis dan akademik mereka. Dataset yang terbentuk terdiri dari 4.327.256 baris dan 15 kolom, memberikan gambaran lengkap mengenai pola belajar dan karakteristik mahasiswa secara holistik. Proses ini penting untuk menganalisis hubungan antara faktor-faktor tersebut dalam memprediksi gaya belajar mahasiswa.

### 4.3 Mapping Activity

Tabel 1. Mapping Activity

activity_type	learning_style
oucollaborate	Visual
oucontent	
questionnaire	
repeatactivity	
resource	
url	
htmlactivity	Auditory
forumng	
ouelluminate	
ouwiki	
sharedsubpage	Kinesthetic
subpage	
glossary	
dataplus	
dualpane	

Pada tahap ini, nilai pada kolom *activity\_type* dipetakan menjadi *learning\_style* berdasarkan klasifikasi gaya belajar VAK. Beberapa tipe aktivitas, seperti quiz, externalquiz, folder, page, dan homepage, dihapus karena tidak memiliki informasi terkait pemetaan aktivitas tersebut. Aktivitas yang tersisa kemudian dikategorikan ke dalam tiga gaya belajar utama: Visual, Auditory, dan Kinesthetic (Rashad Sayed et al., 2024) sebagaimana ditampilkan pada Tabel 1. Proses ini bertujuan untuk mengklasifikasikan jenis aktivitas yang relevan dengan gaya belajar mahasiswa untuk analisis lebih lanjut.

Dengan demikian, kolom pada dataset setelah penggabungan bertambah satu, yaitu *learning\_style*, yang akan menjadi target pada pemodelan yang ditampilkan pada Tabel 2.

### 4.4 Filtering Data

Proses ini diterapkan dengan menggunakan metode menghapus data duplikat pada dataset gabungan untuk mempertahankan satu entri per kombinasi unik dari kode mata kuliah (*code\_module*), jenis aktivitas (*activity\_type*), dan ID mahasiswa (*id\_student*). Setelah penghapusan data duplikat dilakukan, kolom *id\_student* dan *id\_site* dihapus karena tidak relevan dengan pemodelan yang akan dilakukan. Dengan langkah ini, data yang tersisa hanya mencakup satu entri per aktivitas unik yang diikuti setiap mahasiswa pada setiap mata kuliah, menghasilkan total 171.449 baris dan 16 kolom. Hal ini bertujuan untuk mengurangi bias yang mungkin timbul dari pengulangan data dan menjaga integritas analisis selanjutnya, sehingga pola gaya belajar dapat dievaluasi lebih akurat, hasil *filtering* akan ditampilkan pada Tabel 2.

### 4.5 Handling missing value

Pada tahap ini, penanganan nilai missing dilakukan untuk memastikan kelengkapan data sebelum digunakan dalam analisis. Pada dataset OULAD yang digunakan, terdapat beberapa entri dengan nilai kosong pada kolom *imd\_band* sebanyak 7.217 baris, yang berisi informasi mengenai tingkat deprivasi sosial dan ekonomi mahasiswa. Untuk mengatasi hal ini, nilai missing pada kolom *imd\_band* diisi menggunakan modus dari kolom tersebut, yaitu nilai yang paling sering muncul.

Tabel 2 Final Dataset

code_module	code_presentation	activity_type	sum_click	gender	region	highest_education
AAA	2013J	forumng	4	F	Scotland	HE Qualification
AAA	2013J	oucontent	2	F	North Western Region	A Level or Equivalent
GGG	2014J	resource	1	M	South Region	A Level or

code_module	code_presentation	activity_type	sum_click	gender	region	highest_education
AAA	2013J	subpage	3	F	South East Region	Equivalent
FFF	2014J	htmlactivity	1	M	London Region	A Level or Equivalent Lower Than A Level
imd_band	age_band	final_result	studied_credits	disability	num_of_prev_attempts	learning_style
20-30%	35-55	Pass	60	N	0	Auditory
30-40%	35-55	Withdrawn	60	Y	0	Visual
20-30%	35-55	Pass	30	N	0	Visual
50-60%	35-55	Pass	60	N	0	Kinesthetic
10-20%	35-55	Withdrawn	180	N	0	Auditory

#### 4.6 Label Encoding

Pada Tabel 2, dilakukan proses label encoding pada sejumlah kolom kategori dalam dataset untuk mengonversi data kategorikal menjadi bentuk numerik yang dapat digunakan oleh algoritma pembelajaran mesin. Kolom yang di-encode meliputi `code_module`, `code_presentation`, `activity_type`, `gender`, `region`, `highest_education`, `imd_band`, `age_band`, `final_result`, dan `disability`. Proses ini bertujuan agar variabel-variabel tersebut dapat diolah lebih efektif dalam model yang akan digunakan. Sedangkan untuk kolom `learning_style` tetap mempertahankan nilai aslinya untuk menjaga keterkaitannya dengan variabel gaya belajar (Visual, Auditory, Kinesthetic). Hasil dari proses label encoding ini disajikan dalam Tabel 3, yang memperlihatkan dataset dengan nilai kategori yang telah dikonversi menjadi bentuk numerik.

#### 4.7 Feature Transformation

Pada tahap ini, diterapkan beberapa metode feature transformation untuk mengoptimalkan performa model, yaitu log transformation dan power transformation.

Kedua teknik ini digunakan sebelum proses modeling untuk mengatasi masalah skewness atau distribusi data yang tidak normal, terutama pada fitur yang memiliki nilai ekstrim, seperti jumlah klik (`sum_click`), jumlah upaya sebelumnya (`num_of_prev_attempts`), dan kredit yang dipelajari (`studied_credits`). Log transformation berguna untuk meratakan distribusi data dengan menekan efek outlier dan mengurangi ketimpangan dalam nilai fitur. Namun, karena log transformation tidak selalu menghasilkan distribusi normal yang sempurna, ditambahkan power transformation untuk lebih memperbaiki normalisasi distribusi fitur-fitur tertentu.

Dalam tahap modeling, digunakan standardization pada fitur-fitur untuk model Multilayer Perceptron (MLP), karena model ini

sensitif terhadap skala data. Standarisasi mengubah fitur menjadi skala yang sama, membantu meningkatkan performa MLP dalam menangani data yang distandarisasi. Sebaliknya, Random Forest tidak memerlukan standarisasi karena berbasis pohon keputusan yang tidak dipengaruhi oleh skala data. Distribusi data sebelum dan sesudah transformasi logaritmik dapat dilihat Tabel 3.

#### 4.8 Stratified Split

Pada tahap ini, dataset dibagi menjadi dua bagian, yaitu data latih (*training*) dan data uji (*testing*), dengan proporsi 80:20. Untuk memastikan distribusi yang seimbang antara kelas-kelas yang ada, terutama pada kolom target *learning\_style*, dilakukan stratified splitting. Sebelum melakukan pemisahan, dibuat sebuah kolom sementara *stratify\_col*, yang menggabungkan informasi dari kolom *disability* dan *learning\_style* untuk menciptakan kategori gabungan. Dengan menggunakan kolom ini sebagai dasar stratifikasi, data dibagi sehingga distribusi kelas pada data latih dan uji tetap seimbang. Pembagian ini dilakukan dengan menggunakan fungsi *train\_test\_split* dari pustaka *scikit-learn*, yang memastikan bahwa setiap subset memiliki representasi yang proporsional dari kelas-kelas yang ada, sehingga meningkatkan kualitas evaluasi model.

#### 4.9 Pembuatan Model

Pembuatan model menggunakan dua algoritma machine learning yang berbeda, yaitu Random Forest dan Multilayer Perceptron (MLP), untuk melakukan klasifikasi gaya belajar mahasiswa berdasarkan data yang telah diproses. Kedua algoritma ini dipilih karena kemampuannya yang berbeda dalam menangani masalah klasifikasi yang kompleks, terutama dengan data yang memiliki banyak fitur dan kemungkinan hubungan yang tidak linier (Sathe & Adamuthe, 2021).

Random Forest merupakan metode ensemble yang menggabungkan beberapa pohon keputusan untuk menghasilkan prediksi yang lebih stabil dan akurat. Algoritma ini unggul dalam menangani masalah dengan banyak variabel dan data yang bervariasi, serta dapat mengatasi overfitting dengan

efektif (Komputer & Informasi, 2020). Di sisi lain, Multilayer Perceptron (MLP) adalah salah satu jenis jaringan saraf tiruan yang dapat menangkap hubungan non-linier antar fitur, dengan menerapkan

struktur lapisan tersembunyi yang memungkinkan pemodelan pola yang lebih kompleks (Cotrim et al., 2022).

Tabel 3. Labeling dan Feature Transformasi Data

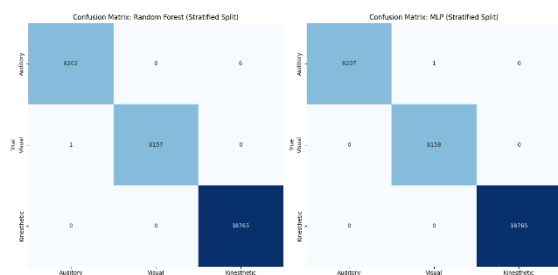
code_module	code_presentation	activity_type	sum_click	gender	region	highest_education
0	1	2	1.520945	0	6	1
0	1	6	0.925609	1	8	0
6	3	12	-0.753990	0	11	0
0	1	14	1.349039	1	3	0
5	3	11	-0.753990	0	7	4
imd_band	age_band	final_result	studied_credits	disability	num_of_prev_attempts	learning_style
1	1	1	1.57176849	0	-0.36788353	Auditory
2	1	2	1.57176849	1	-0.36788353	Visual
1	1	1	0.14600139	0	-0.36788353	Visual
4	1	1	1.57176849	0	-0.36788353	Kinesthetic
0	1	2	2.27739288	0	-0.36788353	Auditory

#### 4.10 Evaluasi Performa

Evaluasi model dilakukan untuk menilai kinerja algoritma dalam mendeteksi gaya belajar. Dalam penelitian ini, model Random Forest dan Multilayer Perceptron (MLP) dievaluasi menggunakan classification report dengan metrik precision, recall, F1-score, dan accuracy. Tabel 4 menunjukkan hasil model pada stratified split, yang menjaga keseimbangan distribusi kelas.

Table 4. Evaluasi Model Menggunakan Stratified Split

Evaluasi	Random Forest	MLP
Precision	99%	100%
Recall	99%	100%
F1-Score	99%	100%
Accuracy	99%	100%



Gambar 2. Confusion Matrix

Hasil evaluasi model menunjukkan bahwa algoritma Random Forest dan Multilayer Perceptron (MLP) memiliki performa yang sangat baik dalam mendeteksi gaya belajar. Berdasarkan metrik precision, recall, F1-score, dan accuracy, MLP mencapai kinerja sempurna dengan skor 100% pada seluruh metrik, menunjukkan bahwa model ini mampu memprediksi setiap sampel dalam data uji

dengan benar tanpa kesalahan. Sementara itu, model Random Forest juga menunjukkan hasil yang mendekati sempurna dengan skor 99% pada setiap metrik evaluasi, dengan hanya beberapa kesalahan kecil.

Pada Gambar 2. menunjukkan bahwa kesalahan prediksi Random Forest terjadi pada beberapa sampel Auditory yang salah diklasifikasikan sebagai Kinesthetic, dan satu sampel Kinesthetic yang diprediksi sebagai Auditory. Hasil ini mengindikasikan bahwa MLP lebih unggul dalam akurasi dibandingkan Random Forest untuk tugas klasifikasi gaya belajar pada dataset ini, menjadikannya pilihan yang efektif untuk prediksi yang sangat akurat.

#### 5. KESIMPULAN DAN SARAN

Penelitian ini berhasil mengklasifikasikan gaya belajar mahasiswa menggunakan algoritma Random Forest dan Multilayer Perceptron (MLP) dengan data dari Open University Learning Analytics Dataset (OULAD), yang diproses melalui tahapan preprocessing, pembuatan fitur, dan stratifikasi berdasarkan status disabilitas. Hasil evaluasi menunjukkan bahwa model MLP mencapai kinerja sempurna pada semua metrik, sedangkan Random Forest mendekati hasil sempurna, dengan sedikit kesalahan prediksi. Ini menunjukkan efektivitas kedua model dalam mendeteksi gaya belajar, terutama dengan mempertimbangkan fitur disabilitas. Berdasarkan hasil ini, disarankan untuk mengeksplorasi algoritma lain, seperti Support Vector Machine (SVM) atau 1D Convolutional Neural Network (CNN), guna membandingkan performanya dalam mendeteksi gaya belajar. Selain itu, penelitian lanjutan dapat menambahkan variabel baru, seperti kebiasaan belajar dan tingkat keterlibatan, untuk memperkaya wawasan tentang faktor-faktor yang memengaruhi gaya belajar

mahasiswa. Memperluas jumlah dan variasi responden serta menggunakan teknik penyeimbangan data lain juga disarankan untuk meningkatkan generalisasi dan ketangguhan model yang dihasilkan.

## DAFTAR PUSTAKA

- AKBAR, R. R., & NASUTION, E. S. 2021. Correlation learning style with grade point average fourth year medical student. *Open Access Macedonian Journal of Medical Sciences*, 9(T3), 358–361. <https://doi.org/10.3889/oamjms.2021.7728>
- CHILDS-KEAN, L., MLIS, M. E. E., & SMITH, M. D. 2020. Use of learning style frameworks in health science education. In *American Journal of Pharmaceutical Education* (Vol. 84, Issue 7, pp. 1–9). American Association of Colleges of Pharmacy. <https://doi.org/10.5688/ajpe7885>
- COTRIM, L. P., BARREIRA, R. A., SANTOS, I. H. F., GOMI, E. S., COSTA, A. H. R., & TANNURI, E. A. 2022. Neural Network Meta-Models for FPSO Motion Prediction from Environmental Data with Different Platform Loads. *IEEE Access*, 10, 86558–86577. <https://doi.org/10.1109/ACCESS.2022.3199009>
- DAHLIANA, L. H., JAELANI, A., & ROKHMAH, U. N. 2023. The impact of the visual, auditory, and kinesthetic model on motivation and learning outcomes of Islamic Elementary School students. *Journal of Integrated Elementary Education*, 3(2), 150–163. <https://doi.org/10.21580/jieed.v3i2.16867>
- DESIANI, A., DEWI, N. R., FAUZA, A. N., RACHMATULLAH, N., ARHAMI, M., & NAWAWI, M. 2021. Handling Missing Data Using Combination of Deletion Technique, Mean, Mode and Artificial Neural Network Imputation for Heart Disease Dataset. In *Science and Technology Indonesia* (Vol. 6, Issue 4). <https://doi.org/11.26554/sti.2221.6.4.333-312>
- DESTRIANI, R., SETIYADI, B., LAMPUNG, U., SUMANTRI BROJONEGORO, J., & LAMPUNG CORRESPONDENCE, B. 2021. *The comparative study in reading comprehension achievement on students with visual, auditory, and kinesthetic learning styles* (Vol. 10, Issue 2).
- DEVY, N. K., HALIM, A., SYUKRI, M., YUSRIZAL, Y., NUR, S., KHALDUN, I., & SAMINAN, S. 2022. Analysis of Understanding Physics Concepts in terms of Students' Learning Styles and Thinking Styles. *Jurnal Penelitian Pendidikan IPA*, 8(4), 2231–2237. <https://doi.org/10.29303/jppipa.v8i4.1926>
- GOMEDE, E., DE BARROS, R. M., & DE SOUZA MENDES, L. 2020. Use of deep multi-target prediction to identify learning styles. *Applied Sciences* (Switzerland), 10(5). <https://doi.org/10.3390/app10051756>
- HASSANZADEH, S., MOONAGHI, H. K., DERAKHSHAN, A., HOSSEINI, S. M., & TAGHIPOUR, A. 2019. Preferred learning styles among ophthalmology residents: An Iranian sample. *Journal of Ophthalmic and Vision Research*, 14(4), 483–490. <https://doi.org/10.18502/jovr.v14i4.5457>
- HIDAYATI, E. N., HIDAYAT, A., & UKIT, U. 2021. Relationship Between Learning Style With Student Learning Outcomes on Structure and Function of Plants. *Pedagonal: Jurnal Ilmiah Pendidikan*, 5(2), 94–105. <https://doi.org/10.33751/pedagonal.v5i2.3794>
- IRMANDA, H. N., SANTONI, M. M., & ASTIRATMA, R. (n.d.). *Cased Based Reasoning untuk Menentukan Gaya Belajar Mahasiswa*.
- JURCZYK, A., SZTURC, J., OTOP, I., OŚRÓDKA, K., & STRUZIK, P. 2020. Quality-based combination of multi-source precipitation data. *Remote Sensing*, 12(11). <https://doi.org/10.3390/rs12111709>
- KAUR, P., & KUMAR, R. (n.d.). Wireless Capsule Endoscopy Video Summarization using Transfer Learning and Random Forests. In *IJACSA) International Journal of Advanced Computer Science and Applications* (Vol. 14, Issue 9). [www.ijacsa.thesai.org](http://www.ijacsa.thesai.org)
- KOMPUTER, D. S., & INFORMASI, F. T. 2020. Perbandingan Kinerja Algoritma untuk Prediksi Penyakit Jantung dengan Teknik Data Mining. In *Journal of Applied Informatics and Computing (JAIC)* (Vol. 4, Issue 1). <http://jurnal.polibatam.ac.id/index.php/JAIC>
- MAULANA BAIHAQI, W., EKO SAPUTRO, R., & SETYO UTOMO, F. 2024. Novel Predictive Framework for Student Learning Styles Based on Felder-Silverman and Machine Learning Model. *Journal of Applied Data Sciences*, 5(4), 1726–1737. <https://doi.org/10.47738/jads.v5i4.408>
- NORINDER, U. 2023. Traditional Machine and Deep Learning for Predicting Toxicity Endpoints. *Molecules*, 28(1). <https://doi.org/10.3390/molecules28010217>
- NYOMAN PARDOMUAN, G., NYOMAN PARWATI, N., & AGUSTINI, K. 2020. Sistem Personalisasi E-Learning Berorientasi Felder Silverman Learning Style Model Pada Mata Pelajaran Teknik Pengambilan Gambar. In *Jurnal EDUTECH Universitas Pendidikan Ganesha* (Vol. 8, Issue 1). <https://ejournal.undiksha.ac.id/index.php/JEU>
- ÖZDEMİR, A., YAVUZ, U., & DAEL, F. A. 2019. Performance evaluation of different



- classification techniques using different datasets. *International Journal of Electrical and Computer Engineering*, 9(5), 3584–3590. <https://doi.org/10.11591/ijece.v9i5.pp3584-3590>
- RASHAD SAYED, A., HELMY KHAFAGY, M., ALI, M., & HUSSEIN MOHAMED, M. 2024. Predict student learning styles and suitable assessment methods using click stream. *Egyptian Informatics Journal*, 26. <https://doi.org/10.1016/j.eij.2024.100469>
- SATHE, M. T., & ADAMUTHE, A. C. 2021a. Comparative study of supervised algorithms for prediction of students' performance. *International Journal of Modern Education and Computer Science*, 13(1), 1–21. <https://doi.org/10.5815/ijmecs.2021.01.01>
- SATHE, M. T., & ADAMUTHE, A. C. 2021b. Comparative study of supervised algorithms for prediction of students' performance. *International Journal of Modern Education and Computer Science*, 13(1), 1–21. <https://doi.org/10.5815/ijmecs.2021.01.01>
- SINGH, V., PENCINA, M., EINSTEIN, A. J., LIANG, J. X., BERMAN, D. S., & SLOMKA, P. 2021. Impact of train/test sample regimen on performance estimate stability of machine learning in cardiovascular imaging. *Scientific Reports*, 11(1). <https://doi.org/10.1038/s41598-021-93651-5>
- SUBRAMANIAM, G., SAFARI BIN JAAFAR, M., & BINTI SAPRI, R. 2023. Exploring Learning Style and Learning Outcome of Students at Polytechnic Sultan Idris Shah. *International Journal of Academic Research in Progressive Education and Development*, 12(3). <https://doi.org/10.6007/ijarped/v12-i3/19287>
- SUSANTI, E., ISMET, & WIYONO, K. 2023. Analysis of Student Learning Style Profiles to Support Learning in class VII of SMP Negeri 01 Jaya Pura. *Jurnal Penelitian Pendidikan IPA*, 9(12), 10650–10655. <https://doi.org/10.29303/jppipa.v9i12.5858>
- SYARIFUDDIN, M., HADI WIJOYO, S., SUKMO WARDHONO, W., Studi Pendidikan Teknologi Informasi, P., Ilmu Komputer, F., Brawijaya Jl Veteran, U., Lowokwaru, K., Malang, K., & Timur, J. (n.d.). *Pengaruh Online Learning Experience Dan Gaya Belajar Terhadap Motivasi Belajar Peserta Didik SMKN 2 Malang Saat Pembelajaran Jarak Jauh (PJJ)*.
- VAN KESSEL, R., WALSH, S., RUIGROK, A. N. V., HOLT, R., YLIHERVA, A., KÄRNÄ, E., MOILANEN, I., HJÖRNE, E., JOHANSSON, S. T., SCHENDEL, D., PEDERSEN, L., JØRGENSEN, M., BRAYNE, C., BARON-COHEN, S., & ROMAN-URRESTARAZU, A. 2019. Autism and the right to education in the EU: Policy mapping and scoping review of Nordic countries Denmark, Finland, and Sweden. *Molecular Autism*, 10(1). <https://doi.org/10.1186/s13229-019-0290-4>
- ZHANG, J., QIU, F., WU, W., WANG, J., LI, R., GUAN, M., & HUANG, J. 2023. E-Learning Behavior Categories and Influencing Factors of STEM Courses: A Case Study of the Open University Learning Analysis Dataset (OULAD). *Sustainability (Switzerland)*, 15(10). <https://doi.org/10.3390/su15108235>

*Halaman ini sengaja dikosongkan*