

## IMPLEMENTASI MODEL HYBRID *MACHINE LEARNING* UNTUK PREDIKSI KELULUSAN PESERTA PELATIHAN KOMPUTER DI LKP MITRA PRESTASI

Ifdhea Finesti<sup>a\*</sup>, Maryam Ulfa Ningtyas<sup>b</sup>, Tia Apriyani<sup>c</sup>, Wardianto<sup>d</sup>, Thoha Firdaus<sup>e</sup>

<sup>a</sup> Sains dan Teknologi / Informatika, [afriresti123@gmail.com](mailto:afriresti123@gmail.com), Universitas Nurul Huda, Kab. OKU Timur, Sumatera Selatan

<sup>b</sup> Sains dan Teknologi / Informatika, [maryamulfaningtyas@gmail.com](mailto:maryamulfaningtyas@gmail.com), Universitas Nurul Huda, Kab. OKU Timur, Sumatera Selatan

<sup>c</sup> Sains dan Teknologi / Informatika, [apriyanitia068@gmail.com](mailto:apriyanitia068@gmail.com), Universitas Nurul Huda, Kab. OKU Timur, Sumatera Selatan

<sup>d</sup> Sains dan Teknologi / Informatika, [wardianto@unuha.ac.id](mailto:wardianto@unuha.ac.id), Universitas Nurul Huda, Kab. OKU Timur, Sumatera Selatan

<sup>e</sup> Fakultas Ilmu Pendidikan / Pendidikan Fisika, [thoha@unuha.ac.id](mailto:thoha@unuha.ac.id), Universitas Nurul Huda, Kab. OKU Timur, Sumatera Selatan

\*Korespondensi

### ABSTRACT

*This study examines the problem of the success rate of computer course program participants at LKP Mitra Prestasi which is still an obstacle for institutional administrators. By applying a hybrid model approach in machine learning, which combines Random Forest and Long Short-Term Memory (LSTM) algorithms through the assembly merging technique, this study aims to develop a predictive model that is able to estimate the success rate of participants and recognize variables that contribute to the achievement of learning outcomes. The probability outputs from both models are used as an input feature for the Logistic Regression meta-classifier that studies the optimal combination of predictions to produce final graduation decisions. This approach leverages the Random Forest's advantages in recognizing static features and LSTM in recognizing temporal patterns. Research data was obtained from 100 course participants which included demographic information, academic performance, as well as data on the order of attendance and assignment submission. The research findings show that the hybrid model with the stacking ensemble built succeeded in achieving 95% accuracy, 95% precision, 100% recall, and 97.44% F1 score, with the main variables affecting graduation being final exam scores, attendance percentage, and accuracy of assignment submission.*

**Keywords:** *Hybrid Machine learning, Random Forest, LSTM, Graduation Prediction*

### ABSTRAK

Penelitian ini mengkaji permasalahan tingkat keberhasilan peserta program kursus komputer di LKP Mitra Prestasi yang masih menjadi kendala bagi pengelola lembaga. Dengan menerapkan pendekatan model hybrid pada *machine learning* yang menggabungkan algoritma *Random Forest* dan Long Short-Term Memory (*LSTM*) melalui teknik *assembly merging*, penelitian ini bertujuan untuk mengembangkan model prediktif yang mampu mengestimasi tingkat keberhasilan peserta dan mengenali variabel-variabel yang berkontribusi terhadap pencapaian capaian pembelajaran. Output probabilitas dari kedua model tersebut digunakan sebagai fitur input bagi *meta-classifier Logistic Regression* yang mempelajari kombinasi prediksi yang optimal untuk menghasilkan keputusan kelulusan akhir. Pendekatan ini memanfaatkan keunggulan *Random Forest* dalam mengenali fitur statis dan *LSTM* dalam mengenali pola temporal. Data penelitian diperoleh dari 100 peserta kursus yang meliputi informasi demografi, kinerja akademik, serta data urutan kehadiran dan penyerahan tugas. Hasil penelitian menunjukkan bahwa model hybrid dengan ensemble stacking yang dibangun berhasil mencapai akurasi 95%, presisi 95%, *recall* 100%, dan skor *F1-score* 97,44%, dengan variabel utama yang mempengaruhi kelulusan adalah nilai ujian akhir, persentase kehadiran, dan akurasi penyerahan tugas.

**Kata Kunci:** Hybrid Machine learning, Random Forest, LSTM, Prediksi Kelulusan

## 1. PENDAHULUAN

Era digital telah meningkatkan permintaan akan keterampilan komputer di berbagai sektor. Perkembangan pesat teknologi informasi mendorong institusi pendidikan, termasuk Lembaga Kursus dan Pelatihan (LKP), untuk memanfaatkan sistem berbasis data dalam pengambilan keputusan. LKP Mitra Prestasi, sebagai penyedia program pelatihan komputer, memiliki peran strategis dalam membekali masyarakat dengan keterampilan digital yang diperlukan dunia kerja. Namun, tingkat kelulusan peserta pelatihan menjadi tantangan tersendiri bagi pengelola institusi. Kondisi ini terjadi karena pengaruh langsung persentase keberhasilan pada performa institusi, kredibilitas organisasi, dan tingkat keyakinan publik terhadap standar program pelatihan yang disediakan. Tingkat kelulusan yang rendah tidak hanya merugikan peserta secara individual, tetapi juga berdampak pada keberlanjutan operasional dan pengembangan program pelatihan secara keseluruhan.

Permasalahan utama yang dihadapi LKP Mitra Prestasi adalah kesulitan dalam melakukan identifikasi dini terhadap peserta yang berpotensi tidak lulus. Sistem prediksi yang ada saat ini masih mengandalkan pendekatan manual yang melibatkan penilaian subjektif dari instruktur berdasarkan pengamatan sehari-hari. Pendekatan ini memiliki beberapa kelemahan signifikan, antara lain ketidakkonsistenan dalam penilaian antar instruktur, keterlambatan dalam deteksi peserta berisiko, dan keterbatasan dalam menganalisis pola kompleks dari berbagai faktor yang mempengaruhi kelulusan. Keterlambatan dalam identifikasi peserta berisiko mengakibatkan intervensi yang diberikan seringkali tidak tepat waktu. Padahal, intervensi dini seperti bimbingan tambahan, penyesuaian metode pembelajaran, konseling akademik, atau modifikasi pendekatan pengajaran dapat meningkatkan peluang keberhasilan peserta secara signifikan. Oleh karena itu, diperlukan sistem prediksi yang objektif, akurat, dan mampu memberikan peringatan dini untuk memungkinkan tindakan korektif yang tepat waktu.

Tingkat kelulusan merupakan indikator kinerja utama bagi institusi pelatihan yang secara langsung memengaruhi reputasi dan keberlanjutan program. Identifikasi dini peserta yang berisiko tidak lulus menjadi hal yang esensial untuk memungkinkan implementasi strategi intervensi yang tepat sasaran. Lebih lanjut, pemahaman mendalam terhadap faktor-faktor yang mempengaruhi kelulusan peserta dapat membantu institusi mengembangkan strategi pembelajaran yang lebih efektif dan program intervensi yang disesuaikan dengan karakteristik spesifik setiap peserta [1,2]. Penerapan metode *machine learning* di bidang pendidikan telah menunjukkan capaian yang memuaskan dalam hasil yang positif dalam berbagai konteks penggunaan. Penelitian oleh [3] mengimplementasikan algoritma Decision Tree untuk prediksi kelulusan mahasiswa pada konteks pendidikan tinggi, mencapai tingkat akurasi sebesar 91% dengan menggunakan *dataset* yang terdiri dari 14 fitur utama termasuk demografi, performance akademik, dan karakteristik pembelajaran. Penelitian tersebut mendemonstrasikan bahwa fitur-fitur seperti usia, gender, program studi, dan indeks prestasi semester awal dapat menjadi prediktor signifikan untuk deteksi dini potensi kelulusan mahasiswa.

Penelitian sebelumnya juga menunjukkan keunggulan implementasi model hybrid dalam meningkatkan akurasi prediksi. Metode hybrid yang mengintegrasikan *Random Forest* dan *Logistic Regression* telah terbukti mampu meningkatkan akurasi prediksi hingga 84,48%, mengalami peningkatan sebesar 1,32% dibandingkan dengan implementasi model tunggal. Hasil ini mengindikasikan bahwa kombinasi algoritma dengan karakteristik komplementer dapat mengoptimalkan performa prediksi secara keseluruhan.

Namun demikian, penerapan model hybrid yang secara spesifik menggabungkan algoritma untuk menangani data terstruktur dan data sekuensial dalam konteks lembaga pelatihan vokasi masih sangat terbatas. Mayoritas penelitian yang ada berfokus pada implementasi model tunggal atau konteks pendidikan formal seperti universitas, sementara karakteristik unik dari lembaga pelatihan vokasi memerlukan pendekatan metodologis yang berbeda dan lebih spesifik [4]. Model *machine learning* tunggal seringkali memiliki limitasi dalam menangani data yang heterogen dan kompleks. *Random Forest*, meskipun menunjukkan keunggulan dalam klasifikasi data terstruktur dan memiliki resistensi terhadap outlier, memiliki limitasi dalam menangani data yang memiliki karakteristik temporal atau sekuensial [5]. Algoritma ini sangat efektif untuk menganalisis fitur-fitur statis seperti demografi peserta, nilai akademik, dan karakteristik pembelajaran, namun kurang optimal untuk memahami pola dinamis seperti konsistensi kehadiran atau trend performa selama periode pelatihan.

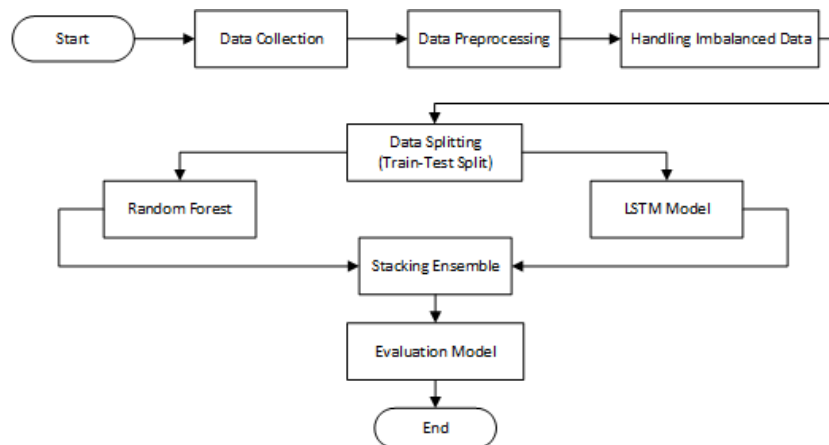
Sebaliknya, Long Short-Term Memory (*LSTM*), yang merupakan arsitektur khusus dari Recurrent Neural Network, memiliki kemampuan superior dalam mempelajari pola dari data sekuensial dan temporal. *LSTM* dapat menangkap dependensi jangka panjang dalam data *time series* dan mengidentifikasi pola kompleks yang berkembang sepanjang waktu [6]. Namun, algoritma ini menunjukkan efisiensi yang lebih rendah ketika diterapkan pada data kategorikal terstruktur yang tidak memiliki dimensi temporal. Pendekatan hybrid yang menggabungkan kekuatan kedua algoritma ini berpotensi mengatasi keterbatasan masing-masing algoritma dan meningkatkan akurasi prediksi secara signifikan. Dengan mengoptimalkan kemampuan *Random Forest* dalam menangani data terstruktur dan kemampuan *LSTM* dalam menganalisis pola sekuensial, model hybrid dapat memberikan prediksi yang lebih komprehensif dan akurat. Mengingat pertimbangan dan analisis gap penelitian tersebut, penelitian ini bertujuan mengembangkan dan mengimplementasikan model hybrid *machine learning* yang menggabungkan *Random Forest* dan *LSTM* untuk prediksi kelulusan peserta pelatihan di LKP Mitra Prestasi.

Kontribusi utama penelitian ini meliputi: (1) Merancang arsitektur model hybrid yang sesuai dengan karakteristik unik data peserta pelatihan vokasi; (2) Mengevaluasi dan membandingkan performa model hybrid dengan model tunggal untuk mendemonstrasikan keunggulan pendekatan yang diusulkan; (3) Memberikan solusi praktis dan implementabel untuk identifikasi dini peserta yang berisiko tidak lulus, yang dapat langsung diterapkan oleh institusi pelatihan. Berdasarkan tinjauan literatur dan analisis teoretis, penelitian ini mengajukan hipotesis bahwa implementasi model hybrid yang mengkombinasikan *Random Forest* dan *LSTM* akan menghasilkan akurasi prediksi yang superior dibandingkan dengan implementasi model tunggal dalam memprediksi kelulusan peserta pelatihan komputer di LKP Mitra Prestasi.

## 2. METODOLOGI PENELITIAN

Penelitian ini menggunakan pendekatan kuantitatif berbasis eksperimen komputasional untuk merancang dan mengevaluasi model prediksi kelulusan peserta pelatihan komputer di LKP Mitra Prestasi. Pendekatan yang diterapkan merupakan gabungan *machine learning*, yaitu pengombinasian dua algoritma berbeda yang masing-masing menangani karakteristik data yang berbeda: *Random Forest* untuk data terstruktur dan Long Short-Term Memory (*LSTM*) untuk data sekuensial. Hasil dari kedua model tersebut kemudian digabungkan menggunakan Teknik *stacking ensemble*, yang memungkinkan pembelajaran tingkat lanjut melalui model meta-klasifikasi.

Seluruh proses penelitian dilaksanakan secara bertahap dalam beberapa langkah utama, mulai dari pengumpulan data hingga evaluasi model. Tahapan-tahapan tersebut dapat dilihat secara visual pada Gambar 1.



Gambar 1. Flowchart Model Penelitian

Flowchart pada Gambar 1 menunjukkan tahapan penelitian yang dimulai dari pengumpulan data, *preprocessing*, pemisahan data, pembangunan model *Random Forest* dan *LSTM* secara terpisah, penggabungan hasil melalui *stacking ensemble*, dan diakhiri dengan evaluasi performa model. Setiap tahapan memiliki output spesifik yang menjadi input untuk tahapan selanjutnya, memastikan alur penelitian yang sistematis dan terstruktur.

### 2.1. Pengumpulan Data

Data yang digunakan dalam penelitian ini merupakan data sekunder yang diperoleh langsung dari sistem informasi akademik LKP Mitra Prestasi. *Dataset* yang dikumpulkan terdiri atas informasi lengkap dari 100 peserta pelatihan komputer yang telah menyelesaikan program pelatihan dalam periode satu tahun terakhir. Data ini dipilih karena representatif dan mencakup variasi yang cukup untuk membangun model prediksi yang robust.

*Dataset* terdiri atas dua jenis data utama dengan karakteristik yang berbeda [7]. Tipe pertama adalah data terstruktur yang berisi aneka ragam atribut demografis dan akademis peserta. Komponen ini mencakup data umur peserta, jenis kelamin, background pendidikan terakhir, tingkat engagement dalam kegiatan kelompok, rata-rata nilai assignment selama pelatihan, nilai ujian tengah semester dan ujian akhir, persentase attendance keseluruhan, dan rasio punctuality dalam pengumpulan tugas. Jenis data yang kedua adalah data sekuensial yang berupa catatan kehadiran peserta selama 20 sesi pelatihan yang direpresentasikan dalam format vektor biner, di mana nilai 1 menunjukkan kehadiran peserta dan nilai 0 menunjukkan ketidakhadiran. Data sekuensial ini penting untuk memahami pola konsistensi dan komitmen peserta selama proses pelatihan berlangsung. *Dataset* kemudian disimpan dalam format .csv dan diolah menggunakan pustaka *Pandas* di *Python*.

### 2.2. Preprocessing Data

*Preprocessing* merupakan tahap fundamental dalam memastikan data memenuhi standar dan format yang dibutuhkan oleh algoritma *machine learning* [7,8]. Proses ini dilakukan secara bertahap melalui pembersihan data yang mencakup pemeriksaan nilai kosong, identifikasi data duplikat, dan deteksi kesalahan input. Mengingat data bersumber dari LKP Mitra Prestasi sebagai institusi resmi, kualitas data menunjukkan standar yang relatif baik tanpa ditemukan nilai yang memerlukan imputasi.

Implementasi pembersihan data dilaksanakan melalui lima tahapan sistematis menggunakan pustaka *Pandas* dalam *Python*, yang menyediakan fungsi-fungsi komprehensif untuk manipulasi dan analisis data terstruktur. Tahap pertama melakukan pemeriksaan nilai kosong (*missing values*) melalui fungsi `df.isnull().sum()` untuk mengidentifikasi keberadaan data yang hilang dalam *dataset* [8]. Hasil evaluasi menunjukkan bahwa tidak ditemukan nilai kosong pada seluruh variabel, yang mengindikasikan integritas data yang optimal dari sumber LKP Mitra Prestasi.

Selanjutnya, verifikasi duplikasi data diimplementasikan melalui fungsi `df.duplicated()` untuk mendeteksi keberadaan record yang identik. Prosedur ini memiliki peran krusial dalam mencegah redundansi data yang berpotensi mengganggu performa dan akurasi model *machine learning*. Tahapan berikutnya melibatkan pemeriksaan kesalahan input menggunakan fungsi `df.describe()` yang menyajikan ringkasan statistik deskriptif untuk mengidentifikasi keberadaan anomali atau *outlier* dalam *dataset*. Analisis statistik mengindikasikan bahwa data memiliki distribusi yang konsisten tanpa ditemukan anomali signifikan, mencerminkan standarisasi dalam proses pengumpulan data dari institusi resmi.

Pengolahan variabel kategorikal kemudian dilakukan untuk memastikan kompatibilitas dengan algoritma *machine learning*. Fitur jenis kelamin dikonversi menggunakan teknik *Label Encoding* melalui *LabelEncoder*, di mana nilai kategorik "L" dan "P" ditransformasi ke bentuk numerik ( $L = 0$ ,  $P = 1$ ). Sementara itu, variabel dengan multiple kategori seperti pendidikan terakhir dan partisipasi kelompok diproses menggunakan *One-Hot Encoding*, menghasilkan kolom biner terpisah untuk setiap kategori unik.

Tahapan akhir melibatkan standarisasi fitur numerik yang meliputi variabel usia, persentase kehadiran, nilai tugas, nilai ujian tengah semester, nilai ujian akhir semester, serta rasio ketepatan pengumpulan tugas. Proses normalisasi diimplementasikan menggunakan *MinMaxScaler* dari pustaka *scikit-learn*, yang mentransformasi seluruh nilai ke dalam rentang  $[0, 1]$ . Tujuan normalisasi ini adalah untuk mengoptimalkan proses pelatihan model dan mencegah bias yang disebabkan oleh dominasi fitur dengan skala nilai yang besar.

### 2.3. Handling Imbalanced data

Class imbalance merupakan tantangan signifikan dalam masalah klasifikasi dunia nyata yang dapat sangat memengaruhi kinerja model dan utilitas praktis [9,10]. Distribusi *label* pada data menunjukkan ketidakseimbangan antara peserta yang lulus dan tidak lulus. Untuk mengatasi hal ini, digunakan teknik *oversampling* pada kelas minoritas dengan pendekatan *RandomOverSampler* dari pustaka *imblearn*.

Tujuannya adalah agar model memiliki performa yang lebih baik dalam mengenali kedua kelas secara proporsional dan tidak berat sebelah.

#### 2.4. Pemisahan Data

Sesudah data dinormalisasi dan diseimbangkan, partisi data dilakukan dengan memanfaatkan fungsi *train\_test\_split* dari library *scikit-learn* dengan proporsi 80:20. Parameter stratify diterapkan pada *label* untuk menjaga proporsi kelas tetap seimbang antara data pelatihan dan data evaluasi [11]. Data pelatihan digunakan untuk membangun model, sedangkan data evaluasi dimanfaatkan untuk mengukur kinerja model.

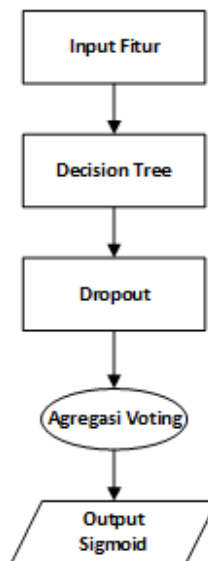
Tabel 1. Data setelah *Preprocessing*

Skala Usia	SMA	SMK	Diploma	Sarjana	Nilai Kehadiran	Persentase Kehadiran	Status kelulusan
0.42	1	0	0	0	0.85	0.78	Lulus
0.65	0	1	0	0	0.93	0.92	Tidak Lulus
...	...	...	...	...	...	...	...

#### 2.5. Random Forest Model

Model *Random Forest* digunakan untuk menangani data terstruktur karena kemampuannya dalam mengolah fitur kategorikal dan numerik secara bersamaan, serta keunggulannya dalam memberikan informasi *feature importance* dan ketahanannya terhadap *overfitting*.

Model ini dibangun menggunakan pustaka *scikit-learn* dengan parameter utama berupa jumlah pohon (*n\_estimators*) sebanyak 100 dan kedalaman maksimum (*max\_depth*) sebanyak 10. Fitur-fitur hasil *preprocessing* seperti usia, nilai-nilai akademik, partisipasi kelompok, dan persentase kehadiran digunakan sebagai input. Model dilatih pada data latih dan menghasilkan keluaran berupa probabilitas kelulusan peserta, bukan *label* langsung, sehingga dapat dimanfaatkan pada tahap integrasi model berikutnya.

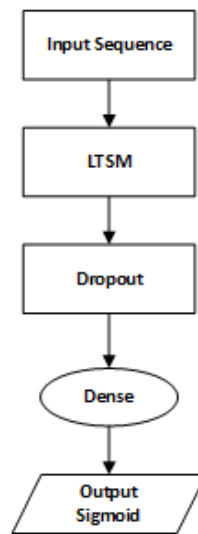


Gambar 2. Skema Arsitektur *Random Forest*

#### 2.6. LSTM Model

Model Long Short-Term Memory (*LSTM*) diterapkan untuk mengolah data sekuensial yang mencakup pola kehadiran peserta selama pelatihan. *LSTM* dipilih karena kemampuannya dalam mendeteksi dependensi temporal dan pola dalam urutan data yang panjang. Data sekuensial dikonversi ke dalam format tiga dimensi (*samples, timesteps, features*) yang sesuai dengan struktur input pada jaringan *LSTM*. Model dibangun dengan satu layer *LSTM* berisi 64 unit, dilanjutkan dengan *dropout* sebesar 0.3 untuk mengurangi *overfitting*, dan satu *dense layer* dengan aktivasi *sigmoid* untuk prediksi biner. Model ini dikompilasi

menggunakan fungsi loss *binary\_crossentropy* dan *optimizer Adam*, serta dilatih selama 30 *epoch* dengan *batch size* 16 Hasil dari model *LSTM* berupa probabilitas kelulusan, yang kemudian digunakan dalam tahap *stacking*.

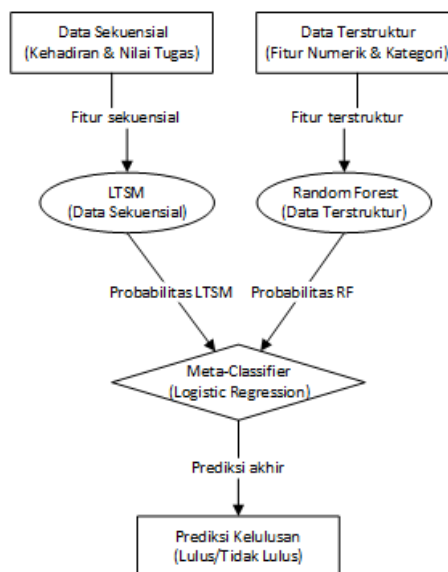


Gambar 3. Arsitektur Model *LSTM*

### 2.7. Hybrid Model with Stacking Ensemble

Output dari model *Random Forest* dan *LSTM* digabungkan menggunakan pendekatan *stacking ensemble*. Nilai probabilitas dari kedua model dijadikan fitur baru yang digunakan sebagai input untuk *model meta-classifier*, yaitu *LogisticRegression*.

*Meta-model* ini dilatih untuk mempelajari kombinasi terbaik dari prediksi kedua model dasar guna menghasilkan keputusan akhir mengenai kelulusan peserta. Teknik *stacking* ini memungkinkan model untuk menggabungkan keunggulan dari dua model dasar, di mana *Random Forest* unggul dalam mengenali fitur statis dan *LSTM* unggul dalam mengenali pola temporal.



Gambar 4. Diagram Arsitektur Hybrid Model (*Stacking Ensemble*)

### 2.8. Evaluasi Model

Evaluasi model dilakukan untuk mengukur seberapa baik performa dalam memprediksi kelulusan peserta. Evaluasi dilakukan menggunakan lima metrik utama, yaitu *accuracy*, *precision*, *recall*, *F1-score*, dan *ROC-AUC* [12]. Akurasi menunjukkan proporsi prediksi yang benar dibandingkan dengan keseluruhan prediksi yang dilakukan. Presisi menilai sejauh mana model tepat dalam mengidentifikasi peserta yang benar-benar lulus, sedangkan *recall* mengevaluasi kemampuan model dalam menangkap seluruh peserta yang seharusnya dinyatakan lulus. *F1-score* digunakan sebagai ukuran keseimbangan antara *precision* dan *recall*, sementara *ROC-AUC* menggambarkan performa klasifikasi model secara umum terhadap nilai ambang prediksi. Evaluasi dilakukan pada data uji yang tidak pernah digunakan dalam proses pelatihan model, untuk memastikan bahwa hasil prediksi mencerminkan kemampuan generalisasi model terhadap data baru. Seluruh implementasi dilakukan menggunakan *Google Colaboratory*.

### 3. HASIL DAN PEMBAHASAN

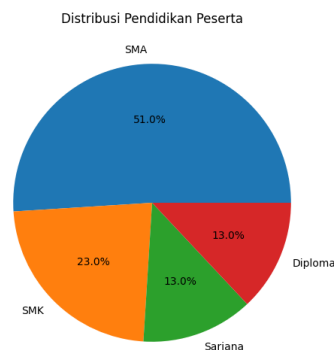
Berdasarkan eksplorasi awal terhadap data terstruktur peserta pelatihan, diperoleh statistik deskriptif sebagaimana disajikan pada Tabel 2.

Tabel 2. Statistik Deskriptif Dataset

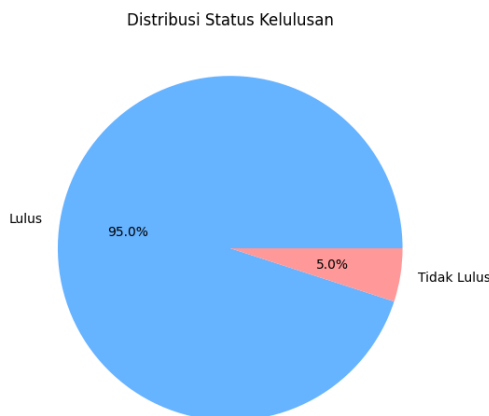
Statistik	Usia	Persentase Kehadiran	Rata-rata Assignment	Nilai UTS	Nilai Ujian Akhir	ASTM
Mean	23.85	91.36	84.22	83.82	87.19	0.8556
Std	3.55	4.67	6.69	8.63	8.09	0.0867
Min	18	81	67	60	68	0.7
25%	21	88	80	78.75	83	0.78
50%	24	91	84.5	85	88.5	0.85
75%	27	94	88.25	89	92	0.93
Max	29	100	98	98	100	1

Tabel 2 menyajikan karakteristik statistik deskriptif yang mengungkap dinamika pembelajaran dalam konteks pendidikan vokasi. Dataset menampilkan heterogenitas demografis dengan rentang usia 18-29 tahun ( $M=23,85$ ), mencerminkan dominasi kelompok usia produktif yang aktif dalam transformasi digital, sementara tingkat partisipasi mencapai 91,36% dengan konsistensi kehadiran yang menunjukkan engagement tinggi peserta terhadap kurikulum pelatihan. Dimensi akademik memperlihatkan progresivitas pembelajaran melalui peningkatan performa dari nilai assignment (84,22) ke ujian tengah semester (83,82) hingga evaluasi akhir (87,19), mengindikasikan efektivitas metodologi pembelajaran yang diterapkan, sedangkan parameter ASTM dengan nilai 0,8556 dan deviasi standar minimal (0,0867) mengkonfirmasi kultur disiplin dan manajemen waktu yang terbangun dalam lingkungan pembelajaran, di mana keseluruhan distribusi data menunjukkan normalitas tanpa outlier signifikan sehingga memberikan fondasi empiris yang robust untuk pengembangan model prediktif berbasis machine learning dalam mengantisipasi tingkat keberhasilan peserta pelatihan vokasi.

Distribusi latar belakang pendidikan peserta program pelatihan di LKP Mitra Prestasi menunjukkan komposisi SMA (51%), SMK (23%), Sarjana (13%), dan Diploma (13%). Dari 100 peserta program pelatihan di LKP Mitra Prestasi, 95% berhasil lulus dan 5% tidak lulus dari program pelatihan.



Gambar 5. Distribusi Latar Belakang Pendidikan Peserta  
 Dari 100 peserta program pelatihan di LKP Mitra Prestasi, 95% berhasil lulus dan 5% tidak lulus dari program pelatihan komputer, seperti yang ditunjukkan pada Gambar 6.



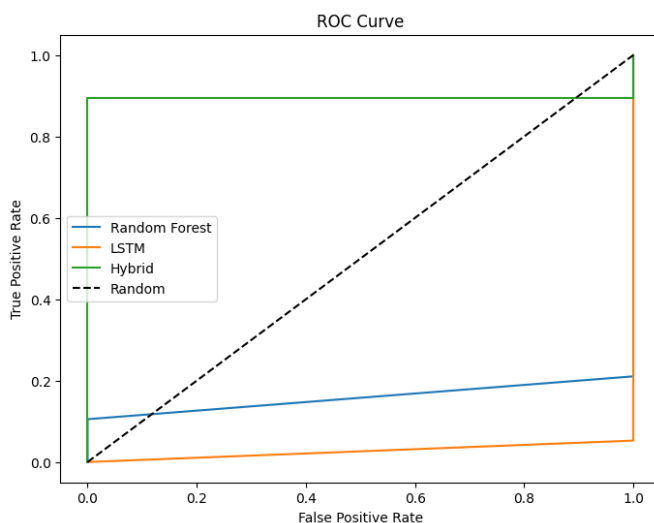
Gambar 6. Distribusi Status Kelulusan Peserta

**3.1 Performa Hybrid**

Hasil evaluasi performa model hybrid dan perbandingannya dengan model Tunggal (*Random Forest* dan *LSTM*) ditunjukkan pada Tabel 3.

Tabel 3. Perbandingan Performa Model Hybrid dengan Model Tunggal

Metric	Random Forest	LSTM	Hybrid Model
Accuracy	90.00%	85.00%	95.00%
Precision	92.31%	83.33%	95.00%
Recall	92.31%	95.24%	100.00%
F1-score	92.31%	88.89%	97.44%
ROC-AUC	58.46%	47.62%	50.00%



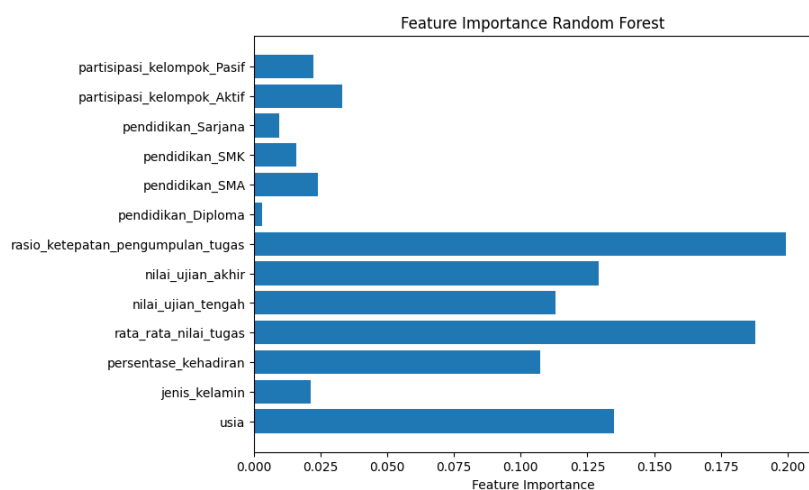
Gambar 7. ROC Curve Model *Random Forest*, *LSTM*, dan Hybrid

Performa model hybrid menunjukkan peningkatan signifikan dibandingkan model tunggal di seluruh metrik utama, mencapai akurasi 95%, *precision* 95%, *recall* 100%, dan *F1-score* 97.44%. Model hybrid berhasil mengidentifikasi seluruh peserta yang tidak lulus (*recall* 100%), yang sangat penting untuk sistem peringatan dini. Namun, nilai *ROC-AUC* yang relatif rendah (50%) di seluruh model mengindikasikan area yang perlu diperbaiki pada penelitian selanjutnya, terutama dalam hal kemampuan diskriminatif model.

Hasil dari analisis tersebut adalah rasio ketepatan pengumpulan tugas memiliki importance yang cukup tinggi (18.7%), mengindikasikan bahwa kedisiplinan dan time management merupakan faktor kritis dalam menentukan keberhasilan. Sebaliknya, faktor demografis seperti usia dan jenis kelamin memiliki pengaruh yang minimal, menunjukkan bahwa program pelatihan bersifat inklusif dan dapat diakses oleh berbagai kalangan.

### 3.2 Analisis Feature Importance

Evaluasi *feature importance* pada model *Random Forest* menghasilkan pemahaman mendalam mengenai faktor-faktor dominan yang mempengaruhi tingkat kelulusan peserta LKP Mitra Prestasi. Hasil analisis ini divisualisasikan dalam Gambar 7.

Gambar 8. Feature importance *Random Forest*

### 3.3 Pola Kehadiran dan Implikasinya

Analisis pola kehadiran dari model *LSTM* mengungkapkan beberapa temuan penting. Dari analisis pola kehadiran, ditemukan bahwa ketidakhadiran berturut-turut lebih dari tiga sesi pelatihan berkorelasi tinggi dengan risiko tidak lulus. Pola ketidakhadiran yang tersebar (non-consecutive) sepanjang program memiliki dampak yang lebih kecil dibandingkan ketidakhadiran yang terkonsentrasi di awal atau akhir program. Selain itu, peserta yang menghadiri lebih dari 90% dari sepuluh sesi pertama memiliki probabilitas kelulusan yang lebih tinggi. Temuan ini menekankan pentingnya pemantauan kehadiran yang konsisten dan menunjukkan bahwa intervensi dini untuk peserta yang menunjukkan pola kehadiran yang mengkhawatirkan dapat menjadi krusial dalam meningkatkan tingkat kelulusan.

Hasil penelitian menunjukkan bahwa pendekatan hybrid *machine learning* yang menggabungkan *Random Forest* dan *LSTM* berhasil meningkatkan akurasi prediksi kelulusan peserta pelatihan. Peningkatan performa ini disebabkan oleh kemampuan model untuk menangani baik data terstruktur maupun data sekuensial secara optimal. *Random Forest* terbukti handal dalam mengolah informasi statis berupa pencapaian akademik dan karakteristik demografis, sedangkan *LSTM* lebih efektif dalam memahami pola yang berubah seiring waktu dari catatan kehadiran. Interpretasi *feature importance* memberikan insight praktis bagi pengelola LKP Mitra Prestasi. Dominasi faktor akademik dan kedisiplinan dalam menentukan kelulusan menunjukkan bahwa fokus pada peningkatan kualitas pembelajaran dan sistem monitoring

kehadiran dapat memberikan dampak signifikan pada tingkat kelulusan. Temuan bahwa latar belakang pendidikan tidak memberikan pengaruh signifikan mengindikasikan inklusivitas program pelatihan.

#### 4. KESIMPULAN DAN SARAN

Model hybrid *machine learning* yang mengintegrasikan algoritma *Random Forest* dan *LSTM* berhasil mencapai tujuan penelitian dengan performa optimal, yaitu akurasi 95%, *precision* 95%, *recall* 100%, dan *F1-score* 97,44%. Penelitian ini membuktikan keunggulan pendekatan hybrid dibandingkan model tunggal dalam konteks prediksi kelulusan peserta pelatihan vokasi. Temuan utama penelitian mengindikasikan bahwa faktor akademik, khususnya nilai ujian akhir dan rata-rata tugas, serta aspek kedisiplinan yang meliputi tingkat kehadiran dan ketepatan pengumpulan tugas merupakan prediktor dominan dalam menentukan kelulusan. Sebaliknya, faktor demografis tidak menunjukkan pengaruh signifikan terhadap hasil prediksi. Implementasi komponen *LSTM* terbukti efektif dalam mengekstraksi pola kehadiran sekuensial yang memberikan informasi prediktif berharga, terutama untuk deteksi dini peserta berisiko tidak lulus.

Keterbatasan penelitian meliputi ukuran dataset yang terbatas (100 sampel), distribusi target yang tidak seimbang dengan rasio 95:5, dan nilai *ROC-AUC* yang masih memerlukan perbaikan. Untuk penelitian selanjutnya, disarankan menggunakan dataset dengan ukuran sampel yang lebih besar dan distribusi yang lebih seimbang, mengintegrasikan variabel sekuensial tambahan ke dalam arsitektur model *LSTM* untuk meningkatkan kemampuan prediksi temporal, serta menerapkan teknik *resampling* atau *cost-sensitive learning* untuk mengatasi permasalahan ketidakseimbangan kelas dan mengoptimalkan performa model secara keseluruhan.

#### Ucapan Terima Kasih

Penulis mengucapkan terima kasih yang tulus kepada LKP Mitra Prestasi atas kesediaannya menyediakan akses data yang menjadi dasar penelitian ini, serta kepada Program Studi Informatika Universitas Nurul Huda yang telah memberikan dukungan berupa bimbingan akademis selama proses penelitian berlangsung.

#### DAFTAR PUSTAKA

- [1] F. Indra Kusuma Budiyanto, I. Hermadi, and M. K. D. Hardhienata, "Prediksi Performa Akademik Mahasiswa untuk Kelulusan Predikat Cum Laude dengan Pendekatan Machine Learning," *J. Ilmu Komput. dan Agri-Informatika*, vol. 11, no. 1, pp. 39–49, 2024, [Online]. <https://doi.org/10.29244/jika.11.1.39-49>
- [2] Satrio Junaidi, R. Valicia Anggela, and D. Kariman, "Klasifikasi Metode Data Mining untuk Prediksi Kelulusan Tepat Waktu Mahasiswa dengan Algoritma Naïve Bayes, Random Forest, Support Vector Machine (SVM) dan Artificial Neural Network (ANN)," *J. Appl. Comput. Sci. Technol.*, vol. 5, no. 1, pp. 109–119, 2024, [Online]. <https://doi.org/10.52158/jacost.v5i1.489>
- [3] B. N. Aisyah and I. Gunawan, "Algoritma Decision Tree Application of Machine Learning to Predict Student Graduation Using the Decision Tree Algorithm," pp. 1–6, 2024.
- [4] N. S. S. Silmi Ath Thahirah Al Azhima, D. Darmawan, N. Fahmi Arief Hakim, I. Kustiawan, M. Al Qibtiya, "Hybrid Machine Learning Model Untuk Memprediksi Penyakit Jantung Denga Metode Logistic Regression Dan Random Forest," *J. Teknol. Terpadu*, vol. 8, no. 1, pp. 40–46, 2022, [Online]. <https://doi.org/10.54914/jtt.v8i1.539>
- [5] M. R. P. Putra, "Model Pembelajaran Mesin Hibridisasi RFXGB Untuk Prediksi Drop Out Siswa MOOC," *G-Tech J. Teknol. Terap.*, vol. 8, no. 1, pp. 2485–2494, 2024, [Online]. <https://doi.org/10.70609/gtech.v8i4.5179>
- [6] D. Sudiana, M. Rizkinia, and N. Tristan, "Development of Long Short-Term Memory (LSTM) Bayesian Network Method for Predicting Wind Power Potential in a Wind Power Plant in Indonesia," *17th Int. Conf. Qual. Res. QIR 2021 Int. Symp. Electr. Comput. Eng.*, no. October 2021, pp. 85–89, 2021, [Online]. <https://doi.org/10.1109/QIR54354.2021.9716204>
- [7] M. A. Setiawan, Kusriani, and A. D. Hartono, "Menggunakan Metode Machine Learning Untuk Memprediksi Nilai Mahasiswa Dengan Model Prediksi Multiclass," vol. 10, no. 1, pp. 190–204, 2025, [Online]. <https://doi.org/10.30591/jpit.v10i1.8334>
- [8] S. Fitriana, Rinianty, R. Laila, S. A. Pratama, and C. A. Lamasitudju, "Prediksi Siswa Putus Sekolah dan Keberhasilan Akademik Menggunakan Machine Learning," vol. 14, no. 2, pp. 10207–10220, 2025, [Online]. <https://doi.org/10.30591/jpit.v10i1.8334>

- [9] D. Kurniasari, L. Humairoso, Warsono, and Notiragayu, "Implementation of Random Forest Method for Customer Churn Classification," vol. 17, no. 1, pp. 187–202, 2025, [Online]. <https://jsi.ejournal.unsri.ac.id/index.php/jsi/article/view/202>
- [10] G. Airlangga, "Anemia Classification Using Hybrid Machine Learning Models : A Comparative Study of Ensemble Techniques on CBC Data," vol. 5, no. 4, pp. 1108–1117, 2024, [Online]. <https://doi.org/10.47065/josyc.v5i4.5848>
- [11] G. A. Syafarina and Z. Zaenuddin, "Implementasi Framework Streamlit Sebagai Prediksi Harga Jual Rumah Dengan Linear Regresi," *Metik J.*, vol. 7, no. 2, pp. 121–125, 2023, [Online]. <https://doi.org/10.47002/metik.v7i2.608>
- [12] M. Ali, M. Shahroz, M. F. Mushtaq, S. Alfarhood, M. Safran, and I. Ashraf, "Hybrid Machine Learning Model for Efficient Botnet Attack Detection in IoT Environment," *IEEE Access*, vol. 12, no. January, pp. 40682–40699, 2024, [Online]. <https://doi.org/10.1109/ACCESS.2024.3376400>