

Perancangan Sistem Prediksi Pendapatan BRT Menggunakan Penerapan *Hybrid Random Forest* dan LSTM

Diah Pita Loka^{1,*}, Lena Magdalena², Mesi Febima³

^{1,2,3}Teknologi Informasi, Sistem Informasi, Universitas Catur Insan Cendekia, Cirebon, Indonesia
Email: ¹pitalokadiah18@gmail.com, ²lena.magdalena@cic.ac.id, ³mesi.febima@cic.ac.id

Abstrak

Penelitian ini berfokus pada penerapan pendekatan hybrid machine learning yang menggabungkan algoritma *Random Forest* dan *Long Short-Term Memory* (LSTM) untuk memprediksi pendapatan *Bus Rapid Transit* (BRT) berdasarkan pola perilaku pelanggan. Dengan semakin pentingnya manajemen transportasi umum, prediksi pendapatan yang akurat telah menjadi krusial untuk perencanaan operasional dan pembuatan kebijakan. Penelitian ini mengatasi tantangan yang dihadapi oleh Dinas Perhubungan Kota Cirebon dalam mengelola operasi BRT tanpa sistem prediksi pendapatan yang memadai untuk mendukung kebijakan tarif transportasi publik yang berkelanjutan. Menggunakan data operasional dari Januari hingga Desember 2024, penelitian ini mengusulkan pendekatan *hybrid* baru di mana *Random Forest* digunakan untuk pemilihan fitur dan analisis kepentingan, sementara LSTM menangani pengenalan pola temporal untuk prediksi pendapatan. Model *hybrid* memproses berbagai variabel termasuk data pelanggan (dewasa dan pelajar), data operasional (rute, jadwal), dan faktor eksternal (hari libur nasional, hari libur sekolah). Algoritma *Random Forest* berhasil mengidentifikasi 5 fitur kunci dari 15 fitur awal, dengan total penumpang (korelasi: 0,996) dan 3 penumpang stasiun teratas (korelasi: 0,996) menunjukkan pentingnya tertinggi. Model LSTM mencapai akurasi prediksi dengan MAE sebesar 8,90%, MSE sebesar 0,0134, RMSE sebesar 11,59%, dan R² sebesar 75,97%. Pendekatan hybrid menunjukkan peningkatan yang signifikan dalam akurasi prediksi pendapatan dibandingkan dengan metode tradisional, memberikan wawasan berharga bagi otoritas transportasi di Kota Cirebon untuk merumuskan kebijakan tarif transportasi publik yang berkelanjutan.

Kata Kunci: *Machine Learning, Random Forest, LSTM, Prediksi, Transportasi Umum*

Abstract

This study focuses on implementing a hybrid machine learning approach combining Random Forest and Long Short-Term Memory (LSTM) algorithms to predict Bus Rapid Transit (BRT) revenue based on customer behavior patterns. With the increasing importance of public transportation management, accurate revenue prediction has become crucial for operational planning and policy-making. The research addresses the challenge faced by the Department of Transportation Cirebon City in managing BRT operations without adequate revenue prediction systems for supporting sustainable public transport tariff policies. Using operational data from January to December 2024, this study proposes a novel hybrid approach where Random Forest is utilized for feature selection and importance analysis, while LSTM handles temporal pattern recognition for revenue prediction. The hybrid model processes various variables including customer data (adult and student), operational data (routes, schedules), and external factors (national holidays, school holidays). The Random Forest algorithm successfully identified 5 key features from 15 initial features, with total customers (correlation: 0.996) and top 3 station customers (correlation: 0.996) showing the highest importance. The LSTM model achieved prediction accuracy with MAE of 8.90%, MSE of 0.0134, RMSE of 11.59%, and R² of 75.97%. The hybrid approach demonstrates significant improvement in revenue prediction accuracy compared to traditional methods, providing valuable insights for transportation authorities in Cirebon City to formulate sustainable public transport tariff policies.

Keywords: Machine Learning, Random Forest, LSTM, Revenue Prediction, Public Transportation

1. PENDAHULUAN

Evolusi teknologi kecerdasan buatan dalam bidang *Artificial Intelligence* (AI) merupakan teknologi yang memungkinkan komputer untuk

berpikir dengan kecerdasan yang setara atau lebih baik daripada kecerdasan manusia, yang dikenal sebagai kecerdasan buatan (AI), dan *Machine Learning* adalah cabang dari kecerdasan buatan yang fokus pada pembelajaran dari data,

mengembangkan sistem yang dapat belajar "secara mandiri" tanpa perlu diprogram ulang oleh manusia [1]. Salah satu aplikasi dari teknologi ini adalah pada sistem prediksi berbasis data yang dapat membantu meningkatkan efisiensi dan efektivitas operasional. Penerapan *machine learning* dapat digunakan dalam transportasi publik, salah satunya adalah *Bus Rapid Transit* (BRT).

Bus Rapid Transit (BRT) adalah sistem transportasi publik yang menggunakan bus dalam operasionalnya, umumnya dioperasikan sebagai fasilitas mobilitas tinggi untuk masyarakat. Mengurangi kemacetan adalah salah satu tujuan dari pengadaan bus ini dan menekan penggunaan transportasi pribadi yang dapat menimbulkan berbagai masalah seperti polusi, kemacetan, dan sebagainya. Bus beroperasi sesuai dengan rute yang telah ditentukan dan tidak dapat menurunkan penumpang sembarangan, melainkan hanya di halte yang disediakan[2]. Tarif yang dikenakan juga sangat murah karena dengan tarif yang terjangkau dan layanan serta fasilitas yang disediakan oleh pemerintah, BRT dipercaya menjadi salah satu hal yang dapat menarik minat masyarakat untuk menggunakan moda transportasi publik.

Di Kota Cirebon, sistem BRT diuji coba oleh Pemerintah Kota Cirebon pada tanggal 12 April 2021. Saat pertama kali diluncurkan, BRT gratis selama periode dari 19 April 2019 hingga 25 April 2021. Operasional koridor 1 jalur utara dimulai dari persimpangan berlian-bypass-kedawung-pilang-krucuk-kesenden-pegambiran dengan tarif Rp. 5.000 dan untuk pelajar Rp. 3.500, tidak hanya pembayaran non-tunai seperti menggunakan OVO, Gopay, Shopee Pay dan sejenisnya melalui QRIS, naik Trans Cirebon juga dapat membayar tunai yang tentu saja dilengkapi dengan wifi dan fasilitas yang baik. Pada 19 Juli 2023, tepat pada ulang tahun Kota Cirebon yang ke-654, Gubernur Jawa Barat meresmikan operasi BRT di koridor 2 menggunakan armada 3 bus. Sistem BRT dikelola oleh Dinas Perhubungan (Dishub) Kota Cirebon bekerja sama dengan PD Pembangunan dan PT Bima Inti Global sebagai operator teknis[3].

Sebagai regulator transportasi publik, Dinas Perhubungan Kota Cirebon menghadapi tantangan yang kompleks dalam merumuskan kebijakan tarif yang berkelanjutan. Pendapatan operasional BRT adalah aspek krusial yang sangat bergantung pada berbagai faktor, baik internal maupun eksternal. Beberapa variabel utama yang mempengaruhi pendapatan termasuk jumlah penumpang harian, tarif yang berlaku, rute perjalanan, dan jam operasional. Selain itu, faktor eksternal seperti hari libur nasional dan libur sekolah juga berperan dalam menentukan fluktuasi jumlah penumpang dan, secara langsung, pendapatan operasional. Namun, menganalisis hubungan antara variabel-variabel ini bukanlah tugas yang mudah, mengingat pola pergerakan penumpang cenderung dinamis dan

dipengaruhi oleh banyak faktor yang sulit diprediksi secara konvensional[3].

Saat ini, Dinas Perhubungan Kota Cirebon masih menghadapi berbagai tantangan dalam pengambilan keputusan strategis terkait kebijakan tarif BRT karena belum memiliki sistem prediksi pendapatan yang akurat. Salah satu masalah utama yang dihadapi adalah fluktuasi jumlah penumpang yang signifikan dari waktu ke waktu, yang dipengaruhi oleh berbagai faktor seperti perubahan musim, kebijakan pemerintah, tren mobilitas masyarakat, dan kondisi ekonomi secara umum. Ketidaktersediaan sistem prediksi pendapatan ini menjadi masalah ketika Dinas Perhubungan Kota Cirebon berencana untuk menerapkan kebijakan tarif gratis atau disubsidi pada periode tertentu tanpa simulasi dampak keuangan yang akurat terhadap keberlanjutan operasional BRT.

Ketidakpastian ini menyebabkan kesulitan dalam perencanaan kebijakan yang matang, karena perkiraan pendapatan yang tidak akurat dapat berdampak pada alokasi anggaran transportasi publik yang tidak efisien. Kondisi ini juga berpotensi menciptakan ketidakseimbangan antara kepentingan publik untuk mendapatkan transportasi yang terjangkau dan kebutuhan untuk menjaga kelangsungan keuangan operator dalam memastikan keberlanjutan layanan BRT dalam jangka panjang. Penelitian ini akan membangun sistem prediksi pendapatan untuk menganalisis pola historis jumlah penumpang dan faktor eksternal yang mempengaruhinya, sehingga dapat menghasilkan perkiraan pendapatan yang lebih akurat untuk mendukung perumusan kebijakan tarif transportasi publik yang berkelanjutan.

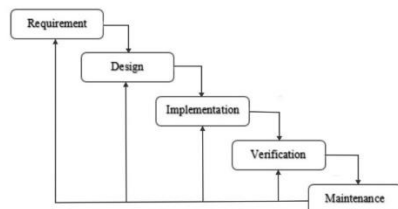
Penelitian ini mengusulkan pendekatan *hybrid* yang menggabungkan algoritma *Random Forest* dan LSTM untuk memprediksi pendapatan BRT berdasarkan pola penumpang. *Random Forest* berfungsi sebagai mekanisme pemilihan fitur untuk mengidentifikasi variabel yang paling penting yang mempengaruhi pendapatan, sementara LSTM menangkap ketergantungan temporal dan tren jangka panjang dalam data perjalanan BRT. Dengan menggabungkan *Random Forest* untuk ekstraksi fitur dan identifikasi pola utama serta LSTM untuk menangkap tren dan ketergantungan jangka panjang dalam data perjalanan BRT, sistem prediksi yang dihasilkan diharapkan lebih akurat, adaptif terhadap perubahan pola penumpang, dan mampu memberikan perkiraan pendapatan yang lebih optimal untuk mendukung pengambilan keputusan strategis dalam manajemen keuangan.

Penelitian ini bertujuan untuk memberikan Dinas Perhubungan Kota Cirebon sebuah sistem prediksi yang akurat yang dapat memberikan informasi yang tepat kepada pemerintah untuk merumuskan kebijakan tarif transportasi publik yang berkelanjutan yang tidak merugikan operasi BRT, menerapkan metode *Hybrid Random Forest*

dan LSTM dalam sistem prediksi pendapatan BRT, serta mengukur dan mengevaluasi tingkat akurasi metode *Random Forest* dan LSTM dalam prediksi pendapatan BRT sebagai dasar ilmiah untuk mengembangkan sistem prediksi yang dapat diandalkan dalam mendukung keputusan strategis.

2. METODOLOGI PENELITIAN

Metode yang digunakan dalam penelitian ini adalah *Waterfall*, sebuah pendekatan pengembangan perangkat lunak yang bersifat sistematis dan bertahap. Metode ini dianalogikan seperti air terjun, di mana setiap tahapan harus diselesaikan sebelum melanjutkan ke tahap berikutnya. Kelebihan metode ini terletak pada alur kerja yang terstruktur dan terdokumentasi dengan baik, sehingga memudahkan dalam pengawasan perkembangan proyek dan pemeliharaan di masa mendatang. Metode ini juga banyak diadopsi dalam pengembangan aplikasi dengan tahapan yang jelas, seperti pada sistem pakar dan sistem informasi [4].

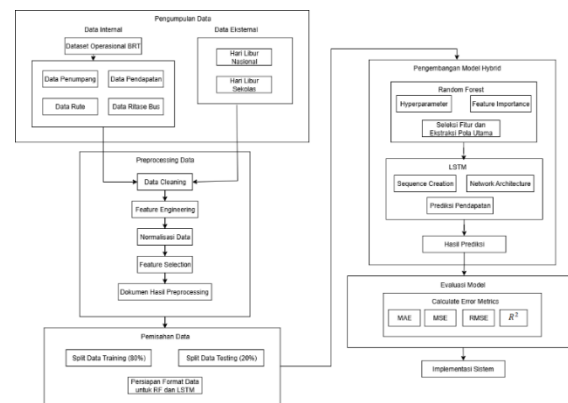


Gambar 1. Langkah Metode *Waterfall*

Pada gambar 1. Langkah Metode *Waterfall* di atas merupakan ilustrasi tahapan pengembangan perangkat lunak yang digunakan dalam penelitian ini [11]. Secara garis besar, metode *Waterfall* mempunyai langkah-langkah sebagai berikut. Tahap pertama adalah analisis kebutuhan, yang dilakukan untuk mengumpulkan informasi terkait proses prediksi pendapatan BRT dan variabel-variabel yang memengaruhinya, seperti data jumlah penumpang dewasa, penumpang pelajar, ritase harian, data halte, serta faktor eksternal seperti hari libur nasional dan libur sekolah yang diperoleh dari Dinas Perhubungan Kota Cirebon. Selanjutnya, pada tahap perancangan, dibuat rancangan sistem prediksi pendapatan berbasis web yang mencakup arsitektur sistem, desain antarmuka pengguna, serta proses *preprocessing data* seperti *data cleaning*, *feature engineering*, *normalisasi*, dan *feature selection* menggunakan algoritma *Random Forest*. Pada tahap implementasi, dibangun model *Hybrid Random Forest* dan LSTM, di mana *Random Forest* digunakan untuk menentukan fitur-fitur penting yang berpengaruh terhadap pendapatan dan LSTM dimanfaatkan untuk memprediksi pendapatan berdasarkan pola data historis penumpang, serta dikembangkan aplikasi berbasis *web* menggunakan *database MySQL* untuk menampilkan hasil prediksi dalam bentuk tabel dan grafik. Selanjutnya dilakukan tahap verifikasi untuk

menguji kinerja model menggunakan metrik evaluasi MAE, MSE, RMSE, dan R^2 . Tahap terakhir adalah pemeliharaan, yaitu pembaruan data, penyesuaian model terhadap pola terbaru, serta perbaikan sistem secara berkala agar aplikasi dapat berjalan optimal dan memberikan hasil prediksi yang akurat dalam mendukung kebijakan transportasi publik di Kota Cirebon.

Penelitian ini mengusulkan pendekatan *hybrid machine learning* dengan menggabungkan teknik prediksi berbasis fitur dan berbasis urutan untuk memprediksi pendapatan dari operasi *Bus Rapid Transit* (BRT). Algoritma *Random Forest* digunakan untuk mengidentifikasi dan memilih fitur-fitur yang paling berpengaruh dari data historis BRT, sementara model *Long Short-Term Memory* (LSTM) diterapkan untuk melakukan peramalan deret waktu berdasarkan fitur-fitur yang terpilih. Model *hybrid* ini dirancang untuk menangkap pola statis maupun ketergantungan temporal dalam data penumpang dan pendapatan.



Gambar 2. Alur Penelitian

Gambar 2. Kerangka penelitian untuk memprediksi pendapatan BRT menggunakan model hybrid *Random Forest* dan LSTM. Proses dimulai dengan pengumpulan data dari sumber internal (operasional BRT) dan eksternal (hari libur). Preprocessing mencakup pembersihan data, rekayasa fitur, normalisasi, dan pemilihan fitur. Dataset kemudian dibagi menjadi set pelatihan dan pengujian. *Random Forest* digunakan untuk mengidentifikasi fitur-fitur penting, yang kemudian digunakan sebagai input untuk model LSTM untuk melakukan prediksi pendapatan secara berurutan. Kinerja model dievaluasi menggunakan metrik MAE, MSE, RMSE, dan R^2 .

2.1 Dataset

Dataset ini mencakup data operasional BRT dari Dinas Perhubungan Kota Cirebon yang mencakup periode Januari hingga Desember 2024, dengan total 366 entri data. Data yang termasuk di dalamnya adalah:

- Data Penumpang

Data penumpang terdiri dari catatan harian yang dikategorikan menjadi penumpang dewasa, penumpang pelajar, dan total jumlah penumpang yang diangkut oleh sistem BRT.

b. Data Pendapatan

Data pendapatan dihitung berdasarkan struktur tarif tetap, di mana penumpang dewasa dikenakan biaya Rp 5.000 dan penumpang pelajar Rp 3.500, yang menghasilkan angka pendapatan total harian.

c. Data Operasional

Data operasional mencakup informasi detail mengenai rute BRT dan lokasi halte bus yang mempengaruhi aliran penumpang dan cakupan layanan.

d. Data Eksternal

Faktor eksternal seperti hari libur nasional dan libur sekolah dimasukkan, karena faktor ini secara signifikan mempengaruhi pola perjalanan dan fluktuasi volume penumpang.

2.2 Preprocessing Data

Preprocessing data merupakan tahap dasar dalam mengembangkan model *machine learning* yang berkualitas. Dalam penelitian ini, *preprocessing* data dilakukan secara sistematis melalui beberapa sub-tahap yang saling terkait untuk memastikan data operasional BRT siap digunakan dalam model prediksi[5].

a. Data Cleaning

Pembersihan data melibatkan proses untuk mengidentifikasi dan memperbaiki kesalahan, inkonsistensi, dan ketidakakuratan dalam dataset untuk memastikan kualitas data sebelum dilakukan pemodelan.

b. Feature Engineering

Feature engineering adalah proses pembuatan fitur baru dari data yang ada untuk meningkatkan kinerja model *machine learning*. Ringkasan dari *feature engineering* berdasarkan kategori dapat dilihat pada Tabel 1.

Tabel 1. Hasil Feature Engineering

Kategori	Fitur Awal	Fitur Baru	Total Fitur	Metode Utama
Temporal	1	11	12	Ekstraksi datetime, transformasi siklik
Penumpang	2	7	9	Rasio, proporsi, agregasi
Ritase	6	8	14	Proporsi, rata-rata
Bus stop	44	10	54	Kategorisasi, agregasi berdasarkan lalu lintas
Intreaksi	0	2	2	Perkalian antar fitur

Lag	0	2	2	Fitur pergeseran waktu
Efisiensi	0	3	3	Rasio operasional

Tabel 1. menunjukkan hasil *feature engineering* yang menghasilkan 96 fitur total dari 53 fitur awal, dengan kategori bus stop memiliki kontribusi terbesar (54 fitur) melalui kategorisasi dan agregasi berdasarkan lalu lintas, diikuti ritase (14 fitur), temporal (12 fitur), penumpang (9 fitur), serta fitur baru interaksi, lag, dan efisiensi (masing-masing 2-3 fitur) menggunakan berbagai metode transformasi untuk meningkatkan kinerja model prediksi.

c. Normalization

Normalisasi dilakukan untuk memastikan semua fitur numerik berada pada skala yang sama, menggunakan persamaan Min-Max Normalization.

d. Feature Selection

Pemilihan fitur menggunakan metode Univariate Selection dengan koefisien korelasi Pearson.

2.3 Split Data

Dataset dibagi menjadi subset pelatihan dan pengujian menggunakan rasio pemisahan temporal 80%-20% untuk memastikan evaluasi model yang andal sambil mempertahankan sifat kronologis dari data operasional BRT. Subset pelatihan berisi 293 hari data dari 1 Januari hingga 19 Oktober 2024, memberikan cakupan musiman yang komprehensif dan pola operasional yang beragam untuk pembelajaran model. Subset pengujian terdiri dari 73 hari data dari 20 Oktober hingga 31 Desember 2024, yang mewakili data masa depan yang belum terlihat untuk evaluasi kinerja[6].

2.4 Random Forest

Dalam konteks penelitian ini, *Random Forest* digunakan untuk mengidentifikasi fitur-fitur penting yang mempengaruhi pendapatan *Bus Rapid Transit* (BRT) di Kota Cirebon. Fitur-fitur tersebut mencakup berbagai variabel seperti jumlah penumpang harian, tarif yang dikenakan, rute perjalanan, waktu operasional, serta faktor eksternal seperti hari libur nasional dan libur sekolah. Dengan menggunakan *Random Forest*, model dapat menilai kontribusi relatif dari masing-masing fitur terhadap prediksi pendapatan[7].

a. Menghitung *impurity reduction* untuk setiap split

Evaluasi kualitas split dilakukan dengan menghitung *variance reduction* sebagai *proxy* untuk *impurity reduction*, yang merupakan

metrik fundamental dalam algoritma CART (*Classification and Regression Trees*), dengan rumus (1).

$$\Delta I(t, s) = \text{var}(t) - \frac{n_{tL}}{n_t} \text{var}(t_L) - \frac{n_{tR}}{n_t} \text{var}(t_R) \quad (1)$$

b. Normalisasi *feature importance*

Feature importance dinormalisasi untuk memudahkan interpretasi dan perbandingan relatif antar fitur, dengan rumus normalisasi *importance* (2).

$$\text{Norm_Imp_Decrease}_i = \frac{\text{Imp_Decrease}_i}{\sum_{j=1}^m \text{Imp_Decrease}_j} \quad (2)$$

c. Menghitung *feature importance* rata-rata

Sesuai dengan prinsip fundamentah *Random Forest* untuk agregasi hasil dari *multiple decision trees*, *feature importance* dari ketiga *tree* diagregasi menggunakan *weighted average* (3).

$$\text{Avg_Importance}_i = \frac{1}{T} \sum_{j=1}^T \text{Importance}_{i,j} \quad (3)$$

d. *Threshgold selection* untuk *feature importance*

Seleksi fitur final dilakukan dengan pendekatan *elbow point* pada grafik *feature importance*, dilandasi oleh prinsip *pareto efficiency*. Berikut adalah rumus *formula cumulative importance* (4).

$$\text{Cum_Importance}_k = \sum_{i=1}^k \text{Sorted_avg_importance}_i \quad (4)$$

2.5 LSTM

Dalam konteks penelitian ini, LSTM digunakan untuk memprediksi pendapatan BRT berdasarkan urutan data historis dari faktor-faktor yang mempengaruhi operasional BRT, seperti jumlah penumpang, tarif, waktu perjalanan, dan faktor eksternal seperti libur nasional dan libur sekolah. Data diolah dalam bentuk urutan dengan jendela waktu 7 hari, di mana setiap fitur dari hari-hari sebelumnya digunakan untuk memprediksi pendapatan pada hari berikutnya[7].

a. Konversi fitur terpilih ke nilai *sequence*

Fitur-fitur yang telah terseleksi dikonversi ke dalam bentuk *sequence* dengan *time window* 1-7, yang merepresentasikan data 7 hari sebelumnya. Konversi ini penting untuk memungkinkan LSTM mengenali pola temporal dalam data. Secara matematis, untuk setiap fitur X dan timestep t , *sequence* dibentuk sesuai rumus (5).

$$X_{\text{seq},t} = [X_{t-7}, X_{t-6}, X_{t-5}, X_{t-4}, X_{t-3}, X_{t-2}, X_{t-1}] \quad (5)$$

Dimana $X_{\text{seq},t}$ = Sequence data untuk fitur X pada timestep t , X_{t-i} = Nilai fitur X pada i hari sebelum timestep t . Penggunaan *sequence* 1-7 memberikan konteks temporal yang cukup untuk model untuk mempelajari pola pendapatan BRT berdasarkan data historis.

b. Penentuan parameter bobot (w)

Untuk menangkap pengaruh temporal yang berbeda dari setiap timestep dalam *sequence*, bobot (W) ditetapkan untuk setiap komponen

sequence 1-7. Bobot ini mencerminkan tingkat kepentingan relatif dari data pada hari-hari sebelumnya terhadap prediksi hari ini. Proses penentuan bobot dilakukan dengan rumus (6).

$$W_i = W_1 \times (1 - \alpha)^{i-1} \quad (6)$$

c. Prediksi

Nilai prediksi pendapatan BRT dihitung dengan mengalikan nilai *sequence* yang telah diolah dengan bobot yang sesuai, kemudian menjumlahkannya dengan bias (7).

$$\hat{y}_t = \sum_{i=1}^7 \sum_{j=1}^P W_{\text{seq},i} \times W_{j,t-i} + b \quad (7)$$

Dimana \hat{y}_t = Nilai prediksi pendapatan pada timestep t , $W_{\text{seq},i}$ = Bobot untuk *sequence* ke- i , $W_{j,t-i}$ = Nilai fitur ke- j pada timestep $t-i$, P = Jumlah fitur terpilih, dan b = Bias.

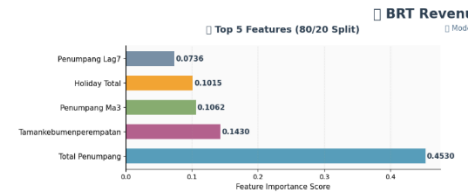
2.6 Evaluasi Model

Untuk mengukur kinerja model prediksi pendapatan BRT, empat metrik evaluasi utama digunakan, yaitu *Mean Absolute Error* (MAE), *Mean Squared Error* (MSE), *Root Mean Squared Error* (RMSE), dan *Coefficient of Determination* (R^2)[8]. Masing-masing metrik ini memberikan wawasan yang berbeda mengenai akurasi dan ketepatan model dalam memprediksi pendapatan BRT.

3. HASIL DAN PEMBAHASAN

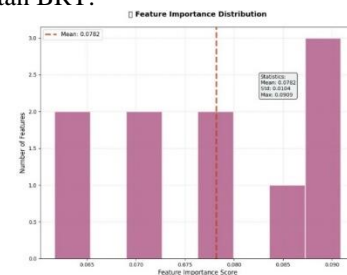
3.1. Hasil *Random Forest*

Algoritma *Random Forest* diterapkan dengan penyetelan hyperparameter untuk mencapai kinerja yang optimal[9]. Model dilatih menggunakan 80% dari dataset dan divalidasi pada 20% sisanya.



Gambar 3. Peringkat *Feature Importance*

Pada gambar 3 diatas menunjukkan peringkat pentingnya fitur yang diperoleh dari algoritma *Random Forest*. Fitur yang paling berpengaruh adalah Total Penumpang, Penumpang Top3 Halte, dan Penumpang Rit 3, yang memiliki skor pentingnya tertinggi dalam memprediksi pendapatan BRT.

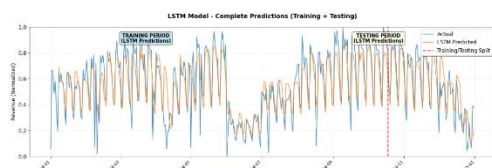


Gambar 4. Distribusi *Feature Importance*

Pada gambar 4 diatas menampilkan distribusi skor pentingnya fitur, dengan skor rata-rata sebesar 0,0782. Fitur dengan skor di atas rata-rata ini dianggap signifikan dan diprioritaskan untuk input model.

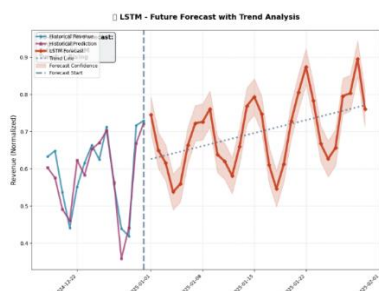
3.2 Hasil *LSTM*

Long Short-Term Memory (LSTM) diterapkan untuk menangkap ketergantungan temporal dalam data berurutan. Arsitektur model dioptimalkan melalui penyetelan hyperparameter yang ekstensif[10].



Gambar 5. Nilai Prediksi dan Aktual

Pada gambar 5 menunjukkan hasil prediksi lengkap dari model LSTM untuk periode pelatihan dan pengujian. Garis biru mewakili pendapatan ternormalisasi yang aktual, sementara garis oranye menunjukkan nilai prediksi. Garis putus-putus merah menandakan pemisahan antara data pelatihan dan pengujian. Model ini sangat mendekati tren aktual, menunjukkan kemampuannya dalam menangkap pola temporal pada pendapatan BRT.



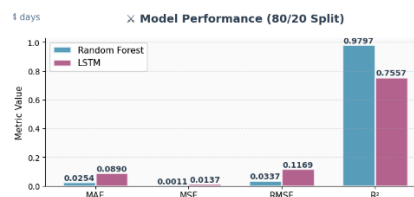
Gambar 6. Prediksi Masa Depan

Pada gambar 6 menyajikan perkiraan pendapatan masa depan menggunakan model LSTM beserta analisis tren. Garis merah menunjukkan nilai pendapatan yang diprediksi, sementara area yang diarsir mewakili interval kepercayaan. Garis tren putus-putus biru menunjukkan tren kenaikan bertahap pada pendapatan. Visualisasi ini menyoroti kemampuan model untuk memproyeksikan nilai masa depan sambil mempertimbangkan ketidakpastian.

3.3 Hasil Evaluasi Model

Evaluasi terhadap model *Random Forest* dan LSTM, berdasarkan pembagian data 80/20 (80% pelatihan, 20% pengujian), memberikan penilaian komprehensif terhadap kinerja mereka dalam memprediksi pendapatan BRT untuk PT. Bima Inti Global Cirebon. Metrik yang digunakan untuk

evaluasi mencakup *Mean Absolute Error* (MAE), *Mean Squared Error* (MSE), *Root Mean Squared Error* (RMSE), dan R^2 Score, yang dihitung pada set data pengujian. Hasil evaluasi dijelaskan di bawah ini pada Gambar 7.

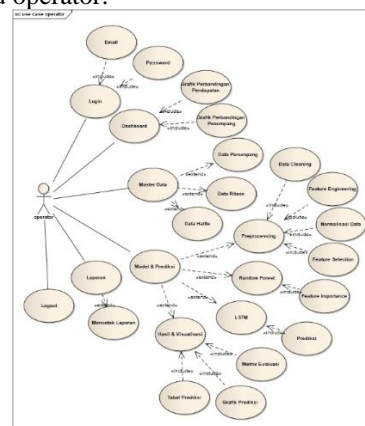


Gambar 7. Evaluasi Model

Pada gambar 7. menampilkan grafik batang perbandingan yang menggambarkan kinerja model *Random Forest* dan *LSTM* berdasarkan empat metrik evaluasi utama: *Mean Absolute Error* (MAE), *Mean Squared Error* (MSE), *Root Mean Squared Error* (RMSE), dan R^2 Score. Grafik ini menggunakan batang biru untuk *Random Forest* dan batang merah muda untuk *LSTM*.

3.4 Hasil Perancangan Sistem

Dalam perancangan aplikasi sistem prediksi pendapatan pada BRT, langkah awal yang dilakukan adalah dengan membuat sebuah diagram use case. Tujuan dari diagram use case ini adalah untuk menggambarkan secara umum proses yang terjadi dalam aplikasi dan bagaimana interaksi antara pengguna dengan sistem[2]. Pada sistem prediksi pendapatan yang akan dibangun, pengguna sistem yaitu operator.



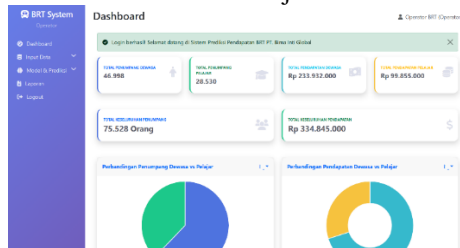
Gambar 8. *Use Case* Operator

Pada gambar 8. *Use case* operator menunjukkan interaksi operator dengan sistem yang mencakup login, pengelolaan data, pemodelan prediksi menggunakan *Random Forest* dan *LSTM*, serta pembuatan laporan dan visualisasi hasil prediksi dalam bentuk tabel dan grafik. Semua fitur saling terhubung melalui hubungan *include* dan *extend*.

3.5 Hasil Implementasi Sistem

Dalam penerapan hasil analisis, dilakukan perancangan sistem yang mencakup pembuatan

antarmuka pengguna dan pengelolaan data menggunakan basis data MySQL untuk mendukung aplikasi sistem prediksi pendapatan berbasis web. Sistem ini dirancang untuk memprediksi pendapatan BRT dengan memanfaatkan algoritma *Random Forest* dan *LSTM*. Proses ini melibatkan pengumpulan data penumpang dan ritase, preprocessing data, serta penerapan model prediksi untuk menghasilkan proyeksi pendapatan yang akurat. Berikut adalah hasil penerapan sistem prediksi pendapatan yang dikembangkan, dengan berbagai fitur untuk mendukung perencanaan operasional dan evaluasi kinerja.



Gambar 9. Halaman *Dashboard*

Pada gambar 9. menunjukkan tampilan utama dashboard sistem prediksi pendapatan BRT, yang menampilkan data total penumpang dewasa, pelajar, dan keseluruhan penumpang, serta pendapatan yang dihasilkan. Grafik perbandingan antara penumpang dewasa dan pelajar juga tersedia untuk analisis lebih lanjut.

The table shows passenger data with columns: No, Tanggal, Status, Penumpang Dewasa, Penumpang Pelajar, Total Penumpang, Pendapatan Dewasa, Pendapatan Pelajar, Total Pendapatan, Hari Libur, and Aksi. It lists 6 entries of passengers.

Gambar 10. Halaman *Kelola Data Penumpang*

Gambar 10. menunjukkan tampilan data penumpang dengan informasi jumlah penumpang dewasa, pelajar, dan total pendapatan per hari. Pengguna juga dapat melihat data berdasarkan status hari libur dan menambahkan atau mengimpor data.

The table shows route data with columns: No, Tanggal, Ri 1, Ri 2, Ri 3, Ri 4, Ri 5, Ri 6, Total Ritase, and Aksi. It lists 6 entries of routes.

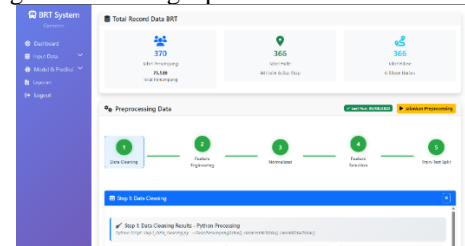
Gambar 11. Halaman *Kelola Data Ritase*

Gambar 11. menunjukkan tabel data ritase BRT, termasuk informasi jumlah ritase per hari dan rincian untuk setiap ritase. Operator dapat mengelola data ritase dan memantau kinerja armada dengan metrik harian dan ritase tertinggi.

The table shows station data with columns: Tanggal, Halte, Total Penumpang, Halte Terpadu, Halte Terpisah, and Aksi. It lists 10 entries of stations.

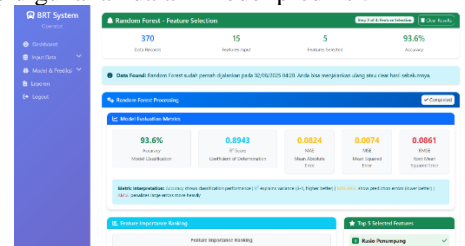
Gambar 12. Halaman *Kelola Data Halte*

Pada gambar 12. menampilkan tabel data halte dengan informasi terkait tanggal, jumlah penumpang, dan lokasi halte. Operator dapat melihat data halte yang terpadat dan kosong, serta mengedit atau menghapus data sesuai kebutuhan.



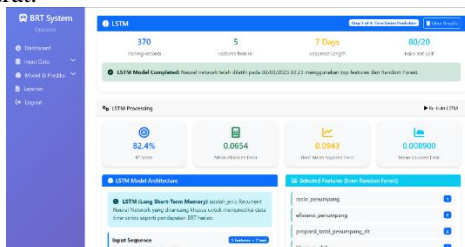
Gamabr 13. Halaman *Preprocessing*

Pada gambar 13. menunjukkan status preprocessing data, yang meliputi tahap pembersihan data, rekayasa fitur, normalisasi, dan pemisahan data untuk pelatihan dan pengujian model. Proses ini penting untuk memastikan data siap digunakan dalam model prediksi.



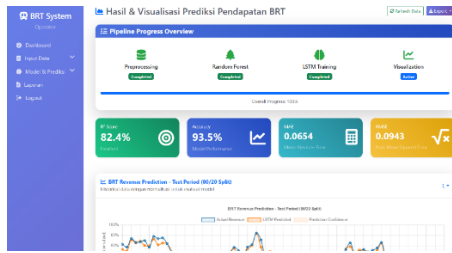
Gambar 14. Halaman *Random Forest*

Pada gambar 14. menampilkan hasil dari proses seleksi fitur menggunakan *Random Forest*, termasuk akurasi model, R^2 score, dan metrik evaluasi lainnya. Fitur yang dipilih akan digunakan untuk membangun model prediksi yang lebih akurat.



Gambar 15. Halaman *LSTM*

Pada gambar 15. menampilkan hasil pemrosesan dan evaluasi model *LSTM*, termasuk akurasi, *Mean Absolute Error* (MAE), dan *Root Mean Squared Error* (RMSE) untuk prediksi pendapatan BRT. *LSTM* digunakan untuk memprediksi pendapatan berdasarkan data historis dengan input sequence tertentu.



Gambar 16. Halaman Hasil & Visualisasi

Pada gambar 16. menampilkan visualisasi hasil prediksi pendapatan BRT dengan menggunakan model *Random Forest* dan *LSTM*. Grafik ini menunjukkan perbandingan antara pendapatan aktual dan prediksi, serta tingkat kepercayaan prediksi.

4 KESIMPULAN

Penelitian ini berhasil mengimplementasikan pendekatan *hybrid Random Forest* dan *LSTM* untuk prediksi pendapatan BRT berdasarkan pola penumpang. Algoritma *Random Forest* secara efektif mengidentifikasi lima fitur kunci dari data operasional, dengan metrik volume penumpang menunjukkan korelasi terkuat dengan pendapatan. Model *LSTM* menunjukkan kemampuannya dalam menangkap pola temporal dengan rata-rata kesalahan prediksi sebesar 6,54%. Pendekatan *hybrid* ini memberikan keuntungan signifikan dibandingkan metode prediksi tradisional, menawarkan manfaat praktis dalam mendukung pengambilan keputusan berbasis data untuk penyusunan kebijakan transportasi publik yang berkelanjutan. Penelitian mendatang sebaiknya mengintegrasikan faktor eksternal tambahan dan mengeksplorasi arsitektur deep learning yang lebih canggih untuk meningkatkan akurasi prediksi.

5 REFERENCES

- [1] M. Valipour, H. Khoshkam, S. M. Bateni, C. Jun, and S. S. Band, "Hybrid machine learning and deep learning models for multi-step-ahead daily reference evapotranspiration forecasting in different climate regions across the contiguous United States East North Central region," *Agric. Water Manag.*, vol. 283, no. April, p. 108311, 2023, doi: 10.1016/j.agwat.2023.108311.
- [2] H. Wijaya, S. Feliyanti, L. Magdalena, and R. Ilyasa, "Perancangan Sistem Informasi Rute Bus Rapid Transit (Brt) Di Kota Cirebon Berbasis Website (Studi Kasus Pt. Bima Inti Global)," *J. Digit.*, vol. 11, no. 2, p. 200, 2021, doi: 10.51920/jd.v11i2.205.
- [3] S. ANWAR, "Analisis Bukti Empiris Tentang Dampak Ekuitas Sosial Ekonomi Operasional Bus Rapid Transit (Brt) Di Negara Berkembang," *Eduka J. Pendidikan, Hukum, dan Bisnis*, vol. 4, no. 2, 2020, doi: 10.32493/eduka.v4i2.3851.
- [4] P. A. Duran, A. V. Vitianingsih, M. S. Riza, A. L. Maukar, and S. F. A. Wati, "Data Mining

Untuk Prediksi Penjualan Menggunakan Metode Simple Linear Regression," *Teknika*, vol. 13, no. 1, pp. 27–34, 2024, doi: 10.34148/teknika.v13i1.712.

- [5] Z. Yang, X. Zheng, and Y. Wu, "Optimization of Complex Systems Based on LSTM Neural Networks and Random Forests," *2024 IEEE 2nd Int. Conf. Electr. Autom. Comput. Eng. ICEACE 2024*, pp. 791–796, 2024, doi: 10.1109/ICEACE63551.2024.10898251.
- [6] H. Saputro, U. Baturaja, and J. A. Yani, "Jurnal Informatika dan Komputer(JIK)," *Jik*, vol. 12, no. 2, p. 83, 2021.
- [7] J. Li, D. Zhu, and C. Li, "Comparative analysis of BPNN, SVR, LSTM, Random Forest, and LSTM-SVR for conditional simulation of non-Gaussian measured fluctuating wind pressures," *Mech. Syst. Signal Process.*, vol. 178, no. July 2021, p. 109285, 2022, doi: 10.1016/j.ymssp.2022.109285.
- [8] M. M. Lucini, P. J. Van Leeuwen, and M. Pulido, "Model error estimation using the expectation maximization algorithm and a particle flow filter," *SIAM-ASA J. Uncertain. Quantif.*, vol. 9, no. 2, pp. 681–707, 2021, doi: 10.1137/19M1297300.
- [9] M. S. Efendi and A. K. Zyen, "Penerapan Algoritma Random Forest Untuk Prediksi Penjualan Dan Sistem Persediaan Produk," vol. 5, no. 1, pp. 12–20, 2024, doi: 10.30865/resolusi.v5i1.2149.
- [10] J. Du, Y. Yang, S. Geng, X. Liu, and Y. Wang, "Predicting Incidence Time Series Based on LSTM Recurrent Neural Network and Key Feature Analysis Based on Random Forest Algorithm," *Proc. - 2024 Int. Conf. Electron. Devices, Comput. Sci. ICEDCS 2024*, pp. 661–665, 2024, doi: 10.1109/ICEDCS64328.2024.00124.