 ISSN:2302-8149
e-ISSN:2540-9719

Penerapan Machine Learning dalam Memprediksi Tarif Tiket Bus AKAP pada Musim Liburan

Applying Machine Learning to Predict Intercity Bus Ticket Prices During the Holiday Season

[1]Rifandi Almanda*, [2]Tety Elida
[1,2]Magister Manajemen Sistem Informasi, Universitas Gunadarma
[1,2]Jl. Margonda Raya No.100, Depok, Jawa Barat, Indonesia, 16424
*e-mail: rifandialmanda@gmail.com

(***received***: 26 May 2025, ***revised***: 18 June 2025, ***accepted***: 23 June 2025)

Abstrak

Adanya Tol Trans Jawa telah meningkatkan persaingan di industri bus AKAP, mendorong perusahaan otobus (PO) untuk merumuskan strategi guna memaksimalkan pendapatan. Selama musim liburan dan hari raya, PO umumnya menaikkan tarif tiket berdasarkan harga kompetitor, padahal terdapat faktor penting lainnya seperti biaya perawatan armada, biaya operasional perjalanan (BBM, tol, komisi pengemudi, uang makan, insentif), serta komisi agen dan biaya makan yang seringkali ikut naik, bahkan dapat melonjak bersamaan jika harga BBM naik. Oleh karena itu, penelitian ini bertujuan untuk menganalisis penumpang, potensi trayek, melakukan tindakan preventif pada armada, dan menganalisis kesesuaian tarif tiket dengan contribution margin dan *cost ratio* menggunakan *machine learning* dengan algoritma *Random Forest Regression*. Model *machine learning* yang dibangun dengan data master trayek, armada, biaya perjalanan, harga tiket, hari libur, serta transaksi pendapatan dan biaya operasional, berhasil mencapai nilai RMSE 3%, R2 score 77%, MAPE 9.56%, dan akurasi 90.44%, kemudian diimplementasikan dalam antarmuka web berbasis Flask dan ASP.NET untuk membandingkan harga realisasi dengan harga prediksi.

Kata kunci: prediksi tarif, transportasi AKAP, machine learning, analisis random forest

Abstract

The construction of the Trans Java Toll Road has intensified competition in the intercity bus (AKAP) industry, prompting bus operators (PO) to formulate strategies to maximize revenue. During holiday and festive seasons, operators typically raise ticket prices based on competitor pricing. However, other critical factors are often overlooked, such as fleet maintenance costs, travel operational expenses (fuel, tolls, driver commissions, meal allowances, and incentives), as well as agent commissions and passenger meals, all of which tend to increase—sometimes drastically—especially when fuel prices rise. This study aims to analyze passenger behavior, route potential, conduct preventive maintenance planning, and evaluate the alignment between ticket pricing, contribution margin, and cost ratio using a machine learning approach with the Random Forest Regression algorithm. The machine learning model was developed using data on route masters, fleet details, travel costs, ticket prices, holidays, and transaction records of revenue and operational expenses. The model achieved an RMSE of 3%, R^2 score of 77%, MAPE of 9.56%, and accuracy of 90.44%. It was then implemented in a web-based interface using Flask and ASP.NET to compare actual ticket prices with predicted prices.

Keywords: *fare prediction, AKAP transportation, machine learning, random forest analysis*

1 Pendahuluan

Perkembangan industri transportasi bus AKAP dengan adanya tol trans jawa turut meningkatkan persaingan antar PO (perusahaan otobus). Persaingan yang ketat mendorong perusahaan untuk merancang strategi yang efektif dalam meningkatkan pendapatannya. Salah satu pendekatan yang digunakan adalah manajemen permintaan, yang mencakup kebijakan, metode, dan sistem yang dirancang secara khusus untuk mengoptimalkan penerimaan pendapatan. Pendekatan ini dikenal

 ISSN:2302-8149
e-ISSN:2540-9719

sebagai manajemen pendapatan (*revenue management*) (Talluri & Ryzin, 2004). Penelitian ini difokuskan pada isu-isu yang terjadi dalam sebuah perusahaan otobus (PO) yang beroperasi di sektor jasa transportasi darat. Perusahaan tersebut menyediakan layanan angkutan darat bagi masyarakat, dan penelitian ini bertujuan untuk mengkaji permasalahan yang timbul dalam pelaksanaan operasionalnya. Dalam penentuan kebijakan untuk menaikan tarif tiket, perusahaan harus melihat persaingan harga dengan kompetitor, karena harga tiket yang diterbitkan harus kompetitif dengan harga kompetitor. Akan tetapi permasalahan yang terjadi saat ini adalah harga tiket tersebut dapat ditentukan dan diterbitkan hari H-7. Padahal permintaan penjualan tiket sudah mulai tinggi mulai di H-14. Dan apabila penjualan sudah terjadi, maka penumpang akan dikenakan tarif sisa (ditagih) dari harga awal yang sudah dibeli sebelumnya. Menurut (Abadi et al., 2020), di sektor transportasi, potensi perpindahan pelanggan ke penyedia layanan lain sangat tinggi. Oleh karena itu, perusahaan harus memiliki langkah antisipatif yang tepat dalam menentukan strategi bisnisnya. Mempertahankan loyalitas pelanggan menjadi faktor krusial agar perusahaan dapat terus bertahan dan bersaing secara efektif di tengah ketatnya kompetisi industri ini.

Saat ini faktor utama penentuan kenaikan harga tarif tiket tersebut hanya berdasarkan harga kompetitor saja. Padahal dalam persiapan menjelang musim liburan dan hari raya, ada beberapa faktor yang harus dikaji. Pertama adalah setiap armada yang akan beroperasi harus dilakukan service maintenance keseluruhan agar mendapatkan hasil maksimal. Kedua adalah biaya perjalanan standar untuk diberikan kepada pengemudi yang berisi BBM, biaya tol, komisi pengemudi, uang makan pengemudi dan insentif apabila armada terisi penuh penumpang (*full seat*). Ketiga adalah komisi agen dan *service* makan. Dari ketiga faktor tersebut pada saat musim liburan dan hari raya, salah satu faktor diantara mengalami kenaikan juga. Bahkan ketiga faktor tersebut dapat naik secara bersamaan jika BBM (solar) mengalami kenaikan dari pemerintah.

Kecerdasan Buatan (AI) merupakan representasi dari mesin yang mampu menampilkan kemampuan menyerupai kecerdasan manusia dan kini semakin luas diterapkan dalam berbagai layanan serta menjadi pendorong utama inovasi (Huang & Rust, 2018). AI terdiri atas berbagai teknologi yang bekerja secara terpadu untuk mengumpulkan, mengolah, dan merespons data dengan cara yang meniru proses berpikir manusia. Seperti halnya manusia, AI mampu menerapkan aturan, belajar dari pengalaman melalui data baru, serta menyesuaikan diri terhadap perubahan yang terjadi di lingkungannya (Russel & Norvig, 2010). Untuk mempermudah pemecahan masalah dan pencapaian tujuan penelitian secara tepat, penulis menetapkan batasan-batasan tertentu dalam ruang lingkup penelitian ini. Adapun batasan-batasan tersebut dijelaskan sebagai berikut:

1. Perancangan model *machine learning* dalam memprediksi lonjakan jumlah penumpang transportasi bus akap pada musim liburan, trayek yang paling ramai, jumlah bus yang di siapkan dan analisis rasio biaya untuk menentukan harga tiket menggunakan data *history* dari data transaksi masa lalu *(history transaction data)* pada saat PO ini mulai beroperasi yaitu pada Desember 2016 sampai Juli 2023.
2. *Machine Learning* dalam penelitian ini menggunakan teknik *supervised learning* dengan metode algoritma *random forest regression* (RFR).
3. Data transaksi masa lalu *(history transaction data)* yang dikumpulkan dalam penelitian ini adalah data data trayek, jumlah penumpang, total biaya uang perjalanan baik yang standar maupun yang penyesuaian, nominal harga tiket sebelumnya dan total biaya perbaikan *(service maintenance)* dalam persiapan menjelang *high season* tersebut.

Tujuan dilakukan penelitian ini antara lain sebagai berikut:

1. Membuat sebuah model machine learning menggunakan algoritma *random forest regression* yang akan digunakan sebagai dasar pembuat sebuah keputusan.
2. Membuat sebuah interface berbasis *website* yang berguna sebagai alat untuk melakukan prediksi tarif tiket.

Kegunaan dari penelitian ini diharapkan dapat berguna di industri transportasi, khusus nya bisnis AKAP (antar kota antar provinsi). Management PO dapat langsung mengambil keputusan untuk menerapkan harga yang seharusnya diterbitkan dan dipublikasikan kepada penumpang. Hal itu sangat berguna dalam bisnis untuk memperoleh keuntungan yang maksimal dan menekan biaya yang akan terjadi. Management PO juga dapat melihat potensi hasil keuntungan yang maksimal karena dilihat dari jumlah penumpang di setiap periode. Selanjutnya, menjalin kerja sama dengan agen perjalanan online (OTA) seperti Traveloka akan menjadi lebih praktis. Hal ini juga akan mempercepat proses

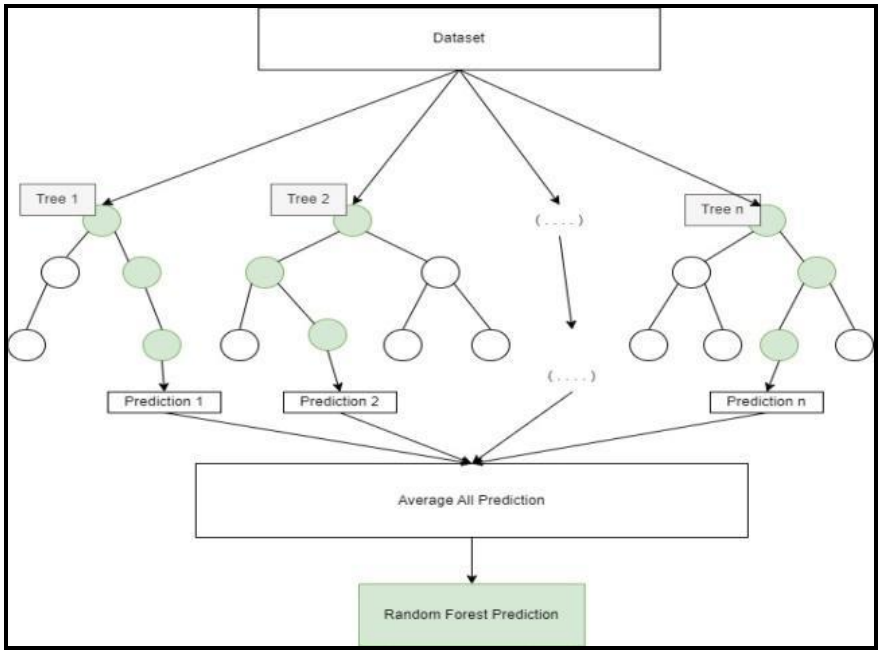
 ISSN:2302-8149
e-ISSN:2540-9719

penjualan secara daring karena tidak perlu lagi khawatir terhadap perubahan harga yang mungkin terjadi setelah harga diterbitkan. Penggunaan machine learning juga sudah diterapkan di industri lain seperti pesawat, taksi *online* dan pengiriman barang (logistik). Penggunaan teknologi *machine learning* dengan algortima *random forest regression* dalam penelitian ini diharapkan juga akan membantu untuk melihat potensi lonjakan dan optimalisasi harga tiket yang di terapkan di perusahaan otobus (PO).

2 Tinjauan Literatur

2.1 Algoritma *Random Forest Regression*

Algoritma *Random Forest* pertama kali dikembangkan oleh Breiman pada tahun 2001. Algoritma ini memiliki peran penting dalam dua jenis tugas utama, yaitu klasifikasi dan regresi. *Random Forest* dapat digunakan untuk berbagai tipe data, termasuk data diskrit, data kontinu, gabungan multivariat, serta data survival. Salah satu keunggulan utama dari algoritma ini adalah kemampuannya dalam mengidentifikasi interaksi antara variabel dependen dan independen, sekaligus menawarkan fleksibilitas tinggi dalam proses eksplorasi data. Selain itu, metode ini tidak mensyaratkan adanya asumsi khusus yang harus dipenuhi sebelum analisis dilakukan (Wulansari, 2018). Metode ini dikembangkan dari *classification and regression tree* (CART) dengan mengintegrasikan teknik bagging (*bootstrap aggregating*) serta penerapan pemilihan fitur secara acak. *Bagging* sendiri merupakan teknik yang dapat meningkatkan akurasi algoritma klasifikasi dengan menggunakan metode ensemble (Samudra, 2019). Algoritma Random Forest terdiri dari dua tahapan utama, yaitu pembentukan sejumlah pohon ‘k’ untuk membangun hutan acak, dan tahap prediksi menggunakan hutan acak tersebut. Hubungan antara fitur dan target direpresentasikan dalam bentuk rangkaian kondisi yang tersusun seperti struktur pohon dari atas ke bawah (Baron, Reis & Shahaf, 2018). Sebuah penelitian oleh Darst dan Malecki (2017) mengaplikasikan metode recursive feature elimination untuk seleksi fitur pada model Random Forest dalam konteks peramalan.



Gambar 1. Tree algoritma random forest

Gambar 1 adalah logika dalam implementasi algoritma *random forest*. Pembuatan data sampel dengan cara pengambilan acak nilai yang dihasilkan dari dataset. Pengambilan sample data digunakan untuk membangun pohon ke ke i (i=1, 2, 3, ..., n), setelah pohon terbentuk, ulangi langkah 1 dan 2 sebanyak k kali.

Dalam proses algoritma *random forest* untuk mendapatkan akurasi yang baik ada beberapa variable yang digunakan antara lain :

a. Variable n_estimators dan max_depth.

 ISSN:2302-8149
e-ISSN:2540-9719

Variable ini digunakan untuk meningkatkan akurasi. Parameter n_estimators menentukan jumlah total pohon keputusan (decision tree) yang akan dibangun dalam model, sementara max_depth mengatur seberapa dalam setiap pohon tersebut dapat berkembang atau berlapis, yang memengaruhi kompleksitas regresi linier di setiap tingkatnya.

Table 1. Table variable n_estimator dan max_depth

HyperParameter	Nilai
max_depth	10
n_estimators	100

Table 1 mendefinisikan berapa estimasi *decision tree* yang akan dibuat dari *variable n_estimator* dengan jumlah 100, setelah itu tingkat kedalaman yang akan dibuat sebanyak 10 kedalaman (*variable max_depth*).

b. *Mean Absolute Percentage Error* (MAPE)

MAPE adalah sebuah metode statistik yang digunakan untuk mengukur tingkat keakuratan suatu model prediksi atau peramalan. Model dianggap memiliki prediksi yang baik apabila nilai MAPE yang dihasilkan berada di bawah 10% (Indarwati et al., 2019). Rumus MAPE akan didefinisikan melalui persamaan (1).

Rumus MAPE adalah : $$\text{MAPE} = \frac{1}{n}\sum_{t=1}^{n}\left|\frac{A_t - F_t}{A_t}\right| \times 100\% \quad (1)$$

Keterangan :
At : nilai aktual pada waktu ke- t
Ft : nilai prediksi pada waktu ke- t
n : jumlah data

Table 2. Table range MAPE

Range MAPE	Arti Nilai
< 10%	Kemampuan Model Peramalan Sangat Baik
10 - 20%	Kemampuan Model Peramalan Baik
20 - 50 %	Kemampuan Model Peramalan Layak
>50%	Kemampuan Model Peramalan Buruk

Tabel 2 menunjukkan klasifikasi tingkat akurasi prediksi berdasarkan nilai MAPE yang diperoleh. Jika akurasi berada di bawah 10%, model prediksi dianggap sangat baik. Model dengan akurasi antara 10% hingga 20% masuk dalam kategori baik, sedangkan akurasi antara 20% hingga 50% menunjukkan bahwa model tersebut masih layak digunakan. Namun, jika akurasi melebihi 50%, maka model prediksi tersebut dinilai kurang baik.

c. *Root Mean Square Error* (RMSE)

Root mean square error (RMSE) adalah salah satu metode yang digunakan untuk menilai keakuratan teknik peramalan. Metode ini mengukur seberapa besar kesalahan yang terjadi dalam hasil prediksi sebuah model. RMSE dihitung sebagai rata-rata kuadrat dari selisih antara nilai yang diprediksi dan nilai sebenarnya, sehingga dapat menggambarkan tingkat kesalahan secara keseluruhan. Semakin kecil nilai RMSE, semakin baik model tersebut karena menunjukkan bahwa prediksi yang dihasilkan sangat mendekati data observasi yang sebenarnya, salah satu ukuran kesalahan dalam peramalan adalah nilai tengah akar kuadrat *root mean square error* (RMSE) (Makridakis et al., 2018). Nilai RMSE yang kecil atau mendekati nol menunjukkan bahwa hasil prediksi sangat sesuai dengan data sebenarnya. Meskipun RMSE juga bisa bernilai negatif, semakin kecil nilai RMSE menandakan tingkat ketepatan prediksi yang lebih tinggi. Berdasarkan hal ini, model tersebut layak untuk diterapkan dalam penelitian ini serta dapat digunakan pada masa mendatang. Rumus RMSE dapat dituliskan melalui persamaan (2) sebagai berikut:

 ISSN:2302-8149
e-ISSN:2540-9719

Rumus RMSE : $$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \quad (2)$$

Keterangan :
N : jumlah data (observasi)
yi : nilai aktual (actual value)
^yi : nilai prediksi (predicted value)
yi - ^yi : error (selisih antara nilai aktual dan prediksi)

d. R-squared (R2)

Nilai *R squared* berkisar antara 0 sampai 1 dan menggambarkan seberapa besar variabel-variabel independen secara bersama-sama memberikan pengaruh terhadap variabel dependen. Ukuran ini berfungsi untuk menilai seberapa kuat variabel laten independen tertentu memengaruhi variabel laten dependen dalam sebuah model analisis.

Table 3. Table kategori range R2

Range R2	Kategori
>=75%	Kuat
50% – 74%	Moderat
0% – 49%	Lemah

Tabel 3 menunjukkan pembagian kategori berdasarkan nilai R square, yang terdiri dari kategori kuat, kategori moderat, dan kategori lemah (Hair et al., 2011).

R-squared tidak terbatas penggunaannya hanya pada regresi, melainkan juga bisa diterapkan pada berbagai jenis model untuk menilai sejauh mana kualitas atau kecocokan model tersebut. Dengan kata lain, R-squared berfungsi sebagai alat evaluasi yang umum digunakan untuk mengetahui seberapa baik model menjelaskan data yang ada. Selain metode perhitungan, *random forest* juga mempunya beberapa variable yang digunakan dalam penerapan prediksi.

2.2 Penelitian Menggunakan Random Forest Regression

Dalam penelitan ini, penulis akan membahas hasil dari beberapa penelitian yang pernah dilakukan, sehingga akan didapatkan keterkaitan dengan penelitian yang penulis lampirkan berikut. Adapun penelitian yang akan dilampirkan terdiri dari 7 penelitian pembuatan mengenai penggunaan *random forest regression*.

Pertama, penelitian yang dilakukan oleh Zebua. et al. (2022) Penelitian ini fokus pada prediksi penetapan tarif penerbangan dengan memanfaatkan teknologi Auto-ML dan metode random forest. Beberapa variabel yang dianalisis meliputi jenis layanan, rute penerbangan, jadwal penerbangan, serta jumlah pemberhentian pesawat pada masing-masing maskapai. Dengan adanya penelitian ini, diharapkan perusahaan maskapai dapat memperoleh prediksi yang akurat terkait penentuan tarif sehingga mereka bisa menetapkan harga yang lebih tepat dan kompetitif sesuai kondisi pasar. Penelitian ini mencoba membandingkan penggunaan model *random forest, logistic* dan *gradient boosting*. Tetapi, dari ketiga model tersebut *random forest* mempunyai nilai prediksi yang sangat baik.

Table 4. Perbandingan model RF, LR, GB, regression

Model	R2 Score	RMSE
1. **RF Regressor**	83.91%	**$175.9**
2. **Log Regressor**	44.25%	**$327.6**
3. **GB Regressor**	**81.68%**	**$187.8**

Table 4 merupakan perbandingan dari penelitian yang dilakukan oleh Zebua. et al. (2022). Random *Forest* mendapatkan R2 *Score* 83% dan RMSE 175.9. Maka dari itu, *random forest* dapat dinyatakan lebih baik dari model *logistic regression* dan *gradient boosting*.

Kedua, penelitian yang dilakukan oleh Nur. et al. (2023), penelitian ini membahas penerapan algoritma random forest regression dalam memprediksi hasil panen padi di Desa Minanga. Fokus utama dari studi ini adalah mengembangkan model prediksi yang akurat berdasarkan berbagai faktor yang memengaruhi produksi padi, seperti luas lahan, jumlah bibit yang digunakan, jenis pupuk, tingkat curah hujan, keberadaan hama dan gulma, upaya pengendalian hama dan gulma, serta metode penanaman padi. Dengan memanfaatkan algoritma random forest regression, penelitian ini bertujuan

 ISSN:2302-8149
e-ISSN:2540-9719

memberikan bantuan dalam memperkirakan hasil panen secara lebih tepat berdasarkan data historis dan variabel-variabel terkait. Simulasi yang dilakukan adalah mencari nilai MAPE dari pengisian variable n_estimator dengan nilai 5, 9, 19, 29, 39, 50 dan 100.

Table 5. Table MAPE berdasarkan jumlah pohon

Jumlah Pohon	MAPE (%)
5	5.126
9	4.884
19	5.155
29	5.139
39	5.072
50	5.093
100	5.207

Table 5 adalah simulasi yang dilakukan oleh Nur. et al. (2023). Model dengan parameter n_estimator=9 menunjukkan performa yang sangat baik. Evaluasi menghasilkan MAPE sebesar 4,884%, menandakan kesalahan prediksi relatif di bawah 5% dan akurasi tinggi. RMSE sebesar 0,250 mengindikasikan selisih prediksi dan nilai aktual yang rendah. Selain itu, model ini memiliki nilai R-kuadrat sebesar 0,99, yang berarti model tersebut mampu menjelaskan 99% dari variasi yang ada dalam data. Hal ini menandakan bahwa model sangat efektif dalam mengidentifikasi dan merepresentasikan pola yang terdapat pada data tersebut. Secara keseluruhan, model ini andal dan layak digunakan untuk prediksi lebih lanjut. Nilai akurasi yang dihasilkan *random forest regression* pada penelitian ini sebesar 95,11%.

Ketiga, Penelitian ini memanfaatkan sekitar 324 fitur dalam penerapan model Random Forest untuk melakukan analisis. Berdasarkan hasil penelitian ini, didapatkan nilai Root Mean Square Error (RMSE) sebesar 0,04, yang mengindikasikan tingkat kesalahan prediksi yang cukup rendah. Sebagai perbandingan, Luong dan Dokuchaev melakukan penelitian serupa pada tahun 2018 dengan menggunakan dataset dari Reuters, yang mencakup periode 1 Januari 2008 hingga 31 Desember 2014. Penelitian mereka meliputi beberapa langkah, dimulai dari pemilihan model yang paling cocok untuk peramalan, dilanjutkan dengan pengukuran volatilitas data, kemudian menentukan model yang tepat untuk memodelkan volatilitas tersebut, dan akhirnya melakukan prediksi nilai volatilitas berdasarkan model yang sudah dipilih. Hasil evaluasi kesalahan dalam penelitian tersebut juga diukur menggunakan RMSE, dan didapatkan nilai sebesar 0,0996, yang menunjukkan tingkat akurasi model dalam memprediksi volatilitas dengan kesalahan yang cukup kecil namun masih lebih tinggi dibandingkan penelitian pertama. Dengan demikian, kedua penelitian ini memberikan gambaran mengenai penggunaan model dan fitur yang berbeda serta metode evaluasi yang serupa untuk peramalan dan pemodelan volatilitas dalam data finansial.

Keempat, Penelitian (Pavlyshenko, 2018) menggunakan metode Random Forest untuk melakukan peramalan terhadap data penjualan. Dalam studi tersebut, dataset yang dipakai berasal dari Rossman Store Sales yang diambil dari platform Kaggle, sebuah sumber data populer untuk keperluan analisis dan pembelajaran mesin. Tahapan awal dalam penelitian ini melibatkan analisis deskriptif yang bertujuan untuk memahami karakteristik data secara mendalam. Analisis ini dilakukan dengan memvisualisasikan data menggunakan berbagai jenis diagram agar pola-pola yang terdapat dalam data bisa lebih mudah dikenali dan dianalisis. Selanjutnya, hasil dari penerapan model Random Forest ini menunjukkan bahwa tingkat kesalahan atau error dalam peramalan yang dihasilkan adalah sebesar 13,6%, yang menjadi indikator kinerja model dalam memprediksi data penjualan tersebut.

Kelima, penelitian yang dilakukan Haryanto, et al. (2023) Studi ini berfokus pada perbandingan efektivitas dua algoritma pembelajaran mesin regresi linier berganda dan regresi hutan acak dalam memprediksi harga rumah. Analisis tersebut mempertimbangkan beberapa variabel penting, termasuk luas tanah, luas bangunan, jumlah kamar tidur, kamar mandi, dan garasi, yang semuanya dianggap sebagai faktor signifikan yang memengaruhi nilai pasar rumah. Tujuan utamanya adalah untuk mengevaluasi seberapa dekat harga jual rumah yang sebenarnya selaras dengan harga yang diprediksi oleh model regresi linier berganda dan regresi hutan acak. Setelah melakukan pengujian dan menganalisis hasilnya, ditemukan bahwa model *multiple linear regression* mencapai akurasi 78,5%, sedangkan model *random forest regression* menunjukkan akurasi yang lebih tinggi yaitu 81,6%. Hal ini menunjukkan bahwa, dalam konteks studi ini, algoritma hutan acak memberikan prediksi harga

 ISSN:2302-8149
e-ISSN:2540-9719

rumah yang lebih tepat dan andal dibandingkan dengan model regresi linier berganda, yang menunjukkan kesesuaiannya yang lebih unggul untuk tugas-tugas regresi yang terkait dengan penilaian real estat.

Keenam, penelitian yang dilakukan oleh Fachid dan Triayudi (2022), Penelitian ini membandingkan algoritma regresi linier dan regresi random forest dalam menganalisis data masyarakat terdampak Covid-19, seperti kasus terinfeksi, kematian, dan pasien pulih. Hasilnya, regresi random forest lebih unggul dengan RMSE 1886.555, MAPE 14.85, dan akurasi 97,7%, dibandingkan regresi linier yang memiliki RMSE 3031.127, MAPE 47.66, dan akurasi 94%. Kesimpulannya, regresi random forest lebih efektif untuk analisis data Covid-19.

Ketujuh, Fitri (2023) penelitian ini membandingkan tiga metode prediksi harga rumah: regresi linier, random forest regression, dan gradient boosted trees regression, menggunakan data harga rumah dari Jakarta Selatan dan Tebet. Hasil menunjukkan random forest regression paling unggul dengan error terendah (RMSE 0,440) dan akurasi tertinggi (81,5%), dibandingkan regresi linier dan gradient boosted trees. Kesimpulannya, random forest adalah algoritma yang sangat efektif untuk memprediksi harga properti dengan tingkat kesalahan prediksi yang lebih rendah. Algoritma ini juga mampu menangani data training berukuran besar dengan efektif, serta menawarkan fleksibilitas dan kemudahan dalam penggunaannya (Suliztia, M. L. ,2020). Akan tetapi nilai yang dihasilkan dari *random forest* mulai RMSE , MAPE, dan R2 mempunyai hasil dengan nilai yang baik dibandingkan dengan algortima regresi yang lain nya sehingga penulis memilih untuk menggunakan *random forest* dalam penelitian ini.

3 Metode Penelitian

Bagian ini menjelaskan pendekatan yang digunakan dalam penelitian untuk menganalisis tarif tiket dan biaya operasional perusahaan otobus selama periode liburan dan hari besar. Penelitian ini memanfaatkan data historis yang diperoleh dari transaksi selama periode Hari Raya Idul Fitri, Natal, dan Tahun Baru, serta dibandingkan dengan periode normal. Pengumpulan data mencakup informasi tarif, jumlah penumpang, biaya perjalanan, biaya maintenance, serta pola pergerakan penumpang dalam arus mudik dan balik. Seluruh data tersebut digunakan untuk memahami dinamika permintaan dan biaya yang memengaruhi penentuan harga tiket selama periode high season. Berikut ini adalah penjelasan metode yang dilakukan :

3.1 Objek Penelitian

Penelitian ini mengambil fokus pada sebuah perusahaan yang beroperasi di sektor transportasi. PT. MA merupakan salah satu perusahaan otobus (PO) yang menyediakan layanan transportasi bisnis antar kota antar provinsi (AKAP). Perusahaan ini mempunyai 372 armada yang beroperasi di 43 trayek aktif khususnya di pulau Jawa. Armada yang dimiliki PT.MA terdiri dari beberapa kelas, mulai dari VIP, bisnis, eksekutif, super eksekutif dan double decker. PT.MA juga mempunyai bengkel sendiri yang dapat digunakan oleh armada untuk melakukan perawatan rutin maupun perbaikan mesin dan body.

3.2 Sumber Data dan Teknik Pungumpulan Data.

Penelitian ini memanfaatkan data transaksi yang berasal dari periode Hari Raya Idul Fitri, Natal, dan Tahun Baru. Dari kumpulan data transaksi selama masa liburan tersebut, sejumlah informasi diambil, seperti harga tiket untuk setiap rute perjalanan, jumlah penumpang yang menggunakan layanan, tarif dasar perjalanan, serta biaya tambahan atau potongan terkait perjalanan. Selain itu, data juga mencakup biaya pemeliharaan layanan armada yang terjadi antara Januari 2016 hingga Juni 2023. Untuk mengumpulkan data tersebut, digunakan beberapa metode pengumpulan data, antara lain:

1. Mengambil *history* tarif tiket penumpang yang dikelompokan per trayek yang terjadi pada periode Hari Raya Idul Fitri, Natal dan Tahun Baru yaitu Desember 2016 sampai Juni 2023. Setelah itu dibandingkan dengan data history harga tiket yang terjadi di periode normal (selain Hari Raya Idul Fitri, Natal dan Tahun Baru).
2. Mengambil data jumlah penumpang yang terjadi pada periode Hari Raya Idul Fitri, Natal dan Tahun Baru dibandingkan dengan jumlah penumpang yang terjadi selain di periode normal.
3. Mengambil data jumlah armada yang beroperasi, baik armada regular atau armada tambahan, statistik data digunakan untuk menentukan prediksi trayek yang akan mendapatkan lonjakan

 ISSN:2302-8149
e-ISSN:2540-9719

penumpang yang tinggi. Dari data tersebut manajemen PO dapat menempatkan armada bantuan pada trayek tersebut.

4. Mengambil data biaya tetap dan biaya variable pada periode Hari Raya Idul Fitri, Natal dan Tahun baru dibandingkan dengan biaya tetap dan biaya variable periode normal.
5. Melakukan rekapitulasi biaya perbaikan per armada yang di kelompokan berdasarkan nomor armada (nomor body) dan tahun perbaikan yang digunakan oleh PT.MA.
6. Membagi beberapa kategori mengenai arus mudik dan arus balik dalam periode Hari Raya Idul Fitri, Natal dan Tahun Baru. Contoh dalam libur Hari Raya Idul Fitri ada istilah "arus mudik" dan "arus balik". Istilah tersebut dipakai juga untuk penelitian ini dalam membagi kategori pada saat Hari Raya Natal dan Tahun Baru, Pada saat menjelang Hari Raya Natal isitilah yang digunakan adalah "arus mudik", untuk Tahun baru istilah yang digunakan adalah "arus balik". Asumsi ini diambil dikarenakan pada saat Hari Raya Natal , masyarakat akan mengambil cuti atau liburan sampai dengan tahun baru. Untuk harga yang di ambil sesuai dengan kenaikan pada saat kondisi arus mudik dan arus balik. Didalam periode arus mudik dan arus balik tersebut juga ada beberapa kali kenaikan sampai hari H. Sebagai contoh Hari Raya Idul Fitri tahun 2022 tepat di tanggal 2 dan 3 Mei 2022. Dimulai dari tanggal 15 April 2022, harga tiket sudah mulai naik bertahap dan puncak nya di tanggal 3 Mei 2022 untuk arus mudik.

3.3 Variabel Penelitian

Variabel penelitian yang digunakan dalam penelitian ini antara lain data transaksi yang telah terjadi di masa lalu. Ada 6 variabel yang akan digunakan sebagai dasar dalam penelitian ini. Antara lain sebagai berikut :

Table 7. Table variable penelitian

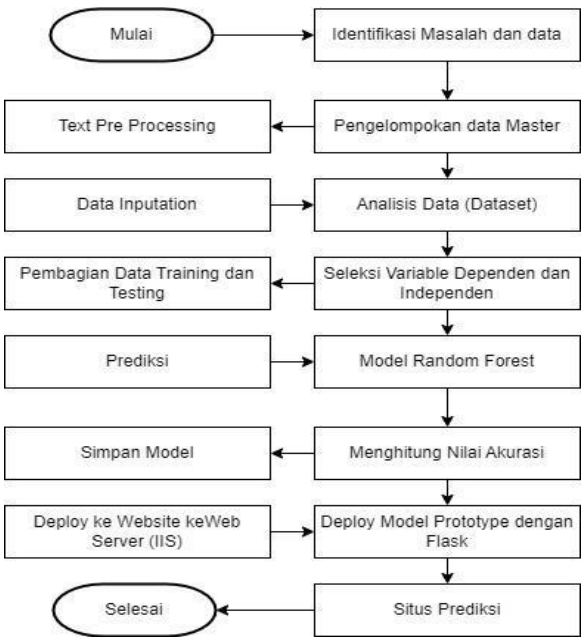
No	Nama Variabel	Kategori	Keterangan	Skala/Satuan	Contoh
1	Harga Tiket	-	Suatu nilai yang akan di bayar oleh penumpang dan akan diterima oleh perusahaan	Numerik	200,000
2	Trayek	0 = BALARAJA - BOBOTSARI 1 = BALARAJA - WONOSOBO 2 = BEKASI - BOBOTSARI 3 = BEKASI - WONOSOBO …… 108 = YOGYAKARTA - BALARAJA 109 = YOGYAKARTA - DEPOK 110 = YOGYAKARTA - JAKARTA 111 = YOGYAKARTA - MERAK 112 = YOGYAKARTA - TANGGERANG	Rute Perjalanan 2 Trip (Berangkat) dan Pulang	Kategori	YOGYAKARTA - JAKARTA
3	Uang Jalan	1 = Biaya Solar 2 = Biaya Premi Kru 3 = Biaya Lain 4 = Biaya FullSeat 5 = Biaya Perpal 6 = Biaya Komisi	Biaya yang dikeluarkan armada untuk menyelesaikan 1 Rute (2 Trip Berangkat dan Pulang)	Numerik	2,163,000
4	Tambahan Biaya Uang	1 = Biaya Solar 2 = Biaya Tol	Tambahan Biaya yang dikeluarkan armada	Numerik	100,000

 ISSN:2302-8149
e-ISSN:2540-9719

	Jalan	3 = BiayaUang Makan Kru	untuk menyelesaikan 1 Rute (2 Trip Berangkat dan Pulang)		
		4 = Biaya Premi Kru			
		5 = Biaya TPR Mel PP			
		6 = Biaya Full Seat			
		7 = Biaya Over Seat			
		8 = Biaya Perpal			
		9 = Biaya Lain Lain			
5	Jumlah Penumpang	1 = Berangkat	Jumlah penumpang arah berangkat atau pulang	Numerik	1,500
		2 = Pulang			
6	Biaya Persiapan Service - Maintenance	-	Kenaikan biaya maintenance yang dapat diambil persentase nya secara periodik	Numerik	300%

Table 7 adalah variabel yang digunakan dalam penelitian. Pertama adalah *variabel* harga tiket, yaitu variabel yang akan dicari nilai nya berdasarkan data variabel pendamping yang dimasukan. Kedua adalah variabel trayek, yaitu rute perjalanan yang dilalui oleh armada. Jumlah trayek dari periode Desember 2016 – Juni 2023 adalah 43 trayek. Trayek ini dibagi menjadi beberapa bagian, mulai dari kelas bus Eksekutif, bus *Double Decker*, dan bus *VIP* . Jumlah kursi untuk armada bus eksekutif adalah 32 kursi, Double Decker ada 50 dan 52 kursi, *VIP* ada 38 sampai 42 kursi. Pada variable trayek ini terdapat penghubung dengan biaya tetap dan biaya variable. Variabel ketiga adalah variable biaya perjalanan standar yaitu biaya yang dikeluarkan untuk armada yang akan beroperasi. Biaya ini dilihat dari trayek yang akan dilalui armada. Keempat adalah variabel penyesuaian biaya perjalanan standar. Variable ini adalah jenis biaya tidak tetap (*variable cost*). Apabila uang perjalanan standar tidak mencukupi saat armada beroperasi, maka dilakukan penyesuaian uang perjalanan standar dengan cara menambahkan biaya, sebaliknya apabila uang perjalanan standar berlebih, maka dilakukan penyesuaian uang perjalanan standar dengan cara mengurangi biaya. Kelima adalah variable jumlah penumpang yang dibagi menjadi 2 kategori, yaitu jumlah penumpang berangkat dan jumlah penumpang pulang. Variable keenam adalah biaya service maintenance adalah jenis variable cost yang fluktuatif sesuai dengan kondisi yang terjadi dalam proses persiapan menjelang musim liburan. Pada variable biaya *service maintenance*, penulis menggunakan *average* kenaikan setiap periode sampai *high season.*

3.4 Tahapan Penelitian



Gambar 2. Flow tahapan penelitian

 ISSN:2302-8149
e-ISSN:2540-9719

Penelitian ini dilakukan dengan beberapa tahapan sesuai dengan *flowchart* yang telah terbentuk pada gambar 2. Tahapan-tahapan tersebut antara lain:

1. Menentukan masalah serta objek penelitian yang akan dijadikan fokus merupakan langkah awal dalam penelitian ini.
2. Dilakukan pengumpulan berbagai data penting, seperti data master trayek, master armada, biaya perjalanan standar, harga tiket, serta data master hari libur seperti Idul Fitri, Natal, dan Tahun Baru. Selain itu, dikumpulkan pula data transaksi terkait pendapatan dan biaya SPJ, penyesuaian biaya standar perjalanan, serta biaya perawatan armada di PT. MA. Dalam proses penelitian ini, semua data yang diperlukan diperoleh melalui SQL Server, kemudian diolah dengan pengelompokan dan pembersihan menggunakan bahasa pemrograman Python agar siap untuk analisis lebih lanjut.

Table 8. data master pada penelitian

No	Master Data
1	Master Trayek
2	Master Armada
3	Master Biaya Perjalanan Standar
4	Master Harga Tiket
5	Master Liburan (Idul Fitri, Natal dan Tahun Baru)
6	Transaksi Pendapatan dan Biaya SPJ (Surat Perintah Jalan)
7	Transaksi Penambahan / Pengurangan Biaya Standar Perjalanan
8	Transaksi Service Maintenance

Table 8 adalah table yang akan dikelompokan setelah data selection dan data pre-proccessing selesai dilakukan. Data tersebut diambil dari data *history* perjalanan armada.

3. Proses awal yang dilakukan adalah *pre-processing* data mentah agar menjadi data yang lebih mudah dipahami dan sesuai dengan kebutuhan penelitian. Tahapan ini meliputi pembersihan data untuk menghapus informasi yang tidak relevan serta normalisasi untuk mengelompokkan data trayek secara konsisten.
4. Data yang mengandung nilai hilang pada variabel kategorik dan numerik diimputasi menggunakan metode modus agar data tetap lengkap.
5. Analisis deskriptif kemudian dilakukan untuk memberikan gambaran umum mengenai kondisi data yang digunakan, dengan hasil yang divisualisasikan dalam bentuk diagram batang.
6. Variabel-variabel yang akan dianalisis juga dipilih dan didefinisikan, dimana harga tiket menjadi variabel dependen, sedangkan trayek, jumlah penumpang, biaya perjalanan standar, penyesuaian biaya perjalanan, dan biaya service maintenance dijadikan variabel independen.
7. Data dibagi menjadi dua bagian utama: data pelatihan yang lebih besar untuk mengembangkan model, dan data pengujian untuk mengevaluasi akurasi model dalam prediksi atau klasifikasi.
8. Metode *random forest* diaplikasikan untuk mendapatkan tingkat pentingnya fitur serta membangun model prediksi harga tiket.
9. Prediksi dilakukan pada data testing dan keseluruhan data penelitian,
10. Keakuratan model dihitung dengan menggunakan mean absolute percentage error (MAPE), dimana nilai akurasi diperoleh dari 100% dikurangi MAPE.
11. Menyimpan model prediksi yang telah dibentuk.
12. Model prediksi yang telah dibuat kemudian disimpan dan dikembangkan menjadi sebuah prototipe menggunakan Flask di lingkungan PyCharm, dengan file utama app.py untuk pengolahan data dan tampilan antarmuka pengguna dibangun menggunakan website berbasis ASP.Net.

3.5 Teknik Analisis Data.

Setelah tahap persiapan dan pengumpulan data selesai, langkah berikutnya dalam penelitian ini adalah melakukan analisis data. Proses analisis bertujuan untuk mengubah data mentah menjadi

 ISSN:2302-8149
e-ISSN:2540-9719

informasi yang berguna, yang kemudian akan dipakai dalam tahap pemodelan. Teknik analisis dimulai dengan proses pembersihan dan pengelompokan data master. Pada tahap ini, data transaksi penjualan, baik dari periode normal maupun periode *high season*, diperiksa untuk mengidentifikasi dan menghapus data ganda maupun data yang memiliki nilai kosong (*null*). Selanjutnya, tarif tiket dikelompokkan berdasarkan arah keberangkatan (arah berangkat dan arah pulang), diklasifikasikan sesuai dengan tahun, trayek, dan jenis kelas bus.

Proses kalkulasi juga dilakukan terhadap data uang perjalanan standar dan penyesuaiannya, yang kemudian dikelompokkan berdasarkan tahun, trayek, dan kelas bus. Data jumlah penumpang disusun berdasarkan trayek, jenis kelas bus, dan tahun operasional, sedangkan data biaya maintenance dihimpun dan dikelompokkan berdasarkan nomor body bus, tahun perbaikan, serta kelas bus masing-masing.

Results Messages

	Tahun	TrayekID	NamaTrayek	HargaTiket	JumlahBus	JumlahPenumpang	Gross	BasicRate	AdjusmentBasicRate	BiayaMaintenance
1	2023	TH012	CILACAP - MERAK	250000	14	17837	3037480000	977524800	165675071	144912070.360000
2	2021	TH022	YOGYAKARTA - TANGERANG	500000	74	123653	23106530000	9950661400	1464698550	972059368.675900
3	2019	TH013	CILACAP - TANGGERANG	350000	42	77872	8499620000	4278650000	321096950	516327488.901650
4	2017	TH025	TEGAL - JAKARTA	95000	3	1339	100925000	76549200	390000	28071986.995400
5	2022	TH011	CILACAP - JAKARTA	450000	32	34844	5414452000	2033306400	290063922	391576146.692300
6	2022	TH029	YOGYAKARTA via Selatan - JAKARTA	450000	22	13380	3149940000	1285108900	47607390	74383947.915000
7	2023	TH021	YOGYAKARTA - BOGOR (VIP)	430000	45	20420	4372160000	1460833500	93411300	76600019.880000
8	2019	TH025	TEGAL - JAKARTA	170000	7	11237	923145000	620210350	15676400	47131121.545000
9	2018	TH008	YOGYAKARTA - BOGOR	330000	21	22229	3671905000	1551500000	80412000	250664320.437145
10	2019	TH002	WONOSOBO - BALARAJA	280000	28	48777	5425325000	2973882625	125806800	906031543.070268
11	2016	TH038	YOGYAKARTA - TANGGERANG	165000	32	13263	1785575000	661117200	45367200	296884048.158460
12	2023	TH042	CILACAP - BOGOR	150000	12	13293	2280850000	935055901	128873190	97314896.130000
13	2019	TH012	CILACAP - MERAK	300000	12	14654	1953940000	1015968050	67152000	236355817.397300
14	2023	TH015	WONOGIRI - BOGOR	480000	18	9047	2491550001	908734800	191606800	25102639.560000
15	2017	TH022	YOGYAKARTA - TANGERANG	320000	36	35809	5127295000	2709400350	73927100	421443218.301240
16	2020	TH037	YOGYAKARTA - DEPOK (DD)	280000	2	1854	459950000	240568500	219000	59075216.190000
17	2017	TH031	KLATEN - MERAK	350000	16	21817	3666660000	1282627450	64600000	183043280.972600

Query executed successfully. (local)\SQLEXPRESS (11.0

Gambar 3. Penyajian dataset

Gambar 3 tahapan analisis data master dan penyusunan *dataset*. Dari hasil pembersihan dan pengelompokan sebelumnya, dihasilkan beberapa tabel master seperti tabel trayek, riwayat tarif tiket, riwayat *basic rate*, riwayat penyesuaian *basic rate,* riwayat perawatan armada (*service summary*), dan riwayat jumlah penumpang. Data dari tabel-tabel tersebut kemudian diproses menggunakan *query* untuk menghasilkan dataset akhir yang berisi informasi penting seperti tahun, ID trayek, nama trayek, harga tiket, jumlah armada yang beroperasi, jumlah penumpang, pendapatan kotor (*gross*), biaya perjalanan standar (*basic rate*), penyesuaian biaya perjalanan (*adjustment basic rate*), dan biaya maintenance. *Dataset* ini selanjutnya digunakan sebagai dasar dalam pemodelan prediksi tarif menggunakan *machine learning.*

4 Hasil dan Pembahasan

Pada langkah ini, dilakukan pembangunan model machine learning menggunakan algoritma *Random Forest Regression*. Model ini bertujuan untuk memperkirakan tarif tiket dengan mempertimbangkan berbagai variabel yang telah dipersiapkan sebelumnya. Setelah dataset akhir siap, proses pemodelan dan pelatihan dilakukan dengan menggunakan bahasa pemrograman Python, yang dibantu oleh beberapa *library* terkenal seperti pandas, numpy, sklearn, serta modul RandomForestRegressor. Langkah awal dalam proses ini adalah membaca data dari file berformat *CSV* yang berisi informasi lengkap mengenai trayek, harga tiket, jumlah armada, jumlah penumpang, biaya perjalanan, dan biaya maintenance. Dataset ini kemudian menjadi dasar untuk membangun model prediksi yang akan dianalisis lebih lanjut dalam bagian berikutnya.

4.1 Pembuatan Model Machine Learning Random Forest Regression

Setelah dataset yang diperlukan sudah siap, tahap berikutnya adalah mengembangkan model prediksi dengan memanfaatkan algoritma Random Forest Regression. Proses ini dilakukan dengan menggunakan bahasa pemrograman Python dan beberapa *library* pendukung seperti pandas, numpy, sklearn, serta RandomForestRegressor.

 ISSN:2302-8149
e-ISSN:2540-9719

```
In [98]: import pandas as pd

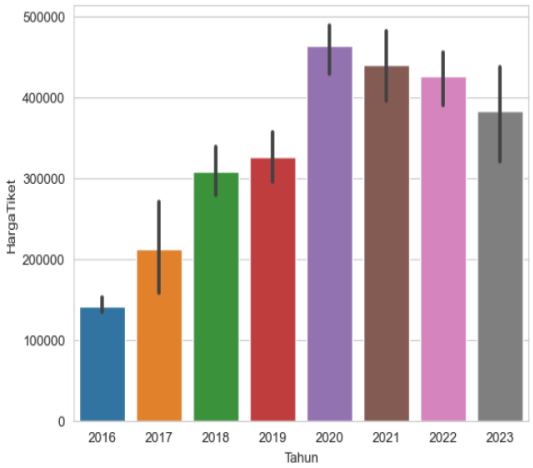
         #features = pd.read_csv('data-train.csv')
         features = pd.read_csv('revenue.csv')
         features.head(5)
```

Out[98]:

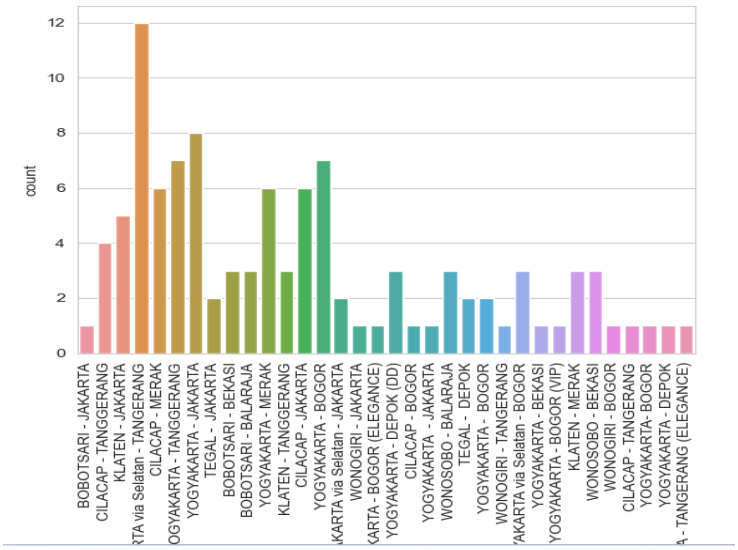
	Tahun	TrayekID	NamaTrayek	HargaTiket	JumlahBus	JumlahPenumpang	Gross	BasicRate	AdjusmentBasicRate	BiayaMaintenance
0	2020	TH007	BOBOTSARI - JAKARTA	450000	9	5388	504550000	251867200	9485000	4.823304e+07
1	2020	TH013	CILACAP - TANGGERANG	450000	49	34707	4220280000	2098405975	164348540	4.668373e+08
2	2017	TH030	KLATEN - JAKARTA	320000	74	68806	9508635000	4167143825	197330075	6.203512e+08
3	2017	TH027	YOGYAKARTA via Selatan - TANGERANG	320000	36	35809	5127295000	2709400350	73927100	4.214432e+08
4	2021	TH013	CILACAP - TANGGERANG	400000	30	33485	4646490000	1988217444	202901995	2.671990e+08

Gambar 4. Proses membaca data set .csv

Gambar 4 adalah langkah awal dimulai dengan membaca file dataset berformat CSV yang berisi seluruh data yang telah diproses sebelumnya. Selanjutnya, dilakukan pemeriksaan struktur data menggunakan fungsi *shape* untuk mengetahui dimensi atau ukuran dari dataset tersebut.



Gambar 5. Proses visualisasi pengelompokan dataset revenue per tahun



Gambar 6. Proses visualisasi pengelompokan data set per trayek.

Gambar 5 dan 6 adalah proses visualisasi data. Setelah itu, dilakukan visualisasi, data dikelompokkan berdasarkan tahun dan trayek guna memahami distribusi pendapatan dan jumlah penumpang pada setiap kategori tersebut. Visualisasi ini membantu dalam melihat *trend* data secara umum sebelum masuk ke proses modeling.

 ISSN:2302-8149
e-ISSN:2540-9719

Tahapan berikutnya adalah melakukan pembersihan data dengan menghapus beberapa kolom yang tidak relevan terhadap pemodelan, yaitu kolom jumlahbus, gross, dan trayekid. Setelah kolom-kolom yang tidak dibutuhkan dihapus, dilakukan proses *label encoding* terhadap kolom-kolom yang diperlukan dalam pelatihan model, seperti tahun, namatrayek, hargatiket, jumlahpenumpang, basicrate, adjusmentbasicrate, dan biayamaintenance.

Untuk mengetahui hubungan antar variabel dalam dataset, digunakan fungsi corr dari pandas guna melihat korelasi antar kolom. Karena kolom hargatiket adalah target yang akan diprediksi oleh model, maka kolom ini dihapus dari fitur sebelum proses pelatihan dimulai.

```
In [51]: #We are splitting the data in to two parts one is used to train the model and another is used to evaluate the model
         x_train, x_test, y_train, y_test = train_test_split(x, y, test_size=0.2, random_state=101)
```

Gambar 7. Proses pembagian model training dan model testing.

Gambar 7 adalah proses melakukan pembagian data menjadi dua bagian, yaitu data pelatihan (training) sebesar 80% dan data pengujian (*testing*) sebesar 20%. Proses ini dilakukan menggunakan modul sklearn.model_selection. Setelah data terbagi, model RandomForestRegressor diimpor dari library *sklearn.ensemble*, dan dilakukan pelatihan model (*training*) dengan parameter *n_estimators* sebanyak 100 pohon keputusan.

```
In [67]: print('RMSE:', np.sqrt(metrics.mean_squared_error(y_test, predictions))/100)
         print('r2_score:', (metrics.r2_score(y_test, predictions)))

         RMSE: 0.03787392362706947
         r2_score: 0.7748228788271925
```

Gambar 8. Hasil RMSE dan R2 Score.

Gambar 8 adalah hasil pelatihan yang dilakukan , Hasil ini menunjukkan bahwa model menghasilkan nilai *Root Mean Square Error* (*RMSE*) sebesar 3% dan nilai R^2 (*R2 Score*) sebesar 77%, yang menunjukkan bahwa model cukup baik dalam menjelaskan variabilitas data.

```
In [112]: # Calculate mean absolute percentage error (MAPE)
          mape = 100 * (errors / test_labels)
          # Calculate and display accuracy
          accuracy = 100 - np.mean(mape)
          #print('MAPE:', np.mean(mape))
          print('Accuracy:', round(accuracy, 2), '%.')

          Accuracy: 90.44 %.
```

Gambar 9. Hasil akurasi model machine learning.

Gambar 9 adalah hasil dari proses prediction training model. Nilai yang dihasilkan menunjukkan tingkat akurasi sebesar 90,44%.

Model yang telah dilatih ini kemudian digunakan untuk melakukan prediksi tarif tiket, yang hasilnya dibandingkan dengan tarif realisasi aktual. Perbandingan dilakukan dengan mengacu pada kombinasi nilai trayek dan tahun. Berdasarkan hasil perbandingan tersebut, ditemukan bahwa dari total data, terdapat 28 baris data di mana nilai prediksi berada di bawah tarif realisasi, 78 baris data menunjukkan nilai prediksi di atas tarif realisasi, dan hanya 1 trayek yang memiliki nilai prediksi yang sama dengan tarif realisasi. Hal ini menunjukkan bahwa secara umum tarif realisasi yang ditetapkan perusahaan sudah cukup selaras dengan nilai yang dihasilkan oleh model prediksi menggunakan algoritma *Random Forest Regression*.

4.2 Pembuatan Website Prediksi Harga Tiket

Dalam rangka mendukung visualisasi data dan pemodelan prediksi tarif tiket, penulis mengembangkan sebuah website berbasis ASP.NET MVC dengan menggunakan Integrated Development Environment (IDE) Visual Studio 2022. Website ini bertujuan untuk menyajikan hasil analisis data dan prediksi secara interaktif, serta mempermudah pengguna dalam memahami pola-pola historis dan tren harga tiket pada periode high season. Website ini memiliki beberapa menu utama, yaitu *Dashboard* Analisis, *History Trayek* List, History Biaya *Maintenance*, dan Proses *Machine Learning*.

a. Dashboard Analisis

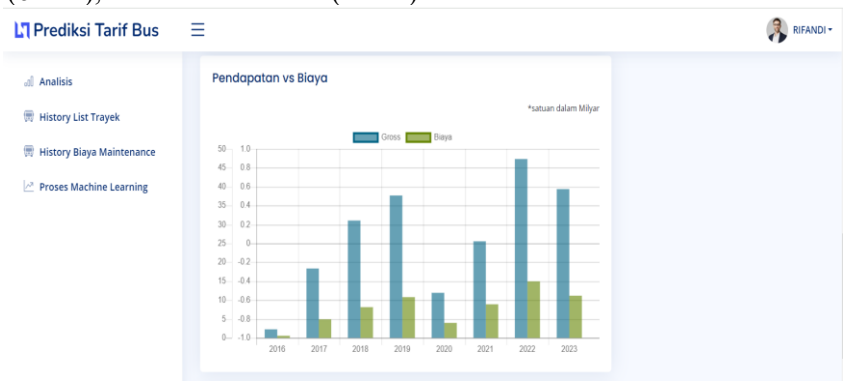
 ISSN:2302-8149
e-ISSN:2540-9719

Menu Dashboard Analisis menyajikan hasil olahan data dari proses pembersihan dan pengelompokan data (*data cleaning*). Informasi yang ditampilkan antara lain: pendapatan (*revenue*) per tahun, harga tiket, serta daftar lima trayek dengan jumlah penumpang terbanyak.



Gambar 10. Dashboard analisis revenue per tahun & Kelas bus

Gambar 10 memperlihatkan grafik yang menggambarkan analisis pendapatan PT. MA selama periode Hari Raya Idul Fitri, Natal, dan Tahun Baru dari tahun 2016 sampai 2022. Terlihat bahwa pendapatan kotor dan bersih mengalami tren kenaikan dari tahun 2016 hingga 2019. Namun, pada tahun 2020, pendapatan tersebut menurun secara signifikan sebagai dampak dari pandemi COVID-19. Setelah itu, pendapatan mulai kembali naik di tahun 2021 dan 2022. Selain itu, data juga menunjukkan jumlah unit bus berdasarkan kelas: eksekutif (287 unit), VIP (270 unit), bisnis (71 unit), super eksekutif (6 unit), dan double decker (4 unit).



Gambar 11. Dashboard analisis gross dan cost per tahun

Gambar 11 menampilkan total penumpang per tahun dari Desember 2016 hingga Juli 2023. Puncak tertinggi jumlah penumpang terjadi pada tahun 2019 dan kembali meningkat di tahun 2022 setelah sempat menurun karena kebijakan PPKM selama pandemi. Grafik ini juga menunjukkan distribusi lintasan bus: utara (melewati tol Trans Jawa), selatan (keluar tol di Pejagan), dan utara DD (lintasan utara khusus untuk bus double decker).



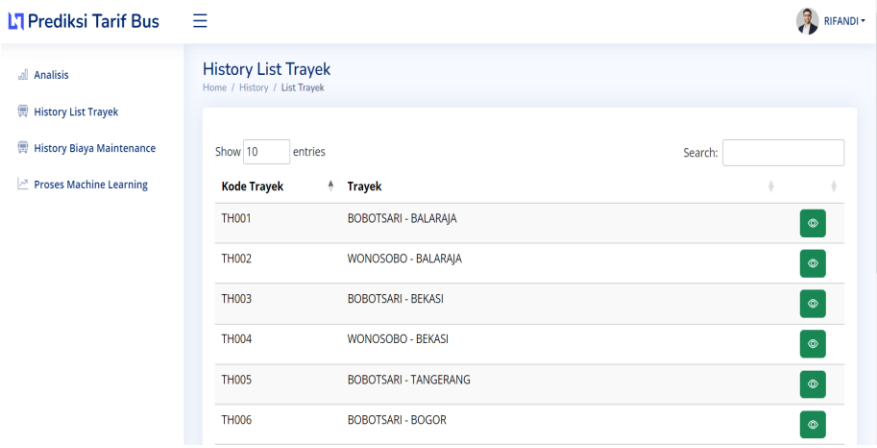
Gambar 12. Dashboard analisis top 5 trayek paling ramai

 ISSN:2302-8149
e-ISSN:2540-9719

Gambar 12 menunjukkan lima trayek dengan jumlah penumpang terbanyak, di mana trayek Yogyakarta – Jakarta menjadi yang paling ramai. Secara keseluruhan, kota Yogyakarta merupakan penyumbang penumpang terbesar selama *periode high season.*

b. Menu History Trayek List

Menu ini menampilkan daftar seluruh trayek yang pernah dioperasikan oleh PT. MA selama periode libur besar. Total terdapat 43 trayek yang beroperasi di wilayah Pulau Jawa. Setiap trayek memiliki empat tab informasi historis:

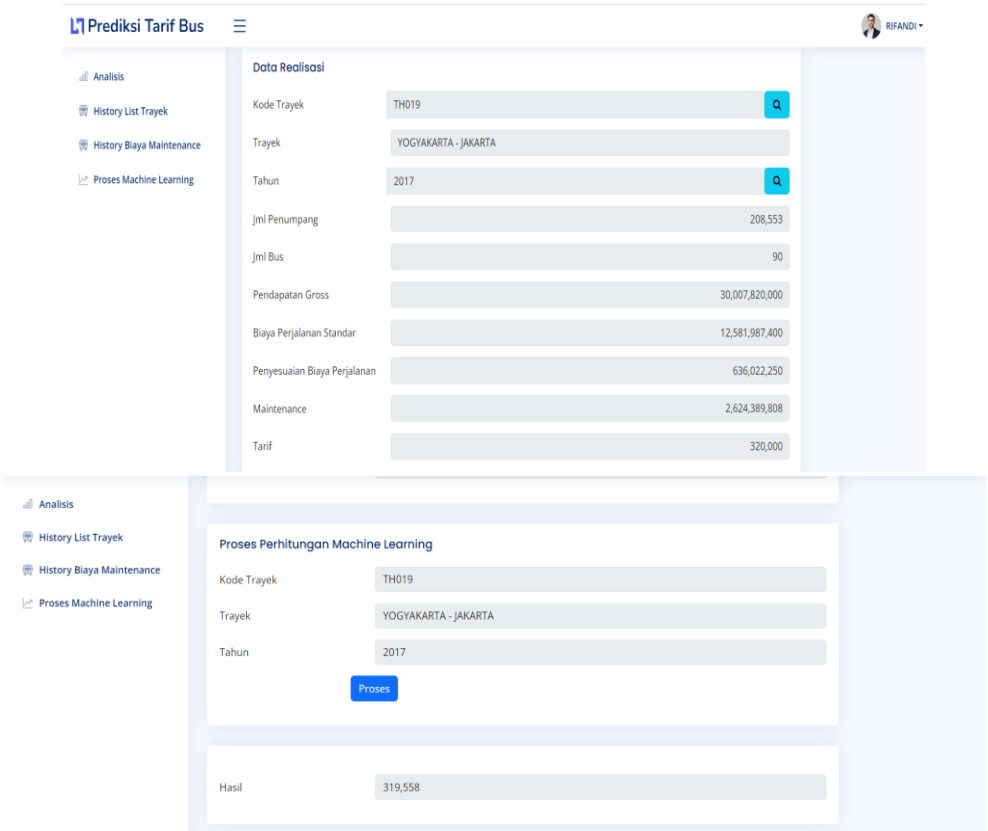


Gambar 13. History trayek list

Gambar 13 adalah list data history trayek yang pernah ada dan digunakan untuk beroperasi selama *periode high season.*

d. Menu Proses *Machine Learning*

Menu ini menyajikan data tarif historis yang digunakan untuk proses prediksi dengan algoritma *Random Forest Regression*. Di dalamnya terdapat fitur untuk melakukan perbandingan antara tarif realisasi dan tarif hasil prediksi model.



Gambar 14. Proses perbandingan harga realisasi dengan harga prediksi

 ISSN:2302-8149
e-ISSN:2540-9719

Gambar 14 menampilkan contoh prediksi tarif pada trayek Yogyakarta – Jakarta untuk tahun 2017. Saat tombol proses ditekan, sistem akan menampilkan hasil prediksi tarif berdasarkan data historis. Pada kasus ini, tarif realisasi adalah sebesar Rp320.000, sedangkan hasil prediksi dari model adalah sebesar Rp319.558 - menunjukkan akurasi yang sangat tinggi. Melalui *website* ini, pengguna tidak hanya dapat melihat histori tarif dan operasional, tetapi juga melakukan simulasi dan analisis prediktif terhadap harga tiket di masa mendatang. Hal ini diharapkan dapat membantu pengambilan keputusan strategis perusahaan secara lebih berbasis data.

5 Kesimpulan dan Saran

Berdasarkan hasil pengujian yang telah dilakukan, dapat disimpulkan bahwa penerapan metode Random Forest pada website prediksi tarif tiket mampu menghasilkan estimasi harga yang sangat mendekati nilai aktual. Model yang dibangun dengan rasio data 80% untuk pelatihan dan 20% untuk pengujian berhasil mencapai nilai RMSE sebesar 3%, R^2 *score* sebesar 77%, dan *MAPE* sebesar 9,56%, dengan akurasi keseluruhan mencapai 90,44%. Hasil ini menunjukkan bahwa model memiliki tingkat akurasi tinggi dan sangat baik dalam melakukan prediksi tarif tiket, sehingga dapat dijadikan dasar pengambilan keputusan harga, khususnya pada periode high season. Untuk pengembangan lebih lanjut, disarankan agar penentuan tarif tidak hanya mengandalkan hasil prediksi model, tetapi juga mempertimbangkan faktor eksternal seperti lonjakan penumpang, biaya operasional, serta perbandingan dengan tarif kompetitor. Dengan menggabungkan analisis prediktif dan faktor pasar aktual, perusahaan dapat membuat keputusan harga yang lebih tepat, kompetitif, dan adaptif terhadap perubahan kondisi.

Referensi

[1] K. Talluri and G. Van Ryzin, "*Revenue Management under a General Discrete Choice Model of Consumer Behavior*," Management Science, Vol. 50, No. 1, pp. 15–33, 2004.

[2] T. M. Oshiro, P. S. Perez, and J. A. Baranauskas, "*How many Trees in a Random Forest?*," *in Machine Learning and Data Mining in Pattern Recognition: 8th International Conference*, MLDM 2012, Berlin, Germany, July 13-20, 2012. Proceedings 8, Springer Berlin Heidelberg, 2012, pp. 154–168.

[3] A. Y. Samudra, "Pendekatan Random *Forest* untuk Model Peramalan Harga Tembakau Rajangan di Kabupaten Temanggung," Bachelor Thesis, Universitas Sanata Dharma, Yogyakarta, 2019.

[4] M. J. Wulansari, "Analisis Faktor-faktor yang mempengaruhi Seseorang terkena Penyakit Diabetes Melitus menggunakan *Regresi Random Forest* (Studi Kasus: Data Diabetes di Virginia Amerika Serikat)," 2018.

[5] N. Nur, F. Wajidi, S. Sulfayanti, and W. Wildayani, "Implementasi *Algoritma Random Forest Regression* untuk memprediksi Hasil Panen Padi di Desa Minanga," Jurnal Komputer Terapan, Vol. 9, No. 1, pp. 58–64, 2023.

[6] C. Haryanto, N. Rahaningsih, and F. M. Basysyar, "Komparasi Algoritma *Machine Learning* dalam memprediksi Harga Rumah," JATI (Jurnal Mahasiswa Teknik Informatika), Vol. 7, No. 1, pp. 533–539, 2023.

[7] S. Fachid and A. Triayudi, "Perbandingan Algoritma Regresi Linier dan Regresi *Random Forest* dalam memprediksi Kasus Positif Covid-19," Jurnal Media Informatika Budidarma, Vol. 6, No. 1, pp. 68–73, 2022.

[8] Y. A. Zebua et al., "Prediksi Penetapan Tarif Penerbangan menggunakan Auto-Ml dengan *Algoritma Random Forest*," Jurnal Tekinkom (Teknik Informasi dan Komputer), Vol. 5, No. 1, pp. 115–122, 2022.

[9] M. L. Suliztia, "Penerapan Analisis *Random Forest* pada *Prototype* Sistem Prediksi Harga Kamera Bekas menggunakan *Flask*," 2020.

[10] I. Reis, D. Baron, and S. Shahaf, "*Probabilistic Random Forest: A Machine Learning Algorithm for Noisy Data Sets*," The Astronomical Journal, Vol. 157, No. 1, pp. 1–12, 2018.

[11] B. F. Darst, K. C. Malecki, and C. D. Engelman, "*Using Recursive Feature Elimination in Random Forest to Account for Correlated Variables in High Dimensional Data*," BMC Genetics, Vol. 19, No. 1, pp. 1–6, 2018.

[12] C. Luong and N. Dokuchaev, "*Forecasting of Realised Volatility with the Random Forests*

 ISSN:2302-8149
e-ISSN:2540-9719

Algorithm," Journal of Risk and Financial Management, Vol. 11, No. 4, pp. 1–15, 2018, doi: 10.3390/jrfm11040061.
[13] B. M. Pavlyshenko, "*Machine-Learning Models for Sales Time Series Forecasting*," Data, Vol. 4, No. 1, p. 15, 2019, doi: 10.3390/data4010015.
[14] R. R. Abadi, I. Nursyamsi, and A. R. Syamsuddin, "*Effect of Customer Value and Experiential Marketing to Customer Loyalty with Customer Satisfaction as Intervening Variable (Case Study on Gojek Makassar Consumers)*," The Asian Journal of Technology Management, Vol. 13, No. 1, pp. 82–97, 2020.
[15] M. Huang and R. T. Rust, "*Artificial Intelligence in Service*," Journal of Service Research, pp. 2–6, 2018, doi: 10.1177/1094670517752459.
[16] S. Russel and P. Norvig, *Artificial Intelligence: A Modern Approach*, New Jersey: CreateSpace Independent Publishing Platform, 2010.