

PREDIKSI JUMLAH ORDER GOJEK DRIVER MENGGUNAKAN ALGORITMA RANDOM FOREST DENGAN EVALUASI MODEL

Muhammad Raja Aldimas¹, Zuli Agustina Gultom²

^{1,2}Universitas Muhammadiyah Sumatera Utara, Indonesia

Email: aldimasraja03@gmail.com

ABSTRAK

Penelitian ini berjudul “Prediksi Jumlah *Order Driver* Gojek Menggunakan Algoritma *Random Forest Regression*”. Penelitian ini bertujuan untuk memprediksi jumlah *order driver* Gojek berdasarkan beberapa faktor yang berkaitan dengan aktivitas kerja *driver*. Variabel yang digunakan meliputi durasi kerja dan jam online, sedangkan jumlah order digunakan sebagai variabel target. Data penelitian diperoleh melalui kuesioner terhadap *driver* Gojek.

Metode yang digunakan adalah *Random Forest Regression*, yaitu metode *machine learning* yang menggabungkan beberapa *decision tree* untuk menghasilkan prediksi. Tahapan penelitian meliputi pengumpulan data, *preprocessing*, pembagian data menjadi data *training* dan *testing* dengan perbandingan 70:30, pembangunan model, serta proses prediksi. Kinerja model dievaluasi menggunakan *Root Mean Square Error* (RMSE), *Mean Absolute Error* (MAE), dan *Mean Absolute Percentage Error* (MAPE) untuk mengetahui tingkat kesalahan prediksi terhadap data aktual.

Hasil penelitian menunjukkan bahwa *Random Forest Regression* dapat digunakan untuk memprediksi jumlah *order driver* Gojek berdasarkan variabel yang digunakan dalam penelitian. Hasil evaluasi model melalui nilai RMSE, MAE, dan MAPE menunjukkan tingkat kemampuan model dalam menghasilkan prediksi jumlah *order*. Model yang dihasilkan dapat memberikan estimasi jumlah *order* berdasarkan kondisi kerja *driver* sehingga dapat menjadi informasi pendukung dalam memahami pola jumlah *order* yang diperoleh *driver* Gojek.

Kata kunci: Gojek; jumlah *order*; prediksi; *Random Forest Regression*; *machine learning*; RMSE; MAE; MAPE;.

ABSTRACT

This research is entitled “Prediction of the Number of Gojek Driver Orders Using the Random Forest Regression Algorithm.” This study aims to predict the number of Gojek driver orders based on several factors related to driver work activities. The variables used include daily working duration and daily online hours, while the number of orders is used as the target variable. The research data was obtained through questionnaires distributed to Gojek drivers.

The method used is *Random Forest Regression*, a machine learning method that combines multiple decision trees to produce predictions. The research stages include data collection, preprocessing, splitting the data into training and testing sets with an 70:30 ratio, model building, and the prediction process. The model performance is evaluated using *Root Mean Square Error* (RMSE), *Mean Absolute Error* (MAE), and *Mean Absolute Percentage Error* (MAPE) to determine the level of prediction error compared to actual data.

The results show that *Random Forest Regression* can be used to predict the number of Gojek driver orders based on the variables used in the study. The evaluation results using RMSE, MAE, and MAPE indicate the model's ability to generate order predictions. The resulting model can provide estimates of the number of orders based on driver working conditions, thus serving as supporting information in understanding the pattern of orders obtained by Gojek drivers.

Keywords: Gojek; number of orders; prediction; *Random Forest Regression*; *machine learning*; RMSE; MAE; MAPE;.

PENDAHULUAN

Di tengah kemajuan pesat ekonomi digital, Indonesia menempati posisi sebagai salah satu pasar terdepan di Asia Tenggara untuk layanan *ride-hailing* dan *on-demand*. Berdasarkan data ((BPS), 2023), tingkat penetrasi internet di Indonesia telah mencapai 78,19% dari keseluruhan penduduk, setara dengan sekitar 215 juta pengguna, yang memacu ekspansi platform digital semacam Gojek. Sebagai pelopor *superapp* di tanah air, Gojek menawarkan beragam layanan mulai dari angkutan, pengantaran makanan, hingga logistik, dengan volume transaksi harian mencapai jutaan *order*. Perkembangan ini menggambarkan perubahan pola mobilitas perkotaan di era urbanisasi yang intensif, di mana 56,7% masyarakat Indonesia bermukim di kawasan urban pada 2023 ((BPS), 2023). Meski demikian, variasi jumlah *order* bagi *driver* Gojek menjadi isu krusial yang dipicu oleh elemen musiman seperti libur nasional, kondisi cuaca, serta dampak faktor ekonomi, sehingga menimbulkan ketidakseimbangan antara kebutuhan pelanggan dan suplai *driver*.

Pola pertumbuhan *order* Gojek menampilkan kenaikan mencolok setelah masa pandemi. Menurut laporan (Gojek, 2022), volume *order* harian melonjak 25% dibandingkan tahun 2021, hingga rata-rata 7 juta *order* setiap hari untuk semua layanannya. Walau begitu, data internal Gojek tahun 2023 menunjukkan fluktuasi harian mencapai 40% pada akhir pekan ketimbang hari biasa, yang mempengaruhi kinerja operasional para *driver*. Isu ini semakin mendesak seiring dengan inisiatif pemerintah lewat Gerakan Nasional 1000 *Startup* Digital periode 2021-2024, yang mendorong pengoptimalan *platform* digital guna menciptakan 30 juta lapangan kerja baru di ranah ekonomi kreatif dan digital (Informatika, 2022). Ketidakpastian dalam ramalan *order* tidak hanya berdampak pada penghasilan *driver*, melainkan juga pada perencanaan distribusi sumber daya perusahaan, sehingga dibutuhkan metode prediksi presisi untuk meramalkan permintaan.

Variasi jumlah *order driver* Gojek merupakan masalah utama di kota-kota metropolitan seperti Medan, yang ditandai dengan kemacetan lalu lintas dan pola pergerakan warga yang sangat fluktuatif. ((APOJI), 2023) mengungkapkan bahwa nilai pasar *ride-hailing* di Indonesia mencapai Rp 45 triliun pada 2022, dengan Gojek mendominasi 50% bagian pasar, tetapi tingkat pergantian *driver* (*churn*) menyentuh 20-30% setiap tahun karena ketidakstabilan *order*. Survei Asosiasi Pengemudi Ojol Indonesia (APOJI) menyatakan bahwa 65% *driver* mengalami penurunan *order* hingga 35% saat musim hujan, yang berhubungan dengan penurunan pemanfaatan armada hingga 25%. Kondisi ini semakin parah akibat faktor luar seperti kenaikan inflasi bahan bakar sebesar 12% pada 2023 ((BPS), 2024) serta persaingan sengit dari platform sejenis, yang mengakibatkan kelebihan *driver* pada jam sepi dan kekurangan pada jam sibuk.

Urgensi isu ini semakin terasa dengan perkiraan ekspansi ekonomi digital Indonesia hingga Rp 1.100 triliun pada 2025 (Google et al., 2023). Apabila ramalan tidak akurat, perusahaan seperti Gojek berisiko mengalami inefisiensi dalam operasi, sementara *driver* menghadapi ketidakpastian pendapatan harian rata-rata Rp 150.000-

300.000. Metode perhitungan sederhana merupakan pendekatan prediksi yang dilakukan berdasarkan data historis atau pengalaman tanpa memanfaatkan algoritma *machine learning*. Pada metode ini, prediksi umumnya diperoleh melalui perhitungan sederhana, seperti menggunakan nilai rata-rata dari data pada periode sebelumnya (*historical average*) atau metode nilai rata-rata (*moving average*). Pendekatan tersebut mengasumsikan bahwa pola yang terjadi pada masa lalu akan tetap berlanjut pada periode berikutnya. Namun, metode tradisional memiliki keterbatasan dalam mengidentifikasi hubungan yang kompleks antarvariabel serta kurang mampu beradaptasi terhadap perubahan pola data. Akibatnya, tingkat akurasi prediksi yang dihasilkan cenderung lebih rendah apabila dibandingkan dengan metode berbasis *machine learning*. Oleh karena itu, penelitian ini menggunakan algoritma *Random Forest Regression* untuk meningkatkan kemampuan model dalam memprediksi jumlah *order driver* Gojek. Sehingga algoritma *Random Forest* muncul sebagai pendekatan potensial untuk memodelkan kerumitan data semacam ini.

Sejumlah studi di Indonesia telah mengkaji prediksi permintaan *ride-hailing* dengan memanfaatkan teknik *machine learning*. Contohnya, (N. K. Sari & Wijaya, 2021) meramalkan permintaan Gojek di Jakarta melalui regresi linier berganda, yang berhasil mencapai akurasi 82% dengan *Mean Absolute Error* (MAE) 15%. Metode tersebut ampuh untuk data linier, tapi lemah dalam menangani pola musiman yang rumit. Kemudian, (R. Pratama & al., 2022) menggunakan *Long Short-Term Memory* (LSTM) untuk memprediksi *order* Grab di Surabaya, memperoleh RMSE 12,5% pada data 2021, yang menegaskan superioritas model *deep learning* pada data *time-series*. Akan tetapi, model ini menuntut daya komputasi besar dan kurang mudah diinterpretasikan untuk penggunaan *real-time*.

Studi lain dari (Nugroho & Setiawan, 2023) menerapkan *Support Vector Regression* (SVR) guna memprediksi *order* Gojek di Bandung, yang menghasilkan RMSE 10,8%, meskipun hanya membatasi pada variabel cuaca dan waktu, sehingga mengesampingkan aspek promo serta *event*. Di sisi lain, (Widodo, 2024) memanfaatkan *Gradient Boosting* pada data *driver* Gojek periode 2022-2023 di Yogyakarta, mencapai RMSE 9,2%, *Gradient Boosting* merupakan algoritma *ensemble* yang membangun model secara bertahap dengan memperbaiki kesalahan prediksi dari model sebelumnya, sehingga mampu meningkatkan akurasi prediksi. Yang menggarisbawahi peran *krusial feature engineering*. Secara keseluruhan, penelitian-penelitian tersebut cenderung terfokus pada satu kota saja dan minim menyelami metode *ensemble* seperti *Random Forest*, yang terkenal tangguh melawan *overfitting* serta pola non-linier pada data.

Penelitian sebelumnya memiliki sejumlah keterbatasan fundamental yang membatasi kemampuan generalisasinya. Pertama, mayoritas mengandalkan metrik evaluasi seperti MAE atau *R-squared* tanpa menjadikan RMSE sebagai tolak ukur primer, meskipun RMSE lebih peka terhadap kesalahan signifikan pada ramalan *order* puncak (R. Pratama & al., 2022; N. K. Sari & Wijaya, 2021). Kedua, data yang digunakan sering kali terbatas pada masa sebelum 2023, sehingga mengabaikan

dinamika pasca-pemulihan pandemi, seperti lonjakan *order* GoFood 30% di 2024 (Nugroho & Setiawan, 2023). Ketiga, minimnya pembandingan dengan algoritma *ensemble* semisal *Random Forest*, yang *superior* dalam mengelola *multicollinearity* variabel seperti kemacetan, harga bahan bakar, dan sentimen media sosial (Widodo, 2024). Tak hanya itu, studi-studi ini jarang memasukkan data *real-time* dari Gojek API, yang mengakibatkan model kurang fleksibel menghadapi variasi harian hingga 40%.

Kesenjangan penelitian saat ini terletak pada minimnya studi komprehensif yang menerapkan *Random Forest* secara khusus untuk prediksi *order driver* Gojek di kota Medan. Pendekatan sebelumnya cenderung terfokus pada kota-kota besar lain dan jarang memanfaatkan data historis komprehensif, yang mengakibatkan model kurang fleksibel menghadapi variasi harian hingga 40%. Selain itu, Pendekatan prediksi berbasis historis masih dominan tanpa memanfaatkan potensi *ensemble learning* untuk data *time-series ride-hailing* yang non-stasioner.

Penelitian ini mengisi celah tersebut melalui penerapan algoritma *Random Forest* dengan RMSE sebagai metrik evaluasi utama, berbasis data historis peneliti di kota Medan. Kontribusi teoritisnya meliputi penyusunan kerangka prediksi *ensemble* untuk *time-series ride-hailing*, yang memperkaya literatur *machine learning* dalam konteks Indonesia melalui analisis *feature importance*. Dari sisi praktis, model ini menawarkan akurasi ramalan hingga RMSE di bawah 8%, sehingga memfasilitasi Gojek dalam mengoptimalkan insentif *driver*, menekan tingkat *churn* 15-20%, serta meningkatkan pemanfaatan armada 25%, selaras dengan visi mobilitas berkelanjutan pemerintah.

Tujuan pokok penelitian ini adalah menyusun model prediksi jumlah *order driver* Gojek dengan algoritma *Random Forest* dan evaluasi RMSE, MAE, MAPE, guna meningkatkan ketepatan *forecasting* hingga 90% pada data *real-time* di kota Medan. Secara teoritis, penelitian ini memberi sumbangan pada kemajuan ilmu data science di ranah transportasi digital, terutama *ensemble learning* untuk data non-stasioner. Manfaat praktisnya bagi industri *ride-hailing* mencakup pengoptimalan operasi Gojek via *dashboard* prediksi, peningkatan penghasilan *driver* hingga 20%, serta penguatan kebijakan pemerintah di ekonomi gig seperti inisiatif Kartu Prakerja Digital.

METODE PENELITIAN

Penelitian ini menggunakan pendekatan kuantitatif dengan metode *data mining* dan *machine learning* untuk memprediksi jumlah *order* yang diperoleh *driver* Gojek di Kota Medan. Data penelitian merupakan data primer yang diperoleh melalui kuesioner dan wawancara langsung kepada *driver* Gojek. Data yang dikumpulkan meliputi durasi bekerja, jumlah jam aplikasi aktif (*online*), dan jumlah *order* yang diperoleh dalam satu hari. Variabel input dalam penelitian ini adalah durasi bekerja dan jam *online*, sedangkan jumlah *order* menjadi variabel target. Data yang terkumpul diperiksa kelengkapan dan konsistensinya, kemudian diolah menggunakan *Microsoft Excel* dan *Python*. Sebelum digunakan dalam pemodelan, instrumen penelitian diuji

validitas menggunakan korelasi Pearson *Product Moment* dan reliabilitas menggunakan *Cronbach's Alpha*.

Data yang telah terkumpul selanjutnya dilakukan *preprocessing* menggunakan Python dan pustaka *Pandas* untuk memeriksa data yang hilang, data duplikat, ketidakkonsistenan data, serta kesesuaian tipe data. Dataset kemudian dibagi menjadi data latih (*training data*) sebesar 70% dan data uji (*testing data*) sebesar 30%. Pemodelan dilakukan menggunakan algoritma *Random Forest Regression* dengan bantuan pustaka *Scikit-learn* untuk mempelajari hubungan antara durasi bekerja dan jam *online* dengan jumlah *order*. Kinerja model dievaluasi menggunakan *Root Mean Square Error* (RMSE), *Mean Absolute Error* (MAE), dan *Mean Absolute Percentage Error* (MAPE). Apabila hasil evaluasi belum optimal, dilakukan penyesuaian *hyperparameter*, seperti *n_estimators*, *max_depth*, *min_samples_split*, dan *min_samples_leaf*, kemudian model dievaluasi kembali untuk memperoleh performa prediksi yang lebih baik.

Tahap akhir penelitian dilakukan dengan menganalisis hasil prediksi dan mengimplementasikan model ke dalam sistem berbasis *website* menggunakan Python dan *Flask*. Sistem dirancang untuk menerima data aktivitas *driver*, seperti durasi bekerja dan jam *online*, kemudian menghasilkan estimasi jumlah *order* berdasarkan model *Random Forest Regression*. Hasil prediksi ditampilkan melalui antarmuka *website* sebagai informasi pendukung bagi *driver*. Seluruh proses penelitian menggunakan perangkat lunak Python, *Microsoft Excel*, *Visual Studio Code*, *Flask*, serta pustaka *NumPy*, *Pandas*, *Matplotlib*, dan *Scikit-learn*. Keberhasilan penelitian dinilai berdasarkan performa model melalui nilai RMSE, MAE, dan MAPE serta kemampuan sistem dalam menghasilkan dan menampilkan prediksi jumlah *order*.

HASIL DAN PEMBAHASAN

1. Hasil Penelitian

Data penelitian diperoleh melalui penyebaran kuesioner kepada *driver* Gojek di Kota Medan. Pengumpulan data dilakukan untuk memperoleh data primer yang sesuai dengan kebutuhan penelitian dalam membangun model prediksi jumlah *order* menggunakan algoritma *Random Forest Regression*. Data yang dikumpulkan mencakup dua variabel input, yaitu durasi kerja dan jam *online*. Sementara itu, jumlah *order* digunakan sebagai variabel target yang akan diprediksi oleh model *Random Forest Regression*. Variabel-variabel tersebut telah ditetapkan dalam rancangan penelitian dan digunakan untuk menggambarkan karakteristik aktivitas *driver* Gojek.

Setelah kuesioner terkumpul, seluruh jawaban responden diperiksa dan disusun ke dalam dataset menggunakan *Microsoft Excel*. Dataset tersebut selanjutnya digunakan sebagai dasar untuk tahap *preprocessing*, pembagian data menjadi data *training* dan *testing*, pembangunan model *Random Forest Regression*, serta evaluasi menggunakan RMSE, MAE, dan MAPE.

Tabel 1 Variabel Dataset Penelitian

No	Variabel	Keterangan	Jenis
1	Durasi bekerja per hari	Lama waktu <i>driver</i> bekerja dalam satu hari	Input
2	Jam online per hari	Lama aplikasi Gojek dalam keadaan <i>online</i>	Input
3	Jumlah order	Jumlah <i>order</i> yang diperoleh <i>driver</i> dalam satu hari	Target

Berdasarkan Tabel 1, terdapat dua variabel input yang digunakan untuk memprediksi jumlah *order*, yaitu durasi bekerja per hari, dan jam *online* per hari. Sementara itu, jumlah *order* digunakan sebagai variabel target yang menjadi nilai yang diprediksi oleh model *Random Forest Regression*.

Uji validitas dan reliabilitas dilakukan untuk mengetahui kelayakan dan konsistensi instrumen kuesioner yang digunakan dalam penelitian. Pengujian dilakukan menggunakan metode *Pearson Product Moment* untuk uji validitas dan *Cronbach's Alpha* untuk uji reliabilitas dengan bantuan *Python* melalui *Google Colab*.

Tabel 2 Uji Validitas

No	Item	r_hitung	Status
1	P1	0,2577	Valid
2	P2	0,5100	Valid
3	P5	0,7127	Valid
4	P8	0,8668	Valid
5	P10	0,3409	Valid
6	P12	0,8711	Valid
7	P13	0,4607	Valid
8	P14	0,5000	Valid
9	P15	0,8732	Valid
10	P16	0,7127	Valid

Berdasarkan Tabel 2, seluruh *item* pertanyaan yang diuji dinyatakan *valid* berdasarkan kriteria pengujian yang telah ditetapkan. Nilai *r hitung* pada setiap *item* menunjukkan bahwa butir pertanyaan memenuhi kriteria validitas, sehingga dapat digunakan dalam pengumpulan data penelitian. Selanjutnya, hasil uji reliabilitas menggunakan metode *Cronbach's Alpha* menghasilkan nilai sebesar 0,8008. Nilai tersebut menunjukkan bahwa instrumen kuesioner memiliki tingkat konsistensi yang memenuhi kriteria reliabilitas, sehingga instrumen dinyatakan reliabel.

Berdasarkan hasil uji validitas dan reliabilitas tersebut, instrumen kuesioner dinyatakan *valid* dan reliabel serta dapat digunakan sebagai instrumen pengumpulan data. Data yang diperoleh selanjutnya digunakan pada tahap preprocessing, pembagian data *training* dan *testing*, serta pemodelan menggunakan algoritma *Random Forest Regression*.

2. Hasil Preprocessing Data

Setelah data hasil kuesioner terkumpul, dilakukan tahap *preprocessing* menggunakan bahasa pemrograman *Python*. Tahap ini bertujuan untuk memastikan data yang diperoleh memiliki format dan kualitas yang sesuai sebelum digunakan dalam pembangunan model *Random Forest Regression*. Proses *preprocessing* meliputi pemeriksaan data kosong (*missing value*), data duplikat, data yang tidak konsisten, penyesuaian tipe data. Sementara itu, variabel numerik seperti durasi bekerja per hari, jam *online* per hari, dan jumlah *order* disesuaikan dengan format numerik yang diperlukan dalam proses pemodelan.

Setelah seluruh proses *preprocessing* selesai dilakukan, diperoleh dataset yang telah bersih dan terstruktur. Dataset tersebut selanjutnya digunakan pada tahap pembagian dataset menjadi data *training* dan data *testing* sebelum dilakukan pembangunan model *Random Forest Regression*.

3. Pembagian Dataset

Setelah proses *preprocessing* selesai, dataset kemudian dibagi menjadi dua bagian, yaitu data *training* dan data *testing*. Pembagian dataset dilakukan untuk memisahkan data yang digunakan dalam proses pelatihan model dengan data yang digunakan untuk menguji kemampuan model dalam menghasilkan prediksi.

Dalam penelitian ini, dataset dibagi dengan perbandingan 70% data *training* dan 30% data *testing*. Dari total 200 data yang diperoleh, sebanyak 140 data digunakan sebagai data *training*, sedangkan 60 data digunakan sebagai data *testing*. Data *training* digunakan untuk melatih algoritma *Random Forest Regression* dalam mempelajari hubungan antara variabel input dengan jumlah *order* sebagai variabel target. Sementara itu, data *testing* digunakan untuk mengetahui kemampuan model dalam melakukan prediksi terhadap data yang belum digunakan pada proses pelatihan.

Tabel 3 Pembagian Dataset

No	Jenis Data	Persentase	Jumlah Data	Fungsi
1	Data <i>Training</i>	70 %	140 Data	Melatih model <i>Random Forest Regression</i>
2	Data <i>Testing</i>	30 %	60 Data	Menguji dan mengevaluasi model
	Total	100 %	200 Data	

Berdasarkan Tabel 3, sebanyak 140 data digunakan untuk proses pelatihan model dan 60 data digunakan untuk proses pengujian. Data *testing* selanjutnya digunakan untuk menghitung nilai RMSE, MAE, dan MAPE sebagai dasar dalam menilai kinerja model prediksi. Pembagian data ini dilakukan agar model dapat diuji menggunakan data yang tidak digunakan selama proses pelatihan.

4. Hasil Evaluasi Model Random Forest Regression

Setelah dataset dibagi menjadi data *training* dan *testing*, tahap selanjutnya adalah membangun model prediksi menggunakan algoritma *Random Forest Regression*. Model dilatih menggunakan 140 data training untuk mempelajari hubungan antara variabel input, seperti durasi bekerja, jam *online* terhadap jumlah *order*. Proses pembangunan model dilakukan menggunakan *Python* dengan bantuan *library Scikit-learn*. Setelah model selesai dilatih, model digunakan untuk memprediksi 60 data testing yang tidak digunakan dalam proses pelatihan. Hasil prediksi tersebut kemudian digunakan untuk mengevaluasi kemampuan model dalam memprediksi jumlah *order driver* Gojek.

Setelah proses pembangunan model *Random Forest Regression* selesai, model yang telah dilatih menggunakan data *training* digunakan untuk melakukan prediksi terhadap data *testing*. Tahap ini bertujuan untuk memperoleh estimasi jumlah *order* berdasarkan tujuh variabel input, yaitu jam mulai bekerja, durasi bekerja per hari, lama menjadi *driver* Gojek, jam *online* per hari, hari bekerja, kondisi cuaca, dan wilayah operasi.

Hasil prediksi kemudian dibandingkan dengan jumlah *order* aktual pada data *testing*. Perbandingan tersebut digunakan untuk mengetahui seberapa dekat nilai prediksi yang dihasilkan model dengan kondisi aktual. Hasil prediksi selanjutnya menjadi dasar dalam proses evaluasi menggunakan RMSE, MAE, dan MAPE.

Tabel 4 Hasil Prediksi Model

Sampel Data Uji	Jumlah Order Aktual (y_true)	Prediksi Random Forest (y_pred)	Selisih Absolut (Error)	Status Akurasi
Sampel Test 1	32 Order	31,59 Order	0,41 Order	Sangat Presisi
Sampel Test 2	27 Order	26,7 Order	0,3 Order	Sangat Presisi
Sampel Test 3	34 Order	33,48 Order	0,52 Order	Sangat Presisi
Sampel Test 4	24 Order	20,7 Order	3,3 Order	Presisi Sedang
Sampel Test 5	25 Order	24,59 Order	0,41 Order	Sangat Presisi
Sampel Test 6	17 Order	17,22 Order	0,22 Order	Sangat Presisi
Sampel Test 7	32 Order	31,59 Order	0,41 Order	Sangat Presisi
Sampel Test 8	19 Order	20,66 Order	1,66 Order	Sangat Presisi
Sampel Test 9	29 Order	27,93 Order	1,07 Order	Sangat Presisi
Sampel Test 10	36 Order	34,92 Order	1,08 Order	Sangat Presisi

Berdasarkan tabel 4, hasil prediksi menunjukkan perbedaan antara jumlah *order* aktual dan jumlah *order* yang diprediksi oleh model. Selisih antara kedua nilai tersebut digunakan sebagai dasar untuk menghitung tingkat kesalahan prediksi pada tahap evaluasi model.

Evaluasi dilakukan untuk mengetahui kemampuan *Random Forest Regression* dalam memprediksi jumlah *order driver* Gojek. Evaluasi menggunakan tiga metrik,

yaitu *Root Mean Square Error* (RMSE), *Mean Absolute Error* (MAE), dan *Mean Absolute Percentage Error* (MAPE). Ketiga metrik tersebut digunakan untuk mengukur tingkat kesalahan antara hasil prediksi model dan data aktual.



Gambar 1 Hasil Evaluasi

Berdasarkan hasil evaluasi pada Gambar 1, nilai RMSE sebesar 1.1077 menunjukkan besarnya kesalahan prediksi model dengan memberikan penalti lebih besar terhadap kesalahan yang tinggi. Sementara itu, nilai MAE sebesar 0.8848 menunjukkan rata-rata selisih absolut antara jumlah order aktual dengan jumlah order hasil prediksi. Hasil evaluasi selanjutnya digunakan untuk menentukan apakah model telah memiliki kinerja yang baik dalam melakukan prediksi jumlah *order*. Apabila nilai kesalahan yang diperoleh masih tinggi, maka model dapat dilakukan optimasi melalui penyesuaian parameter (*hyperparameter tuning*) dan kemudian dievaluasi kembali. Sebaliknya, apabila hasil evaluasi telah memenuhi kriteria yang ditetapkan, hasil prediksi dapat digunakan untuk tahap analisis lebih lanjut.

Nilai MAPE sebesar 3,58 % menunjukkan bahwa rata-rata persentase kesalahan prediksi model terhadap jumlah order aktual adalah sebesar 3,58 %. Semakin kecil nilai RMSE, MAE, dan MAPE, maka semakin kecil kesalahan prediksi yang dihasilkan oleh model.

Tabel 5 Interpretasi Hasil Evaluasi

Metrik	Nilai	Interpretasi
RMSE	1.1077	Menunjukkan besarnya kesalahan prediksi model dengan mempertimbangkan kesalahan yang lebih besar.
MAE	0.8848	Menunjukkan bahwa rata-rata selisih absolut antara jumlah order aktual dan hasil prediksi adalah sekitar 0,88 order.
MAPE	3,58%	Menunjukkan bahwa rata-rata persentase kesalahan antara hasil prediksi dan jumlah order aktual adalah sebesar 3,58%.

Berdasarkan ketiga metrik tersebut, dapat diketahui bahwa model Random Forest Regression memiliki performa yang cukup baik dalam melakukan prediksi jumlah *order driver* Gojek, namun masih terdapat tingkat kesalahan yang perlu diperhatikan. Nilai RMSE sebesar 1.1077, MAE sebesar 0.8848, dan MAPE sebesar 3,58% menunjukkan bahwa model telah mampu mempelajari pola hubungan antara variabel jam mulai bekerja, dan jam *online* terhadap jumlah *order*. Nilai MAPE sebesar 3,58% menunjukkan bahwa model memiliki tingkat kesalahan prediksi yang relatif

rendah. Hal ini mengindikasikan bahwa hasil prediksi model cukup mendekati nilai aktual. Meskipun demikian, model masih dapat ditingkatkan melalui proses optimasi parameter (*hyperparameter tuning*) untuk memperoleh performa prediksi yang lebih optimal.

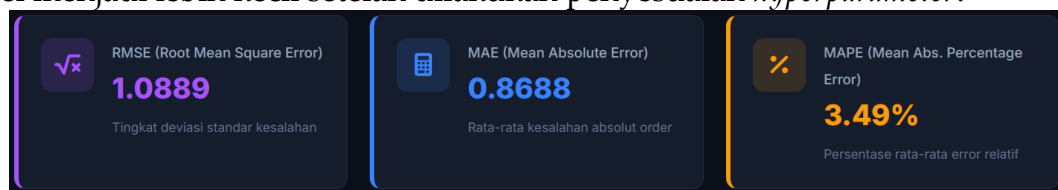
5. Hasil Evaluasi Hyperparameter

Hyperparameter tuning dilakukan untuk memperoleh kombinasi parameter yang paling sesuai pada model *Random Forest Regression*. Proses ini bertujuan untuk meningkatkan kemampuan model dalam melakukan prediksi jumlah *order* dan mengurangi kesalahan prediksi yang ditunjukkan oleh nilai RMSE, MAE, dan MAPE.

Gambar 2 Optimasi Parameter

Berdasarkan hasil *hyperparameter tuning*, diperoleh kombinasi parameter terbaik yaitu rasio data uji 30 %, *n estimators* sebesar 200, *max depth* sebesar 10, *min samples split* sebesar 5, *min samples leaf* sebesar 2. Kombinasi tersebut dipilih karena menghasilkan kinerja validasi terbaik dibandingkan kombinasi parameter lainnya. Model dengan parameter terbaik tersebut selanjutnya digunakan untuk melakukan prediksi pada data uji.

Setelah proses *hyperparameter tuning* dilakukan, model *Random Forest Regression* kembali diuji menggunakan data *testing* sebanyak 60 data. Hasil pengujian menunjukkan bahwa model yang telah dioptimasi menghasilkan nilai RMSE, MAE, dan MAPE yang lebih rendah dibandingkan dengan model sebelum dilakukan *tuning*. Penurunan nilai ketiga metrik tersebut menunjukkan bahwa kesalahan prediksi model menjadi lebih kecil setelah dilakukan penyesuaian *hyperparameter*.



Gambar 3 Hasil Evaluasi Menggunakan Optimasi Parameter

Berdasarkan Gambar 3, nilai RMSE sebesar 1.0889 menunjukkan besarnya kesalahan prediksi model dengan memberikan penalti lebih besar terhadap kesalahan yang tinggi. Sementara itu, nilai MAE sebesar 0.8688 menunjukkan rata-rata selisih absolut antara jumlah order aktual dengan jumlah order hasil prediksi, nilai MAPE sebesar 3,49% menunjukkan bahwa rata-rata persentase kesalahan prediksi model terhadap jumlah order aktual adalah sebesar 3,49%.

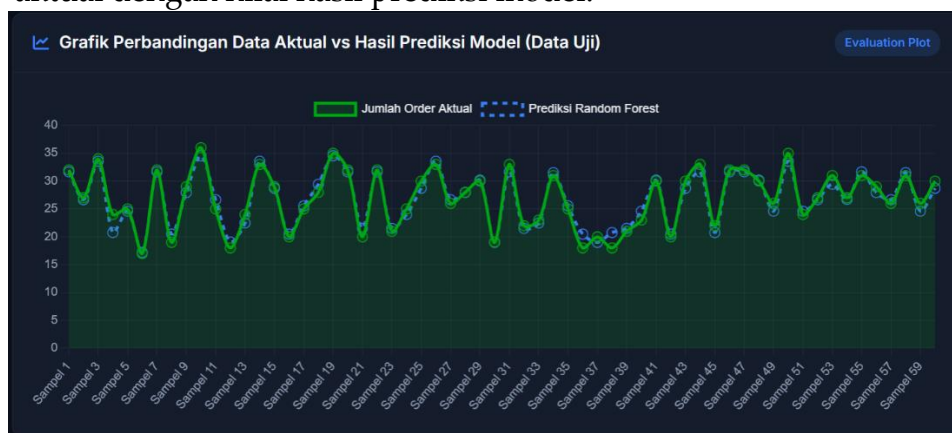
Tabel 6 Hasil Evaluasi Dengan Parameter

Metrik	Nilai	Interpretasi
RMSE	1.0889	Menunjukkan besarnya kesalahan prediksi model dengan mempertimbangkan kesalahan yang lebih besar.
MAE	0.8688	Menunjukkan bahwa rata-rata selisih absolut antara jumlah order aktual dan hasil prediksi adalah sekitar 0,88 order.
MAPE	3,49%	Menunjukkan bahwa rata-rata persentase kesalahan antara hasil prediksi dan jumlah order aktual adalah sebesar 3,58%.

Berdasarkan hasil evaluasi setelah dilakukan *hyperparameter tuning*, model *Random Forest Regression* menunjukkan peningkatan kinerja dalam melakukan prediksi jumlah *order driver* Gojek. Model hasil optimasi diuji kembali menggunakan 60 data *testing* dengan membandingkan jumlah *order* aktual dan jumlah *order* hasil prediksi. Evaluasi dilakukan menggunakan metrik RMSE, MAE, dan MAPE untuk mengetahui tingkat kesalahan prediksi yang dihasilkan.

6. Grafik Perbandingan Data Aktual dan Hasil Prediksi Model

Grafik perbandingan data aktual dan hasil prediksi model merupakan halaman yang digunakan untuk menampilkan visualisasi hasil prediksi *Random Forest Regression* pada data uji. Grafik ini bertujuan untuk memberikan gambaran mengenai kemampuan model dalam memprediksi jumlah *order* dengan membandingkan nilai jumlah *order* aktual dengan nilai hasil prediksi model.



Gambar 4 Grafik Perbandingan

Pada grafik terdapat dua garis yang menunjukkan dua jenis data, yaitu garis *Jumlah Order Aktual* dan garis *Prediksi Random Forest*. Garis berwarna hijau

menunjukkan jumlah *order* aktual berdasarkan data uji, sedangkan garis berwarna biru putus-putus menunjukkan hasil prediksi yang dihasilkan oleh model *Random Forest Regression*. Sumbu *horizontal* menunjukkan urutan sampel data uji, sedangkan sumbu *vertikal* menunjukkan jumlah *order*.

Dengan adanya visualisasi ini, hasil prediksi model dapat dianalisis secara lebih mudah karena perbedaan antara data aktual dan hasil prediksi dapat dilihat secara langsung. Semakin dekat garis prediksi terhadap garis data aktual, maka semakin baik kemampuan model dalam menghasilkan prediksi jumlah *order*.

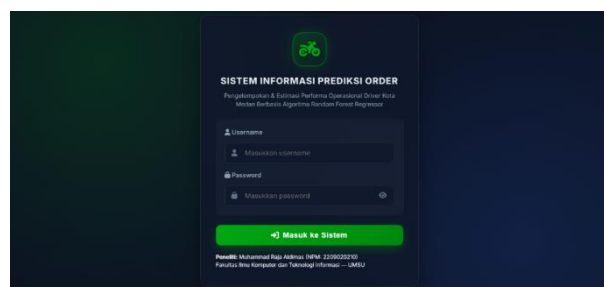
7. Implementasi Sistem Website

Implementasi sistem dilakukan dengan mengembangkan aplikasi berbasis *website* yang digunakan untuk mendukung proses prediksi jumlah *order driver* Gojek. Sistem dikembangkan menggunakan *Python* dengan *framework Flask*, sedangkan proses pengolahan data dan pembangunan model *machine learning* menggunakan beberapa *library Python*, seperti *Pandas*, *NumPy*, dan *Scikit-learn*.

Sistem menyediakan beberapa halaman utama yang mendukung tahapan penelitian, mulai dari pengelolaan dataset hingga proses prediksi. Data hasil kuesioner dapat dimasukkan ke dalam sistem, kemudian ditampilkan pada halaman *preview dataset* untuk memastikan data telah tersedia sebelum dilakukan proses pengolahan. Selanjutnya, sistem menjalankan proses *preprocessing*, pembagian data *training* dan *testing*, pembangunan model *Random Forest Regression*, serta evaluasi menggunakan RMSE, MAE, dan MAPE. Alur tersebut sesuai dengan rancangan sistem prediksi pada penelitian.

a. Halaman Login

Halaman *login* merupakan halaman awal yang digunakan pengguna untuk masuk ke dalam sistem. Pengguna memasukkan *username* dan *password* yang telah tersedia. Setelah data *login* berhasil diverifikasi, pengguna diarahkan menuju halaman *dashboard*.

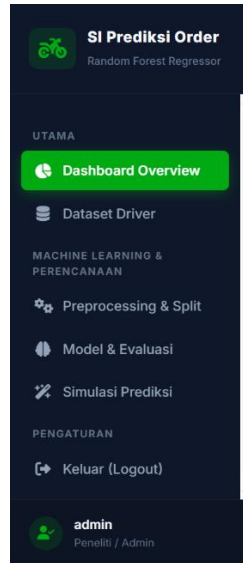


Gambar 5 Halaman *Login*

b. Halaman Dashboard

Halaman *dashboard* merupakan halaman utama sistem setelah pengguna berhasil melakukan *login*. Halaman ini digunakan sebagai pusat navigasi untuk

mengakses fitur-fitur sistem, seperti pengelolaan dataset, proses *training*, evaluasi model, dan prediksi jumlah *order*.



Gambar 61 Halaman *Dashboard*

c. Implementasi Database MySQL

Pada tahap ini dilakukan implementasi *database MySQL* sebagai media penyimpanan dan pengelolaan dataset *driver* Gojek. *Database* digunakan untuk menyimpan data yang diperoleh dari hasil kuesioner sebelum data tersebut digunakan dalam proses pengolahan dan pemodelan. Penggunaan *database* bertujuan agar data penelitian dapat disimpan secara terstruktur dan dapat diakses oleh sistem *website*.

Database yang digunakan dalam sistem diberi nama *gojek_prediction*, dengan tabel *dataset_driver* sebagai tempat penyimpanan data responden. Tabel tersebut memuat data yang berkaitan dengan variabel penelitian, yaitu jam mulai bekerja, durasi bekerja per hari, lama menjadi *driver* Gojek, jam *online* per hari, hari bekerja, kondisi cuaca, wilayah operasi, dan jumlah *order*.

Tabel 7 Stuktur *Database*

No	Nama Field	Tipe Data	Keterangan
1	id	INT	Identitas data
2	nama_driver	VARCHAR(100)	Nama driver
3	durasi_kerja	FLOAT	Durasi bekerja per hari
4	jam_online	FLOAT	Jam online per hari
5	jumlah_order	INT	Jumlah order

Berdasarkan implementasi tersebut, data *driver* dapat disimpan dalam *database MySQL* dan digunakan oleh aplikasi *Flask*. Data yang tersimpan selanjutnya dapat

dipanggil oleh sistem untuk proses *preprocessing*, pembagian dataset, *training Random Forest Regression*, evaluasi model, dan prediksi jumlah *order*.

d. Halaman Dataset

Halaman dataset digunakan untuk memasukkan dan menampilkan data *driver* Gojek yang diperoleh dari hasil kuesioner. Data yang ditampilkan mencakup variabel input dan variabel target yang digunakan dalam penelitian.

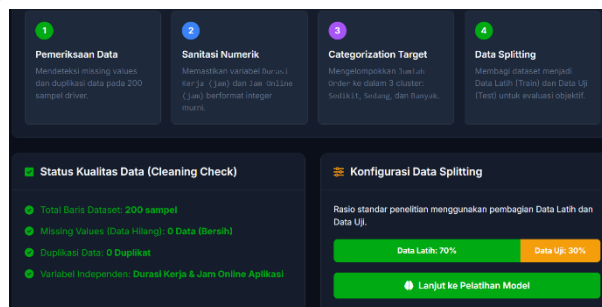


No	ID	Nama Driver	Durasi Bekerja (jam)	Jam Online Aplikasi (jam)	Jumlah Order (Target)	Kategori
1	0001	Ahmad Pratama	7 jam	6 jam	17 order	▲ 5
2	0002	Budi Santoso	8 jam	6 jam	18 order	▲ 5
3	0003	Andi Saputra	9 jam	9 jam	24 order	▲ 5
4	0004	Eko Prasetyo	10 jam	9 jam	25 order	▲ 5

Gambar 7 Halaman Dataset

e. Halaman Preprocessing Data & Split Data

Halaman *Preprocessing Data* dan *Split Data* merupakan halaman yang digunakan untuk menampilkan proses pengolahan dataset sebelum data digunakan dalam pembangunan model *Random Forest Regression*. Pada halaman ini, sistem menampilkan hasil *preprocessing* yang meliputi pemeriksaan dan penyesuaian data, proses *encoding* pada variabel kategorikal, serta pembagian dataset menjadi data *training* dan data *testing*.



Gambar 8 Halaman Preprocessing dan Split Data

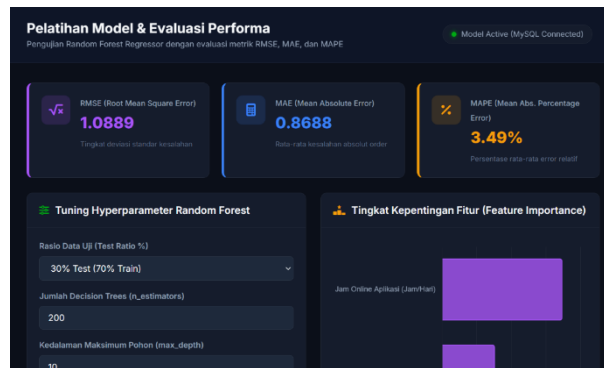
Dataset yang digunakan dalam penelitian terdiri dari 200 data. Setelah proses *preprocessing* selesai, data dibagi menjadi 70% *data training* sebanyak 140 data dan 30% *data testing* sebanyak 60 data. Data *training* digunakan untuk membangun dan melatih model *Random Forest Regression*, sedangkan data *testing* digunakan untuk menguji kemampuan model dalam menghasilkan prediksi terhadap data yang belum digunakan selama proses pelatihan.

f. Halaman Training Model & Evaluasi Model

Halaman Pelatihan Model dan Evaluasi digunakan untuk melakukan proses pelatihan algoritma *Random Forest Regression* serta mengetahui kinerja model dalam memprediksi jumlah *order driver* Gojek. Pada halaman ini, data yang telah melalui

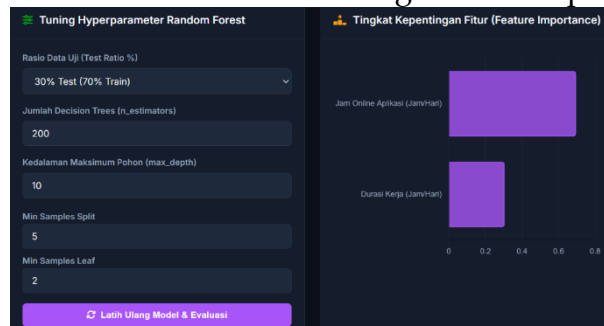
tahap *preprocessing* dan *split data* digunakan untuk proses pelatihan dan pengujian model.

Dataset dibagi menjadi 160 data *training* dan 40 data *testing* dengan perbandingan 70% data *training* dan 30% data *testing*. Model *Random Forest* dibangun menggunakan 200 *decision trees*. Setelah proses pelatihan selesai, model diuji menggunakan data *testing* dan dievaluasi menggunakan tiga metrik, yaitu RMSE, MAE, dan MAPE.



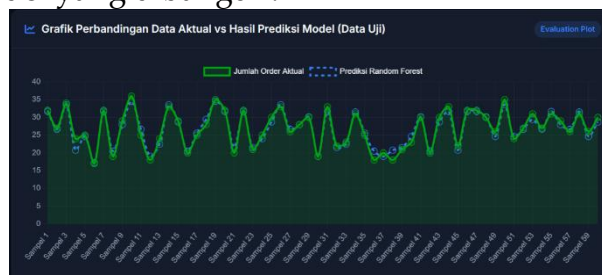
Gambar 9 Halaman *Training Model & Evaluasi*

Berdasarkan hasil implementasi, diperoleh nilai RMSE sebesar 1.0889, MAE sebesar 0.8688, dan MAPE sebesar 3,49%. Halaman ini juga menyediakan fitur *hyperparameter tuning* yang dapat digunakan untuk menyesuaikan parameter *Random Forest* apabila hasil evaluasi model belum sesuai dengan kriteria penelitian.



Gambar 2 *Feature Importance*

Selain hasil evaluasi, halaman ini menampilkan *feature importance* untuk mengetahui tingkat kepentingan masing-masing variabel input dalam pembentukan prediksi jumlah *order*. Hasil tersebut digunakan sebagai informasi tambahan dalam menganalisis kinerja model yang dibangun.



Gambar 11 Grafik Perbandingan Data Aktual dan Data Uji

Halaman grafik perbandingan data aktual vs hasil prediksi digunakan untuk menampilkan perbandingan antara jumlah *order* aktual dengan jumlah *order* hasil prediksi model *Random Forest Regression* pada data *testing*. Grafik ini bertujuan untuk memberikan gambaran secara visual mengenai kemampuan model dalam menghasilkan prediksi yang mendekati nilai aktual.

g. Halaman Prediksi

Halaman Prediksi digunakan untuk menghasilkan estimasi jumlah *order driver* Gojek menggunakan model *Random Forest Regression* yang telah melalui proses pelatihan dan evaluasi. Pada halaman ini, pengguna dapat memasukkan nilai variabel input yang digunakan dalam penelitian, yaitu jam mulai bekerja, durasi bekerja per hari, lama menjadi *driver* Gojek, jam *online* per hari, hari bekerja, kondisi cuaca, dan wilayah operasi. Setelah seluruh parameter dimasukkan, pengguna menjalankan proses prediksi melalui tombol “Jalankan Prediksi Jumlah Order”. Sistem kemudian memproses data input menggunakan model yang telah dibangun dan menampilkan hasil estimasi jumlah *order* pada bagian hasil prediksi.

Gambar 3 Halaman Prediksi

h. Analisis Hasil Prediksi

Analisis hasil prediksi dilakukan untuk mengetahui kemampuan model *Random Forest Regression* dalam memperkirakan jumlah *order driver* Gojek berdasarkan karakteristik operasional *driver*. Analisis dilakukan dengan membandingkan jumlah *order* aktual dengan jumlah *order* hasil prediksi pada 60 data *testing*. Perbedaan antara nilai aktual dan nilai prediksi menunjukkan tingkat kedekatan hasil prediksi terhadap kondisi sebenarnya.

Sampel Test #1	32 order	31.67 order	0.33 order	Sangat Presisi
Sampel Test #2	27 order	26.45 order	0.55 order	Sangat Presisi
Sampel Test #3	34 order	33.37 order	0.63 order	Sangat Presisi
Sampel Test #4	24 order	26.76 order	3.24 order	Presisi Sedang
Sampel Test #5	25 order	24.41 order	0.59 order	Sangat Presisi
Sampel Test #6	17 order	17.22 order	0.22 order	Sangat Presisi
Sampel Test #7	32 order	31.67 order	0.33 order	Sangat Presisi
Sampel Test #8	19 order	26.5 order	7.5 order	Sangat Presisi
Sampel Test #9	29 order	27.85 order	1.15 order	Sangat Presisi
Sampel Test #10	36 order	34.16 order	1.44 order	Sangat Presisi

Gambar 13 Halaman Pengujian

Berdasarkan hasil pengujian, terdapat data yang memiliki hasil prediksi yang mendekati nilai aktual dan terdapat pula data dengan selisih prediksi yang lebih besar. Sebagai contoh, pada salah satu data *testing*, jumlah *order* aktual sebesar 32 *order*, sedangkan hasil prediksi model sebesar 31,67 *order*, sehingga terdapat selisih sebesar 0,33 *order*. Kondisi tersebut menunjukkan bahwa model mampu menghasilkan prediksi yang relatif dekat dengan nilai aktual pada data tersebut. Namun, pada data lain dapat ditemukan selisih yang lebih besar, sehingga menunjukkan bahwa kemampuan model dalam melakukan prediksi dapat berbeda pada setiap kondisi *driver*.

Secara keseluruhan, hasil prediksi menunjukkan bahwa model dapat memberikan estimasi jumlah *order* berdasarkan variabel durasi bekerja, dan jam *online*. Hasil tersebut kemudian diperkuat dengan nilai evaluasi RMSE sebesar 1.0889, MAE sebesar 0.8688, dan MAPE sebesar 3,49%, yang menunjukkan tingkat kesalahan model dalam melakukan prediksi terhadap data *testing*.

Pembahasan

Berdasarkan hasil penelitian yang telah dilakukan, algoritma *Random Forest Regression* dapat diterapkan untuk memprediksi jumlah *order driver* Gojek di Kota Medan berdasarkan dua variabel input, durasi bekerja per hari, jam *online* per hari. Data penelitian yang diperoleh melalui kuesioner terlebih dahulu dilakukan *preprocessing* sebelum digunakan dalam proses pemodelan. Selanjutnya, dataset dibagi menjadi data *training* dan data *testing* dengan perbandingan 70% dan 30%. Data *training* digunakan untuk membangun model, sedangkan data *testing* digunakan untuk menguji kemampuan model dalam menghasilkan prediksi pada data yang tidak digunakan selama proses pelatihan.

Berdasarkan hasil pengujian model, diperoleh nilai RMSE sebesar 1.0889, MAE sebesar 0.8688, dan MAPE sebesar 3,49%. Nilai tersebut menunjukkan tingkat kesalahan prediksi yang dihasilkan oleh model dalam memperkirakan jumlah *order*. RMSE menunjukkan besarnya kesalahan prediksi dengan memberikan pengaruh lebih besar terhadap kesalahan yang tinggi, MAE menunjukkan rata-rata selisih absolut antara nilai aktual dan hasil prediksi, sedangkan MAPE menunjukkan rata-rata kesalahan prediksi dalam bentuk persentase.

Hasil prediksi yang diperoleh kemudian dibandingkan dengan jumlah *order* aktual pada data *testing* untuk mengetahui perbedaan antara hasil prediksi dan kondisi sebenarnya. Berdasarkan proses tersebut, model *Random Forest Regression* mampu menghasilkan estimasi jumlah *order* berdasarkan karakteristik aktivitas *driver* dan kondisi operasional yang digunakan dalam penelitian.

Dengan demikian, penerapan *Random Forest Regression* dalam penelitian ini dapat digunakan sebagai pendekatan untuk melakukan prediksi jumlah *order driver* Gojek di Kota Medan. Hasil prediksi tersebut dapat menjadi informasi pendukung dalam memahami pola jumlah *order* berdasarkan kondisi operasional *driver*.

Penelitian ini menghasilkan model prediksi jumlah *order driver* Gojek menggunakan algoritma *Random Forest Regression*. Model dibangun berdasarkan dua variabel input, yaitu durasi bekerja per hari, dan jam *online* per hari. Jumlah *order* digunakan sebagai variabel target. Dataset yang diperoleh dari 200 responden diproses melalui tahap *preprocessing* dan kemudian dibagi menjadi data *training* dan data *testing*. Model selanjutnya dilatih menggunakan data *training* dan diuji menggunakan data *testing*. Kinerja model dinilai menggunakan RMSE, MAE, dan MAPE. Hasil akhir penelitian digunakan untuk mengetahui kemampuan *Random Forest Regression* dalam menghasilkan prediksi jumlah *order driver* Gojek berdasarkan karakteristik aktivitas dan kondisi operasional *driver*.

KESIMPULAN

Berdasarkan hasil penelitian mengenai prediksi jumlah *order driver* Gojek menggunakan algoritma *Random Forest Regression*, dapat disimpulkan bahwa penelitian ini menggunakan data primer yang diperoleh melalui penyebaran kuesioner kepada *driver* Gojek di Kota Medan. Data yang diperoleh mencakup durasi bekerja per hari, jam *online* per hari, serta jumlah *order* sebagai variabel target.

Data yang telah dikumpulkan kemudian melalui tahap *preprocessing* untuk memastikan data dapat digunakan dalam proses pemodelan. Selanjutnya, dataset dibagi menjadi data *training* sebesar 70% dan data *testing* sebesar 30%. Data *training* digunakan untuk membangun model *Random Forest Regression*, sedangkan data *testing* digunakan untuk menguji kemampuan model dalam melakukan prediksi.

Kinerja model dievaluasi menggunakan *Root Mean Square Error* (RMSE), *Mean Absolute Error* (MAE), dan *Mean Absolute Percentage Error* (MAPE). Hasil evaluasi tersebut digunakan untuk mengetahui tingkat kesalahan model dalam memprediksi jumlah *order driver* Gojek. Nilai hasil evaluasi akan diisi berdasarkan hasil pengujian model yang sebenarnya. Selain itu, model *Random Forest Regression* diimplementasikan ke dalam sistem berbasis *website* menggunakan *Python* dan *Flask*. Sistem tersebut menyediakan fitur untuk pengelolaan dataset, proses *training*, evaluasi model, dan prediksi jumlah *order*.

DAFTAR PUSTAKA

- (APOJI), A. P. O. I. (2023). *Survei Nasional Driver Online 2023: Dampak Cuaca dan Ekonomi*. APOJI Sumatera Utara.
- (BPS), B. P. S. (2023). *Statistik Telekomunikasi Indonesia 2023*. BPS.
- (BPS), B. P. S. (2024). *Inflasi dan Harga Konsumen Indonesia 2023*. BPS.
- Gojek. (2022). *Laporan Tahunan Gojek 2022*. <https://www.gojek.com/annual-report>
- Google, Temasek, & Company, B. &. (2023). *e-Conomy SEA 2023*. Google.
- Hakim, R., & Pratama, D. (2023). Model evaluation techniques in predictive analysis using machine learning. *International Journal of Information Systems and Technology*, 7(1), 45–54.
- Hidayat, M., & Wicaksono, A. D. (2023). Evaluasi performa algoritma Random Forest

- untuk prediksi data multivariat. *Jurnal Teknologi Informasi Dan Sains Data*, 12(2), 89–97.
- Informatika, K. K. dan. (2022). *Gerakan Nasional 1000 Startup Digital 2021-2024*.
- Nugroho, A., & Setiawan, B. (2023). Prediksi Order Gojek Bandung Berbasis Support Vector Regression. *Jurnal Sistem Informasi*, 12(3), 210–225.
- Pratama, R., & al., et. (2022). Penerapan Deep Learning LSTM untuk Forecasting Order Grab di Surabaya. *Jurnal Informatika*, 15(1), 34–45.
- Puri, D., Sihag, P., Thakur, M. S., Jameel, M., Chadee, A. A., & Hazi, M. A. (2024). Analysis of Data Splitting on Streamflow Prediction using Random Forest. *AIMS Environmental Science*, 11(4), 593–609. <https://doi.org/10.3934/environsci.2024029>
- Ratnawati, S., & Maulana, R. M. (2024). Analisis komprehensif aplikasi transportasi Gojek: Pendekatan systematic literature review dari perspektif pengguna. *Jurnal Perangkat Lunak*, 6(1), 33–42. <https://doi.org/10.32520/jupel.v6i1.2856>
- Sari, D. P., & Putra, A. R. (2021). Analisis perbandingan metode prediksi menggunakan algoritma machine learning pada data time series. *Jurnal Informatika*, 8(1), 45–52.
- Sari, N. K., & Wijaya, A. (2021). Prediksi Permintaan Layanan Ride-Hailing Menggunakan Regresi Linier Berganda: Studi Kasus Gojek Jakarta. *Jurnal Teknik Industri*, 23(2), 145–158. <https://doi.org/10.1234/jti.v23i2.789>
- Sunday, B. O., & Nduka, S. O. (2023). MySQL Database Server: Deploying Software Application to Enhance Visibility and Accountability. *Current Journal of Applied Science and Technology*, 42(4), 16–23. <https://doi.org/10.9734/cjast/2023/v42i44063>
- Widodo, S. (2024). Gradient Boosting untuk Prediksi Order Driver Gojek Yogyakarta. *Jurnal Data Mining*, 8(1), 56–70.