

MEMPREDIKSI TINGKAT KECELAKAN JALAN RAYA DI SALATIGA MENGGUNAKAN MACHINE LEARNING

Andika Rossy Febrian *¹⁾, Hendry²⁾

1. Universitas Kristen Satya Wacana, Kota Salatiga
2. Universitas Kristen Satya Wacana, Kota Salatiga

Article Info

Kata Kunci: Peramalan, Kecelakaan Lalulintas,

Keywords: Traffic Accident Forecasting.

Article history:

Received 21 January 2025

Revised 27 February 2025

Accepted 13 March 2025

Available online 2 December 2025

DOI :

<https://doi.org/10.29100/jipi.v10i4.6959>

* Corresponding author.

Corresponding Author

E-mail address:

andika.rossy.febrian@gmail.com

ABSTRAK

Kecelakaan merupakan suatu kejadian di jalan yang tidak terduga yang melibatkan kendaraan dan mengakibatkan korban jiwa maupun kerugian material, kecelakaan juga disebut sebagai kejadian yang memiliki multi faktor atau memiliki banyak penyebab yang mempengaruhi terjadinya kecelakaan. Penelitian ini memiliki tujuan yaitu untuk meramalkan atau memprediksi kecelakaan lalu lintas menggunakan metode *Random Forest* dan *Linear Reggresion* dengan bahasa pemrograman Python. Data yang digunakan dalam penelitian ini yaitu data dari BPS (Badan pusat Statistik) kota Salatiga dengan periode dari Tahun 2017 sampai dengan Tahun 2023 dengan selang waktu perbulan. Penerapan dari kedua model menunjukkan untuk model *Linear Reggresion* memiliki hasil yang lebih baik dari pada model *Random Forest* ini dikarenakan pada perhitungan *Random Forest* selalu menunjukkan hasil decimal dan tidak genap akan tetapi kekurangan tersebut dapat dibenahi dengan menggunakan tools *Difference* yang beeguna untuk membantu mengidentifikasi pola atau kesalahan system dan memberikan arahan untuk perbaikan kedepannya.

ABSTRACT

An accident is an unexpected event on the road that involves vehicles and results in casualties or material damage. Accidents are also referred to as events with multiple factors or many causes that influence their occurrence. The aim of this research is to forecast or predict traffic accidents using the Random Forest and Linear Regression methods with Python programming language. The data used in this research is from BPS (Central Statistics Agency) of Salatiga city, covering the period from 2017 to 2023 with monthly intervals. The application of both models shows that the Linear Regression model produces better results than the Random Forest model. This is because the Random Forest model often produces decimal and non-integer results. However, this shortcoming can be addressed using the Difference tool, which helps identify patterns or system errors and provides guidance for future improvements.

I. PENDAHULUAN

Kecelakaan lalu lintas merupakan salah satu masalah serius di Indonesia yang terus mengalami peningkatan seiring dengan bertambahnya jumlah kendaraan bermotor. Berdasarkan data dari Korps Lalu Lintas (Korlantas) Polri, pada tahun 2023 tercatat sebanyak 120.000 kasus kecelakaan lalu lintas di Indonesia, meningkat sebesar 5% dibandingkan tahun sebelumnya [4]. Peningkatan ini disebabkan oleh berbagai faktor, termasuk jumlah kendaraan bermotor yang terus bertambah. Di Kota Salatiga, data dari Badan Pusat Statistik (BPS) menunjukkan bahwa pada tahun 2023 terdapat peningkatan 8% jumlah kendaraan bermotor dibandingkan tahun sebelumnya, yang turut memengaruhi peningkatan risiko kecelakaan di wilayah tersebut [6]. Dampak dari kecelakaan lalu lintas ini tidak hanya berupa kerugian materiil, tetapi juga menyebabkan cedera ringan, cedera berat, hingga korban jiwa.

Beberapa faktor utama yang memengaruhi terjadinya kecelakaan lalu lintas meliputi kelelahan pengemudi, pelanggaran peraturan lalu lintas, kondisi jalan yang tidak memadai, cuaca, serta kelalaian pengemudi dalam menjaga kelayakan kendaraannya [2]. Oleh karena itu, diperlukan langkah strategis untuk mengurangi angka kecelakaan yang berbasis pada analisis kecenderungan prediksi pola kecelakaan.

Prediksi kecelakaan lalu lintas menjadi langkah penting dalam merencanakan strategi pencegahan. Salah satu pendekatan yang efektif adalah penggunaan metode Machine Learning, karena mampu menangani data dalam

jumlah besar dan memiliki tingkat akurasi yang lebih baik dibandingkan metode statistik tradisional [4]. Dalam penelitian ini, dua algoritma Machine Learning, yaitu Random Forest dan Regresi Linear, digunakan untuk memprediksi angka kecelakaan lalu lintas di Kota Salatiga. Kedua algoritma ini dipilih karena efektivitasnya dalam memodelkan hubungan antara berbagai faktor yang memengaruhi kecelakaan lalu lintas [5].

Penelitian sebelumnya dengan judul **"Prediksi Kecelakaan Lalu Lintas di Bali dengan XGBoost pada Python"** menggunakan algoritma XGBoost, yang merupakan turunan dari *Gradient Boosting Decision Tree* (GBDT). XGBoost dikenal memiliki performa yang sangat baik dalam prediksi, terutama pada data dengan struktur yang kompleks dan variabel yang saling berkorelasi. Namun, penelitian ini hanya menggunakan satu metode prediksi tanpa membandingkannya dengan algoritma lain, sehingga keterbatasannya adalah kurangnya evaluasi komparatif.[1] Selain itu, penelitian ini tidak secara mendalam membahas pengaruh faktor-faktor penyebab kecelakaan, yang penting untuk memberikan konteks strategis dalam pencegahan kecelakaan.

Di sisi lain, penelitian lain berjudul **"Peramalan Jumlah Kecelakaan Lalu Lintas Menggunakan Metode Support Vector Regression"** menggunakan metode *Support Vector Regression* (SVR). SVR unggul dalam mengolah dataset kecil dan kompleks dengan waktu pemrosesan yang relatif singkat. Keunggulan ini menjadikan SVR ideal untuk skenario dengan data terbatas. Namun, kelemahan SVR adalah performanya yang kurang optimal saat diterapkan pada dataset yang lebih besar, karena metode ini memerlukan komputasi yang lebih intensif dan tidak dirancang untuk data dalam skala besar.[6]

Dibandingkan dengan penelitian sebelumnya, penelitian ini menawarkan beberapa keunggulan dan kontribusi unik. Pertama penggunaan dua metode prediksi, *Random Forest dan Regresi Linear*, memungkinkan analisis yang lebih spesifik dan mendalam mengenai faktor-faktor yang memengaruhi kecelakaan [5]. Kedua, penelitian ini mencakup analisis yang komprehensif terhadap faktor-faktor penyebab kecelakaan, yang belum dijelaskan secara rinci dalam penelitian menggunakan XGBoost dan SVR. Ketiga, penelitian ini juga dirancang untuk menangani dataset yang lebih besar, sehingga memberikan model prediksi yang lebih akurat dan dapat diandalkan.

Dengan kontribusi tersebut, penelitian ini diharapkan dapat mengisi gap penelitian sebelumnya, yaitu kurangnya analisis mendalam terhadap faktor-faktor kecelakaan serta keterbatasan dalam penggunaan dataset besar. Melalui pendekatan ini, penelitian ini tidak hanya berfokus pada akurasi prediksi, tetapi juga memberikan wawasan strategis bagi pihak berwenang dalam merencanakan langkah-langkah pencegahan untuk mengurangi angka kecelakaan lalu lintas di masa mendatang [5].

II. METODE PENELITIAN

Pada penelitian ini metode yang akan digunakan adalah Peramalan, dimana metode ini cocok untuk masalah pengolahan data banyak serta banyak fitur dan hubungan yang kompleks. Dalam penelitian ini menggunakan 2 metode yaitu Random Forest dan Regresi Linear[5]. Random Forest di pilih karena dapat menangani hubungan non-linier, tahan terhadap overfitting, dan memiliki kemampuan dalam menangani banyak variable[8]. Regresi Linear dipilih karena metode ini mudah untuk diinterpretasi, komputasi efisien, dan cocok untuk masalah dengan hubungan linier yang cukup jelas[7].

A. Data Penelitian

Data yang akan di gunakan untuk melakukan peramalan adalah jumlah kecelakaan lalu lintas menurut akibatnya untuk tahun 2023 kebawah, yang didapat dari website Badan Pusat Statistik (BPS) kota Salatiga[1]. Data yang dipakai kurang lebih 84 data dari data th 2023 dengan data 2017. Berikut beberapa contoh data yang di pakai dapat di lihat di Table 1.

TABEL I
 DATA KECELAKAAN LALU LINTAS DI KOTA SALATIGA UNTUK TAHUN 2023

Bulan/Tahun	Jumlah Kecelakaan	MD (Meninggal Dunia)	LB (Luka Berat)	LR (Luka Ringan)	Kerugian	Curah Hujan
Januari/2023	27	3	0	28	42500000	281
Februari/2023	21	0	0	26	23250000	194
Maret/2023	27	5	0	36	24100000	377
April/2023	17	2	0	17	10700000	219
Mei/2023	27	1	0	38	38500000	221
Juni/2023	26	2	0	35	48050000	6
Juli/2023	24	7	0	29	39900000	0
Agustus/2023	19	3	0	24	9700000	0
September/2023	34	2	0	38	21300000	0
Oktober/2023	21	1	0	24	41300000	107
Nopember/2023	24	1	2	28	26200000	68
Desember/2023	22	4	0	25	18900000	273

(Gambaran umum data yang akan di gunakan untuk th 2023)

B. Data mining

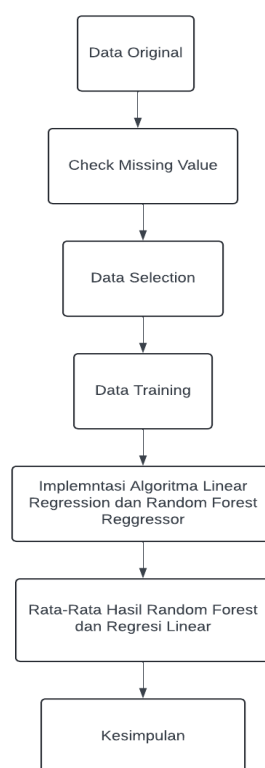
Data mining adalah suatu istilah yang di pakai untuk menganalisis suatu basis data untuk mendapatkan informasi yang berguna, pada proses data mining memanfaatkan berbagai disiplin ilmu pengetahuan seperti matematika, statistik, kecerdasan buatan (AI) dan ilmu komputer untuk mendapatkan informasi data yang telah dikelola. Dengan menggunakan metode data mining ini pengambilan keputusan di masa depan dapat lebih terbantu dan akurat[8].

C. Forecasting

Forecasting adalah proses memperkirakan atau memprediksi nilai atau tren untuk masa depan yang berdasarkan data masa lampau dan analisis pola yang telah tersedia. Proses ini memerlukan berbagai teknik perhitungan dan metode statistik serta algoritma machine learning untuk mendapatkan gambaran apa yang mungkin terjadi di masa depan[6].

D. Kerangka Kerja

Pada penelitian ini menggunakan data mining metode forecasting, dengan memanfaatkan algoritma Linear Regresion dan Random Forest, beikut detail Langkah-Langkah seperti gambar berikut:



Gambar 1 Kerangka Kerja

E. Skenario Pengujian

Skenario penelitian ini mencakup beberapa tahap penting. Pertama, pengumpulan data kecelakaan dari Badan Pusat Statistik (BPS) Kota Salatiga untuk periode 2017 hingga 2023. Data yang digunakan mencakup jumlah kecelakaan, korban meninggal dunia (MD), luka berat (LB), luka ringan (LR), kerugian material, dan curah hujan. Setelah itu, data dilakukan preprocessing dengan pembersihan nilai kosong, outlier. Kedua, dataset dibagi menjadi data latih (75%) dan data uji (25%) menggunakan fungsi *train_test_split* dari Scikit-Learn untuk memastikan bahwa model dapat diuji pada data yang tidak terlihat sebelumnya [5]. Ketiga, model Random Forest dan Regresi Linear diterapkan untuk memproses data dengan fokus pada analisis hubungan linear dan non-linear. Model dievaluasi menggunakan metrik R-squared (R^2), serta divisualisasikan dengan scatter plot untuk membandingkan nilai aktual dengan prediksi. Selain itu, dilakukan eksperimen variabel tambahan, seperti curah hujan, untuk menilai pengaruhnya terhadap hasil prediksi [6]. Hasil penelitian dianalisis untuk memberikan rekomendasi strategis dan menyimpulkan model terbaik antara Random Forest dan Regresi Linear, serta memberikan wawasan bagi pihak berwenang dalam merencanakan langkah-langkah pencegahan kecelakaan lalu lintas di masa mendatang.

III. HASIL DAN PEMBAHASAN PENELITIAN

A. Data Original

Data Original atau dapat di sebut dengan (Raw Data) merupakan data mentah yang belum dilakukan modifikasi pada data tersebut, Data yang digunakan yaitu dengan rentan waktu 7 tahun dari 2017 sampai 2023. Data yang digunakan adalah Jumlah Kecelakaan, Korban Jiwa, luka Berat, Luka Ringan, Total Kerugian, dan Curah Hujan.

TABEL II
DATA ORIGINAL

No	Bulan/Tahun	Jumlah Kecelakaan	Meninggal Dunia (MD)	Luka Berat (LB)	Luka Ringan (LR)	Kerugian	Curah Hujan
1	Januari/2023	27	3	0	28	42500000	281
2	Februari/2023	21	0	0	26	23250000	194
3	Maret/2023	27	5	0	36	24100000	377
4	April/2023	17	2	0	17	10700000	219
5	Mei/2023	27	1	0	38	38500000	221
6	Juni/2023	26	2	0	35	48050000	6
7	Juli/2023	24	7	0	29	39900000	0
8	Agustus/2023	19	3	0	24	9700000	0

B. Data Selection

Dalam Pengambilan data tentunya tidak semua data digunakan oleh karena itu pada data selection data yang di gunakan adalah data yang dapat mewakili dari beberapa data untuk dapat di analisis. Dalam pemilihan data selection menggunakan fitur yaitu '.iloc' dari library pandas dalam pemilihan fitur berbasis index pada dataset kecelakaan lalu lintas.

```
#Index Slicing Using Iloc
x = datacrash.iloc[:, [1,2,3,4,5,6]].values
y = datacrash.iloc[:,1].values
```

Gambar 2 Data Selection Menggunakan '.iloc'

Berikut data yang akan di gunakan :Jumlah Kecelakaan, Korban Jiwa, Korban Jiwa, Luka Berat, Luka Ringan, Total Kerugian, dan Curah Hujan.

TABEL III
DATA ORIGINAL
HASIL DARI PENGGUNAAN 'ILOC'

Jumlah Kecelakaan	Meninggal Dunia (MD)	Luka Berat (LB)	Luka Ringan (LR)	Kerugian	Curah Hujan
5	0	0	8	5000000	635
7	0	0	9	5700000	145
7	0	0	10	7600000	218
5	1	0	8	3500000	235
9	0	0	13	5300000	113
9	1	1	9	4700000	24
6	0	0	8	3800000	17
5	0	0	5	2100000	0
16	1	0	25	7150000	71
4	0	0	7	4700000	86
4	3	1	3	2200000	154

C. Data Training

Pada data *training* setelah data selection, data dilatih (training set) dan data uji (test set) langkah ini perlu dilakukan untuk pengembangan machine learning. Dalam pengujian ini fungsi yang digunakan adalah '*train_test_split*' dari library Scikit-Learn untuk membagi dataset menjadi 2 bagian berdasarkan proporsi yang telah ditentukan.

```
#Splitting the Dataset Into the Training set and Test Set
from sklearn.model_selection import train_test_split
x_train, x_test, y_train, y_test = train_test_split(x, y, test_size = 0.25, random_state = 98)
x_train, x_test, y_train, y_test
```

Gambar 3 Data Training dan Split Menggunakan 'train_test_split'

Proporsi 75% data digunakan sebagai data latih, sedangkan 25% sisanya digunakan sebagai data uji. Penggunaan seed atau state dalam pengacakan data memastikan konsistensi dalam pembagian data, memungkinkan hasil yang dapat direproduksi.

[6,	1,	0,	9,	2900000,	53],
[26,	2,	0,	35,	48050000,	6],
[23,	0,	0,	30,	26600000,	180],
[7,	0,	0,	9,	5700000,	145],
[24,	7,	0,	29,	39900000,	0],
[23,	2,	0,	31,	21500000,	275],
[26,	5,	0,	26,	19100000,	290],
[17,	0,	2,	17,	10900000,	501],
[23,	0,	1,	26,	7400000,	367],
[5,	0,	0,	8,	5000000,	635],
[18,	1,	0,	22,	11350000,	81],
[25,	1,	0,	28,	16650000,	49],
[32,	0,	3,	40,	16900000,	0],

Gambar 4. Hasil Training Dataset

```
array([1, 2, 0, 0, 7, 2, 5, 0, 0, 0, 1, 1, 0, 0, 2, 2, 3, 2, 0, 7, 3, 1,
       2, 2, 1, 0, 1, 4, 1, 0, 0, 2, 0, 1, 1, 0, 4, 0, 0, 1, 1, 1, 1, 1,
       3, 0, 1, 0, 3, 1, 5, 0, 0, 1, 2, 0, 0, 4, 6, 0, 1, 2, 1]),
array([1, 0, 2, 3, 5, 0, 0, 3, 1, 1, 5, 1, 3, 1, 5, 2, 1, 0, 1, 0, 0]))
```

Gambar 5 Hasil Training Dataset

D. Implementasi Algoritma Linear Regression dan Random Forest Regressor

Pada tahap ini adalah tahap yang terpenting dalam pengolahan data mining yaitu implementasi algoritma Linear Regression dan Random forest untuk memproses data, Tools yang digunakan dalam pengolahan data menggunakan IPNY (interactive Python Notebook). berikut adalah hasil untuk implementasi Algoritma *Linear Regression* dan *Random Forest* menggunakan IPNY.

TABEL IV
 HASIL PERHITUNGAN REGRESI LINEAR

No	Actual_Value	Predicted_Value	Difference
1	21.0	21.0	-2.027534e-11
2	5.0	5.0	2.379608e-11
3	17.0	17.0	-3.054623e-11
4	12.0	12.0	2.211209e-11
5	8.0	8.0	2.476952e-11
6	38.0	38.0	-3.514344e-11
7	4.0	4.0	3.056844e-11
8	11.0	11.0	2.187583e-11
9	9.0	9.0	2.308198e-11
10	18.0	18.0	5.929479e-12
11	21.0	21.0	-2.027534e-11
12	5.0	5.0	2.379608e-11
13	17.0	17.0	-3.054623e-11
14	12.0	12.0	2.211209e-11
15	8.0	8.0	2.476952e-11
16	38.0	38.0	-3.514344e-11

TABEL V
 HASIL PERHITUNGAN RANDOM FOREST

No	Actual_Value	Predicted_Value	Difference
1	21.0	20.64	-2.027534e-11
2	5.0	4.90	2.379608e-11
3	17.0	17.08	-3.054623e-11
4	12.0	11.27	2.211209e-11
5	8.0	7.19	2.476952e-11
6	38.0	33.15	-3.514344e-11
7	4.0	4.64	3.056844e-11

8	11.0	10.63	2.187583e-11
9	9.0	9.74	2.308198e-11
10	18.0	17.77	5.929479e-12
11	14.0	13.74	2.235367e-11
12	17.0	17.06	1.824318e-11
13	24.0	23.97	-4.579448e-12
14	21.0	20.63	-6.386713e-11
15	20.0	19.75	6.053824e-12
16	13.0	12.21	2.646594e-11

Berdasarkan hasil implementasi algoritma regresi linear dan random forest memiliki hasil yang terbilang hampir serupa. Walau terbilang serupa pada perhitungan pada random forest selalu menunjukkan hasil decimal, Ini terjadi karena Algoritma ini membagi dataset menjadi sub-set secara acak, lalu membuat pohon keputusan untuk masing-masing sub-set. Setiap pohon menggunakan aturan tertentu untuk memprediksi hasil, dan prediksi akhir adalah rata-rata dari semua pohon. Hal ini membuat Random Forest lebih fleksibel dalam menangkap pola non-linier yang kompleks. Sedangkan algoritma Linear Regression memodelkan hubungan antara variabel dengan menggunakan rumus matematis berbentuk:

$$Y = b_0 + b_1X_1 + b_2X_2 + \dots + b_nX_n$$

Gambar 6 Rumus Perhitungan Regresi Linear

Koefisien b_1, b_2, \dots, b_n dihitung untuk meminimalkan error (selisih antara nilai aktual dan prediksi). Metode ini cocok untuk hubungan linier sederhana. Untuk melihat keakuratan kedua algoritma tersebut diperlukan penggunaan test accuracy untuk dapat melihat apakah hasil yang dihasilkan dari kedua algoritma menunjukkan hasil yang diinginkan.

Train Accuracy: 1.0
 Test Accuracy: 1.0

Gambar 7 Train Accuracy Regresi Linear

Train Accuracy: 0.9991492176391489
 Test Accuracy: 0.9762264244125598

Gambar 8 Train Accuracy Random Forest

Diperlukannya test accuracy ialah untuk dapat menilai seberapa baik model telah mempelajari pola dari data yang dilatih. Dikarenakan model yang di tampilkan kurang presisi dalam menjalankannya maka diperoleh hasil akhir decimal maka diperlukan *Difference*, kegunaannya adalah menganalisis residual untuk membantu menilai kinerja model dengan mengidentifikasi pola atau kesalahan system dan memberikan arah untuk perbaikan kedepannya. Setelah penggunaan *Difference* maka hasil dapat di lihat pada Tabel 4 dan 5. Berikut hasil dari *Difference*:

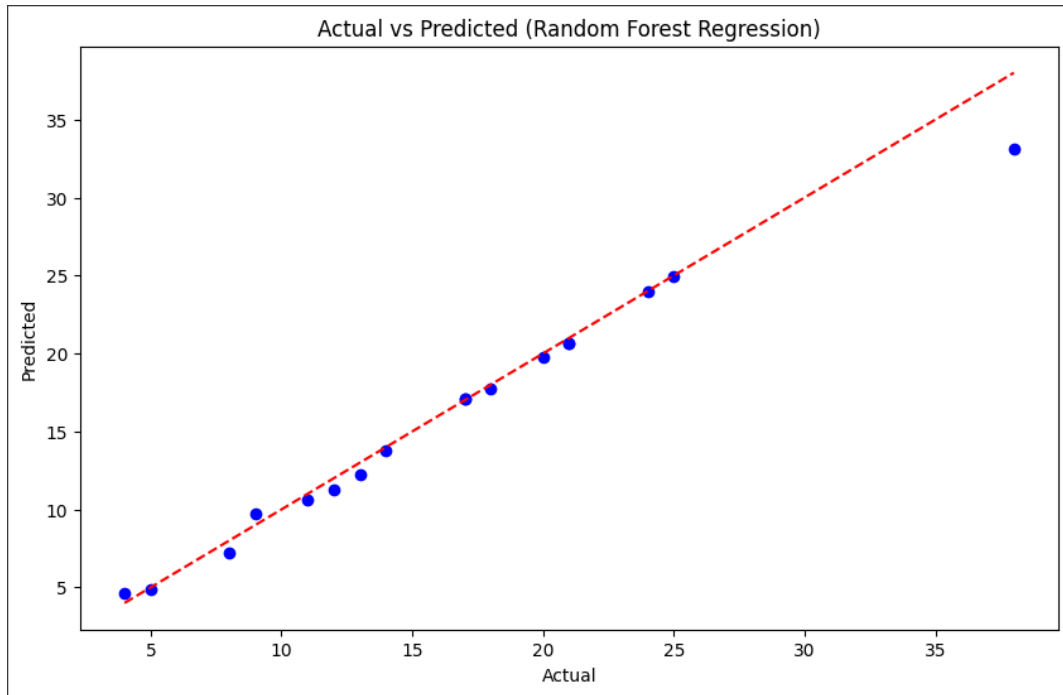
```
array([-2.02753370e-11,  2.37960762e-11, -3.05462322e-11,  2.21120899e-11,
        2.47695198e-11, -3.51434437e-11,  3.05684367e-11,  2.18758345e-11,
        2.30819808e-11,  5.92947913e-12,  2.23536745e-11,  1.82431847e-11,
        -4.57944793e-12, -6.38671338e-11,  6.05382411e-12,  2.64659406e-11,
        -1.25410793e-11])
```

Gambar 9 Hasil Difference Linear Regresion

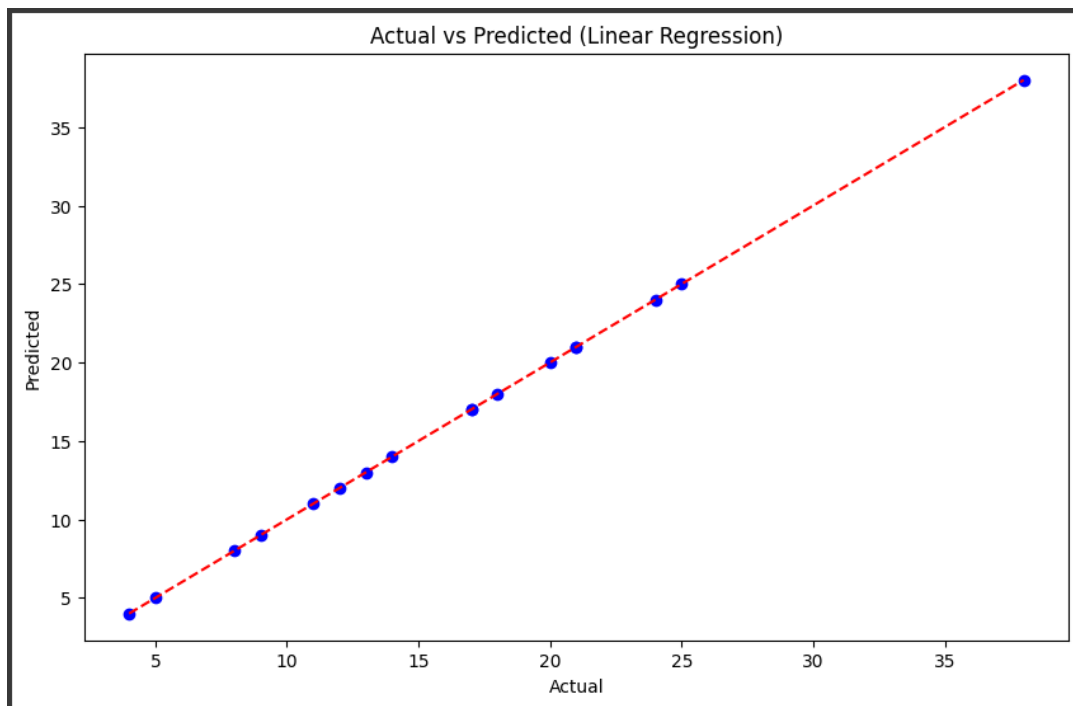
```
array([-0.36, -0.1,  0.08, -0.73, -0.81, -4.85,  0.64, -0.37,  0.74,
        -0.23, -0.26,  0.06, -0.03, -0.37, -0.25, -0.79, -0.02])
```

Gambar 10 Hasil Difference Random Forest

Setelah mengetahui hasil dari perhitungan dilanjut dengan membuat visualisasi data, dilakukannya visualisasi data digunakan untuk dapat melihat seberapa baik model prediksi dalam meperediksi data uji. Berikut hasil visualisasi data:



Gambar 11 Visualisasi data Random Forest



Gambar 12 Visualisasi data Linear Regression

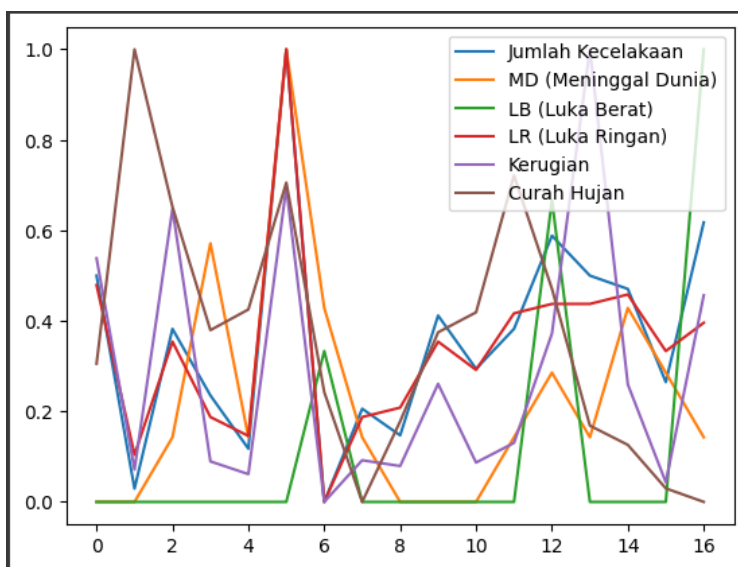
Pada Hasil nilai prediksi vs nilai aktual menggunakan model Random Forest Regression dan Linear Regression, pada sumbu X menampilkan nilai aktual dari data uji, sementara sumbu Y menampilkan nilai prediksi yang dihasilkan oleh setiap model. Setiap titik biru pada plot mewakili satu dari salah satu data uji, dimana posisi horizontal (X) dari titik menunjukkan nilai aktual dan posisi vertikal (Y) menunjukkan nilai prediksi.

Terdapat garis diagonal merah putus putus pada plot yang merupakan referensi Dimana nilai prediksi sama dengan nilai aktual ($Y=X$) jika model sempurna maka titik biru akan sama dengan garis diagonalnya.

Dalam analisis data kecelakaan, kedua model telah digunakan untuk memprediksi jumlah kecelakaan berdasarkan kedua model yaitu Regresi Linear dan Random Forest, hasil dari kedua perhitungan tersebut memiliki persamaan, walau berbeda dalam metode perhitungannya mereka memiliki kesamaan pada hasil. Berikut Hasil dari Pengabungan 2 metode tersebut:

TABEL VI
 HASIL PREDIKSI

Jumlah Kecelakaan	MD (Meninggal Dunia)	LB (Luka Berat)	LR (Luka Ringan)	Kerugian	Curah Hujan
21	0	0	26	23250000	194
5	0	0	8	5000000	635
17	1	0	20	27500000	414
12	4	0	12	5700000	241
8	1	0	10	4600000	270
38	7	0	51	29400000	448
4	3	1	3	2200000	154
11	1	0	12	5800000	0
9	0	0	13	5300000	113
18	0	0	20	12400000	238
14	0	0	17	5600000	266
17	1	0	23	7300000	458
24	2	2	24	16750000	299
21	1	0	24	41300000	107
20	3	0	25	12350000	80
13	2	0	19	3900000	19
25	1	3	22	20050000	0



Gambar 13 Hasil Visualisasi Grafik Perhitungan

Pada gambar 13 terdapat enam garis dari garis-garis tersebut menunjukkan seberapa banyaknya angka kecelakaan yang ada di kota Salatiga. Pada grafik kita dapat melihat bahwa angka yang muncul hanya sampai 1.0, itu dikarenakan pada penggunaan grafik dilakukan normalisasi data, kenapa perlu dilakukan normalisasi data, normalisasi dilakukan untuk memastikan bahwa tidak ada satu fitur dari data yang mendominasi kepada data lain karena skala yang berbeda. Teknik yang digunakan untuk menormalisasi adalah menggunakan teknik Min-Max Scaling, merupakan sebuah teknik untuk mengubah nilai-nilai fitur dalam dataset ke dalam rentang tertentu sebagai contoh 0 dan 1, dan teknik ini menggunakan rumus sebagai berikut:

$$X_{scaled} = \frac{X - X_{min}}{X_{max} - X_{min}}$$

Gambar 14 Rumus Normalisasi

Dapat dilihat pada tabel 6 Hasil perhitungan algoritma random forest dan regresi linear menunjukkan jika prediksi untuk kedepannya mengalami penurunan jika dibandingkan dengan data kecelakaan pada tahun 2023 yang dapat dilihat pada tabel 1, akan tetapi pada prediksi curah hujan menunjukkan untuk kedepannya mungkin akan mengalami hujan berkepanjangan pada setiap bulannya. Hal ini memerlukan langkah dari pihak kepolisian

dan pemerintah setempat untuk melakukan kampanye keselamatan berkendara saat hujan, perbaikan infrastruktur jalan dan penerapan teknologi peringatan dini dikarenakan curah hujan tinggi yang beresiko terjadinya banjir.

Penelitian ini bertujuan membantu pihak kepolisian setempat dalam merencanakan langkah-langkah strategis untuk meminimalkan kecelakaan lalu lintas melalui metode forecasting menggunakan algoritma *random forest* dan *regresi linear*. Sistem forecasting ini dirancang untuk memberikan kemudahan bagi pihak kepolisian dalam menganalisa data histori dan memprediksi potensi kejadian di masa depan. Dengan informasi ini, pihak kepolisian dapat menyusun strategi yang lebih efektif dalam menghadapi tantangan, seperti peningkatan keselamatan jalan, pengelolaan lalu lintas saat musim hujan, dan mitigasi risiko kecelakaan secara menyeluruh. Hasil penelitian akan dibandingkan dengan penelitian terkait sebelumnya untuk menunjukkan kontribusi penelitian ini.

Perbandingan dilakukan dengan penelitian terdahulu dengan judul **"Prediksi Kecelakaan Lalu Lintas di Bali dengan XGBoost pada Python"** Penelitian ini bertujuan untuk menekankan efektifitas metode XGBoost dalam meramalkan kecelakaan, terutama dalam kategori jumlah orang meninggal dan luka berat [1]. Untuk penelitian berikutnya yaitu berjudul **"Peramalan Jumlah Kecelakaan Lalu Lintas Menggunakan Metode Support Vector Regression"**. Pada penelitian ini penggunaan metode Vector Regression (SVR) dengan dua jenis kernel: polinomial dan RBF, penelitian ini menekankan pentingnya tindakan pencegahan berdasarkan hasil peramalan untuk mengurangi kecelakaan [6]. Dibandingkan dengan penelitian terdahulu, penelitian ini lebih berfokus pada penerapan metode *Random Forest* dan *Regresi Linear*, yang memberikan keunggulan dalam analisis data kecelakaan lalu lintas serta curah hujan. *Random Forest* dan *Regresi Linear* lebih mudah diimplementasikan dan memberikan interpretasi yang jelas mengenai pengaruh curah hujan terhadap kecelakaan. Keunggulan *Random Forest* terletak pada kemampuannya menangani data dengan banyak noise dan outlier, serta mengidentifikasi fitur-fitur penting yang mempengaruhi prediksi kecelakaan. Sementara itu, *Regresi Linear* memberikan gambaran yang lebih sederhana namun efektif tentang hubungan antara curah hujan dan kecelakaan. Kedua metode ini sangat berguna dalam merencanakan langkah-langkah pencegahan kecelakaan yang dipengaruhi oleh faktor cuaca, seperti hujan, yang dapat mempengaruhi keselamatan berkendara.

IV. KESIMPULAN

Dari kedua metode yang telah dilakukan untuk memprediksikan tingkat kecelakaan kota Salatiga menggunakan metode Linear Reggression dan metode Random Forest, dapat disimpulkan bahwa penggunaan model Linear Reggression mempunyai keunggulan dalam melakukan pemrosesan data dari pada model Random Forest, hal ini dapat di lihat dengan tingkat keakurasian model Linear Reggression yang sempurna dengan nilai 1.0 berbanding terbalik dengan Random Forest yang hanya memiliki nilai 0,9 walau pun perbedaanya yang tidak terlalu signifikan akan tetapi tetap mempengaruhi hasil.

Maka dari itu setelah melakukan penelitian diperlukan analisis lebih lanjut untuk mengevaluasi kinerja model secara keseluruhan dan memperhatikan faktor-faktor lain seperti kondisi jalan, kondisi kendaraan dan keadaan alam. Selain itu untuk penelitian di masa mendatang dapat mengkomparasi beberapa metode selain metode Linear Reggression dan metode Random Forest.

V. DAFTAR PUSTAKA

- [1] Pandika Pinata, N. N., Sukarsa, I. M., & Dwi Rusjyanthi, N. K. (2020). Prediksi Kecelakaan Lalu Lintas di Bali dengan XGBoost pada Python. *Jurnal Ilmiah Merpati (Menara Penelitian Akademika Teknologi Informasi)*, 8(3), 188. <https://doi.org/10.24843/jim.2020.v08.i03.p04>.
- [2] (Rahmawati et al., 2020)Rahmawati, R., Marsudi, K, D. R., Martono, & Budi, B. S. (2020). Studi Tingkat Kecelakaan Lalu Lintas Jalan di Indonesia Pada Masa Sebelum dan Pada Saat Pandemi Covid Tahun 2020 Berdasarkan Data KKNKT (Komite Nasional Keselamatan Transportasi) Tahun 2015-2020. *Jurusan Teknik Sipil*, 170–181.
- [3] Haryono, E., Widodo, A., & Abusini, S. (2013). Kajian model Automatic Clustering-Fuzzy Time Series-Markov Chain dalam memprediksi data historis jumlah kecelakaan lalu lintas di kota Malang (Study of the Automatic Clustering-Fuzzy Time Series-Markov Chain model to predict historical data of the number . *J. Sains Dasar*, 2(1), 63–71.
- [4] Susanti, A. (2021). PEMODELAN KECELAKAAN LALU LINTAS di KOTA BANDAR LAMPUNG MENGGUNAKAN METODE AUTOREGRESIVE INTEGRATED MOVING AVERAGE (ARIMA). *Repository UIN Raden Intan Lampung*. [http://repository.radenintan.ac.id/16298/1/skripsi 1-2.pdf](http://repository.radenintan.ac.id/16298/1/skripsi%201-2.pdf).
- [5] Aditya Pratama, M., Munawaroh, M., Joko Pranoto, W., Studi Teknik Informatika, P., Sains dan Teknologi, F., & Muhammadiyah Kalimantan Timur, U. (2024). Perbandingan Performa Algoritma Linear Regresi dan Random Forest untuk Prediksi Harga Bawang Merah di Kota Samarinda. *Jurnal Ilmu Teknik*, 1(2), 172–182. <https://doi.org/10.62017/tektonik>
- [6] Apriyanti, N. P. R., Putra, I. K. G. D., & Putra, I. M. S. (2020). Peramalan Jumlah Kecelakaan Lalu Lintas Menggunakan Metode Support Vector Regression. *Jurnal Ilmiah Merpati (Menara Penelitian Akademika Teknologi Informasi)*, 8(2), 72. <https://doi.org/10.24843/jim.2020.v08.i02.p01>.
- [7] Ayuni, G. N., & Fitrihanah, D. (2020). Penerapan Metode Regresi Linear Untuk Prediksi Penjualan Properti pada PT XYZ. *Jurnal Telematika*, 14(2), 79–86. <https://doi.org/10.61769/telematika.v14i2.321>

- [8] Rianto, M., & Yunis, R. (2021). Analisis Runtun Waktu Untuk Memprediksi Jumlah Mahasiswa Baru Dengan Model Random Forest. *Paradigma - Jurnal Komputer Dan Informatika*, 23(1). <https://doi.org/10.31294/p.v23i1.9781>
- [9] Yusuf Supriyanto, M. Ilhamsyah, Ultach Enri; Prediksi Harga Minyak Kelapa Sawit Menggunakan Linear Regression Dan Random Forest. Vol. 8, No.7, Mei 2022.
- [10] Wresti Andriani, Gunawan, Alan Eka Prayoga; Prediksi Nilai Emas Menggunakan Algoritma Regresi Linear; <https://doi.org/10.35760/ik.2023.v28i1.8096> *Jurnal Ilmiah Informatika Komputer*, 28(1): 27-35.
- [11] Segar Napitupulu, Novriadi Antonius Siagian; Prediksi Data Produksi Menggunakan Regresi Linear Sederhana, Vol. 1, No. 2, 2023, p. 95-105.
- [12] Frando Simon Hukum, Ludfi Djakfar dan Muhammad Zainul Arifin; Model Prediksi Kecelakaan Kendaraan Sepeda Motor Pada Ruas Jalan Di Kota Ambon; *REKAYASA SIPIL / Volume 17, No.2 – 2023*.
- [13] Vanissa Wanika Siburian, Ika Elvina Mulyana; Prediksi Harga Ponsel Menggunakan Metode Random Forest; Vol.4 No.1; Prosiding Annual Research Seminar 2018.
- [14] Rahmana Dwi Shaputra; Implementasi Regresi Linear Untuk Prediksi Penjualan Dan Cashflow Pada Aplikasi Point Of Sales Kafe XYZ; 07 Juli 2021.
- [15] Husdi, Hastuti Dalai; Penerapan Metode Regresi Linear Untuk Prediksi Jumlah Bahan Baku Produksi Selai Bilfagi; *Jurnal Informatika*, Vol. 10 No. 2 Oktober 2023.
- [16] Prana, S., Isesa, W., Prayogi, A. S., & Fahrudin, T. M. (2018). PEMODELAN DAN EVALUASI TREND FORECASTING PADA KONDISI KORBAN KECELAKAAN LALU LINTAS MENGGUNAKAN TREND MOMENT DAN LEAST SQUARE. In *Jurnal Sistem Cerdas* (Vol. 01).
- [17] Saadah, S., & Salsabila, H. (2021). Jurnal Politeknik Caltex Riau Prediksi Harga Bitcoin Menggunakan Metode Random Forest (Studi Kasus: Data Acak Pada Awal Masa Pandemic Covid-19). In *Jurnal Komputer Terapan* (Vol. 7, Issue 1). <https://jurnal.pcr.ac.id/index.php/jkt/>
- [18] Fachid, S., & Triayudi, A. (2022). Perbandingan Algoritma Regresi Linier dan Regresi Random Forest Dalam Memprediksi Kasus Positif Covid-19. *JURNAL MEDIA INFORMATIKA BUDIDARMA*, 6(1), 68. <https://doi.org/10.30865/mib.v6i1.3492>
- [19] Fraticasari, S. Y., Ratnawati, D. E., & Wihandika, R. C. (2018). *Optimasi Pemodelan Regresi Linier Berganda Pada Prediksi Jumlah Kecelakaan Sepeda Motor Dengan Algoritma Genetika* (Vol. 2, Issue 5). <http://j-ptiik.ub.ac.id>
- [20] Sofie Soraya, N. (2023). Komparasi linear regression, random forest regression, dan multilayer perceptron regression untuk prediksi tren musik TikTok. *AITI: Jurnal Teknologi Informasi*, 20(Agustus), 191-205.