

## PENERAPAN DECISION TREE REGRESSOR DALAM PENGEMBANGAN BUSINESS INTELLIGENCE UNTUK PREDIKSI PERGERAKAN WISATAWAN JAWA TIMUR

Moh Dzaky Irhab<sup>1</sup>, Aryo Nugroho<sup>2\*</sup>

<sup>1</sup>Fakultas Ilmu Komputer, Sistem Informasi, Universitas Narotama, Surabaya, Indonesia

<sup>2</sup>Fakultas Ilmu Komputer, Universitas Narotama, Surabaya, Indonesia

Email: <sup>1</sup>zakyirhab1@gmail.com, <sup>2</sup>aryo.nugroho@narotama.ac.id

### Abstrak

Kemajuan teknologi dan analisis data sangat penting bagi kemajuan industri pariwisata di Jawa Timur. Penelitian ini bertujuan untuk menerapkan *Decision Tree Regression* (DTR) dalam pengembangan *Business Intelligence* (BI) untuk menganalisis dan meramalkan pergerakan wisatawan di wilayah tersebut. Dengan menggunakan data dari tahun 2018 hingga 2022, termasuk data pergerakan wisatawan, warisan budaya, dan destinasi wisata dari berbagai kota di Jawa Timur, penelitian ini menggunakan DTR untuk analisis, dengan tujuan membuat model prediksi untuk pola wisatawan di masa depan. Pembagian data yang optimal antara data pelatihan dan data pengujian dilakukan dengan rasio 70:30. Hasil penelitian menunjukkan bahwa model DTR untuk pergerakan wisatawan domestik mencapai MAE optimal sebesar 0,843 dan MAPE optimal sebesar 6,11%, sedangkan model untuk pergerakan wisatawan internasional mencapai MAE optimal sebesar 2,309 dan MAPE optimal sebesar 44,20%. Temuan ini menunjukkan potensi dari model DTR yang dikembangkan untuk memberikan wawasan berharga bagi para pemangku kepentingan yang terlibat dalam perencanaan dan manajemen sektor pariwisata di Jawa Timur. Selain itu, penelitian ini juga berkontribusi pada peningkatan BI untuk pariwisata, membentuk dasar untuk strategi pariwisata yang lebih cerdas dan berkelanjutan di wilayah tersebut.

**Kata Kunci:** Business Intelligence, Decision Tree Regressor, Forecasting, Prediksi, Pergerakan Wisatawan

### Abstract

The advancement of technology and data analysis is crucial for the development of the tourism industry in East Java. This study aims to apply *Decision Tree Regression* (DTR) in the development of *Business Intelligence* (BI) to analyze and forecast tourist movements in the region. Utilizing data from 2018 to 2022, including tourist movement data, cultural heritage, and tourist destinations from various cities in East Java, this research employs DTR for analysis, with the objective of creating predictive models for future tourist patterns. The optimal data splitting between training and testing data was performed with a ratio of 70:30. The results indicate that the DTR model for domestic tourist movement achieved an optimal MAE of 0.843 and an optimal MAPE of 6.11%, while the model for international tourist movement achieved an optimal MAE of 2.309 and an optimal MAPE of 44.20%. These findings demonstrate the potential of the developed DTR model to provide valuable insights for stakeholders involved in the planning and management of the tourism sector in East Java. Furthermore, this research contributes to the enhancement of BI for tourism, laying the foundation for smarter and more sustainable tourism strategies in the region.

**Keywords:** Business Intelligence, Decision Tree Regressor, Forecasting, Prediction, Tourist Movement

## 1. PENDAHULUAN

Wisata merupakan sektor ekonomi yang sangat penting di Indonesia, dan Jawa Timur telah menjadi tujuan favorit bagi berbagai macam wisatawan, baik domestik maupun internasional, selama bertahun-tahun. Pada tahun 2018, jumlah kunjungan wisatawan mancanegara ke Jawa Timur mencapai 830.968, mencatat peningkatan sekitar 20,34% dari tahun sebelumnya, yakni 690.509 kunjungan pada tahun 2017. Dari total kunjungan tersebut, sekitar 322.965 wisatawan tiba melalui

Bandara Internasional Juanda, sementara sisanya sebanyak 508.003 kunjungan berasal dari jalur darat atau penerbangan internasional tidak langsung [1]. Untuk mengelola dan meningkatkan pengembangan sektor pariwisata ini secara efisien, pemahaman yang komprehensif tentang visualisasi pola pergerakan wisatawan, tren, dan faktor-faktor yang memengaruhinya menjadi sangat penting. Penggunaan visualisasi data merupakan solusi kontemporer dalam komunikasi visual yang dapat membantu dalam menyajikan informasi secara lebih jelas dan mudah dimengerti [2]. Visualisasi adalah metode untuk membuat gambar, diagram,

atau animasi guna menyajikan informasi. Penggunaan teknik visualisasi informasi sangat bermanfaat dalam menangani masalah pengolahan data besar. Terdapat dua konsep utama dalam visualisasi: visualisasi ilmiah dan visualisasi informasi, dimana keduanya menghasilkan model grafis dan tampilan visual data yang berinteraksi langsung dengan pengguna. Saat menggunakan visualisasi ilmiah untuk mengeksplorasi dan mengumpulkan data, model grafis sering kali dibangun berdasarkan data pengukuran atau simulasi yang merepresentasikan objek atau konsep yang relevan dengan kehidupan nyata [3]. *Business Intelligence* (BI) merupakan sistem yang digunakan untuk menganalisis data dan mengolahnya menjadi informasi untuk membantu pengambilan keputusan proses bisnis [4]. *Business Intelligence* (BI) adalah suatu metode yang mengintegrasikan teknologi, proses, dan teknik analisis untuk mengubah data menjadi informasi yang signifikan dan dapat digunakan untuk mengambil tindakan bagi perusahaan [5]. Pemanfaatan *business intelligence* dalam analisis besar data sangat efisien untuk memahami kondisi bisnis, sehingga menghasilkan informasi yang berarti bagi semua pihak yang terlibat, terutama manajemen, karyawan, pelanggan, mitra bisnis, pemilik, dan pihak-pihak lain yang memiliki kepentingan [6]. *Business Intelligence* (BI) bisa dianggap sebagai penggerak Sistem Manajemen Informasi, dan di lingkungan pendidikan tinggi, BI memberikan kesempatan untuk menyediakan informasi yang signifikan mengenai kinerja [7]. Dalam konteks ini, tujuan pengembangan *business intelligence* (BI) telah terbukti menjadi alat berharga yang mendukung para pemangku kepentingan di sektor pariwisata untuk mengambil keputusan yang cerdas dan strategis, tetapi akan lebih baik lagi jika memanfaatkan teknik seperti *Decision Tree Regression* dalam prediksi data.

Salah satu metode yang digunakan untuk menganalisis dan memprediksi pergerakan wisatawan adalah dengan menggunakan *Decision Tree Regression* (DTR). *Decision Tree* merupakan sebuah struktur berbasis proses yang berurutan, dimulai dari akar dan melibatkan evaluasi fitur serta pilihan di setiap langkah cabangnya, hingga mencapai daun terakhir. Pada titik akhir ini, tujuan atau target yang diinginkan biasanya telah tercapai [8]. Algoritma pohon keputusan mampu menggambarkan urutan beberapa peristiwa dalam bentuk struktur pohon keputusan. Struktur pohon keputusan ini berfungsi untuk membagi dataset yang besar menjadi sejumlah subset data yang lebih kecil [9].

Pada beberapa penelitian terdahulu dalam domain teknologi informasi, khususnya terkait penerapan *business intelligence* (BI) dan prediksi data. Pada penelitian pertama, ini fokus pada pembuatan *dashboard* BI untuk memantau

akreditasi di SMP Negeri 1 Sembawa, menggunakan BI sebagai alat visualisasi [10]. Penelitian selanjutnya membahas pengembangan BI *dashboard* untuk memonitor aktivitas pariwisata di Dinas Pariwisata Provinsi Bali, bertujuan menggantikan penggunaan kertas dengan media komputer untuk memudahkan manajemen data [6]. Selanjutnya hasil penelitian ketiga adalah pengembangan model prediksi prestasi mahasiswa dengan *Data Science. Decision Tree*, dengan akurasi 68%, dipilih sebagai model yang baik untuk memprediksi prestasi mahasiswa [11]. Pada penelitian keempat, fokus yang diinginkan pada penelitian tersebut adalah membandingkan model *Decision Tree, Naïve Bayes*, dan *Random Forest* untuk prediksi klasifikasi penyakit jantung. Hasilnya menunjukkan bahwa *Decision Tree* dengan akurasi 72% dan waktu eksekusi singkat adalah pilihan terbaik [8]. Dan pada penelitian kelima, hasil yang dikeluarkan adalah sebuah prediksi jumlah kunjungan wisatawan mancanegara ke Indonesia dengan menggunakan *Seasonal ARIMA (0,1,1)(0,1,1)12*, yang diklaim memiliki tingkat keakuratan yang baik. Secara keseluruhan, tabel mencerminkan beragam aplikasi teknologi BI dan analisis data dalam berbagai konteks [12]. Dari hasil penelitian terdahulu yang ada, penelitian ini dilakukan dikarenakan tidak adanya/masih belum ada pengembangan *business intelligence* dengan penggunaan model *machine learning* untuk melakukan sebuah prediksi menjadi satu aplikasi atau tambahan dalam fitur aplikasi

Dalam penelitian ini, *Decision Tree Regression* akan digunakan untuk mengembangkan *Business Intelligence* yang dapat menganalisis dan memprediksi pergerakan wisatawan di Jawa Timur, dengan tujuan memberikan pemahaman mendalam tentang bagaimana metode ini dapat mendukung pengambilan keputusan di sektor pariwisata. Penelitian ini akan melibatkan langkah-langkah praktis seperti perencanaan, perancangan, analisis data, implementasi, pengujian, dan review, dengan harapan memberikan dampak positif pada pengembangan sektor pariwisata di Jawa Timur.

## 2. METODOLOGI PENELITIAN



Gambar 1. Alur Penelitian

Penjelasan alur yang digunakan dalam penelitian ini disajikan dalam Gambar 1. Alur penelitian terbagi menjadi beberapa tahapan yang harus dilaksanakan secara menyeluruh agar penelitian ini dapat berjalan secara optimal.

Dalam tahapan metodologi penelitian ini, metode *Decision Tree Regression* (DTR) diterapkan pada tahapan implementasi hasil, setelah melalui proses *preprocessing data* dan feature extraction/selection jika diperlukan. Proses dimulai dengan pembersihan dan pemrosesan data, diikuti oleh pembagian data menjadi set pelatihan dan pengujian dengan rasio yang telah ditentukan. Setelah itu, metode DTR diterapkan untuk membangun model, diakhiri dengan evaluasi kinerja model menggunakan metrik yang relevan.

### 2.1 Perencanaan

Tahap awal dalam penelitian ini adalah perencanaan, di mana peneliti harus mengumpulkan informasi dan data yang relevan terkait dengan tujuan penelitian. Dataset yang digunakan dalam penelitian ini diperoleh dari Opendatajatim, yang merupakan portal data yang mendukung analisis, inovasi, dan transparansi data. Perencanaan juga melibatkan identifikasi kebutuhan untuk pengembangan dashboard dan perkiraan pergerakan jumlah wisatawan di Jawa Timur.

### 2.2 Perancangan

Tahap kedua melibatkan desain berdasarkan rencana yang telah disiapkan sebelumnya, yang penting untuk memastikan kesesuaian dengan kebutuhan. Desain penelitian mencakup pembuatan kasus pengguna dan aliran penggunaan aplikasi dalam bentuk *flowchart* dan juga penggunaan *usecase*.

### 2.3 Analisis dan Pengolahan Data

Tahap ketiga melibatkan analisis dan pemrosesan data yang telah dikumpulkan sebelumnya. Analisis ini bertujuan untuk mengidentifikasi dan mengatasi nilai yang kosong atau tidak relevan yang dapat mempengaruhi hasil akhir. Kesalahan pada tahap ini dapat berdampak signifikan pada model peramalan dan visualisasi dalam *dashboard* bisnis cerdas. Data diproses dengan menggunakan transformasi *log oneplus*, yang bertujuan untuk mengurangi variabilitas yang signifikan dalam data dan stabilisasi distribusi data yang memiliki variasi tinggi.

### 2.4 Implementasi Hasil

Setelah analisis data, tahap berikutnya adalah menerapkan data dalam model peramalan menggunakan metode *Decision Tree Regression* (DTR) dan visualisasi *dashboard* bisnis cerdas. Proses dimulai dengan membagi data yang telah diproses menjadi set pelatihan dan pengujian

dengan rasio yang ditentukan. Model DTR kemudian dilatih menggunakan set pelatihan untuk mengidentifikasi pola, diikuti dengan prediksi terhadap data pengujian untuk mengevaluasi akurasi model dengan metrik MAE (*Mean Absolute Error*) dan MAPE (*Mean Absolute Percentage Error*). Hasil model DTR divisualisasikan dalam dashboard bisnis cerdas. Meskipun penerapan model dan pengembangan dashboard telah selesai, hasil visualisasi melalui Streamlit belum dapat diakses. Hasil akhir digabungkan dalam satu kerangka kerja (*framework*) untuk memudahkan akses pengguna, menggunakan Streamlit.

### 2.5 Pengujian

Pada tahap pengujian, peneliti melakukan pengujian model yang telah disusun. Evaluasi model dilakukan menggunakan matriks pengujian, yaitu MAE (*Mean Absolute Error*) dan MAPE (*Mean Absolute Percentage Error*) guna mengukur bagus tidaknya sebuah model prediksi yang telah dibuat.

*Mean Absolute Error* (MAE) sendiri adalah rata-rata nilai aktual dan nilai prediksi yang bernilai mutlak positif [13]. Berikut persamaan 1 yang merupakan rumus MAE yang diterapkan.

$$\frac{1}{n} \sum_{t=1}^n |A_t - F_t| \quad (1)$$

Sedangkan MAPE (*Mean Absolute Percentage Error*) merupakan rata-rata absolut antara nilai peramalan dan nilai aktual yang dinyatakan sebagai persentase. MAPE adalah pengukuran statistik tentang akurasi perkiraan (prediksi) pada metode peramalan. MAPE digunakan untuk menghitung persentase kesalahan rata-rata secara mutlak (absolut) [14]. Yang dimana semakin kecil persentase yang dihasilkan pada rumus ini maka semakin baik model. Persamaan 2 tersebut merupakan cara perhitungan untuk menghitung MAPE

$$\frac{100}{n} \sum_{t=1}^n \left| \frac{A_t - F_t}{A_t} \right| \quad (2)$$

**Tabel 1.** Range Persentase MAPE

Nilai Mape	Akurasi Prediksi
< 10 %	Sangat Baik
10 – 20 %	Baik
20 – 50 %	Cukup
> 50 %	Buruk

Pada Tabel 1, Nilai MAPE yang kurang dari 10% maka memiliki tingkat akurasi prediksi yang tinggi. Jika nilai MAPE ada di antara 10% sampai 20% maka akurasi prediksinya baik dan bila diantara 20% sampai 50% tingkat akurasi *reasonable*. Jika nilai MAPE diatas 50% tingkat akurasi prediksinya rendah. Maka sedapat mungkin nilai MAPE dari peramalan yang telah dilakukan mencapai nilai dibawah 10% [15].

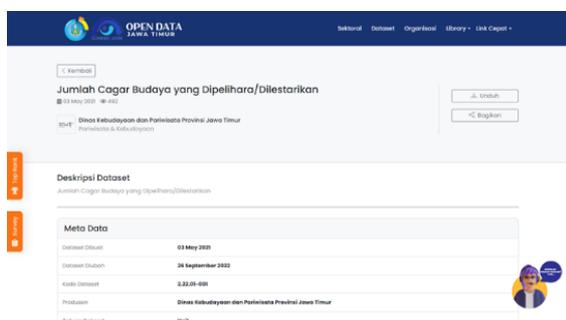
## 2.6 Review

Ini merupakan tahap akhir di mana peneliti mengevaluasi hasil penelitian untuk memastikan kesesuaian dengan tujuan yang telah ditetapkan. Selain itu, pada tahap ini juga dilakukan penyusunan laporan penelitian dengan pendekatan sistematis, yang mencakup hasil yang telah diperoleh dan dokumentasi temuan secara komprehensif.

## 3. HASIL DAN PEMBAHASAN

### 3.1. Perencanaan

Hasil yang dikeluarkan dari tahap perencanaan ini yaitu peneliti berhasil mendapatkan 3 *dataset* yang bisa diolah/digunakan untuk memprediksi pergerakan wisatawan dan berhasil melakukan riset terkait tools serta hal yang dibutuhkan untuk menjalankan penelitian tersebut. Data yang dikumpulkan mencakup pergerakan wisatawan, cagar budaya, dan tempat wisata di berbagai kota di Jawa Timur dari bulan januari tahun 2018 hingga bulan desember 2022. Pemilihan data tersebut didasari karena adanya hubungan besar kecilnya pergerakan wisatawan tersebut dengan sedikit banyaknya tempat wisata yang ada di Jawa Timur. Terlihat pada Gambar 2 merupakan salah satu tampilan dataset yang digunakan dan diambil dari portal data *opendata.jatimprov.go.id*. Seluruh dataset tersebut disimpan dengan format CSV (*Comma Separated Values*).

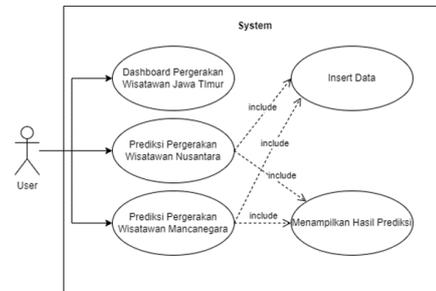


**Gambar 3.** Tampilan Dataset Yang Digunakan Dari Open Data Jatim

Setelah melakukan riset, alat utama yang digunakan dalam penelitian ini mencakup penggunaan *Tableau* dalam pengembangan

*business intelligence*, *Microsoft Excel* untuk menggabungkan tiga *dataset* menjadi satu *dataset* final, *Google Colab* sebagai *platform* untuk program *machine learning* dalam memprediksi pergerakan wisatawan, dan *Streamlit* sebagai *framework* untuk mengintegrasikan *Tableau* dengan model *machine learning* yang telah dibuat sebelumnya.

### 3.2. Perancangan

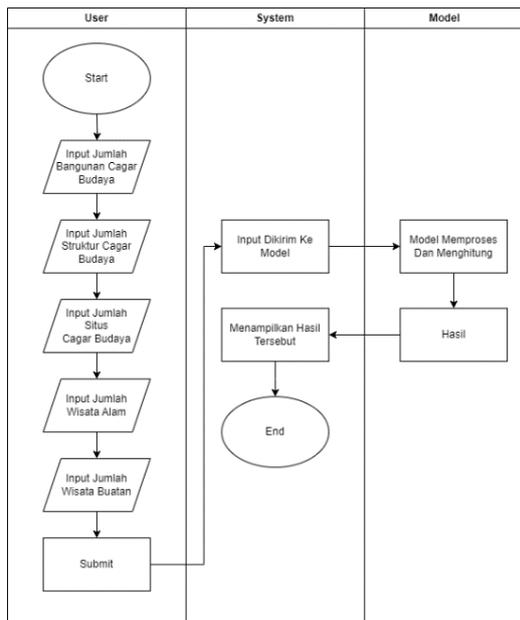


**Gambar 2.** Use Case

Pada tahap ini, peneliti melakukan pembuatan *usecase* dan *flowchart* dalam prediksi pergerakan wisatawan. Terlihat pada Gambar 3 merinci sebuah *use case* yang telah dikembangkan pada tahap perancangan aplikasi. Dalam *use case* ini, tergambar dengan jelas bahwa akses ke aplikasi dibatasi hanya untuk satu aktor, yakni pengguna (*user*). Pengguna ini, khususnya pemerintah dalam konteks pariwisata, memiliki kemampuan untuk melakukan sejumlah tindakan di dalam aplikasi.

Pengguna aplikasi ini memiliki fasilitas untuk menyelami visualisasi inteligeni bisnis pergerakan wisatawan di Jawa Timur. Selain itu, mereka dapat melakukan prediksi terkait pergerakan wisatawan nusantara, serta merancang proyeksi untuk pergerakan wisatawan mancanegara. Dalam konteks ini, "*user*" secara khusus mengacu pada entitas pemerintah yang bergerak di sektor pariwisata.

Tahap prediksi menjadi bagian yang signifikan dalam fungsionalitas aplikasi ini. Pengguna diberi kemampuan untuk menyisipkan data, dan hasil dari penyisipan tersebut akan diolah oleh model yang ada dalam aplikasi. Selanjutnya, aplikasi akan menampilkan hasil prediksi berdasarkan data yang telah dimasukkan oleh pengguna. Proses ini memperkuat keterlibatan pengguna dalam pengambilan keputusan berbasis data, dengan memberikan sarana untuk meramalkan pergerakan wisatawan berdasarkan informasi yang telah dikumpulkan dan diinjeksikan oleh pengguna sendiri.



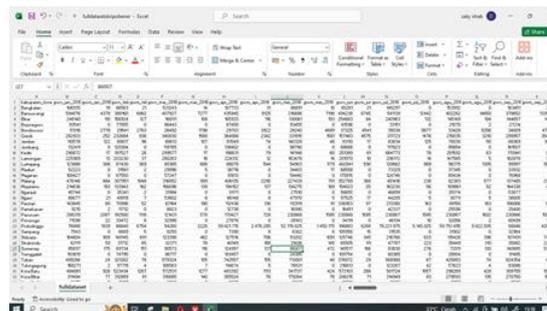
**Gambar 4.** Flowchart Forecasting Pergerakan Wisatawan

Setelah itu Gambar 4 memperlihatkan sebuah *flowchart* yang menggambarkan secara detail alur proses dari tahapan *forecasting* pergerakan wisatawan, baik yang bersifat nusantara maupun mancanegara. Dalam gambaran alur tersebut, pengguna diharuskan untuk memasukkan sejumlah parameter yang menjadi fokus perhitungan, termasuk jumlah bangunan cagar budaya, jumlah struktur cagar budaya, jumlah situs cagar budaya, jumlah wisata alam, dan jumlah wisata buatan. Setelah pengguna menyelesaikan langkah-langkah ini, langkah selanjutnya adalah melakukan submit.

Proses submit tersebut mengaktifkan sistem untuk mengirimkan seluruh input yang telah dimasukkan oleh pengguna ke model yang telah terlatih sebelumnya. Model ini kemudian melakukan perhitungan berdasarkan input tersebut dan menghasilkan output yang mencerminkan prediksi pergerakan wisatawan. Output ini dapat mencakup berbagai informasi terkait tren pergerakan, estimasi jumlah pengunjung, dan variabel-variabel lain yang signifikan.

Hasil prediksi yang diperoleh dari model selanjutnya ditampilkan oleh sistem kepada pengguna. Dengan demikian, pengguna dapat dengan mudah mengakses dan mengevaluasi hasil prediksi tersebut, memberikan mereka wawasan yang berharga dalam pengambilan keputusan terkait perencanaan dan pengelolaan pariwisata. Seluruh alur ini membentuk sistem yang terintegrasi, memungkinkan pengguna untuk berpartisipasi aktif dalam proses *forecasting* dan meraih manfaat dari informasi prediktif yang dihasilkan.

### 3.3. Analisis dan Olah Data



**Gambar 5.** Proses Penyatuan Dataset

Gambar 5 menyoroti tahap kritis di mana terjadi penyatuan beberapa kumpulan data. Kumpulan data ini mencakup rincian pergerakan wisatawan, warisan budaya, dan statistik pariwisata untuk setiap kota. Tujuan utama di balik penggabungan kumpulan data ini adalah untuk mengungkap potensi pola dan tren yang mungkin timbul dari integrasinya.

Penting untuk dicatat bahwa mayoritas periode pembaruan data yang digunakan dalam analisis ini bersifat tahunan. Dalam tahap ini, dilakukan pula pemeriksaan rinci terkait adanya nilai-nilai kosong serta nilai-nilai yang mungkin memiliki tanda titik dan koma. Upaya detil ini bertujuan untuk memastikan kualitas dan integritas data yang dihasilkan.

Setelah tahap penemuan data selesai, dilakukan proses pembersihan yang melibatkan penambahan kolom baru, seperti jumlah pergerakan wisatawan nusantara dan mancanegara, dengan periode pembaruan yang disesuaikan menjadi per tahun. Langkah selanjutnya mencakup penghapusan dan pengisian beberapa nilai kosong yang ditemukan, serta penyesuaian tipe data untuk meminimalkan potensi kehilangan nilai asli dan mengurangi potensi kesalahan data pada saat proses visualisasi dan prediksi masa depan. Semua tindakan ini diarahkan pada pemastian integritas dan keandalan data guna mendukung analisis visualisasi dan peramalan yang akurat.

Dalam Gambar 6 dibawah, terlihat dataset akhir yang merupakan hasil dari penggabungan sejumlah dataset yang sebelumnya telah dilakukan. Sebelum penggabungan, langkah pembersihan data dilakukan untuk mengatasi nilai yang hilang, duplikasi, dan ketidakcocokan format agar dataset siap untuk analisis. Pada tahap ini, penelitian mengalami pendalaman yang sangat mendetail terkait bagaimana setiap kolom dalam dataset tersebut memiliki potensi pengaruh terhadap jumlah pergerakan wisatawan, baik yang bersifat nusantara maupun mancanegara. Pendalaman ini dilakukan oleh peneliti dengan berfokus pada perspektif ketertarikan wisatawan terhadap berbagai aspek.

Gambar 6. Final Dataset Akhir

Dari hasil pendalaman tersebut, sejumlah delapan kolom data terpilih sebagai variabel yang signifikan untuk analisis lebih lanjut. Variabel-variabel tersebut mencakup jumlah bangunan cagar budaya, jumlah struktur cagar budaya, jumlah situs cagar budaya, total wisata alam, total wisata buatan, total wisata budaya, jumlah pergerakan wisatawan nusantara, dan jumlah pergerakan wisatawan mancanegara.

jumlah_bangunan_cagar_budaya	jumlah_struktur_cagar_budaya	jumlah_situs_cagar_budaya	total_wisata_alam	total_wisata_buatan	total_wisata_budaya	jumlah_pergerakan_wisatawan_nusantara	jumlah_pergerakan_wisatawan_mancanegara
1.38254	3.387285	1.859438	2.079442	0.89147	2.079442	16.411944	5.398338
2.832213	1.859438	2.325265	2.893712	2.187225	3.17054	15.412853	11.748335
2.079442	3.387285	2.387595	1.945910	1.388294	2.832212	14.85721	7.889910
1.859438	3.17054	2.464607	1.945910	2.079442	2.187225	13.477392	2.703830
2.772589	0.000000	3.328381	1.945910	1.859438	1.791759	12.309887	10.704153
4.307332	1.945910	0.89147	0.000000	3.851148	1.859438	14.827323	8.942377
2.772589	1.791759	0.000000	0.000000	2.944439	2.325265	14.832769	2.854949
3.325265	1.945910	0.000000	0.000000	2.079442	2.187225	15.88954	4.330733
2.187225	0.89147	0.89147	1.988112	1.945910	1.859438	13.842737	5.483332
3.88992	0.89147	0.000000	1.859438	2.772589	3.325265	15.475241	8.249782

Gambar 7. Transformasi Data Log One Plus

Dalam Gambar 7, terlihat bahwa salah satu tahap kritis dalam persiapan data adalah proses transformasi *log*, yang diimplementasikan untuk model *forecasting* pergerakan wisatawan nusantara dan mancanegara di Jawa Timur. Proses transformasi ini diterapkan secara merata kepada setiap kolom dalam tabel *dataset*. Gambar ini memberikan hasil yang menunjukkan bahwa seluruh nilai dan tipe data pada *dataset* mengalami perubahan menjadi tipe data *float* setelah melalui proses transformasi *log*.

Proses transformasi *log* ini tidak hanya berfokus pada konversi tipe data, tetapi juga memiliki tujuan lebih dalam. Transformasi *log* digunakan untuk mengatasi asimetri dan memitigasi dampak *outlier* pada distribusi data. Dengan menerapkan transformasi ini, *dataset* menjadi lebih sesuai dengan asumsi dasar yang diperlukan oleh beberapa model peramalan. Oleh karena itu, setiap kolom diubah menjadi nilai *float* untuk menunjukkan bahwa proses transformasi *log* telah berhasil diterapkan secara konsisten pada semua variabel dalam *dataset*, mempersiapkan data untuk tahapan selanjutnya dalam analisis dan peramalan pergerakan wisatawan.

### 3.4. Implementasi Hasil

```

from sklearn.model_selection import RandomizedSearchCV
from scipy.stats import randint

# Tentukan distribusi yang ingin digunakan untuk setiap parameter
param_dist = {
    'max_depth': randint(1, 100), # Rentang nilai yang valid untuk max_depth
    'min_samples_split': randint(2, 100), # Rentang nilai yang valid untuk min_samples_split (mulai dari 2)
    'min_samples_leaf': randint(1, 100), # Rentang nilai yang valid untuk min_samples_leaf
    # Tambahkan parameter lain yang valid jika perlu
}

# Inisialisasi DecisionTreeRegressor
algorithm = DecisionTreeRegressor(max_features = 1.0)

# Inisialisasi RandomizedSearchCV dengan model dan parameter distribusi
random_search = RandomizedSearchCV(algorithm, param_distributions=param_dist, n_iter=1000, cv=5, random_state=42)

# Lakukan pencarian parameter terbaik
random_search.fit(X_train, y_train)

# Cetak parameter terbaik
print("Parameter terbaik:", random_search.best_params_)

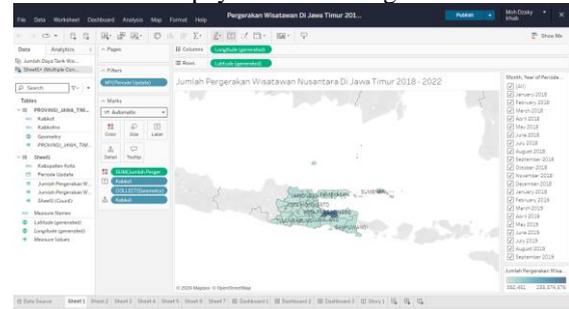
```

Gambar 8. Syntax Machine Learning Prediksi Jumlah Pergerakan Wisatawan

Dalam Gambar 8, terlihat sintaksis yang digunakan dalam proses pemodelan untuk memprediksi jumlah pergerakan wisatawan nusantara dan mancanegara. Algoritma yang digunakan dalam tugas ini adalah *decision tree regression*, dengan peningkatan kinerja melalui penerapan teknik *hyperparameter random search*. Sintaksis pada gambar menggambarkan implementasi algoritma ini dalam bahasa pemrograman yang digunakan.

Parameter *decision tree*, yang memiliki peran kunci dalam membentuk model, dijelaskan dalam *dictionary* yang dinamakan "*param\_dist*". *Dictionary* ini mencakup berbagai parameter yang akan digunakan dalam *decision tree*. Selanjutnya, proses *hyperparameter random search* menggunakan nilai-nilai dalam *dictionary* ini untuk memulai pencarian secara acak dengan seribu iterasi.

Proses pencarian *hyperparameter* secara acak (*random search*) ini bertujuan untuk menemukan kombinasi parameter terbaik yang dapat mengoptimalkan kinerja model. Dengan melakukan seribu iterasi, algoritma ini secara efektif mengeksplorasi ruang parameter untuk mencapai konfigurasi terbaik. Sintaksis ini mencerminkan upaya untuk meningkatkan akurasi.

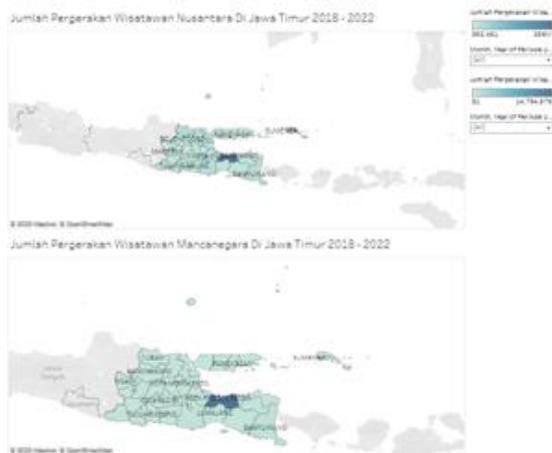


Gambar 9. Tampilan Pembuatan Business Intelligence Tableau Public

Setelah menyelesaikan tahap pemodelan, langkah berikutnya adalah menciptakan *dashboard business intelligence* untuk analisis pergerakan wisatawan nusantara dan mancanegara di Jawa Timur. Pada Gambar 9, terlihat tampilan dari proses pembuatan *dashboard* menggunakan *platform tableau public*. Tahapan ini menandai bagian krusial dalam menghadirkan hasil analisis secara visual dan dapat diakses secara lebih intuitif.

Dalam proses pembuatan *dashboard business intelligence*, langkah pertama melibatkan penggunaan data yang telah dibersihkan sebelumnya. Visualisasi data dilakukan dengan variasi yang cermat, bertujuan untuk menghasilkan wawasan baru dari setiap aspek pergerakan wisatawan. Dalam konteks ini, *tableau public* digunakan sebagai alat utama untuk merancang dan mengeksekusi visualisasi data yang komprehensif.

Proses pembuatan *business intelligence* dimulai dengan memasukkan *dataset* yang ingin divisualisasikan ke dalam *platform*. Selanjutnya, dilakukan pembuatan visualisasi data di setiap *sheet*, dengan fokus pada elemen-elemen yang dianggap krusial untuk memahami karakteristik pergerakan wisatawan. Seluruh *sheet* yang telah dibuat kemudian digabungkan dalam sebuah *dashboard*, menciptakan tampilan yang baik. Langkah terakhir melibatkan penyimpanan hasil *dashboard* ke dalam *cloud*, memungkinkan aksesibilitas yang lebih mudah dan bersama-sama.



**Gambar 10.** Halaman Pertama Dashboard Business Intelligence

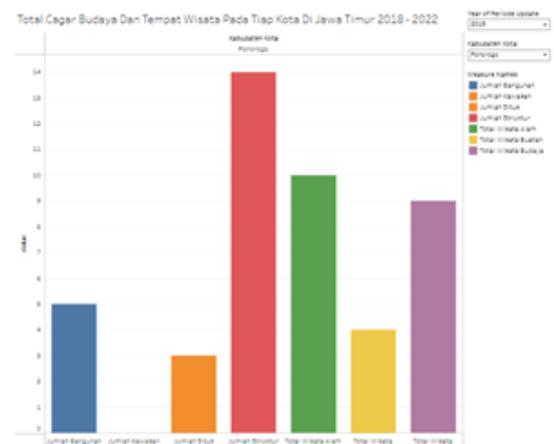
Pada Gambar 10, merupakan halaman pertama hasil dari implementasi *business intelligence dashboard* yang didapatkan dari data yang sudah dikumpulkan dan diolah sebelumnya. Halaman tersebut menampilkan peta persebaran pergerakan wisatawan nusantara dan mancanegara di Jawa Timur. Indikasi tingginya wisatawan terlihat dengan tebal tipis warna biru pada peta, semakin tebal warna biru maka tingkat pergerakan wisatawan tinggi dan sebaliknya. *User* disini juga bisa mengubah periode update bulan dan tahun

pada peta persebaran sesuai kebutuhan dan keinginan *user*.



**Gambar 11.** Halaman Kedua Dashboard Business Intelligence

Pada Gambar 11, merupakan halaman kedua dari hasil dari implementasi *business intelligence dashboard*. Pada halaman ini, ada dua *area chart* diatas yang menampilkan list sepuluh kota yang memiliki jumlah tinggi dalam pergerakan wisatawan nusantara dan mancanegara dan dua *tree chart* yang menampilkan list sepuluh kota yang memiliki jumlah tinggi dalam pergerakan wisatawan nusantara dan mancanegara pada hari libur. Pada *area chart* tersebut *user* bisa melakukan *highlight* pada tiap kota dengan klik salah satu daftar kabupaten kota yang memiliki kotak warna yang berbeda beda. *User* juga bisa mengubah periode update dari *area chart* dengan skala periode tahun. Dan untuk *tree chart* sendiri, *User* bisa klik salah satu kotak yang ada dan melihat detail nama kota dan jumlah pergerakan wisatawan. Indikasi besar kecilnya suatu pergerakan wisatawan pada *tree map* tersebut adalah besar kecilnya kotak dan juga tebal tipisnya warna di setiap kotak.



**Gambar 12.** Halaman Kedua Dashboard Business Intelligence

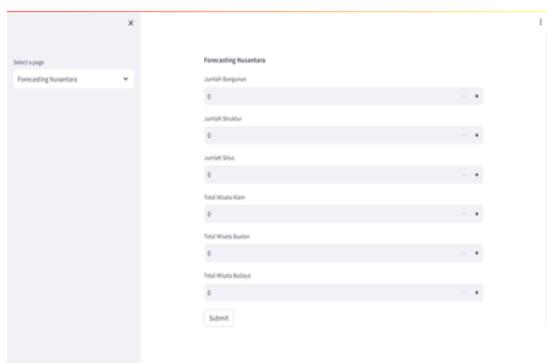
Pada Gambar 12, merupakan halaman ketiga dari hasil dari implementasi *business intelligence dashboard*. Pada halaman ini menampilkan *bar chart* yang dimana menampilkan total cagar budaya dan tempat wisata pada tiap kota di Jawa timur. *Chart* ini memberikan informasi total jumlah bangunan cagar budaya, jumlah kawasan budaya, jumlah situs budaya, jumlah struktur budaya, total wisata alam, total wisata buatan, dan total wisata budaya pada setiap kota. *User* bisa mengubah periode *update* pada chart dengan skala pertahun.

Setelah hasil modelling prediksi dan juga pengembangan *dashboard business intelligence* jumlah pergerakan wisatawan telah dilakukan, hal akhir yang dilakukan pada tahap ini adalah menyatukan dua hasil tersebut menjadi satu. Pada proses ini, peneliti menggunakan *framework streamlit* sebagai sarana penggabungan kedua hasil tersebut dan juga penggunaan *ngrok* yang dimana berfungsi sebagai alat yang digunakan untuk membuat sambungan *tunnel* yang aman dari internet ke *server* lokal atau mesin virtual yang di-*host* di berbagai *platform*.



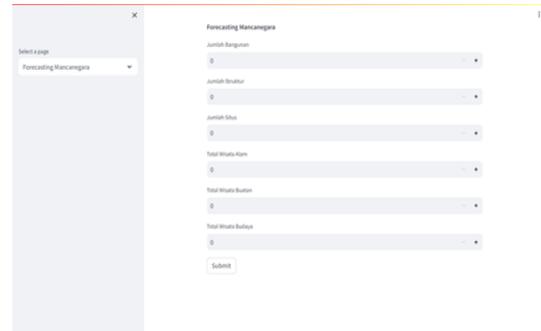
**Gambar 13.** Halaman Utama Web Aplikasi

Berikut pada Gambar 13 merupakan halaman utama *web* aplikasi dari hasil penggabungan *business intelligence* dan juga prediksi pergerakan wisatawan di Jawa timur. *User* bisa melakukan navigasi ke tampilan lain seperti *forecasting* wisatawan nusantara dan mancanegara hanya dengan memilih halaman tersebut pada bagian *sidebar dropdown*



**Gambar 14.** Halaman Forecasting Pergerakan Wisatawan Nusantara

Berikut pada Gambar 14, merupakan tampilan *forecasting* wisatawan nusantara dan mancanegara yang dimana kedua halaman tersebut memiliki tampilan yang sama yang dimana *user* bisa mengisi data yang ada dan nantinya hasil tersebut akan muncul dibawah tombol *submit*.



**Gambar 15.** Halaman Forecasting Pergerakan Wisatawan Mancanegara

Pada Gambar 15, merupakan tampilan *forecasting* wisatawan mancanegara yang dimana halaman ini juga memiliki tampilan yang sama yang dimana *user* bisa mengisi data yang ada dan nantinya hasil tersebut akan muncul dibawah tombol *submit*. Kedua tampilan *forecasting* juga memiliki *sidebar dropdown* yang berfungsi sama seperti yang ada pada tampilan utama yaitu melakukan navigasi ke halaman lainnya.

### 3.5. Pengujian

Pada tahap ini, model prediksi pergerakan wisatawan nusantara dan mancanegara dilakukan pengujian guna melihat apakah model tersebut cukup baik dalam menghasilkan sebuah prediksi. Pada Persamaan 2.1 dan 2.2 yang telah dijelaskan pada bab sebelumnya, persamaan tersebut digunakan untuk menguji dan menghasilkan nilai MAE dan MAPE pada model prediksi yang telah dibuat. Berikut hasil yang uji yang tertera.

**Tabel 2.** Hasil Pengujian Model Prediksi Pergerakan Nusantara

Split Data	Parameter Terbaik	MAE	MAPE
60% Train : 40% Test	'max_depth': 58, 'min_samples_leaf': 5, 'min_samples_split': 6	0.89	6.53%
70% Train : 30% Test	'max_depth': 60, 'min_samples_leaf': 12,	0.84	6.11%

30% Test	'min_samples_split': 33		
80% Train : 20% Test	'max_depth': 95, 'min_samples_leaf': 8, 'min_samples_split': 26	0.87	6.33 %
90% Train : 10% Test	'max_depth': 86, 'min_samples_leaf': 11, 'min_samples_split': 17	0.86	6.35 %

Tabel 2 menunjukkan hasil pengujian model prediksi pergerakan wisatawan nusantara dengan berbagai pembagian data. Hasil terbaik diperoleh pada kombinasi 70% data latih dan 30% data uji, yang menghasilkan MAE sebesar 0.84 dan MAPE sebesar 6.11%. Ini menunjukkan bahwa pembagian ini adalah yang paling optimal untuk model ini.

**Tabel 3.** Hasil Pengujian Model Prediksi Pergerakan Mancanegara

Split Data	Parameter Terbaik	MAE	MAPE
60% Train : 40% Test	'max_depth': 66, 'min_samples_leaf': 9, 'min_samples_split': 24	2.88	56.71%
70% Train : 30% Test	'max_depth': 55, 'min_samples_leaf': 4, 'min_samples_split': 97	2.30	44.20 %
80% Train : 20% Test	'max_depth': 26, 'min_samples_leaf': 8, 'min_samples_split': 64	2.26	53.74 %
90% Train : 10% Test	'max_depth': 3, 'min_samples_leaf': 41, 'min_samples_split': 96	2.80	51.30 %

Tabel 3 menunjukkan hasil pengujian model prediksi pergerakan wisatawan mancanegara. Pembagian data yang memberikan hasil terbaik pada model ini adalah 70% data latih dan 30% data uji, yang menghasilkan MAE sebesar 2.30 dan MAPE sebesar 44.20%. Meskipun nilai MAPE lebih tinggi dibandingkan dengan model nusantara, hasil ini tetap menunjukkan bahwa kombinasi tersebut adalah yang paling optimal untuk model mancanegara.

#### 4. KESIMPULAN

Berdasarkan hasil penelitian yang dilakukan, maka kesimpulan yang bisa diambil pada penelitian ini sebagai berikut:

1. Pada penelitian ini berhasil melakukan pengembangan *business intelligence* dengan menggunakan *decision tree regression* yang bisa digunakan untuk memprediksi pergerakan wisatawan nusantara dan mancanegara pada daerah Jawa Timur.
2. Hasil pengujian menunjukkan bahwa kombinasi *splitting data* terbaik untuk model prediksi wisatawan nusantara adalah 70% data *train* dan 30% data test, yang menghasilkan nilai MAE sebesar 0.84 dan MAPE sebesar 6.11%. Sedangkan untuk model prediksi wisatawan mancanegara, kombinasi *splitting data* terbaik juga adalah 70% data *train* dan 30% data test, menghasilkan nilai MAE sebesar 2.30 dan MAPE sebesar 44.20%.

Untuk penelitian selanjutnya, disarankan untuk mempertimbangkan penggunaan teknik pemodelan yang lebih kompleks, seperti *ensemble methods* (misalnya *Random Forest* atau *Gradient Boosting*), yang dapat meningkatkan akurasi prediksi. Selain itu, eksplorasi variabel tambahan dan analisis fitur lebih mendalam dapat membantu memahami faktor-faktor yang mempengaruhi pergerakan wisatawan dengan lebih baik. Peneliti juga disarankan untuk melakukan pengujian dengan metode *cross-validation* untuk mendapatkan evaluasi model yang lebih *robust*.

#### 5. REFERENCES

- [1] Dhebys Suryani, Rokhimatul Wakhidah, and A. Latifudin, "PERAMALAN JUMLAH PENGUNJUNG WISATAWAN MANCANEGARA MENGGUNAKAN METODE FUZZY TIME SERIES DI JAWA TIMUR," *J. Inform. Polinema*, vol. 8, no. 2, pp. 43–48, Mar. 2022, doi: 10.33795/jip.v8i2.525.
- [2] A. Mulyani, "VISUALISASI DATA TICKETING SERVICEDESK DENGAN DASHBOARD PADA PT BRANTAS ABIPRAYA (PERSERO)," vol. 7, 2023.

- [3] S. Angreini and E. Supratman, "Visualisasi Data Lokasi Rawan Bencana Di Provinsi Sumatera Selatan Menggunakan Tableau," *J. Nas. Ilmu Komput.*, vol. 2, no. 2, pp. 135–147, Nov. 2021, doi: 10.47747/jurnalnik.v2i2.528.
- [4] M. R. Atsani, G. Tyas Anjari, and N. Mega Saraswati, "Pengembangan Business Intelligence Di Rumah Sakit (Studi Kasus: RSUD Prof. Dr. Margono Soekarjo Purwokerto)," *Telematika*, vol. 12, no. 2, pp. 124–138, Aug. 2019, doi: 10.35671/telematika.v12i2.839.
- [5] T. Tumini and E. S. Subekti, "Implementasi Business Intelligence Untuk Menganalisis Data Proses Manufaktur Menggunakan Google Data Studio," *J. Ilm. Tek. Inform. Dan Komun.*, vol. 3, no. 3, pp. 143–151, 2023.
- [6] E. H. Saragih, I. P. A. Bayupati, and G. A. A. Putri, "Pengembangan Business Intelligence Dashboard untuk Monitoring Aktivitas Pariwisata (Studi Kasus: Dinas Pariwisata Provinsi Bali)," *J. Teknol. Inf. Dan Ilmu Komput.*, vol. 8, no. 6, pp. 1159–1168, Nov. 2021, doi: 10.25126/jtiik.2021863755.
- [7] D. Trisnawarman and M. C. Imam, "BUSINESS INTELLIGENCE FRAMEWORK FOR PERFORMANCE MEASUREMENT IN HIGHER EDUCATION STUDY PROGRAMS," *J. Muara Sains Teknol. Kedokt. Dan Ilmu Kesehatan*, vol. 4, no. 2, p. 249, Oct. 2020, doi: 10.24912/jmstkik.v4i2.8877.
- [8] D. H. Depari, Y. Widiastiwi, and M. M. Santoni, "Perbandingan Model Decision Tree, Naive Bayes dan Random Forest untuk Prediksi Klasifikasi Penyakit Jantung," *Inform. J. Ilmu Komput.*, vol. 18, no. 3, p. 239, Dec. 2022, doi: 10.52958/iftk.v18i3.4694.
- [9] D. Septhya *et al.*, "Implementasi Algoritma Decision Tree dan Support Vector Machine untuk Klasifikasi Penyakit Kanker Paru: Implementation of Decision Tree Algorithm and Support Vector Machine for Lung Cancer Classification," *MALCOM Indones. J. Mach. Learn. Comput. Sci.*, vol. 3, no. 1, pp. 15–19, May 2023, doi: 10.57152/malcom.v3i1.591.
- [10] A. S. Wibowo and A. Andri, "Dashboard Business Intelligence Vusialisasi Data Akreditasi Sekolah Pada SMP Negeri 1 Sembawa," *J. Nas. Ilmu Komput.*, vol. 2, no. 4, pp. 249–256, Nov. 2021, doi: 10.47747/jurnalnik.v2i4.536.
- [11] T. Tommy and A. M. Husein, "Model Prediksi Prestasi Mahasiswa Berdasarkan Evaluasi Pembelajaran Menggunakan Pendekatan Data Science," *Data Sci. Indones. DSI*, vol. 1, no. 1, pp. 14–20, 2021.
- [12] A. Meimela, "Prediksi Jumlah Kunjungan Wisatawan Mancanegara ke Indonesia," *Media Wisata*, vol. 19, no. 1, pp. 34–41, May 2021, doi: 10.36276/mws.v19i1.64.
- [13] L. S. Ihzaniah, A. Setiawan, and R. W. N. Wijaya, "Perbandingan Kinerja Metode Regresi K-Nearest Neighbor dan Metode Regresi Linear Berganda pada Data Boston Housing," *Jambura J. Probab. Stat.*, vol. 4, no. 1, pp. 17–29, May 2023, doi: 10.34312/jjps.v4i1.18948.
- [14] A. T. Nurani, A. Setiawan, and B. Susanto, "Perbandingan Kinerja Regresi Decision Tree dan Regresi Linear Berganda untuk Prediksi BMI pada Dataset Asthma," *J. Sains Dan Edukasi Sains*, vol. 6, no. 1, pp. 34–43, May 2023, doi: 10.24246/juses.v6i1p34-43.
- [15] A. Nurfadilah, W. Budi, E. Kurniati, and D. Suhaedi, "Penerapan Metode Moving Average untuk Prediksi Indeks Harga Konsumen," vol. 21, no. 1, 2022.