

## PREDIKSI JUMLAH KEDATANGAN WISATAWAN MANCANEGERA DI INDONESIA BERDASARKAN PINTU MASUK KEDATANGAN UDARA

<sup>1)</sup> Arya Prayuda, <sup>2)</sup> Irfan Pratama

<sup>1,2,3)</sup>Program Studi Sistem Informasi, Fakultas Teknologi Informasi, Universitas Mercu Buana Yogyakarta

<sup>1,2,3)</sup>Gg. Jemb.Merah No.84C, Soropadan, Condongcat, Depok, Sleman – DI Yogyakarta - Indonesia

E-mail : [arya12prayuda07@gmail.com](mailto:arya12prayuda07@gmail.com), [irfanp@mercubuana-yogya.ac.id](mailto:irfanp@mercubuana-yogya.ac.id)

### ABSTRAK

Indonesia memiliki keanekaragaman dan kekayaan alam yang menjadi daya tarik pariwisata di Indonesia. Pariwisata sebagai salah satu industri yang memberikan devisa tertinggi bagi negara karena memberikan dampak yang positif. Namun, adanya COVID-19 mengakibatkan penurunan jumlah kunjungan karena pembatasan wisatawan mancanegara. Pada Januari hingga November 2021, terjadi penurunan drastis sebesar 61,82% pada jumlah kunjungan wisatawan mancanegara dibanding periode yang sama di tahun 2020. Selain COVID-19, serta dukungan dalam membangun sarana yang menunjang peningkatan jumlah wisatawan mancanegara. Dari kondisi tersebut diperlukan prediksi yang digunakan sebagai dasar perencanaan dan membantu pengambilan keputusan. Tujuan dari penelitian ini adalah untuk mengembangkan model prediksi yang lebih akurat pada penelitian sejenis menggunakan data yang sama dalam memprediksi kunjungan wisatawan mancanegara di Indonesia melalui pintu masuk udara dengan menggunakan metode *XGBoost*, *Random Forest*, dan *Catboost* dengan berfokus pada metrik hasil evaluasi akurasi RMSE, MAE, dan MAPE serta melakukan prediksi pada periode 12 bulan kedepan. Dataset yang digunakan diambil dari Badan Pusat Statistik (BPS) yaitu data kunjungan wisatawan mancanegara berdasarkan pintu masuk kedatangan pada periode Januari 2017 sampai November 2021. Data yang digunakan berbentuk *time series* dan tidak stasioner. Dari hasil penelitian dapat dilihat berdasarkan hasil evaluasi akurasi bahwa model *XGBoost* penelitian ini mendapatkan hasil evaluasi akurasi lebih baik dibandingkan dua model lainnya dengan mendapatkan hasil nilai evaluasi akurasi RMSE sebesar 671935.2, MAE 648139.1, dan MAPE 20985.35. Model *XGBoost* lebih baik dengan nilai kesalahan akurasi yang lebih kecil dibandingkan model *Random Forest*, *Catboost*, dan penelitian sejenis yang menggunakan metode ARIMA dengan nilai RMSE 779670.7, MAE 749030.4, dan MAPE 23196.45.

**Kata Kunci:** Pariwisata, Prediksi, *XGBoost*, *Random Forest*, *Catboost*

### ABSTRACT

Indonesia has diversity and natural wealth that attracts tourism in Indonesia. Tourism is one of the industries that provides the highest foreign exchange for the country because it has a positive impact. However, the existence of COVID-19 has resulted in a decrease in the number of visits due to restrictions on foreign tourists. From January to November 2021, there was a drastic decrease of 61.82% in the number of foreign tourist visits compared to the same period in 2020. In addition to COVID-19, as well as support in building facilities that support the increase in the number of foreign tourists. From these conditions, predictions are needed that are used as a basis for planning and helping decision making. The purpose of this study is to develop a more accurate prediction model in similar studies using the same data in predicting foreign tourist arrivals in Indonesia through air entrances using the *XGBoost*, *Random Forest*, and *Catboost* methods by focusing on the accuracy evaluation results metrics RMSE, MAE, and MAPE and making predictions for the next 12 months. The dataset used is taken from the Central Statistics Agency (BPS), namely data on foreign tourist arrivals based on the arrival entrance in the period January 2017 to November 2021. The data used are time series and non-stationary. From the research results, it can be seen based on the accuracy evaluation results that the *XGBoost* model of this study gets better accuracy evaluation results than the other two models by getting the results of the RMSE accuracy evaluation value of 671935.2, MAE 648139.1, and MAPE 20985.35. The *XGBoost* model is better with a smaller accuracy error value than the *Random Forest* model, *Catboost*, and similar research using the ARIMA method with an RMSE value of 779670.7, MAE 749030.4, and MAPE 23196.45.

**Keyword:** Tourism, Prediction, *XGBoost*, *Random Forest*, *Catboost*

### PENDAHULUAN

Indonesia merupakan negara dengan keanekaragaman yang sangat besar[1].

Indonesia memiliki keanekaragaman dan kekayaan alam yang menjadi daya tarik tersendiri di dunia sebagai salah satu tujuan untuk berwisata[2]. Indonesia menjadi salah

satu negara pilihan bagi para wisatawan mancanegara (wisman). Wisatawan mancanegara atau wisman mengunjungi Indonesia menggunakan transportasi udara melalui pintu masuk udara seperti melalui bandara internasional[3].

Kekayaan dan keindahan alam Indonesia yang merupakan komponen penting pada industri wisata di Indonesia dengan wilayah yang luas dan didukung dengan sumber daya alam berpotensi untuk dimanfaatkan menjadi daya tarik wisatawan mancanegara. Kemenparekraf sebagai lembaga yang bertanggung jawab terhadap pariwisata di Indonesia, melakukan promosi bertema *Wonderful* Indonesia atau Pesona Indonesia untuk memperkenalkan kekayaan, keanekaragaman, dan budaya Indonesia[4], [5].

Akan tetapi, dengan munculnya COVID-19 mengakibatkan penurunan jumlah kunjungan wisatawan mancanegara. Ini mengakibatkan sejumlah negara di dunia termasuk Indonesia yang harus memberikan kebijakan pembatasan bagi warga negara asing untuk masuk ke Indonesia. Salah satu sektor yang sangat terdampak adalah pariwisata yang menimbulkan krisis sosial dan ekonomi[6], [7]. Dapat terlihat pada Januari sampai dengan November 2021, kunjungan wisatawan mancanegara ke Indonesia mengalami penurunan yang cukup turun drastis yaitu sebesar 61,82% dibandingkan dengan periode yang sama di tahun 2020. Selain kasus COVID-19 yang cukup berdampak pada industri pariwisata di Indonesia, juga diperlukan dukungan untuk membangun sarana dan prasarana demi menunjang peningkatan jumlah kunjungan wisatawan mancanegara.

Berdasarkan kondisi tersebut maka diperlukan prediksi kunjungan wisatawan mancanegara melalui pintu masuk udara yang nantinya dapat digunakan pemerintah sebagai dasar perencanaan dan pengambilan keputusan

pada industri pariwisata. Salah satu cara melakukan prediksi adalah menerapkan penggunaan *data mining*[3], [8].

*Data Mining* adalah kegiatan pengumpulan dan penggunaan data untuk menemukan pola keteraturan sehingga hasil dari *data mining* dapat digunakan dalam membantu pengambilan keputusan pada masa mendatang[9]. Namun pada buku Algoritma Data Mining dan Pengujian, Data mining juga diartikan sebagai proses penambangan data yang menghasilkan sebuah pengetahuan[10]. Terdapat beberapa metode atau teknik yang ada pada *data mining* yaitu estimasi, *forecasting*, klasifikasi, klustering dan asosiasi. Metode *forecasting* menjadi yang paling sesuai untuk melakukan sebuah prediksi.

*Forcaseting* atau peramalan adalah suatu usaha dalam memprediksi kejadian pada masa yang akan datang dengan cara melakukan suatu pengujian pada masa lalu dengan menggunakan data dalam bentuk mingguan, bulanan atau triwulan dengan memperhatikan suatu nilai berupa data di masa lalu dan data saat ini. Metode *forecasting* yang umum digunakan adalah deret waktu atau *time series*[11], [12].

*Time Series* atau deret waktu adalah rangkaian nilai-nilai suatu variabel yang dicatat dalam jangka waktu berurutan menggunakan analisa pola hubungan antar variable yang diperkirakan dan variable waktu[4]. Pada buku Aplikasi Ekonometrika untuk Data Panel dan *Time Series*[13]. *Time series* atau deret waktu adalah sebuah observasi pada sebuah obyek sepanjang kurun waktu tertentu

Sebelumnya, penelitian sejenis dengan melakukan prediksi jumlah kunjungan wisatawan mancanegara sudah pernah dilakukan dengan melakukan membandingkan beberapa methods yaitu ARIMA, *prophet*, dan glmnet yang diambil dari BPS mulai periode Januari 2017 hingga November 2021,

membandingkan prediksi berdasarkan RMSE (*Root Mean Squared Error*), MAPE (*Mean Absolute Percentage Error*), dan MAE (*Mean Absolute Error*). Hasil dari perbandingan metode ARIMA, *prophet*, dan *glmnet* pada penelitian sebelumnya menunjukkan metode ARIMA menghasilkan nilai akurasi atau RMSE 779670.7, MAE 749030.4, dan MAPE 23196.45 lebih baik jika dibandingkan dengan kedua metode lainnya yaitu *prophet* dan *glmnet*[3].

Penelitian ini bertujuan untuk mengembangkan model prediksi yang lebih akurat untuk memprediksi jumlah kunjungan wisatawan mancanegara berdasarkan pintu masuk udara yang diambil dari BPS mulai periode Januari 2017 hingga November 2021 yang sama pada penelitian sebelumnya diatas, dengan melakukan perbandingan tiga metode yaitu *xgboost*, *random forest*, dan *catboost* sama seperti penelitian sebelumnya diatas. Namun, penelitian ini akan berfokus pada peningkatan akurasi dengan membandingkan tiga metode yang berbeda dari penelitian sebelumnya berdasarkan hasil metrik RMSE, MAPE, dan MAE.

Dalam melakukan penelitian ini, penulis melakukan studi pustaka penelitian terkait untuk menjadi referensi mengenai penelitian ini, sebagai berikut.

1. Penelitian yang dilakukan oleh Diana Tri S, Eka Patriya, dan Rodiah[14] yang melakukan permodelan dengan judul model *random forest regression* peramalan penyebaran COVID-19 di Indonesia dengan membentuk *training* dan *validation set* dengan perhitungan nilai terendah nilai *error* MAE yang menampilkan hasil peramalan dan data aktual tidak berbeda jauh.
2. Dalam penelitian yang dilakukan oleh Adrian Febriansyah I, Asep Id

H, dan Fajri Rakhmat U[15]. Melakukan penelitian dengan judul prediksi curah hujan menggunakan metode *catboost* yang dibagi menjadi data latih dan uji yang dilakukan *pre-processing* yang menghasilkan prediksi curah hujan dengan hasil akurasi cukup tinggi sebesar 94.22% yang menjadikan *catboost* menjadi metode yang cocok digunakan untuk data bersifat kategorikal.

3. Penelitian yang dilakukan Muhammad Dzul AA, Herman, dan Wistiani Astuti[16]. Dengan judul penelitian peramalan kebutuhan obat menggunakan *xgboost* menghasilkan nilai latih pada obat vital dengan skor rata-rata RMSE 17,34 dan MAE 13,03 dan dapat lebih rendah untuk obat *essential* masih dibutuhkan perbaikan dengan model yang lebih optimal.
4. Dengan penelitian yang dilakukan oleh Teri Kincowati, Muhammad Tanzil F, Bayu Rahayudi[1]. Dengan judul prediksi jumlah kunjungan wisatawan mancanegara ke Indonesia menggunakan metode *average-base fuzzy time series models* memberikan hasil dimana jumlah data latih sangat berpengaruh pada hasil MAPE, berdasarkan hasil 6 kali pengujian didapatkan MAPE terbaik sebesar 9,184% dengan jumlah data pengujian sebanyak 24 data dengan prediksi selama 12 bulan ke depan.
5. Pada penelitian yang dilakukan Adelia Ramadhani, Sri Wahyuningsih, dan Mailiyani Singoringo[7]. Disebutkan pada penelitian tersebut bahwa model

intervensi ARIMA (0,1,1) dengan fungsi step adalah model terbaik dengan periode data Januari 2014 – Oktober 2021, dengan hasil MAPE sebesar 9,91% dapat diketahui kunjungan wisatawan mancanegara kan meningkat dari November 2021 hingga November 2022.

6. Penelitian yang dilakukan oleh Muktar Redy Susila[17] . Penelitian ini menggunakan model VARX dengan model terbaik VARX (1,2,3,7) yang didasarkan pada hasil nilai RMSE terkecil dengan pintu udara 72.217, laut 0,175, dan darat 0,092 yang tujuannya adalah melihat jumlah kunjungan wisatawan mancanegara yang dipengaruhi oleh wabah COVID-19.

## METODE

Pada penelitian ini terdapat beberapa tahapan-tahapan yang dilakukan untuk membuat peramalan yaitu:

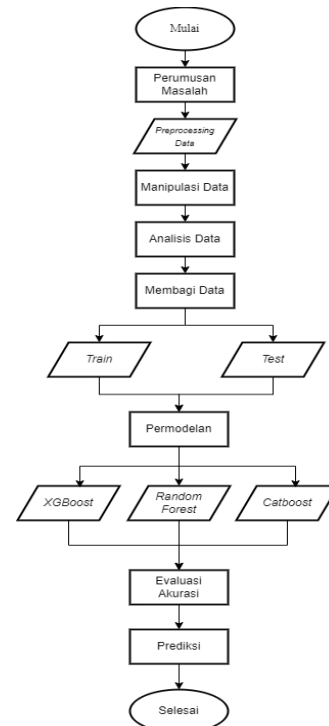
### *Sumber Data*

Pada penelitian ini untuk melakukan prediksi jumlah kunjungan wisatawan mancanegara di Indonesia berdasarkan pintu masuk kedatangan udara menggunakan dataset yang diambil melalui situs *website* Badan Pusat Statistik (BPS) <https://www.bps.go.id/>, yaitu jumlah kunjungan wisatawan mancanegara per bulan ke Indonesia menurut pintu masuk (kunjungan) dari Januari 2017 sampai dengan November 2021 dengan jumlah 59 data.

Dataset yang diambil dalam bentuk bulanan dan hanya berfokus pada pintu masuk udara (Bandara Internasional I Gusti Nurah Rai, Seokarno-Hatta, Juanda, Kualanam, Husein Sastranegara, Adi Sucipto, Banda Internasional Lombok, Sam Ratulangi, Minangkabau, Sultan Syarif Kasim II, Ahmad Yani, Supadio,

Hasanuddin, Sultan Badaruddin II, dan Pintu Udara Lainnya).

### *Alur Penelitian*



Gambar 1 Alur Penelitian

1. **Perumusan Masalah**  
Mengidentifikasi permasalahan pada penelitian berdasarkan latar belakang permasalahan serta melakukan studi literatur untuk mengumpulkan penelitian terkait yang membahas topik penelitian dengan permasalahan yang sama. Melakukan pemahaman terhadap konteks penyebab masalah sehingga mencakup aspek yang memberikan gambaran terhadap situasi dari permasalahan dan mendapatkan landasan teori untuk mengembangkan penelitian yang efektif.
2. **Preprocessing Data**  
Melakukan pembersihan pada data dalam memastikan *missing value* dan data yang digunakan. Dilanjutkan dengan melakukan eksplor terhadap data dengan tujuan untuk memahami pola data historis.

### 3. Manipulasi Data

Tahapan manipulasi data mencakup proses transformasi tipe data yang tidak sesuai agar meningkatkan efisiensi dan memudahkan proses analisis data.

### 4. Menganalisis Data

Melakukan observasi pada tren data deret waktu serta mengidentifikasi data untuk mengetahui informasi dalam data untuk memahami model atau metode selanjutnya. Membuat plot data yang menampilkan pola dan deret waktu pada data.

### 5. Membagi Data

Pada tahapan ini dataset penelitian akan dibagi menjadi dua bagian untuk melakukan prediksi yaitu data pelatihan (*train*) dan data pengujian (*test*). Data pelatihan akan digunakan sebagai permodelan, sedangkan data pengujian akan digunakan untuk evaluasi model prediksi.

### 6. Permodelan

Tahapan ini akan melakukan implementasi permodelan metode *Xgboost*, *Random Forest*, dan *Catboost* pada data. Hasil dari permodelan digunakan menjadi pengujian evaluasi dalam menilai akurasi.

### 7. Pengujian Evaluasi Akurasi

Tahap ini akan dilakukan perbandingan terhadap hasil uji evaluasi akurasi setiap model dengan berfokus pada parameter evaluasi yang mencakup *Root Mean Square Error* (RMSE), *Mean Absolute Error* (MAE), dan *Mean Absolute Percentage Error* (MAPE).

RMSE merupakan metode yang digunakan untuk mengevaluasi teknik peramalan dalam mengukur akurasi perkiraan dari suatu model

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

[18]

MAE salah satu metode mengukur keakuratan model peramalan, MAE menunjukkan rata-rata kesalahan absolut antar hasil dengan rill.

$$MAE = \frac{1}{n} \sum_{i=1}^n |f_i - y_i|$$

[19]

MAPE adalah nilai rata-rata perbedaan yang ada diantara nilai prediksi dan realisasi pada evaluasi hasil peramalan, nilai MAPE dapat di sebutkan sebagai hasil persenan [20].

$$MAPE = \left( \frac{1}{n} \sum_{t=1}^n \left| \frac{Dt - y_t^m}{Dt} \right| \times 100\% \right)$$

[8]

Pada setiap parameter memiliki perspektif yang berbeda pada keakuratan model.

### 8. Hasil Prediksi

Merupakan hasil akhir prediksi yang menampilkan RMSE, MAE, dan MAPE dari setiap model yang di ujikan pada penelitian.

## HASIL

Penelitian ini melakukan prediksi jumlah kunjungan wisatawan mancanegara di Indonesia berdasarkan pintu masuk udara dengan menerapkan *data mining* yang membandingkan tiga metode peramalan yaitu *Xgboost*, *Random Forest*, dan *Catboost*. Tujuan dari penelitian ini adalah untuk meningkatkan nilai akurasi prediksi dari penelitian sejenis dengan metode yang berbeda.

**Analisis Data**

Dataset yang digunakan pada penelitian ini adalah data jumlah kunjungan wisatawan mancanegara per bulan berdasarkan pintu masuk kedatangan yang berfokus pada pintu masuk udara pada Januari 2017 sampai dengan November 2021. Analisis akan dilakukan untuk mendapatkan pola yang ada pada data untuk mempermudah dalam melakukan permodelan, bentuk data yang disajikan pada Tabel 1.

*Table 1 Dataset Pintu Masuk Udara*

No	Bulan	Pintu Udara
1	2017-01-01	746666
2	2017-02-01	720428
3	2017-03-01	749150
4	2017-04-01	812927
5	2017-05-01	809941
6	2017-06-01	763470
.....	....	.....
.....	.....	.....
53	2021-06-01	16234
54	2021-07-01	7175
56	2021-08-01	1609
57	2021-09-01	4748
58	2021-10-01	15722
59	2021-11-01	19909

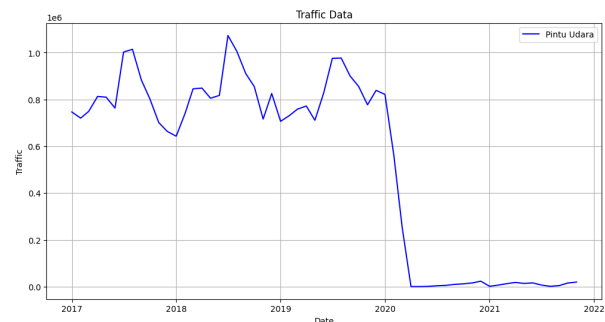
Jumlah kunjungan wisatawan mancanegara di Indonesia melalui pintu udara sebanyak 31.427.170 dengan jumlah kunjungan tertinggi pada bulan Juli tahun 2018 sebanyak 1.073.385, sedangkan jumlah kunjungan terendah adalah sebanyak 506 pada bulan Mei tahun 2020 yang disajikan pada Tabel 2.

*Table 2 Nilai Mean, Min, Max Pintu Udara*

Lokasi	Mean	Max	Min
Pintu Udara	532664	1073385	506

Prediksi dengan menggunakan tiga

model yang berbeda yaitu *XGBoost*, *Random Forest*, dan *Catboost* akan dilakukan dengan dataset yang berbentuk *time series* disajikan pada Tabel 1. dengan tujuan membandingkan antar model untuk menentukan hasil evaluasi akurasi terbaik. Sebelumnya, untuk mengetahui pola data kunjungan wisatawan mancanegara di Indonesia melalui pintu udara maka disajikan plot atau visualisasi data pada Gambar 2.



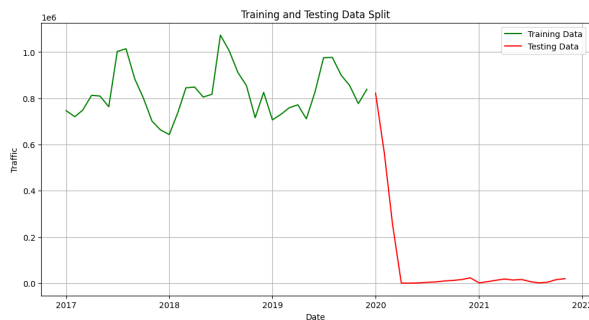
*Gambar 2 Plot Dataset Pintu Udara*

Berdasarkan pada Gambar 2. terlihat jumlah kunjungan wisatawan mancanegara ke Indonesia melalui pintu udara pada Januari 2017 sampai November 2021 berfluktuasi setiap bulannya dengan jumlah kunjungan yang naik turun atau tidak menentu. Gambar 2. juga menampilkan bahwa kunjungan wisatawan mancanegara di Indonesia mulai mengalami penurunan yang signifikan pada awal tahun 2020 yang saat itu merupakan awal masuk COVID-19 ke Indonesia.

**Pemisahan Data**

Pemisahan data akan dilakukan untuk membuat permodelan dan analisis. Pemisahan data diambil dari dataset yang dibagi menjadi dua yaitu data pelatihan (*train*) dan data pengujian (*test*). Data pelatihan akan menggunakan data historis pada Januari 2017 sampai Desember 2019 yang terdiri dari 36 data yang dapat dilihat pada Gambar 3. Data pelatihan akan digunakan untuk membuat permodelan prediksi dengan menyesuaikan parameter pola dan tren. Sedangkan, data pengujian akan menggunakan data pada Januari

2020 sampai November 2021 yang terdiri sebanyak 23 data.



Gambar 3 Pembagian Data Train dan Test

Data pengujian akan digunakan untuk melakukan evaluasi akurasi model dengan menilai model dapat melakukan prediksi dengan baik atau tidak. Maka dengan pembagian ini terdapat total 59 jumlah data asli, dimana 36 data untuk melakukan analisis dan sisanya sebanyak 23 data digunakan untuk evaluasi atau penilaian.

### Permodelan dan Pengujian

Setelah dilakukan analisis data dan pembagian data *train* dan *test* maka selanjutnya akan dilakukan implementasi atau pembuatan permodelan untuk dilakukan pelatihan dengan penerapan tiga model berbeda premalan yaitu *XGboost*, *Random Forest*, dan *Catboost*.

#### 1. XGboost

*eXtreme Gradient Boosting* atau kepanjangan dari *XGBoost* adalah sebuah *Machine learning Algorithm* yang berbasis *gradient tree boosting*. Algoritma dapat digunakan pada *forecasting time series* karena dapat menangani permasalahan *machine learning* dalam ukuran yang cukup besar[16].

#### 2. Random Forest

Dalam melakukan peramalan terdapat beberapa metode yang dapat digunakan, salah satunya adalah *Random Forest*. adalah model regresi dengan teknik *supervised learning* dengan memakai

fitur yang terdapat pada data historis. Fitur dan target akan digabungkan dengan serangkaian kondisi dan diatur dalam sebuah struktur sehingga membentuk seperti pohon dari atas kebawah[14].

#### 3. Catboost

*Catboost* atau *Categorical Boosting* adalah sebuah metode yang termasuk dalam *Gradient Boosting Decision Tree* dengan memanfaatkan *decision tree* sebagai predictor. Metode ini biasanya digunakan untuk data kategorikal serta juga mampu menangani berbagai jenis data, salah satunya adalah *time series* dalam meningkatkan akurasi pada sebuah prediksi[15].

Pada tahap pengujian, setiap model akan di evaluasi menggunakan beberapa metrik akurasi untuk dapat mengetahui hasil kinerja pada setiap model yang fokus ke beberapa metrik yaitu *Root Mean Square Error (RMSE)*, *Mean Absolute Error (MAE)*, dan *Mean Absolute Percentage Error (MAPE)*. Metrik-metrik tersebut digunakan untuk memberikan informasi terkait hasil setiap model dan seberapa baik model dalam melakukan prediksi berdasarkan data.

Table 3 Hasil Evaluasi Akurasi Prediksi

Model	RMSE	MAE	MAPE
<i>Xgboost</i>	<b>671935.2</b>	648139.1	20985.35
<i>Random Forest</i>	676084.2	<b>638950.3</b>	<b>20599.04</b>
<i>Catboost</i>	698083.0	672780.1	21778.72
[3]ARIMA	779670.7	749030.4	23196.45
[3]Glmnet	802060.2	770418.0	24450.08
[3]Prophet	802417.7	771330.4	24853.98

Berdasarkan hasil pada Tabel 3. model *XGBoost* menampilkan hasil lebih baik



dibandingkan dua model lainnya. Model *XGBoost* unggul dengan nilai RMSE yaitu sebesar 671935.2, MAPE 20985.35, dan MAE 648139.1, nilai-nilai tersebut dapat mengindikasikan bahwa model *XGBoost* memiliki kinerja yang baik karena memiliki deviasi standar kesalahan prediksi lebih kecil.

Model *Random Forest* dan *Catboost* memiliki hasil yang kurang baik jika dilihat dari nilai RMSE-nya. Namun, meskipun begitu pada model *Random Forest* memiliki nilai MAPE dan MAE yang lebih rendah dibandingkan dengan *XGBoost*. Hal ini menunjukkan model *Random Forest* memiliki keakuratan dalam mereduksi kesalahan absolut dan presentase rata-rata. *Random Forest* dengan nilai MAPE 20599.04 dan MAE 638950.3, sedangkan *XGBoost* dengan nilai MAPE 20985.35 dan MAE 648139.1.

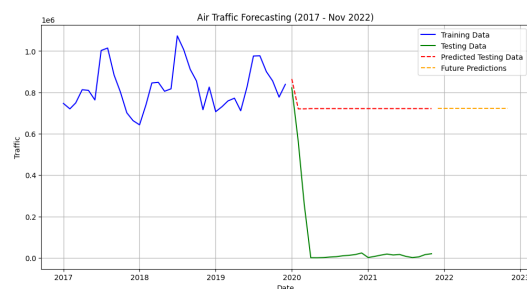
Meskipun model *Random Forest* memiliki keunggulan karena memiliki nilai MAPE dan MAE lebih baik, namun dengan nilai RMSE yang dimiliki model *XGBoost* lebih rendah menunjukkan bahwa model ini lebih efektif, terutama jika dilihat dataset memiliki outliers yang tinggi. Maka model *Random Forest* dan *Catboost* kurang cocok di implementasikan dengan jenis dataset ini.

Bukan hanya pada model *Random Forest* dan *Catboost* saja, berdasarkan pada Tabel 3. terlihat model *XGBoost* juga unggul saat dilakukan perbandingan nilai RMSE, MAPE, dan MAE dengan model ARIMA yang ada pada penelitian sejenis[3] karena memiliki nilai evaluasi akurasi lebih rendah. Model ARIMA pada penelitian tersebut menampilkan nilai RMSE sebesar 779670.7, MAPE 23196.45, dan MAE 749030.4. Hal ini dapat diartikan bahwa penelitian dengan menggunakan model *XGBoost* mendapatkan hasil evaluasi akurasi lebih baik dibandingkan pada penelitian sebelumnya atau sejenis[3] dan memiliki kinerja prediksi lebih baik serta

deviasi kesalahan prediksi lebih kecil.

### Prediksi

Berdasarkan hasil evaluasi akurasi model, *XGBoost* memiliki kemampuan yang lebih baik dalam penanganan data yang bersifat kompleks dan outliers. Model *XGBoost* menunjukkan bahwa nilai RMSE sebesar 671935.2 sehingga dapat diindikasikan model ini lebih baik dan dapat diandalkan dalam menangani konsistensi prediksi, meskipun *Random Forest* memiliki keunggulan pada nilai MAPE dan MAE.



Gambar 4 Plot Prediksi 12 Bulan Kedepan

Pada penelitian ini penulis juga menyertakan plot yang menampilkan prediksi kunjungan wisatawan mancanegara di Indonesia pada 12 bulan kedepan yaitu Desember 2021 sampai dengan November 2022. Dilihat pada plot tersebut dimana hasil prediksi (garis oranye) 12 bulan ke depan selaras atau mengikuti data *train* (garis merah) dan tidak sejalan dengan data asli. Dengan keterbatasan sampel data yang digunakan pada penelitian ini, penelitian masih tetap dapat menampilkan hasil evaluasi akurasi dengan melakukan perbandingan berbagai model serta melakukan prediksi kedatangan wisatawan mancanegara selama 12 bulan kedepan.

## KESIMPULAN

Penelitian ini mengembangkan model prediksi yang lebih akurat berdasarkan hasil evaluasi akurasi RMSE, MAE, dan MAPE dari penelitian sebelumnya yang menggunakan model ARIMA. Penelitian ini menggunakan model *Xgboost*, *Random Forest*, dan *Catboost*



untuk melakukan prediksi kunjungan wisatawan mancanegara di Indonesia yang berfokus pada pintu masuk udara.

Hasil pada penelitian ini menunjukkan bahwa model *XGBoost* lebih baik dengan nilai evaluasi akurasi RMSE 671935.2, MAPE 20985.35, dan MAE 648139.1. Model *XGBoost* unggul jika dibandingkan dengan model perbandingan lain yaitu *Random Forest* dan *Catboost* dalam akurasi prediksi. Meskipun model *Random Forest* memiliki keunggulan pada nilai MAE 638950.3 dan MAPE 20599.04 dibandingkan *XGBoost*, dengan keunggulan nilai RMSE yang didapatkan oleh model *XGBoost* maka model ini lebih efektif dan tangka kesalahan prediksi lebih kecil dibandingkan model lain.

Pada penelitian ini juga membandingkan hasil dari penelitian sejenis sebelumnya yang menggunakan metode ARIMA, *Glmnet*, dan *Prophet* pada penelitiannya. Pada penelitian tersebut model ARIMA unggul karena mendapatkan nilai evaluasi akurasi RMSE sebesar 779670.7, MAPE 23196.45, dan MAE 749030.4.

Maka dapat disimpulkan berdasarkan hasil evaluasi diatas bahwa model *XGBoost* lebih baik dalam melakukan prediksi kunjungan wisatawan mancanegara di Indonesia berdasarkan pintu masuk udara dibandingkan *Random Forest*, *Catboost*, dan ARIMA yang digunakan pada penelitian sejenis sebelumnya. Penggunaan model *XGBoost* direkomendasikan dalam melakukan prediksi kunjungan wisatawan mancanegara dengan tujuan diharapkan dapat mendukung dalam usaha pengambilan keputusan pada industri pariwisata di Indonesia.

## DAFTAR PUSTAKA

- [1] T. Kincowati, M. T. Furqon, Dan B. Rahayudi, "Prediksi Jumlah Kunjungan Wisatawan Mancanegara Ke Indonesia Menggunakan Metode Average-Based Fuzzy Time Series Models," 2019. [Daring]. Tersedia Pada: [Http://J-Ptiik.Ub.Ac.Id](http://J-Ptiik.Ub.Ac.Id)
- [2] H. Mukhtar, R. Muhammad, T. Reny Medikawati, Dan Yoze Rizki, "Peramalan Kedatangan Wisatawan Mancanegara Ke Indonesia Menurut Kebangsaan Perbulannya Menggunakan Metode Multilayer Perceptron," *Jurnal Coscitech (Computer Science And Information Technology)*, Vol. 2, No. 2, Hlm. 113–119, Des 2021, Doi: 10.37859/Coscitech.V2i2.3324.
- [3] F. Mu Dan A. Gunaryati, "Prediksi Kunjungan Wisatawan Mancanegara Melalui Pintu Udara Menggunakan Arima, *Glmnet*, Dan Prophet Prediction Of Foreign Tourist Visits Via Airline Using Arima, *Glmnet*, And Prophet."
- [4] S. A. Salimu Dan Y. Yunus, "Prediksi Tingkat Kedatangan Wisatawan Asing Menggunakan Metode Backpropagation (Studi Kasus: Kepulauan Mentawai)," *Jurnal Informatika Ekonomi Bisnis*, Hlm. 98–103, Des 2020, Doi: 10.37034/Infec.V2i4.50.
- [5] F. Islamiyah, R. Nurafni, E. Kartika, D. Sri, S. Politeknik, Dan N. Malang, "Peramalan Kedatangan Wisatawan Mancanegara Indonesia: Metode Holt's Winter Exponential Smoothing."
- [6] A. Meimela, "Prediksi Jumlah Kunjungan Wisatawan Mancanegara Ke Indonesia," Vol. 19, No. 1, 2021, Doi: 10.36275/Mws.
- [7] A. Ramadhani, S. Wahyuningsih, Dan M. Siringoringo, "Forecasting The Number Of Foreign Tourist Visits To Indonesia Used Intervention Analysis With Step Function," *Jurnal Matematika, Statistika Dan Komputasi*, Vol. 19, No. 1, Hlm. 146–162, Sep 2022, Doi: 10.20956/J.V19i1.21607.
- [8] A. Fitriani, A. Lutfiajati Pradhyani, A.

- Maulana, Dan J. Haerul Jaman, "Prediksi Jumlah Kedatangan Wisatawan Mancanegara Se-Asia Ke Indonesia Berdasarkan Kewarganegaraan Menggunakan Fuzzy Time Series," 2023.
- [9] F. D. Pratama, I. Zufria, Dan T. Triase, "Implementasi Data Mining Menggunakan Algoritma Naïve Bayes Untuk Klasifikasi Penerima Program Indonesia Pintar," *Rabit: Jurnal Teknologi Dan Sistem Informasi Univrab*, Vol. 7, No. 1, Hlm. 77–84, Jan 2022, Doi: 10.36341/Rabit.V7i1.2217.
- [10] Nofriansyah Dicky Dan Nurcahyo Gunadi Widi, *Algoritma Data Mining Dan Pengujian*. Deepublish, 2019.
- [11] D. Nababan, V. Fajar Alexandro Nipu, R. Rizald, B. Baso, Dan D. Arisandi, "Optimalisasi Kinerja Mikrokomputer Raspberry Pi Pada Smart Greenhouse Berbasis Internet Of Things (Iot) Dengan Algoritma Forecasting Moving Average," *Rabit: Jurnal Teknologi Dan Sistem Informasi Univrab*, Vol. 8, No. 2, Hlm. 164–172, Jul 2023, Doi: 10.36341/Rabit.V8i2.3452.
- [12] "Peramalan Jumlah Kunjungan Wisatawan Mancanegara Ke Indonesia Menggunakan Autoregressive Integrated Moving Average (Arima)".
- [13] Firdaus M, *Aplikasi Ekonometrika Untuk Data Panel Dan Time Series*, Vol. Cetakan 7. Bogor: Ipb Press, 2018.
- [14] Diana Tri Susetianingtias, Eka Patriya, Dan Rodiah, "Model Random Forest Regression Untuk Peramalan Penyebaran Covid-19 Di Indonesia," *Decode: Jurnal Pendidikan Teknologi Informasi*, Vol. 2, No. 2, Hlm. 84–95, Sep 2022, Doi: 10.51454/Decode.V2i2.48.
- [15] "Prediksi Curah Hujan Menggunakan Metode Categorical Boosting (Catboost)."
- [16] J. Khatib Sulaiman, M. Dzul Asmi Alhamdi, Dan W. Astuti, "Peramalan Kebutuhan Obat Menggunakan Xgboost Studi Kasus Pada Rumah Sakit Xyz," *Indonesian Journal Of Computer Science Attribution*, Vol. 12, No. 5, Hlm. 2023–2757.
- [17] M. R. Susila, "Pemodelan Multivariate Kunjungan Wisatawan Mancanegara Ke Indonesia Melalui Pintu Udara, Laut, Dan Darat Yang Melibatkan Dampak Wabah Covid-19," *Barekeng: Jurnal Ilmu Matematika Dan Terapan*, Vol. 15, No. 3, Hlm. 467–478, Sep 2021, Doi: 10.30598/Barekengvoll15iss3pp467-478.
- [18] F. Indra Sanjaya Dan D. Heksaputra, "Prediksi Rerata Harga Beras Tingkat Grosir Indonesia Dengan Long Short Term Memory," Vol. 7, No. 2, Hlm. 163–174, 2020, [Daring]. Tersedia Pada: <Http://Jurnal.Mdp.Ac.Id>
- [19] A. A. Suryanto, A. Muqtadir, Dan S. Artikel, "Penerapan Metode Mean Absolute Error (Mea) Dalam Algoritma Regresi Linear Untuk Prediksi Produksi Padi Info Artikel: Abstrak," No. 1, Hlm. 11, 2019.
- [20] I. Nabillah Dan I. Ranggadara, "Mean Absolute Percentage Error Untuk Evaluasi Hasil Prediksi Komoditas Laut," *Joins (Journal Of Information System)*, Vol. 5, No. 2, Hlm. 250–255, Nov 2020, Doi: 10.33633/Joins.V5i2.3900.