Algoritma *Data Mining* untuk Memprediksi Kunjungan Turis Mancanegara sebagai Upaya Membangkitkan Sektor Pariwisata di Indonesia

GEMASTIK 2022



Tim BismillahEmas

Disusun oleh:

Harry Krisna Darmawan 19/446781/TK/49886

Rizky Alif Ramadhan 19/446785/TK/49890

Roby Attoillah 19/444068/TK/49264

UNIVERSITAS GADJAH MADA

2022

SURAT PERNYATAAN ORISINALITAS KARYA GEMASTIK 2022

Yang bertandatangan dibawah ini:

Nama

: 1. Harry Krisna Darmawan

19/446781/TK/49886 (Ketua)

2. Rizky Alif Ramadhan

19/446785/TK/49890 (Anggota 1)

3. Roby Attoillah

19/444068/TK/49264 (Anggota 2)

Nama Tim

: BismillahEmas

Judul Karya

: Algoritma Data Mining untuk Memprediksi Kunjungan Turis

Mancanegara sebagai Upaya Membangkitkan Sektor Pariwisata di

Indonesia

Perguruan Tinggi

: Universitas Gadjah Mada

Dengan ini menyatakan sesungguhnya bahwa karya ini belum pernah dipublikasikan sebelumnya di media manapun dan belum pernah diikutsertakan dalam kompetisi lain dan GEMASTIK sebelumnya, dan/atau tidak pernah digunakan untuk media komunikasi apapun serta tidak mengandung unsur plagiat di dalamnya.

Demikian pernyataan ini saya buat dalam keadaan sadar dan tanpa ada unsur paksaan dari siapapun. Jika di kemudian hari ditemukan ketidakbenaran informasi, maka saya bersedia menerima sanksi sesuai dengan peraturan yang berlaku.

Yogyakarta, 21 Oktober 2022

Mengetahui,

Direktur Kemahasiswaan,

Dr. Sindung Tjahyadi, M.Hum.

NIP 196410151991031003

Ketua TIM

(Harry Krisna Darmawan)

NIM. 19/446781/TK/49886

ORIGIN	IALITY REPORT			
SIMIL	5% ARITY INDEX	14% INTERNET SOURCES	4% PUBLICATIONS	3% STUDENT PAPERS
PRIMA	RY SOURCES			
1	Submitt Student Pape		tas Pelita Harap	1 %
2	ramaht Internet Sour	raveler.com		1 %
3	reposito	ori.usu.ac.id		1 %
4	id.scribe			1 %
5	Sulistyo Compo Data Ku	o. "Penerapan I nent Analysis (unjungan Wisat	Hasan, Gunawar Metode Principlo PCA) untuk Clus tawan Mancane Informatika, 202	e stering gara ke
6	id.123d Internet Sour			1 %
7	123dok Internet Soui			1 %
8	kumpar Internet Soul	ran.com		

a_sebagai_Upaya_Membangkitkan_Sektor_Pariwisata_di_Ind...

		1	%
9	www.liputan6.com Internet Source	1	%
10	repository.unair.ac.id Internet Source	1	%
11	mistermampet.blogspot.com Internet Source	1	%
12	mediaperencana.perencanapembangunan.or.id	1	%
13	www.bps.go.id Internet Source	1	%
14	www.ojk.go.id Internet Source	1	%
15	Stanov Purnawibowo. "Fragmen Keramik Asing: Jejak Hubungan Dagang di Situs Kotacina", Berkala Arkeologi Sangkhakala, 2018 Publication	<1	%
16	laela6.wordpress.com Internet Source	<1	%
17	repositorio.utp.edu.co Internet Source	<1	%
18	adindasutrisno.wordpress.com Internet Source	<1	%

19	j-ptiik.ub.ac.id Internet Source	<1%
20	nanopdf.com Internet Source	<1%
21	repository.pnb.ac.id Internet Source	<1%
22	www.batamnews.co.id Internet Source	<1%

Exclude quotes Off
Exclude bibliography On

Exclude matches

Off

ABSTRAK

Pariwisata merupakan sumber devisa terbesar bagi Indonesia setelah migas. Pandemi COVID-19 yang berlangsung hampir 2 tahun sangat berdampak bagi pariwisata di Indonesia. Kebijakan pemerintah saat pandemi COVID-19, yaitu Pembatasan Sosial Skala Besar (PSBB) dan pembatasan akses keluar masuk negara menyebabkan jumlah kunjungan wisatawan mancanegara (Wisman) ke Indonesia mengalami penurunan yang signifikan. Hal ini tentunya juga memengaruhi perekonomian di Indonesia khususnya pada sektor pariwisata. Mulai April 2022, pemerintah kembali membuka akses keluar masuk negara, hal ini tentunya merupakan titik balik bagi kebangkitan sektor pariwisata di Indonesia. Kebangkitan tersebut tentunya membutuhkan strategi yang baik agar hasil yang diinginkan tercapai, sehingga perekonomian khususnya pada sektor pariwisata dapat membaik dan berangsur-angsur kembali normal. Pada penelitian ini, penulis melakukan eksplorasi terhadap kunjungan wisatawan mancanegara (Wisman) di Indonesia serta memprediksi kunjungan wisatawan mancanegara (Wisman) menggunakan algoritma data mining yang berhubungan dengan peramalan deret waktu seperti ARIMA (Autoregressive Integrated Moving Average). Peneliti akan mengembangkan ARIMA dengan algoritma supervised learning, yaitu XGBoost Regressor. Algoritma tersebut dievaluasi menggunakan standar ukuran pada kasus regresi seperti R2 dan RMSE. Diharapkan dengan penelitian ini, pemerintah mampu membuat kebijakan yang sesuai di masa yang akan datang.

Kata kunci: Pariwisata, Pandemi, Data Mining, Ekonomi, Prediksi, Machine Learning

1. LATAR BELAKANG

Menurut majalah Forbes, Indonesia merupakan negara paling indah di dunia (Tempo, 2022). Hal ini tentunya menjadi daya tarik tersendiri bagi wisatawan mancanegara (Wisman) serta menjadi potensi bagi Indonesia di sektor pariwisata. Menteri Pariwisata dan Ekonomi Kreatif (Parekraf), Sandiaga Salahudin Uno menyampaikan, seperti yang dikutip dari CNBC Indonesia, bahwa pariwisata masih menjadi penyumbang devisa terbesar di Indonesia setelah migas (CNBC, 2021). Tercatat pada tahun 2019, sektor pariwisata mampu menyumbang US\$16,9 miliar atau 4,7% dari PDB Indonesia di tahun tersebut (Data Alinea, 2022).

Sejak pertama kali kasus COVID-19 terdeteksi di Indonesia, pemerintah mengambil berbagai langkah kebijakan untuk menekan laju penyebaran. Di antara kebijakan yang dikeluarkan oleh pemerintah adalah dengan menetapkan Pembatasan Sosial Skala Besar (PSBB). Kebijakan ini tentunya sangat berdampak pada berbagai bidang ekonomi, di antaranya adalah sektor pariwisata. Tercatat pada tahun 2020 jumlah kunjungan Wisman mengalami penurunan drastis sebanyak 75,03 persen jika dibandingkan dengan jumlah kunjungan wisata mancanegara pada tahun 2019 (BPS, 2021). Penurunan ini otomatis membuat pemasukan dan devisa negara menjadi menurun di sektor pariwisata. Pada tahun 2020, jumlah devisa dari sektor pariwisata hanya sebesar \$3,3 miliar dolar AS. jumlah yang jauh lebih rendah dibandingkan tahun sebelumnya yang mencapai \$16,9 miliar dolar AS (Data Alinea, 2022).

Semakin lama dengan kondisi seperti ini, perekonomian negara semakin memburuk. Maka dari itu, pemulihan pada sektor pariwisata pun diperlukan. Dalam rangka itu, pemerintah Indonesia telah membuat kebijakan bahwa mulai 6 April 2022, turis dari 43 negara sudah bisa masuk Indonesia dengan Visa on Arrival (VoA) atau bebas visa kunjungan bagi pemegang paspor negara ASEAN (CNBC, 2022). Dengan kebijakan ini, tercatat bahwa kunjungan turis mancanegara di tahun 2022 terus mengalami perbaikan meskipun jika dibandingkan dengan periode sebelum pandemi masih jauh di bawahnya (Liputan6, 2022).

Dengan adanya momentum pemulihan pariwisata ini, diperlukan strategi untuk bisa memanfaatkannya dengan baik sehingga perekonomian dapat membaik dengan maksimal. Adapun salah satu cara yang bisa dilakukan adalah dengan melakukan analisis terhadap data kunjungan turis sehingga sektor pariwisata dapat melakukan persiapan secara maksimal dan mendapatkan hasil yang maksimal juga.

Penelitian ini dilakukan dengan mengembangkan algoritma data mining untuk dapat melakukan prediksi kunjungan Wisman berdasarkan data yang terdapat di Badan Pusat Statistik dan Kementerian Pariwisata dan Ekonomi Kreatif. Prediksi yang telah dilakukan dapat digunakan sebagai dasar pengambilan keputusan yang dapat diterjemahkan dalam bentuk kebijakan-kebijakan guna mempercepat pemulihan perekonomian nasional.

2. TUJUAN

Penelitian ini bertujuan untuk:

- a. Melakukan eksplorasi terhadap data kunjungan turis mancanegara ke Indonesia
- b. Menemukan algoritma *data mining* yang tepat untuk memprediksi kunjungan turis mancanegara di Indonesia

3. MANFAAT

Manfaat diadakannya penelitian ini adalah untuk :

- a. Mengetahui ringkasan data kunjungan turis mancanegara yang berkunjung ke Indonesia
- b. Dapat digunakan untuk melakukan prediksi kunjungan turis mancanegara di Indonesia sebagai dasar pengambilan keputusan dalam sektor pariwisata

4. BATASAN YANG DIGUNAKAN

Dalam penelitian ini, batasan yang akan digunakan adalah sebagai berikut:

- a. Data yang digunakan adalah data jumlah kunjungan Wisatawan Mancanegara (Wisman) dari Januari 2018 sampai Juli 2022 pada pintu masuk utama Indonesia
- b. Prediksi yang dilakukan adalah prediksi jumlah kunjungan total dari pintu masuk utama Indonesia untuk 1 tahun ke depan, serta prediksi jumlah kunjungan pada setiap pintu masuk pada bulan yang memiliki jumlah kunjungan total terbanyak menurut prediksi

5. STUDI LITERATUR

A. Wisatawan Mancanegara

Wisatawan adalah setiap orang yang melakukan perjalanan ke wilayah geografis suatu negara secara sukarela, yaitu perjalanan lebih dari 24 jam dan kurang dari 12 bulan di wilayah yang menjadi maksud wisata. Wisatawan di Indonesia setidaknya dapat dibagi menjadi dua kategori, yaitu: 1) wisatawan nusantara dan 2) wisatawan mancanegara. Wisatawan mancanegara menurut *United Nations World Tourism Organization* adalah setiap orang yang melakukan perjalanan ke suatu wilayah geografis di luar negara asalnya secara sukarela selain untuk tujuan sekolah atau bekerja.

B. Time-Series Forecasting.

Time-Series Forecasting adalah pembuatan prediksi ilmiah dilakukan berdasarkan data historis dengan titik waktu tertentu. Dalam pembuatan prediksi tersebut, dilakukan pembangunan model terlebih dahulu melalui analisis historis kemudian model tersebut digunakan untuk melakukan pengamatan sehingga dapat mendukung pengambilan keputusan yang tepat untuk waktu mendatang. *Time-Series Forecasting* atau Peramalan Deret Waktu adalah proses menganalisis data deret waktu menggunakan statistik dan pemodelan untuk membuat prediksi dan menginformasikan pengambilan keputusan yang strategis. Semakin lengkap data yang dimiliki, semakin akurat perkiraannya.

C. ARIMA

ARIMA (*Autoregressive Integrated Moving Average*) adalah salah satu metode yang digunakan untuk melakukan prediksi pada data *time series*. Tujuan utama model ini adalah untuk mendeskripsikan *autocorrelation* yang terdapat pada data sehingga model ini cocok digunakan untuk mengolah *stationary time series data*. ARIMA merupakan gabungan model *autoregression* dan *moving average* yang dapat dirumuskan sebagai berikut:

$$y'_{t} = c + \phi_{1}y'_{t-1} + \dots + \phi_{p}y'_{t-p} + \theta_{1}\varepsilon_{t-1} + \dots + \theta_{q}\varepsilon_{t-q} + \varepsilon_{t}$$
 (1)

Dengan y' merupakan differenced series dan ruas kanannya merupakan prediktor yang memiliki nilai lag dan galat lag.

D. XGBoost Regressor

XGBoost regression atau Extreme Gradient Boosting untuk regresi merupakan implementasi salah satu algoritma machine learning untuk klasifikasi dan pemodelan regresi. Model XGBoost melakukan klasifikasi maupun regresi menggunakan metode ensemble pada beberapa decision tree secara bertahap yang selanjutnya dilakukan fitting terhadap galat hasil prediksi. Model ini memiliki performa komputasi yang jauh lebih cepat dibandingkan metode ensemble tradisional pada data yang besar.

6. METODOLOGI

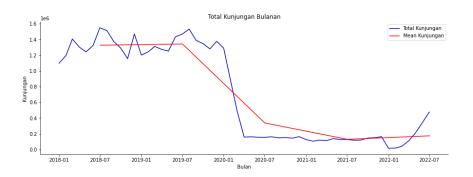
A. Data Preparation

Data yang digunakan adalah data jumlah kunjungan wisatawan mancanegara dari tahun 2018 sampai Juli 2022 pada pintu masuk utama Indonesia yaitu : 'Bali', 'Banten', 'Jatim', 'Sumut', 'Jabar', 'DIY', 'NTB', 'Sulut', 'Sumbar', 'Riau', 'NAD', 'Jateng', 'Kalbar',

'Sulsel', 'Sumsel', 'Kepri', 'Papua', 'NTT', 'Lainnya'. Data yang tersedia di Internet bersumber dari Kemenparekraf dan masih berformat .pdf. Peneliti mengubah dan menjadikan data tersebut menjadi .csv dengan kolom Bulan, Provinsi1, Provinsi2,..., ProvinsiN, serta menambahkan kolom "total_kunjungan" yang merupakan hasil penjumlahan kunjungan pada setiap Provinsi.

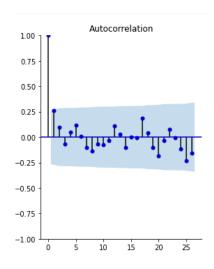
B. Exploratory Data Analysis (EDA)

Berdasarkan gambar 6.1 dapat dilihat bahwa total kunjungan wisatawan di Indonesia cukup stabil pada tahun 2018 sampai awal 2020. Kemudian saat COVID-19 ditetapkan sebagai bencana darurat, jumlah kunjungan wisatawan turun drastis. Setelah Januari 2022 jumlah kunjungan wisatawan kembali menunjukan *trend* peningkatan.



Gambar 6.1 Total Kunjungan Bulanan di Indonesia

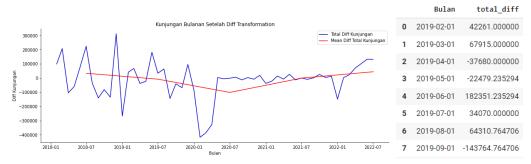
Gambar 6.2 menunjukan bahwa siklus kenaikan/penurunan berlangsung dengan siklus 12 bulan (jarak antara puncak atau jarak antar lembah). Hal ini berguna dalam menentukan jumlah lag pada ekstraksi fitur pada tahap *feature engineering*.



Gambar 6.2 Autokorelasi

C. Feature Engineering

• Melakukan Diff Transformation



Gambar 6.3 Diff Transformation

Diff transformation adalah cara sederhana untuk menghilangkan struktur sistematis dari deret waktu. Tujuannya adalah melihat pola berulang dari data tersebut. Diff transformation mengurangi nilai ke-N dengan nilai sebelumnya. "total_diff" pada Gambar 6.3 menunjukan perbedaan atau selisih kunjungan wisatawan pada bulan tersebut dengan bulan sebelumnya. Data hasil dari diff transformation ini juga dapat disebut sebagai data stasioner.

Melakukan Ekstraksi Fitur

	Bulan	total	total_diff	lag_1	lag_2	lag_3	lag_4
0	2019- 02-01	1.243996e+06	42261.000000	-267988.000000	312240.000000	-134392.000000	-81283.764706
1	2019- 03-01	1.311911e+06	67915.000000	42261.000000	-267988.000000	312240.000000	-134392.000000
2	2019- 04-01	1.274231e+06	-37680.000000	67915.000000	42261.000000	-267988.000000	312240.000000
3	2019- 05-01	1.251752e+06	-22479.235294	-37680.000000	67915.000000	42261.000000	-267988.000000

Gambar 6.4 Data Hasil Ekstraksi Fitur

Data deret waktu pada asalnya tidak dapat dimodelkan menggunakan model *supervised learning* seperti XGBoost. Agar dapat dimodelkan menggunakan XGBoost maka diperlukan ekstraksi fitur dengan skema *time-lag*. *Time-lag* adalah pergeseran waktu yang dilakukan pada data historis. "lag_N" menunjukan nilai "total diff" pada N bulan sebelumnya.

Membuat Fitur "is closed access"

"is_closed_access" menunjukan apakah pada bulan tersebut akses keluar masuk pada pintu Indonesia ditutup atau tidak. Jika bernilai 1 maka akses ditutup dan jika bernilai 0 akses dibuka. Telah diketahui bahwa akses ditutup mulai bulan April 2020 sampai bulan April 2022.

Melakukan Normalisasi

Normalisasi dilakukan menggunakan MinMaxScaler dengan skala -1 sampai 1. Normalisasi dilakukan agar skala antar fitur tidak *imbalance*.

D. Modeling

Terdapat 2 tahap pemodelan:

- A. Memodelkan dan memprediksi "total_diff" pada data stasioner untuk bulan-bulan berikutnya menggunakan <u>ARIMA</u>.
- B. Melakukan ekstraksi fitur (*time-lag*) pada data stasioner tersebut, kemudian memodelkan, serta memprediksi "total_kunjungan" menggunakan <u>XGBoost Regression</u> pada bulan-bulan berikutnya untuk total keseluruhan daerah dan setiap daerah.

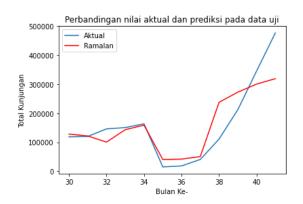
E. Evaluation and Validation

Evaluasi dan validasi dilakukan menggunakan ukuran standar regresi yaitu *Root Mean Squared Error* dan *R2_Score*. Data latih adalah data dari Januari 2019 sampai Juli 2021. Data uji adalah data dari Agustus 2021 sampai Juli 2022.

7. HASIL DAN ANALISIS

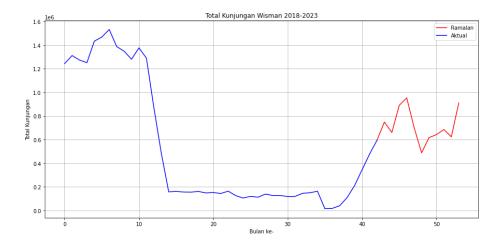
A. Pengujian

Dari hasil pengujian pada data uji didapat RMSE = 64661.827 dan R2_Score = 0.747 untuk total kunjungan. Dapat dilihat pada grafik, model **sudah cukup baik memprediksi jumlah** kunjungan wisatawan pada data uji (12 bulan terakhir), *trend* hasil dari model tersebut sudah sesuai dengan *trend* data aktual atau data latih.



Gambar 7.1 Hasil Pengujian ARIMA+XGBoost Model

B. Prediksi



Gambar 7.2 Hasil Prediksi

Gambar di atas menunjukan bahwa jumlah kunjungan Wisman pada bulan Agustus 2022 sampai Juli 2023 akan mengalami peningkatan. Kunjungan Wisman pada bulan Agustus 2022 diprediksi mencapai 590 ribu kunjungan kemudian meningkat pada bulan November 2022 yang diperkirakan akan mencapai lebih dari 890 ribu pengunjung. Puncaknya adalah pada bulan desember 2022 dengan jumlah kunjungan Wisman mencapai lebih dari 950 ribu kunjungan. Kunjungan Wisman diprediksi akan menurun pada awal tahun 2023 dengan prediksi titik terendahnya pada bulan Februari sebanyak 488 ribu kunjungan. Jumlah kunjungan diprediksi akan kembali meningkat secara signifikan pada Juli 2023 mencapai lebih dari 900 ribu kunjungan Wisman.

C. Lima Daerah dengan Kunjungan Terbanyak pada Bulan Desember 2022

Setelah memprediksi total kunjungan Wisman di seluruh wilayah di Indonesia, penulis juga memprediksi total kunjungan Wisman pada setiap daerah di Indonesia untuk bulan Desember 2022, setelah itu mengurutkannya. Daerah yang memiliki kunjungan wisatawan mancanegara terbanyak pada bulan Desember 2022 adalah Bali dengan 475.952 wisatawan, Kepulauan Riau dengan 257.787 wisatawan, Banten dengan 145.883 wisatawan, Jawa Timur dengan 22.024 wisatawan, serta Sumatera Utara dengan 16.450 wisatawan.

Tabel 7.1 Lima Daerah dengan Kunjungan Terbanyak pada Bulan Desember 2022

 Bali
 475952.406250

 Kepri
 257787.343750

 Banten
 145883.593750

 Jatim
 22024.634766

 Sumut
 16450.718750

8. KESIMPULAN DAN SARAN

- A. Kunjungan Wisman mengalami penurunan drastis semenjak COVID-19 melanda indonesia. Setelah bulan Januari 2022 jumlah kunjungan Wisman mulai mengalami kenaikan seiring dibukanya pintu masuk bagi warga asing.
- B. Model ARIMA + XGBoost cukup baik dalam melakukan prediksi atau peramalan pada jumlah kunjungan wisatawan mancanegara di Indonesia dengan skor evaluasi RMSE = 64661.827 dan R2_Score = 0.747 untuk total kunjungan.
- C. Pemerintah wajib mengantisipasi kenaikan jumlah wisatawan mancanegara dengan memperbaiki dan menghidupkan kembali industri wisata pada daerah-daerah prioritas terutama sebelum bulan Desember 2022 karena menurut prediksi yang dilakukan, lebih dari 900.000 wisatawan mancanegara akan berkunjung ke Indonesia pada bulan tersebut.
- D. Lima prioritas daerah yang akan dikunjungi wisatawan mancanegara pada bulan Desember 2022 adalah Bali, Kepulauan Riau, Banten, Jawa Timur, dan Sumatera Utara.

9. DAFTAR PUSTAKA

Tempo.co. 2022. Inilah Negara Terindah di Dunia versi Forbes, Indonesia Urutan Pertama. https://travel.tempo.co/read/1629629/inilah-negara-terindah-di-dunia-versi-forbes-indonesia-urutan-pertama (diakses pada 21 September 2022)

CNBC Indonesia. 2021. Sandi Uno: Bali Penyumbang Devisa Terbesar di Pariwisata RI. https://www.cnbcindonesia.com/news/20211222193439-4-301297/sandi-uno-bali-penyumbang-devisa-terbesar-di-pariwisata-ri (diakses pada 21 September 2022)

Azzahra, Qonita. 2022. Perkembangan Devisa Pariwisata. https://data.alinea.id/perkembangan-devisa-pariwisata-b2feY9CX9c (diakses pada 21 September 2022)

Badan Pusat Statistik. 2021. Berita Resmi Statistik 1 Februari 2021. https://www.bps.go.id/pressrelease/2021/02/01/1796/jumlah-kunjungan-Wisman-ke-indonesia-bulan-desember-2020-mencapai-164-09-ribu-kunjungan-html (diakses pada 21

- CNBC Indonesia. 2022. Pariwisata Dibuka Lagi, Turis dari 43 Negara Bebas Masuk RI. https://www.cnbcindonesia.com/lifestyle/20220411134829-33-330605/pariwisata-dibuka-lagi-turis-dari-43-negara-bebas-masuk-ri (diakses pada 21 September 2022)
- Liputan6.com. 2022. 477 Ribu Turis Asing Penuhi Indonesia di Juli 2022, Tertinggi selama Pandemi Covid-19. https://www.liputan6.com/bisnis/read/5058059/477-ribu-turis-asing-penuhi-indonesia-di-juli-2022-tertinggi-selama-pandemi-covid-19 (diakses pada 21 September 2022)
- Tableau Software, LLC. 2022. *Time Series Forecasting: Definition, Applications, and Examples*. https://www.tableau.com/learn/articles/time-series-forecasting#:~:text=Time%20series%20 forecasting%20occurs%20when,drive%20future%20strategic%20decision%2Dmaking (diakses pada 3 Oktober 2022)
- Kementerian Pariwisata dan Ekonomi Kreatif. (2021). Statistik Pariwisata dan Ekonomi Kreatif 2020. Jakarta, DKI: Penulis, Diakses dari https://kemenparekraf.go.id/direktori-statistik/statitsik-industri-pariwisata-dan-ekonomi-kreatif-2020 pada 3 Oktober 2022
- International Recommendations for Tourism Statistics 2008. (2008). Department of Economic and Social Affairs. United Nations. https://www.un.org/development/desa/cdpmo/tools/2020/international-recommendations-to-urism-statistics-2008-irts-2008 (diakses pada 3 oktober 2022)
- S. M. Idrees, M. A. Alam and P. Agarwal, "A Prediction Approach for Stock Market Volatility Based on Time Series Data," in IEEE Access, vol. 7, pp. 17287-17298, 2019, doi: 10.1109/ACCESS.2019.2895252.
- M. Chen, Q. Liu, S. Chen, Y. Liu, C. -H. Zhang and R. Liu, "XGBoost-Based Algorithm Interpretation and Application on Post-Fault Transient Stability Status Prediction of Power System," in IEEE Access, vol. 7, pp. 13149-13158, 2019, doi: 10.1109/ACCESS.2019.289344