

Analisis Prediksi Jumlah Pengunjung *Suncity Waterpark* Sidoarjo Menggunakan Metode Regresi Linier Berganda

DOI: <http://dx.doi.org/10.35889/jutisi.v14i2.2699>

Creative Commons License 4.0 (CC BY – NC) 

Mochammad Ilham Ardiansyah^{1*}, Neny Kurniati², Arda Surya Editya³

Informatika, Universitas Nahdlatul Ulama Sidoarjo, Sidoarjo, Indonesia

*e-mail *Corresponding Author*: ilhamardiansyah055@gmail.com

Abstrack

Suncity Waterpark Sidoarjo is a leading destination in Sidoarjo that experiences visitor fluctuations due to seasonal and promotional factors. Accurate visitor prediction is essential for effective resource management, service quality, and planning. This study aims to analyze and predict daily visitor numbers using Multiple Linear Regression, and compares its performance with K-Nearest Neighbors (K-NN) and Decision Tree methods. The dataset consists of 212 daily records over a specific period, including key variables such as holidays, promotions, extracurricular events, and regular visitor numbers. The research process includes data collection, preprocessing, model development, and evaluation using Root Mean Square Error (RMSE). Results show that Multiple Linear Regression achieved the lowest RMSE value of 3.532, outperforming K-NN and Decision Tree. The findings conclude that Multiple Linear Regression is the most effective method for predicting visitor numbers.

Keywords: *Data Mining; Tourism; Visitor Prediction; Multiple linear regression; Root Mean Square Error*

Abstrak

Suncity Waterpark Sidoarjo merupakan destinasi unggulan di Sidoarjo yang mengalami fluktuasi pengunjung akibat faktor musiman dan promosi. Prediksi pengunjung yang akurat sangat penting untuk manajemen sumber daya, kualitas layanan, dan perencanaan yang efektif. Penelitian ini bertujuan untuk menganalisis dan memprediksi jumlah pengunjung harian menggunakan Regresi Linier Berganda, dan membandingkan kinerjanya dengan metode *K-Nearest Neighbors* (K-NN) dan *Decision Tree*. Dataset terdiri dari 212 catatan harian selama periode tertentu, termasuk variabel-variabel utama seperti hari libur, promosi, kegiatan ekstrakurikuler, dan jumlah pengunjung rutin. Proses penelitian meliputi pengumpulan data, praproses, pengembangan model, dan evaluasi menggunakan *Root Mean Square Error* (RMSE). Hasil penelitian menunjukkan bahwa Regresi Linier Berganda mencapai nilai RMSE terendah sebesar 3,532, mengungguli K-NN dan *Decision Tree*. Temuan tersebut menyimpulkan bahwa Regresi Linier Berganda merupakan metode yang paling efektif untuk memprediksi jumlah pengunjung.

Kata kunci: *Data Mining; Pariwisata; Prediksi Pengunjung; Regresi linier berganda; Root Mean Square Error*

1. Pendahuluan

Sektor pariwisata adalah salah satu pilar utama ekonomi yang memberikan kontribusi signifikan terhadap kemajuan ekonomi lokal dan global [1]. Di Indonesia, pariwisata menjadi bagian dari sektor yang sangat penting untuk dikembangkan karena dapat memberi pengaruh pada kehidupan masyarakat di sekitarnya terutama pada sistem perekonomian [2]. pariwisata dianggap sebagai salah satu industri terbesar di Indonesia dan di dunia, serta merupakan sektor yang tumbuh pesat dan menciptakan banyak lapangan kerja. industri pariwisata telah berkembang pesat selama beberapa dekade terakhir, dan terbukti menjadi salah satu industri

paling populer secara global [3]. Jika potensi sumber daya pariwisata ini dikelola dengan baik dan profesional, akan menjadi daya tarik yang luar biasa bagi wisatawan. Selain itu, industri pariwisata sebagai industri jasa yang tergolong industri yang akan memegang peranan penting dalam menentukan kebijakan ketenagakerjaan karena kebutuhan akan lapangan kerja tetap akan semakin meningkat akibat meningkatnya jumlah wisatawan di masa depan.

Suncity Waterpark Sidoarjo merupakan salah satu taman rekreasi air di Jawa Timur, Indonesia. Taman air ini terletak di kota Sidoarjo dan merupakan tujuan wisata yang diminati oleh wisatawan lokal maupun regional. Sebagai salah satu ikon pariwisata di Sidoarjo, Suncity Waterpark berperan penting dalam meningkatkan ekonomi lokal melalui penciptaan lapangan kerja dan peningkatan pendapatan daerah dari sektor pariwisata. Jumlah pengunjung taman air ini bervariasi sepanjang tahun dengan puncak kunjungan biasanya terjadi pada akhir pekan, hari libur, dan musim liburan sekolah. Namun, berdasarkan data kunjungan akhir-akhir ini, jumlah pengunjung Suncity Waterpark menunjukkan fluktuasi yang signifikan dan cenderung sulit diprediksi secara akurat. Hal ini menimbulkan permasalahan dalam hal perencanaan operasional, seperti persiapan tenaga kerja, pasokan logistik, hingga perencanaan promosi musiman. Masalah ini bersifat kuantitatif dan terukur, seperti jumlah pengunjung harian, pengaruh musim liburan, promosi, serta kondisi cuaca yang memengaruhi minat kunjungan. Oleh karena itu, perlu dilakukan pendekatan analitis berbasis data untuk mengatasi ketidakpastian tersebut.

Penelitian ini menawarkan penggunaan metode regresi linier berganda untuk memprediksi jumlah pengunjung Suncity Waterpark Sidoarjo berdasarkan beberapa variabel independen. Algoritma ini merupakan metode prediksi untuk mencari tahu hubungan antara variabel yang mempengaruhi dengan variabel dipengaruhi [4]. Berdasarkan persamaan regresi yang variabel perubahannya berbentuk skala untuk mengetahui nilai prediksi [5], dan melihat tingkat seberapa besar determinasi atau pengaruh pada variabel bebas terhadap variabel terkait atau hasil prediksi [6]. Metode ini juga telah banyak digunakan dalam berbagai penelitian serupa di bidang pariwisata dan peramalan permintaan konsumen, sehingga diyakini relevan dan valid untuk konteks permasalahan ini.

Tujuan analisis prediksi jumlah pengunjung di Suncity Waterpark Sidoarjo bertujuan untuk menganalisis pola kunjungan berdasarkan data historis yang memengaruhi fluktuasi jumlah wisatawan. Dan mengetahui metode mana yang dapat menghasilkan prediksi yang lebih akurat dari membandingkan 3 model metode diantaranya ialah Regresi linier berganda, *K-Nearest Neighbors* (K-NN), *Decision tree regression*. Serta hasil pengujian dilakukan dengan menggunakan metode evaluasi estimasi, yaitu *Root Mean Square Error* (RMSE). Lebih lanjut, hasil dari analisis ini dapat menjadi dasar dalam pengambilan keputusan strategis guna meningkatkan kualitas pelayanan serta pengalaman pengunjung. Dengan memahami pola kunjungan, pengelola dapat mencegah kepadatan yang berlebihan sehingga kenyamanan wisatawan tetap terjaga. Selain itu, prediksi jumlah pengunjung juga dapat membantu pemerintah daerah dan pemangku kepentingan dalam merancang kebijakan pengembangan sektor pariwisata yang lebih berkelanjutan dan kompetitif. Oleh karena itu, analisis ini tidak hanya memberikan manfaat bagi pengelola Suncity Waterpark tetapi juga turut mendukung pertumbuhan industri pariwisata di Sidoarjo secara keseluruhan.

2. Tinjauan Pustaka

Terdapat beberapa penelitian seperti yang dilakukan oleh Muis Nanja dan Hastuti Dalai, yang berjudul "Prediksi Jumlah Kunjungan Wisatawan Asing Menggunakan Algoritma Linier Regresi *Multivariat*". Berdasarkan penelitian yang dilakukan bertujuan untuk menganalisis pola kunjungan wisatawan asing di Gorontalo dan memprediksi jumlah kunjungan di masa depan menggunakan Algoritma Regresi Linier Berganda berbasis *Forward Selection*. Dengan memahami fluktuasi jumlah kunjungan, penelitian ini bertujuan memberikan rekomendasi strategis bagi pemerintah dan dinas pariwisata dalam merencanakan anggaran dan program pariwisata yang lebih efektif, serta meningkatkan kesadaran stakeholder tentang pentingnya stabilitas kunjungan dalam mendukung pengembangan ekonomi daerah. Berdasarkan hasil pengujian metode diperoleh nilai *Root Mean Square Error* (RMSE) yaitu, Regresi *Multivariat* 2660,89, Regresi *Multivariat Forward Selection* 556,49, Regresi *Multivariat Backward Selection* 2377,44. Sehingga dapat disimpulkan bahwa metode yang paling baik digunakan untuk melakukan prediksi yaitu Regresi *Multivariat Forward Selection* [7].

Penelitian selanjutnya oleh Beauty Leony karamoy dkk., yang berjudul "Data Mining Estimasi Pendapatan Tempat Wisata di Kota Tomohon Menggunakan Algoritma Regresi Linear Berganda". Penelitian ini bertujuan untuk memprediksi pendapatan destinasi wisata di Kota Tomohon dengan menggunakan algoritma regresi linier berganda. Hasil penelitian menunjukkan bahwa variabel independen yang terdiri dari harga tiket masuk (X1) dan jumlah pengunjung (X2) berpengaruh signifikan terhadap variabel dependen yaitu pendapatan per bulan (Y). Model persamaan regresi yang diperoleh adalah $Y = -5594 + 250X_1 + 2.26X_2$. Dengan menggunakan model ini, prediksi pendapatan bulan berikutnya untuk masing-masing destinasi wisata di Kota Tomohon dapat ditentukan, yang menunjukkan bahwa pendekatan regresi linear berganda efektif untuk estimasi pendapatan wisata [8].

Penelitian selanjutnya oleh Tamsil Hermawan dkk., yang berjudul "Sistem Prediksi Jumlah Wisatawan di Lambangan Pauno Desa Kendek Menggunakan Metode *Support Vector Machine*". Tujuan penelitian ini adalah untuk menganalisis faktor-faktor yang mempengaruhi peningkatan kunjungan wisatawan ke Pantai Lambangan Pauno di Banggai Laut, serta mengevaluasi dampak kunjungan tersebut terhadap pendapatan lokal dan pengembangan pariwisata. Dengan memahami elemen-elemen yang menarik wisatawan, penelitian ini bertujuan memberikan rekomendasi strategis bagi pengelola objek wisata dan pemerintah daerah untuk meningkatkan kualitas layanan serta infrastruktur, sehingga dapat lebih menarik pengunjung dan mengoptimalkan potensi pariwisata di daerah tersebut. Dari hasil penelitian diperoleh bahwa pengkategorian penelitian yang dihasilkan oleh SVM sudah sangat baik. Hal ini dibuktikan oleh hasil pengujian yang menghasilkan tingkat akurasi 90% [9].

Penelitian selanjutnya oleh Binsar Antoni Manurung dkk., yang berjudul "Penerapan Algoritma Regresi Linear Untuk Memprediksi Jumlah Wisatawan". Tujuan penelitian ini dilakukan untuk memprediksi jumlah wisatawan di suatu destinasi dengan menggunakan analisis regresi linear, serta mengidentifikasi variabel signifikan yang mempengaruhi fluktuasi jumlah wisatawan, khususnya terkait dengan event yang diadakan. Dengan mengembangkan persamaan regresi linear, penelitian ini bertujuan untuk memberikan prediksi yang akurat mengenai jumlah wisatawan di masa depan, sehingga dapat membantu dalam pengoptimalan strategi pengembangan pariwisata dan meningkatkan kestabilan kunjungan wisatawan. Hasil Prediksi menunjukkan bahwa persamaan regresi linear dapat diterapkan untuk memprediksi jumlah wisatawan pada periode berikutnya dengan tingkat akurasi sebesar 95% dan MAPE (*Mean Absolute Percentage Error*) sebesar 5,21%, dari hasil tersebut dapat disimpulkan bahwa peramalan dengan menggunakan metode regresi linear termasuk dalam kategori baik [10].

Berdasarkan tinjauan terhadap beberapa penelitian sebelumnya, diketahui bahwa pendekatan regresi linier berganda telah banyak digunakan dalam memprediksi jumlah kunjungan atau pendapatan pariwisata, seperti pada penelitian [7]. Yang menggunakan pendekatan *forward selection* untuk meningkatkan akurasi, serta penelitian [8]. Yang fokus pada estimasi pendapatan berdasarkan variabel harga tiket dan jumlah pengunjung. Penelitian lain oleh [9]. menggunakan metode SVM untuk klasifikasi wisatawan, sementara [10]. menyoroti pengaruh event terhadap prediksi wisatawan dengan regresi linier sederhana. Adapun kebaruan dari penelitian ini terletak pada pendekatan regresi linier berganda yang diterapkan secara spesifik pada kasus prediksi jumlah pengunjung Suncity Waterpark Sidoarjo, dengan mengintegrasikan variabel Hari Libur, Promosi, Ekstrakurikuler, dan Reguler sebagai parameter input yang merepresentasikan karakteristik unik pengunjung di kawasan wisata tersebut. Pendekatan ini belum banyak dijumpai dalam penelitian terdahulu yang lebih umum dan tidak menyentuh segmentasi aktivitas pengunjung secara eksplisit. Dengan demikian, penelitian ini menawarkan kontribusi baru dalam bentuk model prediktif berbasis segmentasi aktivitas yang dapat membantu pengelola wisata dalam membuat keputusan operasional yang lebih tepat dan efisien.

3. Metodologi

Metodologi penelitian kali ini diawali dengan proses pengumpulan data, selanjutnya data akan dibagi menjadi data training dan data testing dengan proporsi tertentu guna memastikan model yang dibangun dapat diuji secara akurat. Pada tahap berikutnya dilakukan proses data mining dengan menerapkan algoritma Regresi Linier Berganda. Setelah model dibangun tahapan berikutnya adalah evaluasi terhadap kinerja algoritma yang digunakan.

1) Pengumpulan Data

Pada tahap proses pengumpulan data peneliti mendapatkan data dari catatan kasir Suncity Waterpark yang menghimpun informasi historis mengenai jumlah pengunjung Suncity Waterpark Sidoarjo sebagai dasar untuk melakukan prediksi. Dimana data set yang digunakan diproses dengan menerapkan tekning mining yaitu Merupakan bidang ilmu yang bertujuan mengeksplorasi dan eksploitasi pengetahuan dari suatu informasi [11]. Data yang digunakan mencakup periode dari 1 Januari hingga 31 Juli 2024 dengan total sebanyak 212 entri harian. Data ini tersimpan dalam format Excel dan memuat sejumlah variabel penting, seperti tanggal, status hari libur, jumlah pengunjung reguler, pengunjung yang memanfaatkan promosi baik dilakukan melalui media sosial maupun papan informasi di area lobi masuk, pengunjung dari kegiatan ekstrakurikuler, serta total keseluruhan pengunjung setiap harinya. Seluruh variabel tersebut merefleksikan karakteristik pengunjung yang datang dan akan menjadi landasan utama dalam analisis serta prediksi jumlah pengunjung.

2) Pre Prosesing data

Setelah data berhasil dikumpulkan tahap selanjutnya adalah pre prosesing data atau prapengolahan data hal ini dilakukan dengan tujuan untuk memastikan bahwa data yang digunakan dalam proses analisis dan peramalan berada dalam kondisi bersih, konsisten, dan siap diolah. Tahap awal dalam proses *preprocessing* dimulai dengan membersihkan data dari nilai yang hilang, data ganda, maupun data yang tidak wajar yang berpotensi mengganggu akurasi analisis. Penanganan dilakukan sesuai kondisi, seperti mengisi nilai kosong dengan estimasi berdasarkan pola yang ada atau menghapus data yang dianggap tidak relevan. Selanjutnya dilakukan transformasi data untuk memastikan semua format sesuai standar yang dibutuhkan, seperti mengubah tanggal menjadi format yang dapat diproses oleh sistem serta mengonversi kategori seperti hari libur ke dalam bentuk numerik misalnya 1 untuk hari libur dan 0 untuk hari biasa. Pada tahap berikutnya, normalisasi data dilakukan apabila ditemukan perbedaan skala yang signifikan antar variabel guna menghindari dominasi nilai tertentu terhadap hasil analisis. Namun dalam penelitian ini normalisasi hanya diterapkan pada variabel-variabel yang memang memengaruhi keseimbangan model prediksi. Berikut hasil dataset yang diperoleh berjumlah 173 dari data asli berjumlah 212 data. Contoh hasil Preprosesing data set dapat dilihat pada Tabel 1.

Tabel 1. Data set

Tanggal	Hari libur	Pengunjung Promosi	Pengunjung Extrakurikuler	Pengunjung Reguler	Total Pengunjung
01/01/2024	1	0	0	1722	1.722
02/01/2024	0	0	0	215	215
03/01/2024	0	0	0	124	124
04/01/2024	0	0	0	70	70
05/01/2024	0	0	0	80	80
06/01/2024	1	0	570	143	716
07/01/2024	1	0	0	354	354
09/01/2024	0	129	0	81	209
10/01/2024	0	0	315	62	377
11/01/2024	0	130	62	69	261
12/01/2024	0	252	461	68	781
13/01/2024	1	0	1025	224	1.249
14/01/2024	1	60	0	538	598
15/01/2024	0	0	132	2	134
16/01/2024	0	70	342	57	477
17/01/2024	0	0	305	42	347
18/01/2024	0	0	220	38	258
19/01/2024	0	95	397	66	558
20/01/2024	1	0	689	199	890
Σ	60	10617	43924	34451	89449

Pada penelitian ini hari libur sebagai variabel X1, pengunjung promosi sebagai X2, pengunjung ekstrakurikuler X3, pengunjung reguler X4, dan pengunjung harian menjadi Y.

3). Implementasi Prediksi

a) Regresi Linier Berganda

Pada tahap ini implementasi menggunakan metode Regresi Linier Berganda. Regresi Linier Berganda merupakan metode prediksi yang digunakan ketika terdapat satu variabel terikat dan lebih dari satu variabel bebas. Tujuan dari analisis ini adalah untuk memahami arah hubungan antara variabel-variabel tersebut, apakah hubungan yang terbentuk bersifat positif atau negatif, serta untuk memperkirakan perubahan nilai variabel terikat apabila terjadi peningkatan atau penurunan pada variabel bebas. Yang menjadi pembeda antara Regresi Linier Sederhana adalah jumlah variabel bebas. Regresi Linier Sederhana hanya ada satu variabel bebas dan variabel tidak bebas [12]. Adapun rumus perhitungan metode Regresi Linier Berganda sebagai berikut:

$$Y = a + b_1X_1 + b_2X_2 + \dots + b_nX_n \quad (1)$$

Keterangan:

Y= Variabel yang diprediksi (variable dependent)

a = Konstanta

X1,X2,X3,..., Xn= Variabel yang diketahui (variable Independent)

b1, b2, b3, ..., bn = Koefisien regresi

n = Jumlah data

Konstanta a dan koefisien-koefisien regresi b1, b2, dan, b3 dapat dihitung menggunakan rumus:

$$a = \begin{bmatrix} n & \sum x_1 & \sum x_2 & \sum x_3 \\ \sum x_1 & \sum x_1^2 & \sum x_1x_2 & \sum x_1x_3 \\ \sum x_2 & \sum x_1x_2 & \sum x_2^2 & \sum x_2x_3 \\ \sum x_3 & \sum x_1x_3 & \sum x_2x_3 & \sum x_3^2 \end{bmatrix} \quad (2)$$

$$H = \begin{bmatrix} \sum y \\ \sum x_1y \\ \sum x_2y \\ \sum x_3y \end{bmatrix} \quad (3)$$

$$a = \frac{\det M_1}{\det M} \quad b_1 = \frac{\det M_2}{\det M} \quad b_2 = \frac{\det M_3}{\det M} \quad b_n = \frac{\det M_n}{\det M} \quad (4)$$

Diawali dengan menghitung koefisien regresi b1, b2, dan b3, kemudian dilanjutkan menghitung konstanta a dengan mengambil hasil dari perhitungan koefisien. Lalu menghitung persamaan regresi dengan mengambil hasil dari perhitungan konstanta. Dan untuk nilai variabel x1, x2, x3 dan xn menggunakan dataset lalu dapat memperoleh hasil prediksinya [13].

b) K-Nearest Neighbors

K-Nearest Neighbor (KNN) adalah algoritma pembelajaran non-parametrik yang digunakan dalam klasifikasi dan regresi. Untuk kasus regresi, KNN memprediksi nilai output berdasarkan rata-rata nilai dari k data tetangga terdekat terhadap suatu data baru. Prinsip dasar dari KNN adalah bahwa data yang memiliki fitur serupa cenderung memiliki nilai output yang serupa juga. Oleh karena itu, untuk memprediksi nilai dari data baru, KNN mencari sejumlah data (k) di dataset pelatihan yang paling dekat dengan data tersebut berdasarkan jarak tertentu. Berikut persamaan menghitung jarak *euclidean*:

$$d = \sum_{i=1}^p (x_{2i} - x_{1i})^2 \quad (5)$$

Keterangan:

d = jarak eucliden

x_{2i} = nilai pada data uji ke = i

x_{1i} = nilai pada data latih ke = i

p = jumlah atribut yang digunakan

c) Decision Tree

Decision Tree Regression diawali dengan proses perhitungan nilai *entropy*, yaitu ukuran yang menggambarkan tingkat ketidakpastian dalam suatu dataset. Nilai *entropy* digunakan untuk menentukan seberapa homogen data pada suatu node sebelum dan sesudah dilakukan pemisahan. Semakin rendah nilai *entropy*, semakin murni data dalam *node* tersebut. Perhitungan *entropy* membantu dalam memilih atribut terbaik untuk membagi data agar dapat meminimalkan impuritas dan menghasilkan model prediksi yang lebih akurat. Berikut Persamaan mencari nilai *entropy*:

$$Entropy(S) = - \sum_{i=1}^n p_i \log_2(p_i) \quad (6)$$

Keterangan:

S adalah himpunan data,

p_i adalah proporsi data dari kelas ke-iii dalam node tersebut,

n adalah jumlah total kelas dalam dataset.

Setelah menghitung nilai *entropy* dan *gain*, langkah selanjutnya adalah membuat cabang untuk setiap nilai pada atribut akar. Proses ini dilanjutkan secara berulang dengan membagi data di setiap cabang hingga terbentuk pohon keputusan yang utuh. Setelah seluruh pohon keputusan terbentuk, proses prediksi dilakukan dengan menggabungkan hasil prediksi dari masing-masing pohon. Pada kasus klasifikasi, prediksi ditentukan berdasarkan suara terbanyak. Sementara dalam regresi, hasil prediksi diperoleh dari nilai rata-rata atau median dari semua prediksi yang dihasilkan oleh pohon.

4). Evaluasi

Tahapan selanjutnya adalah evaluasi kinerja algoritma yang telah digunakan dalam proses data mining. Evaluasi ini bertujuan untuk menilai seberapa baik model dalam melakukan prediksi dengan membandingkan hasil yang dihasilkan dengan data aktual. Dalam proses evaluasi, dilakukan pengukuran tingkat akurasi dan tingkat error dari setiap model yang dibangun. Pengukuran ini menggunakan *Root Mean Square Error* (RMSE), hal ini merupakan salah satu metrik yang paling sering digunakan dalam evaluasi model Regresi Linier Berganda maupun model regresi lainnya. RMSE mengukur sejauh mana prediksi model mendekati nilai aktual dalam dataset [14]. Rumus RMSE sebagai berikut:

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (7)$$

Keterangan:

n = jumlah data (jumlah observasi)

y_i = nilai aktual pada data ke- i

\hat{y}_i = nilai prediksi pada data ke- i

$(y_i - \hat{y}_i)$ = residual atau kesalahan (selisih antara nilai aktual dan nilai prediksi)

$(y_i - \hat{y}_i)^2$ = kuadrat dari residual = rata-rata kesalahan kuadrat.

4. Hasil dan Pembahasan

Pada Tahap ini menyajikan hasil dari penelitian yang telah dilakukan termasuk proses pengujian yang menyertainya. Selain itu, bagian ini juga memuat pembahasan terkait hasil penelitian dan pengujian tersebut untuk memberikan pemahaman yang lebih mendalam.

4.1 Hasil Prediksi Regresi Linier Berganda

Pada tahap ini disajikan hasil perhitungan prediksi yang diperoleh menggunakan metode Regresi Linier Berganda. Prediksi ini dilakukan berdasarkan data yang telah melalui tahap *preprocessing* dan pemodelan sebelumnya. Tujuan dari proses ini adalah untuk mengetahui sejauh mana model Regresi Linier Berganda mampu memprediksi nilai target berdasarkan variabel-variabel independen yang telah ditentukan. langkah selanjutnya adalah membentuk data matriks yang akan digunakan dalam analisis Regresi Linier Berganda. Pembentukan data matriks ini bertujuan untuk mempermudah proses perhitungan serta memastikan bahwa struktur data sesuai dengan persyaratan model regresi. Data yang telah melalui proses *preprocessing* kemudian diorganisasikan ke dalam format tabel, di mana setiap baris merepresentasikan satu observasi, dan setiap kolom menunjukkan nilai dari masing-masing variabel. Berikut hasil analisis perhitungan data matriks pada Tabel 2.

Tabel 2. Analisis Perhitungan Data Matriks

X1Y	X2Y	X3Y	X4Y	X1X2	X1X3	X1X4	X2X3	X2X4	X3X4	x1^2	x2^2	x3^2	x4^2
1722	0	0	2965284	0	0	1722	0	0	0	1	0	0	2965284
0	0	0	46225	0	0	0	0	0	0	0	0	0	46225
0	0	0	15376	0	0	0	0	0	0	0	0	0	15376
0	0	0	4900	0	0	0	0	0	0	0	0	0	4900
0	0	0	6400	0	0	0	0	0	0	0	0	0	6400
716	0	408120	102388	0	570	143	0	0	81510	1	0	324900	20449
354	0	0	125316	0	0	354	0	0	0	1	0	0	125316
0	26961	0	16929	0	0	0	0	10449	0	0	16641	0	6561
0	0	118755	23374	0	0	0	0	0	19530	0	0	99225	3844
0	33930	16182	18009	0	0	0	8060	8970	4278	0	16900	3844	4761
0	196812	360041	53108	0	0	0	116172	17136	31348	0	63504	212521	4624
1249	0	1280225	279776	0	1025	224	0	0	229600	1	0	1050625	50176
598	35880	0	321724	60	0	538	0	32280	0	1	3600	0	289444
0	0	17688	268	0	0	0	0	0	264	0	0	17424	4
0	33390	163134	27189	0	0	0	23940	3990	19494	0	4900	116964	3249
0	0	105835	14574	0	0	0	0	0	12810	0	0	93025	1764
0	0	56760	9804	0	0	0	0	0	8360	0	0	48400	1444
0	53010	221526	36828	0	0	0	37715	6270	26202	0	9025	157609	4356
890	0	613210	177110	0	689	199	0	0	137111	1	0	474721	39601
42889	69847925	43711784	22249637	3499	16542	22578	2526520	2478620	4668671	60	1968213	36447010	15009517

Setelah menemukan analisis perhitungan data matriks, tahap selanjutnya mencari matriks a dan matriks H. Berikut ini adalah hasil perhitungan matriks a dan matriks H yang dihasilkan pada data yang didapatkan dari jumlah dari masing-masing data analisis di atas bisa dilihat di Tabel 3.

Tabel 3. Matriks a dan Matriks H

a	173	60	10617	43924	34451	H	89449
	60	60	3499	16542	22587		42889
	10617	3499	1968213	2526520	2478620		6987925
	43924	16542	2526520	36447010	4668671		43711784
	34451	22587	2478620	4668671	15009517		22249637

Pencarian matriks a juga menyesuaikan dengan jumlah variabel yang digunakan, di dalam penelitian ini menggunakan 5 variabel yang terdiri dari 4 variabel x dan 1 variabel y maka jumlah masing-masing a (konstanta) yaitu a1, a2, a3, a4, a5, yang bertujuan untuk mencari determinan matriks. Pada matriks a1 akan mengubah nilai dikolom pertama dengan nilai H pada matriks a maka hasilnya bisa dilihat di Tabel 4.

Tabel 4. Matriks a1

a1	89449	60	10617	43924	34451
	42889	60	3499	16542	22587
	6987925	3499	1968213	2526520	2478620
	43711784	16542	2526520	36447010	4668671
	22249637	22587	2478620	4668671	15009517

Pada matriks a2 akan mengubah nilai dikolom kedua dengan nilai H pada matriks a maka hasilnya bisa dilihat di Tabel 5.

Tabel 5. Matriks a2

a2	173	89449	10617	43924	34451
	60	42889	3499	16542	22587
	10617	6987925	1968213	2526520	2478620
	43924	43711784	2526520	36447010	4668671
	34451	22249637	2478620	4668671	15009517

Pada matriks a3 akan mengubah nilai dikolom ketiga dengan nilai H pada matriks a maka hasilnya bisa dilihat di Tabel 6.

Tabel 6. Matriks a3

a3	173	60	89449	43924	34451
	60	60	42889	16542	22587
	10617	3499	6987925	2526520	2478620
	43924	16542	43711784	36447010	4668671
	34451	22587	22249637	4668671	15009517

Pada matriks a4 akan mengubah nilai dikolom keempat dengan nilai H pada matriks a maka hasilnya bisa dilihat di Tabel 7.

Tabel 7. Matriks a4

a4	173	60	10617	89449	34451
	60	60	3499	42889	22587
	10617	3499	1968213	6987925	2478620
	43924	16542	2526520	43711784	4668671
	34451	22587	2478620	22249637	15009517

Pada matriks a5 akan mengubah nilai dikolom kelima dengan nilai H pada matriks a maka hasilnya bisa dilihat di Tabel 8.

Tabel 8. Matriks a5

a5	173	60	10617	43924	89449
	60	60	3499	16542	42889
	10617	3499	1968213	2526520	6987925
	43924	16542	2526520	36447010	43711784
	34451	22587	2478620	4668671	22249637

Kemudian dilakukan pencarian determinan dari masing-masing matriks maka hasilnya bisa dilihat di Tabel 9.

Tabel 9. Perhitungan Determinan

det(A)	9,83546
det(a1)	3,71293
det(a2)	4,52789
det(a3)	9,7562
det(a4)	9,80333
det(a5)	9,76601

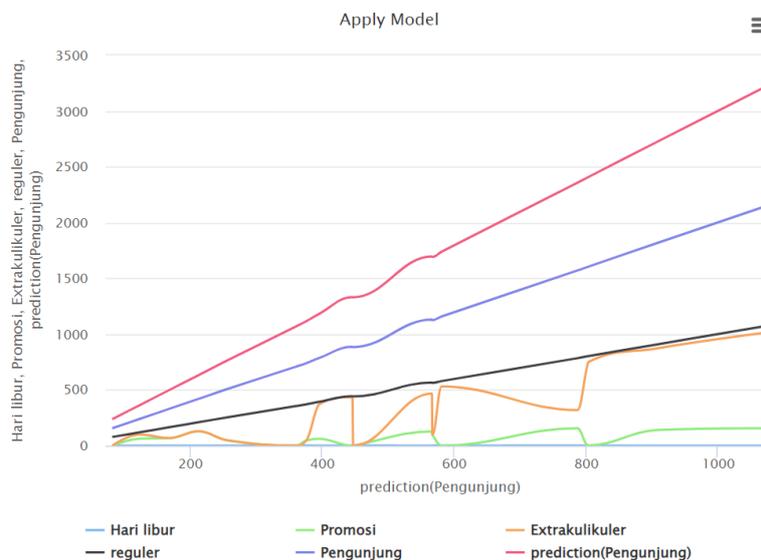
Dari hasil determinan dapat dihasilkan nilai konstanta(a) dan koefisien(b) yang didapatkan dari pembagian masing-masing determinan. Nilai yang dihasilkan adalah a, b1, b2, b3, dan b4 bisa dilihat pada Tabel 10.

Tabel 10. Perhitungan Konstanta Dan Koefisien

a	3,77504756
b1	4,60363984
b2	0,99194144
b3	0,99673334
b4	0,99293897

Dari hasil a dan bn dapat dihasilkan persamaan Regresi Linier Berganda untuk melakukan prediksi pengunjung Suncity Waterpark Sidoarjo. Sebagai contoh perhitungan proses prediksi bisa dilakukan dengan menggunakan nilai x1 (hari libur) = 1, x2 (promosi) = 0, x3 (ekstrakurikuler) = 570, x4 (reguler) = 143. Kemudian memasukkan nilai tersebut kepersamaan Regresi Linier Berganda untuk melihat hasil prediksinya.

$$Y = 3,77504756 + 4,60363984.(1) + 0,99194144.(0) + 0,99673334.(570) + 0,99293897.(143) = 718,506$$

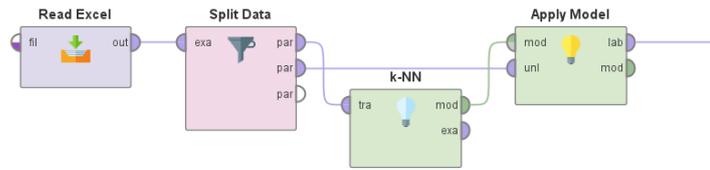


Gambar 1. Grafik Regresi Linier Berganda

Pada Gambar 1 Menunjukkan grafik hasil prediksi jumlah pengunjung menggunakan regresi linier berganda berdasarkan beberapa variabel seperti hari libur, promosi, kegiatan ekstrakurikuler, dan hari reguler. Garis merah muda merepresentasikan hasil prediksi jumlah pengunjung, sementara garis biru tua menunjukkan data pengunjung aktual. Terlihat bahwa hasil prediksi cukup akurat karena mengikuti pola dari data aktual. Variabel reguler menunjukkan tren yang stabil, sedangkan promosi dan ekstrakurikuler menunjukkan fluktuasi yang signifikan, menandakan bahwa keduanya memiliki pengaruh besar saat berlangsung. Hari libur tampak kurang berpengaruh karena nilainya cenderung datar. Secara keseluruhan, model ini dapat digunakan untuk memprediksi jumlah pengunjung secara efektif berdasarkan kombinasi aktivitas yang dilakukan.

4.2. Hasil Prediksi KNN

Penelitian ini menggunakan perangkat lunak RapidMiner dalam membangun model prediksi dengan algoritma *K-Nearest Neighbors*. Dan alurnya ada pada Gambar 2.



Gambar 2. Proses Prediksi KNN

Tahapan pertama dimulai dengan membaca dataset menggunakan komponen *Read Excel*, yang memuat data jumlah pengunjung Suncity Waterpark Sidoarjo. Setelah itu, data dipisahkan menggunakan *Split Data* menjadi dua bagian, yaitu data latih (*training data*) sebanyak 90% dari jumlah data dan data uji (*testing data*) sebanyak 10% dari jumlah data. Data latih digunakan untuk membangun model dengan algoritma *k-NN*, yang bekerja dengan mencari sejumlah tetangga terdekat berdasarkan nilai jarak (umumnya menggunakan *Euclidean distance*) untuk melakukan prediksi. Model yang telah terbentuk kemudian diterapkan pada data uji melalui komponen *Apply Model* guna menghasilkan nilai prediksi. Berikut hasil prediksi pada Gambar 3. Hasil prediksi KNN.

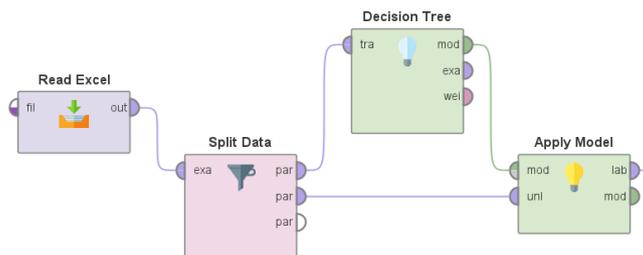
Row No.	Pengunjung	prediction(P...	Tgl	Hari libur	Promosi	Ekstrakulik...	reguler
1	354	289.577	Jan 7, 2024	1	0	0	354
2	209	443.942	Jan 9, 2024	0	129	0	81
3	441	428.333	Jan 21, 2024	1	0	0	441
4	78	669.000	Jan 31, 2024	0	0	0	78
5	907	664.583	Feb 6, 2024	0	140	731	35
6	786	370.833	Feb 10, 2024	1	155	163	465
7	565	372.056	Mar 6, 2024	0	127	340	98
8	562	250.833	Apr 21, 2024	1	97	0	465
9	802	446.444	May 3, 2024	0	3	746	53
10	119	544.972	May 13, 2024	0	62	37	20
11	579	981.556	May 16, 2024	0	0	530	49
12	443	1039.833	May 20, 2024	0	0	437	6
13	392	547.682	Jun 3, 2024	0	62	305	25

ExampleSet (17 examples, 2 special attributes, 5 regular attributes)

Gambar 3. Hasil Prediksi KNN

4.3. Hasil Prediksi Decision tree

Penelitian ini menggunakan perangkat lunak RapidMiner dalam membangun model prediksi dengan algoritma *Decision Tree*.



Gambar 4. Proses Prediksi Decision Tree

Proses pengolahan data menggunakan algoritma *Decision Tree* dalam penelitian ini dilakukan melalui tahapan yang ditunjukkan pada Gambar 4 menggunakan RapidMiner. Pertama, data jumlah pengunjung Suncity Waterpark Sidoarjo diimpor melalui komponen *Read Excel*. Selanjutnya, data dibagi menjadi dua bagian menggunakan *Split Data*, yaitu data latih sebanyak 90% dan data uji sebanyak 10%. Data latih digunakan untuk membangun model prediksi menggunakan algoritma *Decision Tree*, yang secara otomatis menganalisis atribut-atribut dan membentuk struktur pohon berdasarkan nilai *entropy* dan *gain* tertinggi. Setelah model terbentuk, komponen *Apply Model* digunakan untuk menerapkan model tersebut pada

data uji guna memperoleh hasil prediksi. Berikut hasil Prediksi dapat dilihat pada Gambar 5. Hasil prediksi *decision tree*.

Row No.	Pengunjung	prediction(P...	Tgl	Hari libur	Promosi	Extrakulikul...	reguler
1	354	289.577	Jan 7, 2024	1	0	0	354
2	209	443.942	Jan 9, 2024	0	129	0	81
3	441	428.333	Jan 21, 2024	1	0	0	441
4	78	669.000	Jan 31, 2024	0	0	0	78
5	907	664.583	Feb 6, 2024	0	140	731	35
6	786	370.833	Feb 10, 2024	1	155	163	465
7	565	372.056	Mar 6, 2024	0	127	340	98
8	562	250.833	Apr 21, 2024	1	97	0	465
9	802	446.444	May 3, 2024	0	3	746	53
10	119	544.972	May 13, 2024	0	62	37	20
11	579	981.556	May 16, 2024	0	0	530	49
12	443	1039.833	May 20, 2024	0	0	437	6
13	392	547.682	Jun 3, 2024	0	62	305	25

Gambar 5. Hasil Prediksi Decision Tree

4.4. Evaluasi dan Pembahasan Hasil

Untuk mengetahui sejauh mana tingkat akurasi model dalam melakukan prediksi, dilakukan evaluasi menggunakan metode *Root Mean Square Error* (RMSE). RMSE merupakan salah satu metrik evaluasi yang umum digunakan dalam analisis regresi untuk mengukur besarnya kesalahan antara nilai aktual dan nilai prediksi yang dihasilkan oleh model. Semakin kecil nilai RMSE maka semakin baik pula performa model dalam melakukan prediksi terhadap data yang dianalisis. Disini peneliti menggunakan aplikasi rapidminer dengan jumlah data sebanyak 176 data yang dibagi menjadi 90% data *training* dan 10% data *testing*, algoritma yang menggunakan rasio 90:10 memiliki performa yang lebih baik dibandingkan dengan algoritma lainnya dimana tingkat akurasinya dapat diperoleh sebesar 99,40% [15]. Berikut tabel data hasil evaluasi prediksi jumlah pengunjung Suncity Waterpark Sidoarjo menggunakan RSME.

Tabel 11. Hasil Evaluasi

Metode	RMSE
Regresi Linier Berganda	3,532
Decision Tree	91,958
k-NN	315,968

Berdasarkan Tabel 11 hasil perhitungan *Root Mean Square Error* (RMSE) pada tabel di atas, dapat disimpulkan bahwa metode Regresi Linier Berganda menghasilkan nilai RMSE paling kecil, yaitu sebesar 3,532. Hal ini menunjukkan bahwa model Regresi Linier Berganda memiliki tingkat kesalahan prediksi yang paling rendah dibandingkan dengan metode lainnya. Sementara itu, metode *Decision Tree* dan *k-Nearest Neighbors* (k-NN) menghasilkan nilai RMSE yang jauh lebih tinggi, masing-masing sebesar 91,958 dan 315,968. Temuan ini sejalan dengan penelitian oleh Pradito dan Purnia [16], yang menunjukkan bahwa regresi linier memiliki performa prediksi yang lebih baik dibandingkan dengan *Neural Network* dalam memprediksi nilai kurs mata uang, dengan nilai RMSE sebesar 28.012 ± 5.626 .

Performa *Decision Tree* yang berada di peringkat kedua juga didukung oleh penelitian oleh Utami et al. [17], yang menunjukkan bahwa *Decision Tree* dapat digunakan secara efektif dalam klasifikasi penyakit kardiovaskular, meskipun performanya dapat dipengaruhi oleh ukuran dan kualitas dataset.

Adapun algoritma k-NN menunjukkan performa paling rendah dalam penelitian ini, yang konsisten dengan penelitian oleh Anggraini dan Setiawan [18], di mana k-NN mengalami penurunan akurasi dan peningkatan error saat digunakan pada dataset yang memiliki variabel skala besar dan tidak distandarasi. Dengan demikian, hasil penelitian ini menguatkan temuan-temuan sebelumnya bahwa pemilihan algoritma yang tepat harus disesuaikan dengan karakteristik data. Penelitian ini juga berkontribusi dalam memperluas bukti empiris bahwa regresi linier masih menjadi metode yang handal dalam prediksi data numerik dengan korelasi linier.

5. Simpulan

Berdasarkan penelitian yang telah dilakukan maka bisa disimpulkan bahwa metode Regresi Linier Berganda memiliki performa yang sangat baik untuk memprediksi jumlah pengunjung Suncity Waterpark Sidoarjo dibandingkan dengan *Decision tree regression* dan *k-Nearest Neighbor*. Hasil evaluasi menunjukkan bahwa Regresi Linier Berganda memiliki nilai RMSE yang paling rendah yaitu sebesar 3,532 dibandingkan algoritma lainnya. Hal ini mengindikasikan bahwa metode tersebut memiliki tingkat kesalahan yang lebih kecil dan mampu memberikan prediksi jumlah pengunjung Suncity Waterpark Sidoarjo dengan tingkat akurasi yang lebih tinggi. Dengan demikian, Regresi Linier Berganda dapat dianggap sebagai algoritma yang paling efektif dalam memperkirakan jumlah pengunjung berdasarkan variabel yang digunakan dalam penelitian ini. Untuk meningkatkan akurasi prediksi, disarankan mengeksplorasi metode *ensemble learning* atau *hybrid modeling* serta menambahkan variabel seperti tren pariwisata, ekonomi dan cuaca. Penggunaan dataset yang lebih luas juga dapat membantu model menangkap pola lebih kompleks. Selain itu, hasil penelitian ini dapat diterapkan dalam manajemen Suncity Waterpark Sidoarjo guna mendukung strategi pemasaran, alokasi sumber daya, dan peningkatan layanan berdasarkan prediksi jumlah pengunjung.

Daftar Referensi

- [1] E. A. Marlina, S. Nurdiani dan A. Khairunnisa, "Strategi Pengembangan Pariwisata Berbasis 4A Di Desa Haranggaol, Kabupaten Simalungun," *Jurnal Pariwisata Tawangmangu*, vol. 1, no. 3, pp. 18-24, 2025.
- [2] Agnesia. B. Oktaviani dan E. Yuliani, "Dampak Pengembangan Pariwisata Terhadap Kondisi Ekonomi Masyarakat," *Jurnal Kajian Ruang*, vol. 3, no. 1, pp. 1-17, 2023
- [3] Indra. M. Hasibuan, S. Mutthaqin, R. erianto dan I. Harahap, "Kontribusi Sektor Pariwisata Terhadap Perekonomi Nasional," *Jurnal Ekonomi dan Perbankan Syariah*, vol. 8, no. 2, pp. 1177-1277, 2023.
- [4] S. Lailiyah, A. yusnita dan L. Hariri, "Prediksi Persediaan Bahan Baku Untuk Produksi Makanan Olahan "Sanggar Krispi" Menggunakan Metode Regresi Linear berganda," *Jurnal Sistem Informasi dan Sistem Komputer*, vol. 8, no. 2, pp. 84-94, 2022.
- [5] E. D. Sri Mulyani, A. M. Sahrin, D. R. Pratama, D. R. Puspita, E. S. Pamungkas, L. F. Rahman dan M. Ali, "Estimasi Pertumbuhan Penduduk Di Kabupaten Tasikmalaya Menggunakan Metode Regresi Linear Berganda," *Infosys (Information System) Journal*, vol. 6, no. 1, pp. 1-11, 2021.
- [6] H. D. Kesuma, D. Apriadi, H. Juliansa dan E. Etriyanti, "Implementasi Data Mining Prediksi Mahasiswa Baru Menggunakan Algoritma Regresi Linear Berganda," *Jurnal Ilmiah Binary STMIK Bina Nusantara Jaya*, vol. 4, no. 2, pp. 62-66, 2022.
- [7] M. Nanja dan H. Dalai, "Prediksi Jumlah Kunjungan Wisatawan Asing Menggunakan Algoritma *Linier Regresi Multivariat*," *Jurnal Teknik Informatika dan Sistem Informasi*, vol. 8, no. 4, pp. 1844-1853, 2021.
- [8] B. L. Karamoy, V. P. Rantung, S. Kumajas dan Y. F. Ratumbuisang, "Data Mining Estimasi Pendapatan Tempat Wisata Di Kota Tomohon Menggunakan Algoritma Regresi Linier Berganda," *Jurnal Teknik Informatika*, vol. 5, no. 4, pp. 234-240, 2024.
- [9] T. Hermawan, Y. A. Mustofa dan K. C. Pelangi, "Sistem Prediksi Jumlah Wisatawan Di Lambangan Pauno Desa Kendek Menggunakan Metode *Support Vector Machine*," *Jurnal Banthayo Lo Komputer*, vol. 3, no. 1, pp. 9-14, 2024.
- [10] Binsar. A. Manurung, A. Gea, A. P. Silalahi dan N. Samosir, "Penerapan Algoritma Regresi Linear Untuk Memprediksi Jumlah Wisatawan," *Jurnal Ilmiah Sistem Informasi*, vol. 4, no. 1, pp. 1-9, 2024.
- [11] H. D. Kesuma, R. Yanto, alfiarini dan Ahmadi, "Penerapan Data Mining Prediksi Jumlah Wisatawan Menggunakan Algoritma Regresi Linear Berganda," *Jurnal Penelitian Ilmu dan Teknologi Komputer*, vol. 16, no. 1, pp. 227-237, 2024.
- [12] D. Tampubolon, D. Saripurna dan A. , "Implementasi Regresi Linier Berganda Untuk Memprediksi," *Jurnal CyberTech*, vol. 3, no. 1, pp. 176-185, 2020.

- [13] A. Anggara, K. Auliasari dan Y. A. Pranoto, "Metode Regresi Linier Berganda Untuk Prediksi Omset Penyewaan Kamera Di Joe Kamera," *JATI (Jurnal Mahasiswa Teknik Informatika)*, vol. 7, no. 1, pp. 852-858, 2023.
- [14] M. Irsyad, S. Amelia dan Y. M. Ardi, "Prediksi Harga Ponsel Berdasarkan Spesifikasinya Menggunakan Algoritma *Linear Regression*," *Inti Nusa Mandiri*, vol. 19, no. 2, pp. 251-258, 2025.
- [15] A. M. Sarah, B. Kurniadi dan E. Warsini, "Implementasi Metode Regresi Linear Dalam Memprediksi Penyakit Anemia Secara Dini," *Jurnal Teknologi Komputer dan Sistem Informasi*, vol. 3, no. 1, pp. 14-23, 2023.
- [16] B. Pradito dan D. S. Purnia, "Komparasi Algoritma *Linear Regression* dan *Neural Network* untuk Memprediksi Nilai Kurs Mata Uang," *Evolusi: Jurnal Sains dan Manajemen*, vol. 10, no. 2, pp. 64–71, 2022.
- [17] N. Utami, K. A. Baihaqi, E. E. Awal, dan D. Waiddin, "Analisis Kinerja Algoritma Decision Tree dan Random Forest dalam Klasifikasi Penyakit Kardiovaskular," *Building of Informatics, Technology and Science (BITS)*, vol. 6, no. 2, pp. 970–980, 2024.
- [18] D. Anggraini dan A. Setiawan, "Evaluasi Performa Algoritma *K-Nearest Neighbor* dan *Naive Bayes* dalam Prediksi Harga," *Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi)*, vol. 3, no. 2, pp. 283–289, 2019.