

Analisis Perbandingan Algoritma Untuk Prediksi Harga Tiket Pesawat

Arina Sholiha ¹⁾, Hanief Fathurrahman²⁾, Nurazmi Aprilia ³⁾ Muhammad Al-Ghifari ⁴⁾,
Muhammad Nur Jufri T ⁵⁾

^{1,2} Program Studi Sistem Informasi, Sekolah Tinggi Teknologi Terpadu Nurul Fikri.

³ Program Studi Informatika, Fakultas Ilmu Komputer, Universitas Amikom Yogyakarta.

⁴ Program Studi Informatika, Fakultas Teknik, Universitas Malikussaleh.

⁵ Program Studi Akuntansi Terapan, Fakultas Ekonomi dan Bisnis, Universitas Negeri Makassar.

Abstrak

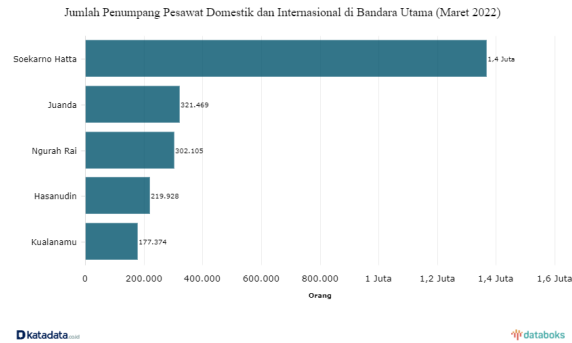
Meningkatnya mobilitas masyarakat yang berdampak pada sektor transportasi khususnya pesawat terbang yang akhirnya meninggi pula statistik penumpang pesawat terbang dapat mempengaruhi harga tiket pesawat. Tujuan dari penelitian ini adalah untuk membandingkan beberapa model algoritma yang dapat memprediksi harga tiket pesawat dengan harapan kedepannya dapat membantu pelanggan yang ingin melakukan pembelian tiket pesawat. Dataset yang digunakan bersumber dari *opendata* (kaggle) yang dilakukan analisis menggunakan 4 algoritma yaitu Regresi Linear, Regresi KNN, Regresi Multi Layer Perceptron, dan Random Forest Regression. Hasil dari penelitian ini didapatkan kesimpulan bahwa data fitur-fitur pendukung dengan fitur harga tiket pesawat dapat diprediksi paling baik menggunakan model algoritma Random Forest Regression dengan nilai koefisien determinasi sebesar 0.9849 (98.49%). Yang kemudian algoritma ini juga diperkuat dengan hasil nilai evaluasi prediksi RMSE sebesar 2783.2705 dan MAPE sebesar 0.0704 yang mana nilai ini merupakan nilai terkecil dibandingkan dengan model algoritma lainnya yang telah diujikan.

Kata kunci: *Perbandingan Algoritma, Prediksi Harga Tiket Pesawat, Random Forest Regression*

I. PENDAHULUAN

Semakin pulihnya wabah virus Covid-19 yang didukung oleh masifnya sosialisasi program vaksinasi yang terjadi di seluruh dunia termasuk Indonesia membuat pemerintah mulai melonggarkan aturan terkait pembatasan pergerakan masyarakat. Hal ini secara signifikan membuat mobilitas masyarakat kembali meningkat dan mempengaruhi salah satu sektor industri transportasi seperti angkutan udara yang sebelumnya memiliki kurva menurun.

Berdasarkan data yang dilansir oleh Badan Pusat Statistik (Maret, 2022), persentase peningkatan penumpang pesawat domestik dan internasional naik hingga 47,8%. Peningkatan jumlah penumpang ini terjadi di semua bandara utama di Indonesia, yaitu Bandara Ngurah Rai-Denpasar 57,3%, Bandara Soekarno Hatta-Tangerang 50,47%, Bandara Juanda-Surabaya 46%, Bandara Hasanuddin-Makassar 40,23%, dan Bandara Kualanamu-Medan sebesar 28,95%.



Gambar 1. Data Penumpang Pesawat Domestik dan Internasional pada Maret 2022

Pertumbuhan penumpang tersebut juga dipengaruhi oleh banyak faktor diantaranya adalah pendapatan pengguna, layanan, kepercayaan dan keamanan. Tidak hanya itu destinasi tujuan tidak kalah penting dalam menentukan peningkatan penggunaan layanan jasa penerbangan.

Meningkatnya statistik penumpang pesawat terbang juga dapat dipengaruhi oleh harga tiket pesawat. Hal ini didukung oleh penelitian (Rahmadanni, M., et al. 2021) yang membahas mengenai pengaruh harga, promosi, dan kualitas pelayanan terhadap keputusan pembelian tiket pesawat secara online di situs Traveloka.com. Berdasarkan hasil pengujian hipotesis dalam penelitian tersebut menunjukkan bahwa variabel harga berpengaruh positif signifikan terhadap keputusan pembelian tiket pesawat. Dalam transaksi bisnis salah satu nya di industri transportasi terdapat beberapa faktor yang mempengaruhi penentuan harga dari suatu produk barang maupun jasa tak terkecuali terkait harga tiket pesawat. Beberapa hal yang berkemungkinan mempengaruhi penentuan harga tiket pesawat di antaranya ialah durasi, maskapai, kota asal dan tujuan, jenis kelas penerbangan, waktu keberangkatan dan kedatangan, serta jarak waktu dalam penjualan dengan tanggal keberangkatan pesawat.

Oleh karena itu, peneliti ingin mencoba beberapa algoritma dalam memprediksi harga tiket pesawat yang harapannya model ini dapat dikembangkan dan diterapkan ke depannya untuk membantu pelanggan yang ingin melakukan pembelian tiket pesawat.

Ada beberapa penelitian terkait dengan penelitian ini di antaranya yaitu penelitian yang dilakukan oleh Evita Fitri dan Dwiza Riana yang melakukan analisis perbandingan model prediction dalam memprediksi harga saham menggunakan metode Linear Regression, Random Forest Regression, dan Multilayer Perceptron. Dalam penelitian tersebut didapatkan hasil bahwa model prediksi Linear Regression mampu memprediksi data historis harga saham lebih baik dengan nilai prediksi kesalahan yang paling rendah yaitu RMSE sebesar 0.010 dan MAPE sebesar 1.2% dan disimpulkan bahwa model Linear Regression merupakan model yang paling baik dibandingkan dengan model Random Forest Regression maupun Multi-Layer Perceptron (Fitri, E., et al. 2022).

Pada penelitian yang dilakukan oleh Yusuf Supriyanto, M. Ilhamsyah, dan Ultach Enri yaitu memprediksi harga minyak kelapa sawit menggunakan Linear Regression dan Random Forest didapatkan kesimpulan bahwa performa algoritma Linear Regression lebih baik dibandingkan Random Forest dengan nilai RMSE yang paling rendah setelah dilakukan tiga skenario (Supriyanto, Y., et al. 2022)

Pada penelitian yang dilakukan oleh Syakirah Fachid dan Agung yang membahas perbandingan algoritma Regresi linear dan Regresi Random Forest dalam memprediksi kasus positif Covid-19 didapatkan hasil bahwa algoritma regresi random forest lebih baik digunakan dengan nilai RMSE yang didapatkan sebesar 1886.555, MAPE 14.85

yang mana nilai ini terkecil dibanding algoritma regresi linear yang kemudian juga didapatkan akurasi sebesar 97.7% yang lebih besar dibandingkan akurasi algoritma regresi linear (Fachid, S., et al. 2022).

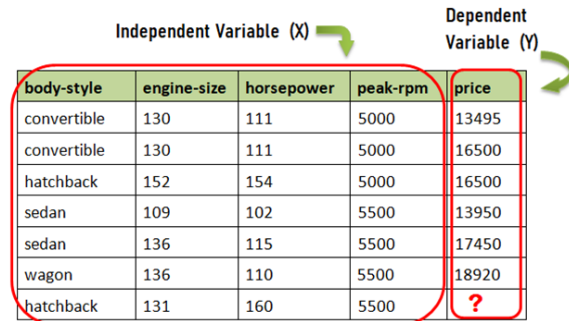
Pada penelitian lainnya yang dilakukan oleh Dapit Virdaus dan Putri Taqwa Prasetyaningrum yang memprediksi harga bawang merah di Yogyakarta menggunakan metode K-Nearest Neighbor didapatkan hasil dan kesimpulan bahwa algoritma ini berhasil diimplementasikan dengan baik dengan hasil akurasi tertinggi yang didapatkan yaitu sebesar 91.67% pada pemodelan data di kondisi uji coba ke 2 dari empat kondisi yang diterapkan (Virdaus, D., et al. 2020).

Berdasarkan hal di atas maka pada penelitian ini, peneliti berfokus mengambil empat algoritma terkait prediksi atau peramalan untuk mengetahui algoritma mana yang terbaik, memiliki tingkat akurasi tertinggi serta memiliki nilai evaluasi RMSE dan MAPE terendah dalam memprediksi harga tiket pesawat. Empat model algoritma ini di antaranya ialah: Linear Regression, K-Nearest Neighbor, Multi-Layer Perceptron, dan Random Forest Regression. Dengan melakukan pengujian antara fitur dengan harga tiket.

II. LANDASAN TEORI

A) Algoritma Prediksi : Regression

Regresi adalah metode statistic yang dipakai memperkirakan hubungan antara sebuah variabel terikat dan satu variabel independen atau lebih. Metode ini juga bisa digunakan untuk menilai kekuatan hubungan antara variabel dengan perkiraan masa depan.



body-style	engine-size	horsepower	peak-rpm	price
convertible	130	111	5000	13495
convertible	130	111	5000	16500
hatchback	152	154	5000	16500
sedan	109	102	5500	13950
sedan	136	115	5500	17450
wagon	136	110	5500	18920
hatchback	131	160	5500	?

Gambar 2. Regresi Linear

Analisis regresi termasuk beberapa variasi, yakni linear, linear majemuk, dan nonlinear. Model yang paling umum adalah linear dan linear majemuk. Sementara itu, nonlinear biasa dipakai untuk kelompok data yang lebih kompleks karena hubungan antar variabel tidak sejalan.

1. Linear Regression

Linear Regression merupakan metode dengan teknik peramalan kuantitatif statistik yang pada umumnya menggunakan data historis yang menitikberatkan pada pola, perubahan pola, dan faktor gangguan yang disebabkan oleh pengaruh acak. (Indarwati et al., 2018)

Rumus untuk Linear Regression dengan metode kuadrat terkecil adalah :

$$a = \frac{(\sum y)(\sum x^2) - (\sum x)(\sum yx)}{n(\sum x^2) - (\sum x)^2}$$

$$b = \frac{n(\sum xy) - (\sum x)(\sum y)}{n(\sum x^2) - (\sum x)^2}$$

$$y = a + b.x_1 + b.x_2$$

Keterangan:

- y : Variabel dependent
- x : Variabel independent
- a : Konstanta
- b : Koefisien regresi

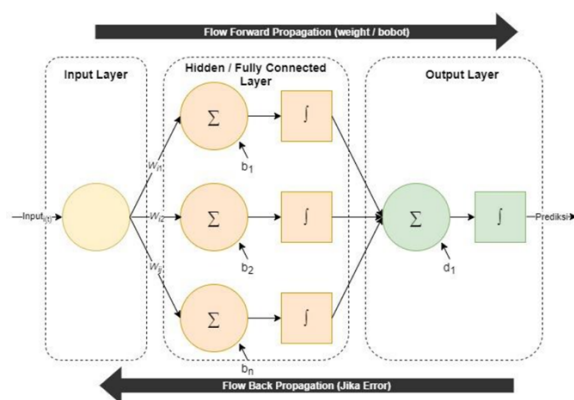
2. K-NN Regression

K-Nearest Neighbor (K-NN) adalah suatu metode yang menggunakan algoritma supervised dimana hasil dari sampel uji yang baru diklasifikasikan berdasarkan mayoritas dari kategori pada K-NN. Tujuan dari algoritma ini adalah mengklasifikasi objek baru berdasarkan atribut dan sampel latih.

Pengklasifikasian tidak menggunakan model apapun untuk dicocokkan dan hanya berdasarkan pada memori. Diberikan titik uji, akan ditemukan sejumlah K objek (titik training) yang paling dekat dengan titik uji. Klasifikasi menggunakan voting terbanyak di antara klasifikasi dari K objek. Algoritma K-NN menggunakan klasifikasi ketetanggaan sebagai nilai prediksi dari sampel uji yang baru. Dekat atau jauhnya tetangga biasanya dihitung berdasarkan jarak Euclidean. (Yahya & Hidayanti, 2020).

3. MLP (Multi-Layer Perceptron) Regression

Multilayer merupakan arsitektur dengan bentuk penggabungan beberapa layer dan neuron tiap layer dalam layer perceptron.



Gambar 3. Multilayer Perceptron (MLP)

Multilayer Perceptron (MLP) merupakan arsitektur yang banyak digunakan untuk neural network. Karena target nilai real

dalam set pelatihan diketahui. Pada Gambar ditunjukkan proses kerja utama dari mulai input dari suatu waktu ($i(t)$) ke hidden layer dan sampai output layer. Kegiatan itu dinamakan feed forward. Sedangkan proses backtracking yakni melakukan alur kembali dari output menuju input, menyesuaikan bobot (w) neural network atau jaringan saraf dengan tujuan mengurangi kesalahan. Proses feed propagation dari input ke output dan backpropagation diulang beberapa kali sampai kesalahan mencapai nilai di bawah ambang batas yang telah ditentukan. (Suradiradja, 2021).

4. Random Forest Regression

Metode random forest (RF) merupakan metode yang dapat meningkatkan hasil akurasi, karena dalam membangkitkan simpul anak untuk setiap node dilakukan secara acak. Metode ini digunakan untuk membangun pohon keputusan yang terdiri dari root node, internal node, dan leaf node dengan mengambil atribut dan data secara acak sesuai ketentuan yang diberlakukan. Root node merupakan simpul yang terletak paling atas, atau biasa disebut sebagai akar dari pohon keputusan. Internal node adalah simpul percabangan, dimana node ini mempunyai output minimal dua dan hanya ada satu input. Sedangkan leaf node atau terminal node merupakan simpul terakhir yang hanya memiliki satu input dan tidak mempunyai output. Pohon keputusan dimulai dengan cara menghitung nilai entropy sebagai penentu tingkat ketidakmurnian atribut dan nilai information gain. Untuk menghitung nilai entropy digunakan rumus seperti pada persamaan 1, sedangkan nilai information gain menggunakan persamaan 2.

Rumus 1:

$$Entropy(Y) = \sum_i p(c|Y) \log_2 p(c|Y) \quad (1)$$

Di mana Y adalah himpunan kasus dan $p(c|Y)$ merupakan proporsi nilai Y terhadap kelas c.

Rumus 2:

$$Information\ Gain(Y, a) = Entropy(Y) - \sum_{v \in Values} \frac{|Y_v|}{|Y|} Entropy(Y_v) \quad (2)$$

Di mana *Values* (a) merupakan semua nilai yang mungkin dalam himpunan kasus a. Y_v adalah subkelas dari Y dengan kelas v yang berhubungan dengan kelas a. Y_a adalah semua nilai yang sesuai dengan a.

B) Model Evaluasi

1. MAPE

Mean Absolute Percentage Error (MAPE) adalah nilai rata-rata perbedaan absolut yang ada diantara nilai dari prediksi dan nilai realisasi yang disebutkan sebagai hasil persenan dari nilai realisasi. Penggunaan Mean Absolute Percentage Error (MAPE) pada evaluasi dari hasil peramalan dapat melihat tingkat akurasi terhadap angka peramalan dan angka realisasi. (Putro et al., 2018)

Nilai Mean Absolute Percentage Error (MAPE) dapat dihitung dengan menggunakan persamaan berikut

$$MAPE = \frac{\sum |x_t - y_t|}{\sum x_t} \times 100\%$$

Keterangan:

x_t = nilai aktual
 y_t = nilai prediksi
 n = jumlah data

Berdasarkan Lewis (1982), nilai MAPE dapat dijabarkan ke dalam empat kategori, yaitu:

MAPE	Keterangan
< 10 %	Sangat Baik
10 - 20 %	Baik
20 - 50 %	Wajar
> 50 %	Buruk

Semakin kecil nilai MAPE yang dihasilkan maka semakin kecil pula kesalahan (*error*) hasil prediksi atau peramalan sehingga hasil suatu metode memiliki kemampuan prediksi yang sangat baik apabila nilai MAPE < 10 % dan begitu pula seterusnya.

2. RMSE

RMSE atau Root Mean Square Error digunakan dalam analisis perbandingan data lapangan dengan data hasil proses yang diolah untuk mengetahui tingkat kesalahan yang terjadi.

Adapun cara menghitungnya adalah dengan cara mengurangi nilai aktual dengan nilai prediksi lalu dikuadratkan dan dibagi sejumlah banyaknya data.

Atau bisa menggunakan rumus persamaan berikut

$$RMSE = \sqrt{\frac{\sum_{t=1}^n (A_t - F_t)^2}{n}}$$

Keterangan:

A = nilai data aktual
 F = nilai hasil prediksi
 t = urutan data
 n = banyaknya data

3. Koefisien Determinasi

Menurut Kuncoro (2013), koefisien Determinasi (R^2) pada intinya mengukur seberapa jauh kemampuan model dalam menerangkan variasi variabel terikat. Formula untuk menghitung koefisien determinasi adalah:

$$R^2 = (TSS - SSE) / TSS = SSR / TSS$$

Nilai Koefisien determinasi adalah di antara nol dan satu. Nilai R^2 yang kecil berarti kemampuan variabel-variabel independen dalam menjelaskan variasi variabel dependen amat terbatas. Uji determinan (R^2) pada intinya mengukur seberapa jauh kemampuan model dalam menerangkan variasi variabel dependen.

Menurut Ghozali (2016:95) koefisien determinasi (R^2) pada intinya mengukur seberapa jauh kemampuan model dalam menerangkan variasi variabel terikat. Nilai adjusted R^2 yang kecil atau mendekati nol berarti kemampuan variabel bebas dalam menjelaskan variabel-variabel terikat sangat terbatas. Semakin tinggi nilai adjusted R^2 maka semakin tinggi variabel bebas dapat menjelaskan variasi variabel terikat.

III. METODOLOGI PENELITIAN

A) Dataset

Dalam penelitian kali ini peneliti menggunakan dataset yang bersumber dari *opendata* (kaggle) yang dikumpulkan oleh SHUBHAM BATHWAL berupa file CSV dengan banyak data sebesar 300261 baris dengan 11 kolom yang meliputi; *airline*, *flight*, *source_city*, *departure_time*, *stops*, *arrival_time*, *destination_city*, *class*, *duration*, *days_left*, dan *price*. 300261 data tersebut sudah bersih sehingga akan dibagi menjadi

80:20 di mana data training sebesar 80% dan data testing sebesar 20%.

B) Algoritma Prediksi

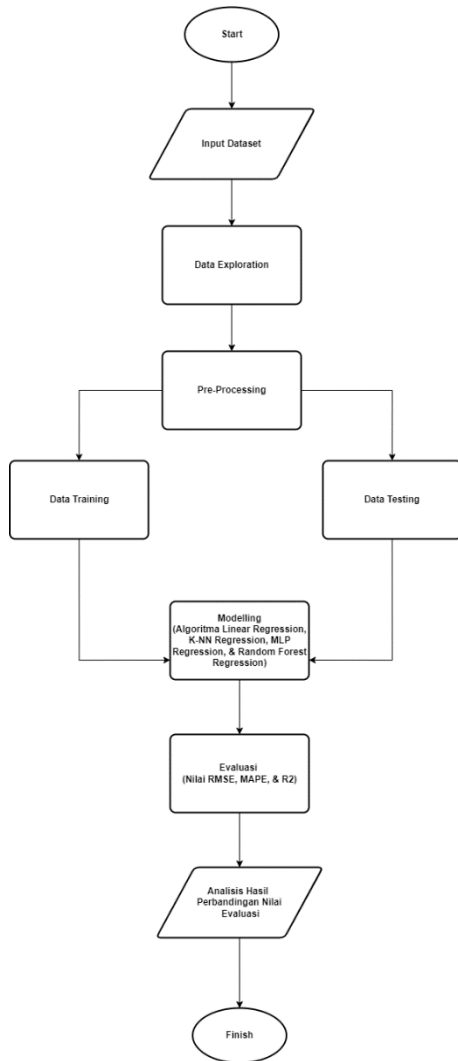
Dalam penelitian ini, peneliti menggunakan beberapa algoritma sebagai perbandingannya yaitu : Linear Regression, K-Nearest Neighbor (K-NN) Regression, Multilayer Perceptron (MLP) Regression dan Random Forest Regression. Algoritma-algoritma ini akan digunakan untuk memprediksi harga tiket pesawat berdasarkan nilai MAPE, RMSE dan Koefisien Determinasi.

C) Model Evaluasi

Dalam penelitian ini untuk mengukur keakuratan keseluruhan dari tiap model peneliti menggunakan koefisien determinasi sebagai metode evaluasi untuk mengukur performa model dalam menerangkan variasi variabel. Selain itu, untuk memperkuat kegiatan analisis perbandingan algoritma peneliti menggunakan model evaluasi Mean Absolute Percentage Error (MAPE) dan RMSE atau Root Mean Square Error yang mana apabila nilai model evaluasi tersebut semakin kecil, maka semakin baik tingkat akurasi prediksinya. Keakuratan dari sebuah model peramalan dalam melakukan prediksi ditentukan oleh nilai terkecil dari masing-masing model evaluasi akurasi data.

IV. HASIL DAN PEMBAHASAN

Secara umum, tahapan perancangan sistem yang dilakukan pada penelitian ini dapat dilihat pada gambar 4 berikut ini.



Gambar 4. Rancangan Sistem

Setelah memasukkan dataset yang telah dikumpul, peneliti terlebih dahulu melakukan eksplorasi data sebelum peneliti melakukan analisis data yang tersedia ke dalam algoritma regresi linear; K-NN regresi; MLP regresi; & random forest regresi. Hal ini dimaksud agar dapat mengetahui apakah data tersebut masih kotor atau tidak. Selain itu eksplorasi data juga dimaksudkan agar mengetahui data *outlier* sehingga dapat diatasi sebelum masuk ke dalam tahapan pre-processing.

A) Data Exploration

Saat melakukan eksplorasi data, didapatkan sebagai berikut:

	duration	days_left	price
count	300153.000000	300153.000000	300153.000000
mean	12.221021	26.004751	20889.660523
std	7.191997	13.561004	22697.767366
min	0.830000	1.000000	1105.000000
25%	6.830000	15.000000	4783.000000
50%	11.250000	26.000000	7425.000000
75%	16.170000	38.000000	42521.000000
max	49.830000	49.000000	123071.000000

Setelah diketahui deskripsi keseluruhan data. Kemudian peneliti melakukan plot antar fitur dengan label. Untuk kasus peneliti label menggunakan data dari harga dan fitur menggunakan data selain data harga.

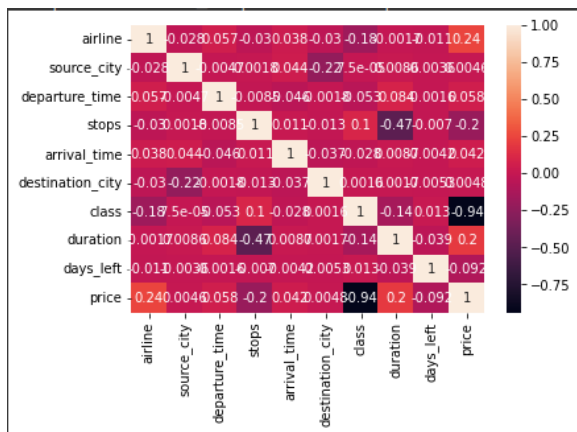
B) Pre-Processing

Pada tahapan ini peneliti membuang data yang tidak dibutuhkan pada kolom Unnamed, karena kolom tersebut tidak memberikan informasi disebabkan tidak memiliki judul yang berhubungan dengan data lainnya.

Kemudian setelah membuang data yang tidak digunakan, langkah berikutnya melakukan transformasi data dengan melakukan label encoding. Hal ini dilakukan agar informasi yang ada dalam table terlihat lebih sederhana. Maksudnya data dalam baris di setiap kolom akan diubah berdasarkan kategori menjadi urutan index. Misal ada 3

kolom dengan 4 baris data yang berbeda, maka data pada baris di kolom tersebut akan diubah berdasarkan kategori kesamaan menjadi nilai index.

Data telah disederhanakan, kemudian melakukan cek korelasi antar data, didapatkan hasil sebagai berikut



Terlihat di atas bahwa tingkat korelasi antar data sangat rendah sekali. Yang artinya data tidak memiliki hubungan satu sama lain. Adapun data yang memiliki hubungan paling tinggi dengan data harga adalah airline di angka 0.24 dan duration di angka 0.2.

Karena peneliti ingin memprediksi data harga menggunakan data-data selain nya maka tahapan selanjutnya menetapkan data fitur dan target. Di mana data target adalah price, sedangkan data fitur adalah data selain data price.

C) Modelling

Pada proses modeling sebelum melakukan training data, peneliti melakukan beberapa tahapan diantaranya :

- PCA proses di mana data akan diringkas secara keseluruhan. Dengan mentransformasi data fitur sehingga menghasilkan data sebagai berikut:

	x	y
0	24.707649	-10.839858
1	24.711322	-10.704047
2	24.699145	-10.868236
3	24.709362	-10.727338
4	24.713930	-10.639059

- Setelah melakukan PCA proses tahapan berikutnya adalah membagi data fitur dan label ke dalam data training dan data testing. Data dibagi menjadi 80:20 dimana data training sebesar 80% dan data testingnya sebesar 20%.

1. Linear Regression

Setelah data dibagi, data dilatih menggunakan model regresi linier guna mengetahui hubungan antar 2 variabel dengan memperkirakan masa depan. Adapun hasil dari test menggunakan model ini didapatkan nilai akurasi sebesar 90.45% dengan nilai MAPE sebesar 0.43 /43%. nilai koefisien determinasi sebesar 0.90 dan nilai RMSE sebesar 7014.30

```
ML model (Linear Regression) Accuracy is 90.4554350718816 %
RMSE Linear Regression = 7014.309680483319
MAPE Linear Regression = 0.4335581089683283
r2_score Linear Regression = 0.904554350718816
```

2. K-NN Regression

Pada model K-NN regresi didapatkan nilai akurasi sebesar 71.45% dengan hasil evaluasi sebagai berikut :

- o Koefisien Determinasi : 0.7145
- o RMSE : 12130.395
- o MAPE : 0.7346 /73.46%

```
ML model (KNN Regression) Accuracy is 71.45459772431661 %
```


RMSE KNN Regression = 12130.395238823565
 MAPE KNN Regression = 0.7346804255154473
 r2_score KNN Regression = 0.7145459772431662

3. MLP Regression

Pada model MLP Regression didapatkan nilai akurasi sebesar 93.75% dengan hasil evaluasi sebagai berikut :

- o Koefisien Determinasi : 0.93
- o RMSE : 5675.0698
- o MAPE : 0.2724

ML model (MLP Regression) Accuracy is 93.75217739794842 %

RMSE MLP Regression = 5675.069867200348
 MAPE MLP Regression = 0.27245158660646834
 r2_score MLP Regression = 0.9375217739794842

4. Random Forest Regression

Pada model Random Forest Regression didapatkan nilai akurasi sebesar 98.49% dengan hasil evaluasi sebagai berikut:

- o Koefisien determinasi : 0.9846
- o RMSE : 2783.27
- o MAPE : 0.0703 / 7.039%

ML model (Random Forest Regressor) Accuracy is 98.49721409419112 %

RMSE Random Forest Regressor = 2783.2705499351273
 MAPE Random Forest Regressor = 0.07039641548186061
 r2_score Random Forest Regressor = 0.9849721409419112

D) Hasil Pengujian Modelling

Hasil evaluasi dari pengujian model dengan algoritma Linear Regression, K-Nearest Neighbor (K-NN) Regression, Multilayer Perceptron (MLP) Regression dan Random Forest Regression dengan membandingkan nilai RMSE, MAPE dan Koefisien Determinasi sebagai berikut :

	Linear regressio n	K-NN	MLP	Random Forest
RMSE	7014.3096 80483319	12130.3 9523882 3565	5675.06 9867200 348	2783.270 5499351 273

MAPE	0.4353581 08968328 3	0.73468 0425515 4473	0.27245 1586606 46834	0.070396 4154818 6061
R2	0.9045543 50718816	0.71454 5977243 1662	0.93752 1773979 4842	0.984972 1409419 112

Berdasarkan dari tabel tersebut bisa disimpulkan bahwa Random Forest Regression merupakan algoritma yang paling baik diantara keempat algoritma lainnya dikarenakan nilai RMSE dan MAPE lebih kecil dan nilai Koefisien Determinasi (R2) nya tertinggi dari keempat algoritma yang diuji.

V. KESIMPULAN

Mobilitas masyarakat yang kembali meningkat dan mempengaruhi sektor transportasi khususnya pesawat. Dengan pertumbuhan yang sangat signifikan dan juga dipengaruhi oleh beberapa faktor seperti pendapatan pengguna, layanan, kepercayaan dan keamanan. Tidak hanya itu destinasi tujuan tidak kalah penting dalam menentukan peningkatan penggunaan layanan jasa penerbangan. Analisa Prediksi sebuah harga tiket pesawat merupakan sebuah hal yang penting dalam melakukan pembelian tiket. Dengan perbandingan ini diharapkan dapat mempermudah serta membantu pelanggan yang ingin melakukan pembelian tiket dengan melakukan pengujian antara fitur dengan harga tiket.

Setelah dilakukan uji coba menggunakan algoritma Linear Regression, K-NN, MLP dan Random Forest. Penelitian ini menghasilkan kesimpulan bahwa Random Forest Regression merupakan algoritma yang paling baik diantara dari keempat algoritma lainnya dengan nilai koefisien determinasi sebesar 0.9849 (98.49%). Yang kemudian algoritma ini juga diperkuat dengan hasil nilai evaluasi prediksi RMSE sebesar 2783.2705 dan MAPE sebesar 0.0704 yang mana nilai ini

merupakan nilai terkecil dibandingkan dengan model algoritma lainnya yang telah diujikan.

Dari penelitian yang telah dilakukan, maka peneliti dapat memberi saran untuk dikembangkannya penelitian ini yaitu dengan mencoba menggunakan algoritma lain untuk melakukan perbandingan sehingga dapat diketahui algoritma yang lebih baik lagi performanya. Selain itu dapat pula dalam pengembangan penelitian selanjutnya dengan melanjutkan penggunaan algoritma *Random Forest Regression* ini untuk diterapkan dalam aplikasi prediksi harga tiket pesawat.

DAFTAR PUSTAKA

- Fachid, S., & Triayudi, A. (2022, Januari). Perbandingan Algoritma Regresi Linier dan Regresi Random Forest Dalam Memprediksi Kasus Positif Covid-19. *JURNAL MEDIA INFORMATIKA BUDIDARMA*, 6(1), 68-73. doi:10.30865/mib.v6i1.3492
- Fitri, E., & Riana, D. (2022, April). ANALISA PERBANDINGAN MODEL PREDICTION DALAM PREDIKSI HARGA SAHAM MENGGUNAKAN METODE LINEAR REGRESSION, RANDOM FOREST REGRESSION AND MULTILAYER PERCEPTRON. *METHOMIKA: Jurnal Manajemen Informatika & Komputerisasi Akuntansi*, 6(1), 69-78. doi:https://doi.org/10.46880/jmika.Vol 6No1.pp69-78
- Ghozali, I. (2016). *Aplikasi Analisis Multivariate*. Semarang: Badan Penerbit Universitas Diponegoro.
- Indarwati, T., Irawati, T., & Rimawati, E. (2018, Oktober). PENGGUNAAN METODE LINEAR REGRESSION UNTUK PREDIKSI PENJUALAN SMARTPHONE. *Jurnal TIKomSiN*, 6(2), 1-6. doi:https://doi.org/10.30646/tikomsin.v6i2.369
- Kuncoro, M. (2013). *Metode Riset Untuk Bisnis dan Ekonomi*. Jakarta: Erlangga.
- Nabillah, I., & Ranggadara, I. (2020, November). Mean Absolute Percentage Error untuk Evaluasi Hasil Prediksi Komoditas Laut. *Journal of Information System*, 5(2), 250-255. doi:10.33633/joins.v5i2.3900
- Rahmadanni, M., Nuringwahyu, S., & Hardati, R. N. (2021, Januari). PENGARUH HARGA, PROMOSI DAN KUALITAS PELAYANAN TERHADAP KEPUTUSAN PEMBELIAN TIKET PESAWAT SECARA ONLINE DI SITUS TRAVELOKA.COM. *JIAGABI*, 102-109.
- Rizaty, M. A. (2022, Mei 10). *Penumpang Pesawat Domestik dan Internasional Naik 47,8% pada Maret 2022*. Diambil kembali dari databoks: https://databoks.katadata.co.id/datapublish/2022/05/10/penumpang-pesawat-domestik-dan-internasional-naik-478-pada-maret-2022
- Schouten, K., Frasincar, F., & Dekker, R. (2016, June). An Information Gain-Driven Feature Study for Aspect-Based Sentiment Analysis. *Natural Language Processing and Information Systems. NLDB 2016. Lecture Notes in Computer Science()*, 9612, 48-59. doi:https://doi.org/10.1007/978-3-319-41754-7_5
- Supriyanto, Y., Ilhamsyah, M., & Enri, U. (2022, Mei). Prediksi Harga Minyak Kelapa Sawit Menggunakan Linear Regression

Dan Random Forest. *Jurnal Ilmiah Wahana Pendidikan*, 8(7), 178-185. doi:10.5281/zenodo.6559603

Suradiradja, K. H. (2021, Desember). Algoritme Machine Learning Multi-Layer Perceptron dan Recurrent Neural Network untuk Prediksi Harga Cabai Merah Besar di Kota Tangerang. *Faktor Exacta*, 14(4), 194-205. doi:10.30998/faktorexacta.v14i4.10376

Virdaus, D., & Prasetyaningrum, P. T. (2020, November 20). Penerapan Data Mining Untuk Memprediksi Harga Bawang

Merah Di Yogyakarta Menggunakan Metode K-Nearest Neighbor. *Journal of Information System and Artificial Intelligence (JISAI)*, 1(1), 1-8.

Yahya, & Hidayanti, W. P. (2020, Juli). Penerapan Algoritma K-Nearest Neighbor Untuk Klasifikasi Efektivitas Penjualan Vape (Rokok Elektrik) pada "Lombok Vape On". *Infotek : Jurnal Informatika dan Teknologi*, 3(2), 104-114.