

**PENERAPAN ALGORITMA *RANDOM FOREST* DENGAN
HYPERPARAMETER TUNING UNTUK MEMPREDIKSI HARGA SEWA
KOST DI KOTA BANDAR LAMPUNG**

(Skripsi)

Oleh

RIZQATASYAA ACHMAD ZAHRA



**JURUSAN MATEMATIKA
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS LAMPUNG
BANDAR LAMPUNG
2023**

ABSTRACT

APPLICATION OF RANDOM FOREST ALGORITHM WITH HYPERPARAMETER TUNING TO PREDICT BOARDING HOUSE RENTAL PRICE IN BANDAR LAMPUNG CITY

By

Rizqatasyaa Achmad Zahra

The large number of immigrants in the city of Bandar Lampung who will continue their studies or want to find work with the status of migrating and have a low economic. Renting a boarding house is an alternative solution for these migrants. However, prospective tenants are often common with problems of boarding house rental prices. How to find out the price will be incurred based on the desired facility. Therefore, information related to predictions of boarding house rental prices is important. There are several methods for predicting boarding house rental prices based on scraping data on mamikos.com, one of which is the Random Forest Regression method. In this study, predictions of boarding house rental prices will be studied by trying several ratios of training data and testing data, as well as increasing accuracy using Hyperparameter Tuning Random Search. The best results were obtained when the ratio was 90% training data and 10% testing data with a prediction accuracy of 79,93%. This accuracy rate increased by 0,75% after tuning the parameters to 80.68%. Based on the level of importance of the variable, it is found that the variable rental time has the greatest influence with a value of 0.726079.

**Keywords : Boarding House Rental Price, Random Forest Regression,
Hyperparameter Tuning, Variable Importance**

ABSTRAK

PENERAPAN ALGORITMA *RANDOM FOREST* DENGAN *HYPERPARAMETER TUNING* UNTUK MEMPREDIKSI HARGA SEWA KOST DI KOTA BANDAR LAMPUNG

Oleh

Rizqatasyaa Achmad Zahra

Banyaknya pendatang di Kota Bandar Lampung yang akan melanjutkan studi ataupun ingin mencari pekerjaan yang berstatus merantau dan memiliki ekonomi rendah. Menyewa kost menjadi solusi alternatif bagi para pendatang tersebut. Namun, calon penyewa sering dihadapkan dengan masalah harga sewa kost. Bagaimana mencari informasi biaya yang dikeluarkan berdasarkan fasilitas yang diinginkan. Oleh karena itu, informasi terkait prediksi harga sewa kost menjadi penting. Ada beberapa metode untuk memprediksi harga sewa kost berdasarkan data hasil *scraping* pada mamikos.com, salah satunya adalah metode *Random Forest Regression*. Pada penelitian ini akan dikaji prediksi harga sewa kost dengan mencoba beberapa rasio pembagian data latih dan data uji, serta meningkatkan akurasi menggunakan *Hyperparameter Tuning Random Search*. Hasil terbaik yang diperoleh yaitu ketika rasio 90% data latih dan 10% data uji dengan akurasi prediksi sebesar 79,93%. Tingkat akurasi ini meningkat sebesar 0,75% setelah dilakukan *tuning* parameter menjadi 80,68%. Berdasarkan tingkat kepentingan variabel, diperoleh bahwa variabel waktu sewa memberikan pengaruh terbesar dengan nilai sebesar 0,726079.

Kata Kunci : Harga Sewa Kost, *Random Forest Regression*, *Hyperparameter Tuning*, Variabel *Importance*

**PENERAPAN ALGORITMA *RANDOM FOREST* DENGAN
HYPERPARAMETER TUNING UNTUK MEMPREDIKSI HARGA SEWA
KOST DI KOTA BANDAR LAMPUNG**

Oleh
RIZQATASYAA ACHMAD ZAHRA
1917031056

Skripsi

Sebagai Salah Satu Syarat untuk Mencapai Gelar
SARJANA MATEMATIKA

Pada

Jurusan Matematika
Fakultas Matematika dan Ilmu Pengetahuan Alam
Universitas Lampung



**FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS LAMPUNG
BANDAR LAMPUNG
2023**

Judul Skripsi : **PENERAPAN ALGORITMA *RANDOM FOREST* DENGAN *HYPERPARAMETER TUNING* UNTUK MEMPREDIKSI HARGA SEWA KOST DI KOTA BANDAR LAMPUNG**

Nama Mahasiswa : **Rizqatasyaa Achmad Zahra**

Nomor Pokok Mahasiswa : **1917031056**

Jurusan : **Matematika**

Fakultas : **Matematika dan Ilmu Pengetahuan Alam**



1. Komisi Pembimbing

Dr. Aang Nuryaman, S.Si., M.Si.
NIP. 197403162005011001

Dr. Ahmad Faisol, S.Si., M.Sc.
NIP. 198002062003121003

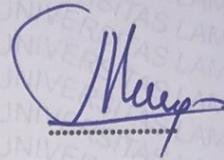
2. Ketua Jurusan Matematika

Dr. Aang Nuryaman, S.Si., M.Si.
NIP. 197403162005011001

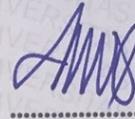
MENGESAHKAN

1. Tim Penguji

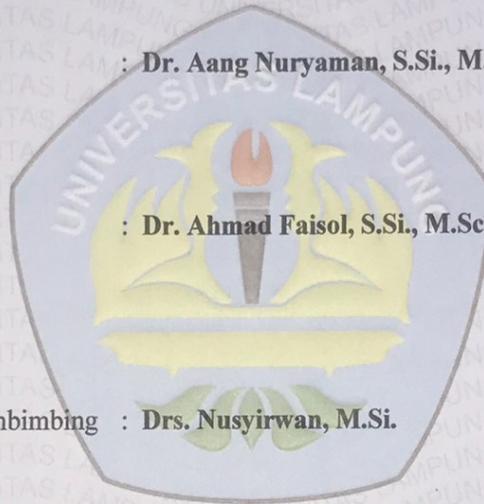
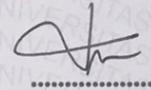
Ketua : **Dr. Aang Nuryaman, S.Si., M.Si.**



Sekretaris : **Dr. Ahmad Faisol, S.Si., M.Sc.**



Penguji
Bukan Pembimbing : **Drs. Nusyirwan, M.Si.**



2. Dekan Fakultas Matematika dan Ilmu Pengetahuan Alam



Dr. Eng. Satripto Dwi Yuwono, S.Si., M.T.
NIP. 197407052000031001

Tanggal Lulus Ujian Skripsi : **8 Februari 2023**

PERNYATAAN SKRIPSI MAHASISWA

Saya yang bertanda tangan di bawah ini:

Nama : **Rizqatasyaa Achmad Zahra**
Nomor Pokok Mahasiswa : **1917031056**
Jurusan : **Matematika**
Judul Skripsi : **PENERAPAN ALGORITMA *RANDOM FOREST* DENGAN *HYPERPARAMETER TUNING* UNTUK MEMPREDIKSI HARGA SEWA KOST DI KOTA BANDAR LAMPUNG**

Dengan ini menyatakan bahwa skripsi ini adalah hasil pekerjaan saya sendiri dan semua tulisan yang tertuang dalam skripsi ini telah mengikuti kaidah karya penulisan ilmiah Universitas Lampung.

Bandar Lampung, 8 Februari 2023

Penulis



Rizqatasyaa Achmad Zahra
NPM. 1917031056

RIWAYAT HIDUP

Penulis Bernama Rizqatasyaa Achmad Zahra, dilahirkan di Kuningan, Jawa Barat pada 17 April 2002. Penulis merupakan anak kedua dari tiga bersaudara, pasangan Bapak Asep Achmad dan Ibu Iyus Rara.

Penulis menempuh pendidikan Sekolah Dasar di MI Darul Hidayah pada tahun 2007-2013. Kemudian melanjutkan pendidikan Sekolah Menengah Pertama di MTsN 1 Kota Tangerang pada tahun 2013-2016 dan melanjutkan pendidikan Sekolah Menengah Atas di MAN 1 Kota Tangerang pada tahun 2016-2019. Pada tahun 2019, penulis terdaftar sebagai mahasiswa S1 Jurusan Matematika Fakultas Matematika dan Ilmu Pengetahuan Alam (FMIPA) Universitas Lampung melalui jalur SBMPTN.

Selama menjadi mahasiswa, penulis aktif dalam organisasi Koperasi Mahasiswa (Kopma) Unila dan Natural FMIPA Unila. Selain aktif dalam organisasi kampus, pada tahun 2021 penulis mengikuti kegiatan Kredensial Mikro Mahasiswa Indonesia (KMMI) di Universitas PGRI Adi Buana Surabaya. Pada tahun 2022 penulis mengikuti kegiatan Magang dan Studi Independen Bersertifikat (MSIB) yaitu Studi Independen *Information Worker* di PT. MariBelajar Indonesia Cerdas pada bulan februari-Juni dan Studi Independen *Introduction to Python for Data Science* pada bulan Agustus-Desember di PT. Hacktivate Teknologi Indonesia.

Penulis juga melaksanakan Kuliah Kerja Nyata (KKN) di Desa Cilangkap, Kabupaten Lebak, Banten pada bulan januari-februari 2022. Dan melaksanakan Kerja Praktik (KP) di Badan Pusat Statistik (BPS) Kota Jakarta Barat pada bulan Juni-Juli 2022.

KATA INSPIRASI

“Dan barang siapa yang bertakwa kepada Allah, niscaya Allah menjadikan baginya kemudahan dalam urusannya.”

(Q.S. At-Talaq : 4)

“Boleh jadi kamu membenci sesuatu, padahal ia amat baik bagimu dan boleh jadi (pula) kamu menyukai sesuatu, padahal ia amat buruk bagimu. Allah maha mengetahui sedangkan kamu tidak mengetahui.”

(Q.S. Al-Baqarah : 216)

“Dan hanya kepada tuhan-mu lah hendaknya kamu berharap.”

(Q.S. Al-Insyirah : 8)

“Janganlah engkau bersedih, sesungguhnya Allah bersama kita.”

(Q.S. At-Taubah : 40)

PERSEMBAHAN

Alhamdulillahirobbil'alamin, dengan mengucapkan rasa syukur atas segala puji dan kehadiran Allah Swt. yang telah melimpahkan nikmat serta hidayah-Nya sehingga skripsi ini dapat diselesaikan dengan baik dan tepat pada waktunya. Saya persembahkan karya ini kepada :

Mama dan Papa

Terima kasih atas segala kasih sayang, motivasi, semangat dan pengorbanan yang telah diberikan. Terima kasih telah memberikan doa dan dukungan yang tiada henti kepada penulis.

Dosen Pembimbing dan Pembahas

Terima kasih telah membantu dan membimbing, serta memberikan arahan dan ilmu yang sangat bermanfaat.

Sahabat-sahabatku

Terima kasih telah memberikan doa dan semangat kepada penulis.

Almamater Tercinta, Universitas Lampung

SANWACANA

Puji dan syukur kehadirat Allah Swt. atas segala rahmat dan karunia-Nya sehingga penulis dapat menyelesaikan Skripsi ini dengan judul “**Penerapan Algoritma *Random Forest* dengan *Hyperparameter Tuning* untuk Memprediksi Harga Sewa Kost di Kota Bandar Lampung**”. Skripsi ini disusun sebagai salah satu syarat memperoleh gelar Sarjana Matematika (S.Mat) pada Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Lampung.

Dalam menyelesaikan Skripsi ini, penulis mendapat dukungan, bimbingan dan bantuan dari beberapa pihak. Oleh karena itu, penulis berterima kasih kepada :

1. Bapak Dr. Aang Nuryaman, S.Si., M.Si., selaku Dosen Pembimbing I dan Ketua Jurusan Matematika yang senantiasa memberikan arahan, bimbingan, bantuan, dan saran yang mendukung sehingga penulis dapat menyelesaikan skripsi ini.
2. Bapak Dr. Ahmad Faisol, S.Si., M.Sc. selaku Dosen Pembimbing II dan Dosen Pembimbing Akademik atas bantuan dan bimbingan kepada penulis selama perkuliahan dan proses penyusunan skripsi ini.
3. Bapak Drs. Nusyirwan, M.Si., selaku Dosen Pembahas yang telah memberikan kritik dan saran yang membangun selama proses penyusunan skripsi.
4. Bapak Dr. Eng. Suropto Dwi Yuwono, S.Si., M.T., selaku Dekan Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Lampung.
5. Seluruh dosen, staf, dan Civitas Akademik Jurusan Matematika Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Lampung.
6. Mama, Papa, Aa Fadhli dan Fathir yang selalu memberikan semangat dan doa kepada penulis.

7. Sahabat Emang_jaya_store, Clara, Vista, dan Nurje, terima kasih telah menemani sejak mahasiswa baru dan memberikan semangat selama perkuliahan.
8. Teman-teman seperjuangan, Dhifa, Rizke, Kamela, Rachma, Dini, Sinta dan Aldiansyah atas kerja sama, bantuan dan dukungan dalam menjalani perkuliahan serta selama proses penyusunan skripsi ini.
9. Seluruh teman Jurusan Matematika 2019, terima kasih atas kebersamaannya.
10. Sahabatku Salsa, terima kasih telah memberikan semangat dan menampung seluruh keluh kesah penulis selama proses penyusunan skripsi ini.
11. Semua pihak yang telah membantu penulis dalam menyelesaikan skripsi ini yang tidak dapat penulis sebut satu persatu.

Semoga kebaikan yang telah diberikan kepada penulis dapat terbalaskan dengan kebaikan. Penulis menyadari penyusunan skripsi ini masih jauh dari kata sempurna dan masih terdapat banyak kekurangan. Oleh sebab itu, saran dan kritikan yang membangun senantiasa penulis harapkan demi menyempurnakan skripsi ini.

Bandar Lampung, 8 Februari 2023

Penulis

Rizqatasyaa Achmad Zahra

NPM. 1917031056

DAFTAR ISI

	Halaman
DAFTAR TABEL	xv
DAFTAR GAMBAR	xvii
I. PENDAHULUAN	1
1.1 Latar Belakang dan Masalah	1
1.2 Tujuan Penelitian.....	5
1.3 Manfaat Penelitian.....	5
II. TINJAUAN PUSTAKA	6
2.1 Kost.....	6
2.2 Data Mining.....	6
2.3 <i>Machine Learning</i>	8
2.4 <i>CART (Classification and Regression Tree)</i>	10
2.5 <i>Bootstrap Aggregating (Bagging)</i>	11
2.6 <i>Random Forest</i>	12
2.7 <i>Hyperparameter Tuning</i>	15
2.8 Evaluasi Model.....	17
2.9 Variabel <i>Importance</i>	19
2.10 <i>Web Scraping</i>	20
2.11 <i>Web Framework Flask</i>	21
III. METODOLOGI PENELITIAN	22
3.1 Waktu dan Tempat Penelitian	22
3.2 Data Penelitian.....	22
3.3 Metode Penelitian	24
IV. HASIL DAN PEMBAHASAN	26
4.1 <i>Preprocessing Data</i>	26

4.1.1	Label <i>Encoder</i>	28
4.2	Analisis Deskriptif.....	29
4.3	Pembagian Data <i>Training</i> dan Data <i>Testing</i>	33
4.4	Analisis <i>Random Forest Regression</i>	34
4.4.1	Pembagian Data 60 <i>Training</i> dan 40 <i>Testing</i>	34
4.4.2	Pembagian Data 70 <i>Training</i> dan 30 <i>Testing</i>	36
4.4.3	Pembagian Data 80 <i>Training</i> dan 20 <i>Testing</i>	38
4.4.4	Pembagian Data 90 <i>Training</i> dan 10 <i>Testing</i>	40
4.5	Simulasi Prediksi <i>Random Forest Regression</i>	42
4.6	<i>Hyperparameter Tuning</i>	50
4.7	Prediksi dan Evaluasi Model	51
4.8	Variabel <i>Importance</i>	57
4.9	Pembuatan <i>Web Application</i> Menggunakan <i>Flask</i>	60
4.9.1	Hasil Tampilan <i>Website</i> Prediksi Harga Sewa Kost.....	63
4.9.2	Prediksi Harga Sewa.....	64
V.	KESIMPULAN	66
	DAFTAR PUSTAKA	67
	LAMPIRAN	

DAFTAR TABEL

Tabel	Halaman
1. Parameter Pada Metode <i>Random Forest</i>	16
2. Variabel Penelitian.....	23
3. Data Sesudah <i>Preprocessing</i>	27
4. Hasil Label <i>Encoder</i>	28
5. Rata-rata Harga Sewa Kost Berdasarkan Waktu Sewa	32
6. Deskriptif Variabel Harga Sewa.....	33
7. Data <i>Training</i> dan Data <i>Testing</i>	33
8. Nilai Aktual dan Nilai Prediksi Rasio 60:40	35
9. Hasil Evaluasi Model RFR dengan Rasio 60:40	35
10. Nilai Aktual dan Nilai Prediksi Rasio 70:30	36
11. Hasil Evaluasi Model RFR dengan Rasio 70:30	37
12. Nilai Aktual dan Nilai Prediksi Rasio 80:20	38
13. Hasil Evaluasi Model RFR dengan Rasio 80:20	39
14. Nilai Aktual dan Nilai Prediksi Rasio 90:10	40
15. Hasil Evaluasi Model RFR dengan Rasio 90:10	40
16. Hasil Percobaan Pembagian Data <i>Training</i> dan <i>Testing</i>	41
17. Data <i>Bootstrap</i> Pohon ke-1.....	42
18. Data <i>Bootstrap</i> dan Variabel <i>Random Selection</i> Pohon ke-1	43
19. Nilai MSE Terkecil dari Ketiga Variabel Terpilih	44
20. Data Percabangan Variabel Waktu Sewa	45
21. Data Percabangan Variabel Waktu Sewa = Bulanan.....	45
22. Nilai MSE Terkecil Percabangan Variabel Waktu Sewa = Bulanan.....	45
23. Percabangan Variabel Waktu Sewa = Bulanan untuk Variabel AC.....	46

24. Percabangan Variabel Waktu Sewa = Bulanan untuk Variabel AC dan Variabel Tipe Kost.....	47
25. Data Percabangan Variabel Waktu Sewa = Tahunan	48
26. Nilai MSE Terkecil Percabangan Variabel Waktu Sewa = Tahunan	48
27. Percabangan Variabel Waktu Sewa = Tahunan untuk Variabel Tipe Kost. 48	
28. Perbandingan Nilai Aktual dan Nilai Prediksi Model <i>Tuning</i>	54
29. Hasil Evaluasi Model RFR dengan <i>Hyperparameter Tuning</i>	55
30. Perbandingan Evaluasi Model Parameter <i>Default</i> dan Model Hasil <i>Hyperparameter Tuning</i>	56
31. Nilai Variabel <i>Importance</i> Data Sampel.....	59

DAFTAR GAMBAR

Gambar	Halaman
1. Tahapan Data <i>Mining</i>	8
2. <i>Training Dataset</i> dan <i>Test Dataset</i>	9
3. Kerangka <i>Artificial Intelligence</i> dan <i>Machine Learning</i>	9
4. <i>Decision Tree</i> Pada CART	11
5. Kontruksi <i>Random Forest Regression</i>	14
6. Ilustrasi Proses <i>Random Search</i>	16
7. Diagram Alir <i>Random Forest</i> dengan <i>Hyperparameter Tuning</i>	25
8. Ubah Nama Variabel Pada <i>Python</i>	26
9. Konversi Tipe Data <i>Category</i>	28
10. Hasil Konversi Variabel <i>Category</i>	29
11. Jumlah Kost Tiap Kecamatan	30
12. Tipe Kost	30
13. Waktu Sewa Berdasarkan Kecamatan	31
14. Model <i>Random Forest Regression</i> dengan Parameter <i>Default</i>	34
15. Hasil Percabangan Variabel Waktu Sewa	44
16. Percabangan Variabel Waktu Sewa = Bulanan untuk Variabel AC	46
17. Percabangan Variabel Waktu Sewa = Bulanan untuk Variabel AC dan Variabel Tipe Kost	47
18. Percabangan Variabel Waktu Sewa = Tahunan untuk Variabel Tipe Kost	49
19. Prediksi <i>Random Forest Regression</i>	49
20. <i>Hyperparameter Tuning</i> Optimal	50
21. Model <i>Random Forest Regression</i> Hasil <i>Hyperparameter Tuning</i>	51
22. Hasil Pohon <i>Random Forest Regression</i>	52

23. Perbandingan Nilai Aktual dan Prediksi.....	53
24. Perbandingan Nilai Aktual dan Prediksi Harga Sewa untuk Data Waktu Sewa = 6 Bulan.....	53
25. Perbandingan Nilai Aktual dan Prediksi Harga Sewa untuk Data Waktu Sewa = Bulanan.....	54
26. Perbandingan Nilai Aktual dan Prediksi Harga Sewa untuk Data Waktu Sewa = Mingguan.....	54
27. Perbandingan Nilai Aktual dan Prediksi Harga Sewa untuk Data Waktu Sewa = Tahunan.....	54
28. Pohon Keputusan Data Sampel <i>Bootstrap</i> ke-1	57
29. Variabel <i>Importance</i>	60
30. <i>Save Model</i> Menggunakan <i>Pickle</i>	61
31. <i>main.html</i>	61
32. <i>result.html</i>	62
33. <i>App.py</i>	63
34. Alamat IP <i>Website</i>	63
35. Tampilan <i>Website</i> Prediksi Harga Sewa Kost	64
36. Hasil Prediksi Harga Sewa Kost.....	65

I. PENDAHULUAN

1.1 Latar Belakang dan Masalah

Kota Bandar Lampung merupakan kota terbesar di Provinsi Lampung, dan merupakan salah satu kota terpadat di Pulau Sumatera dan Luar Jawa. Memiliki luas 197,22 km² yang terdiri dari 20 Kecamatan dan 126 Desa (Kelurahan) (BPS, 2022). Menurut Mustika & Sukarti (2017), banyak pendatang baik pelajar atau mahasiswa yang akan melanjutkan studinya di Kota Bandar Lampung yang berasal dari luar daerah maka akan mencari tempat tinggal seperti rumah kost di daerah tersebut untuk dijadikan tempat tinggal mereka sampai studinya selesai dengan menyewa. Bukan hanya pelajar dan mahasiswa saja, para pekerja yang ingin mencari pekerjaan yang berstatus merantau dan memiliki ekonomi rendah juga akan mencari tempat tinggal sementara seperti kosan untuk menyimpan dan menaruh barang (Widiatmoko & Santoso, 2015). Menyewa kost sebagai solusi alternatif bagi kalangan pelajar, mahasiswa dan pekerja yang sedang merantau dan terkendala terkait ekonomi. Namun, calon penyewa sering dihadapkan dengan masalah harga sewa. Bagaimana mencari informasi biaya yang dikeluarkan berdasarkan fasilitas yang diinginkan (Samhadi, dkk., 2022). Oleh karena itu perlu melakukan prediksi untuk menentukan biaya sewa kost yang akan dikeluarkan, yaitu dengan bantuan analisis data *mining*.

Data *mining* merupakan salah satu teknik analisis yang dapat digunakan untuk memprediksi. Alat analisis yang digunakan dalam data *mining* yaitu *machine learning* yang merupakan cabang ilmu dari *artificial intelligence* (kecerdasan

buatan) yang memungkinkan komputer bertindak cerdas seperti manusia dan membuat keputusan berdasarkan data (Alzubi, *et al.*, 2018). Salah satu teknik pembelajaran yang digunakan adalah *supervised learning* dengan menggunakan *dataset* yang diberi label untuk membuat prediksi (Nurhayati, dkk., 2019). Algoritma yang termasuk kategori *supervised learning* salah satunya adalah *Random Forest*.

Random Forest adalah optimasi dari pohon keputusan yang merupakan metode *ensemble* yang lebih baru dalam *Decision Tree* dan merupakan model *ensemble* yang istimewa sebab memiliki *bagging* di dalamnya (Mambang & Byna, 2017). Algoritma ini dapat menangani data pelatihan dalam jumlah besar dengan baik serta dapat mengasikkan lebih sedikit kesalahan dan menghasilkan hasil yang sangat baik (Agustina, dkk., 2012).

Menurut Azhar, dkk. (2021), masalah yang sering terjadi saat membangun model *Random Forest* yaitu menentukan parameter optimal untuk meningkatkan akurasi dari model yang lebih akurat. Sehingga *hyperparameter tuning* diperlukan untuk memperoleh parameter yang optimal. Menurut Bergstra & Bengio (2012), menggunakan *hyperparameter* berbasis *Random Search* lebih efisien dibanding dengan *Grid Search* dalam meningkatkan performa model *machine learning*, kombinasi parameter yang dihasilkan selama proses optimasi dapat dinyatakan sebagai nilai yang menentukan proses pembelajaran dari model *Random Forest*. Selain itu, *Random Search* hanya membutuhkan waktu yang lebih sedikit dalam pengeksesusiannya (Liashchynskyi & Liashchynskyi, 2019).

Framework merupakan sebuah kerangka kerja atau kumpulan *script* yang dapat digunakan untuk mengembangkan aplikasi untuk menyelesaikan berbagai masalah pemrograman seperti menghubungkan ke *database*, memanggil variabel dan *file*. Memungkinkan pengembang untuk membuat aplikasi lebih cepat (Rosmala, dkk., 2011). Salah satunya *web framework* yang dibuat dengan Bahasa pemograman *Python* yang menyediakan *library flask* untuk pengembangan aplikasi berbasis *website*. Dalam mengaplikasikan model *machine learning* prediksi harga sewa kost

akan dibuat menjadi sebuah aplikasi berbasis *website* sederhana menggunakan *framework flask*.

Untuk mendapatkan informasi mengenai harga sewa kost, beberapa penelitian telah dilakukan terkait dengan sistem informasi pencarian dan pemesanan rumah kost. Diantaranya adalah penelitian yang dilakukan Budiman, dkk. (2019), studi ini mengkaji detail kost seperti lokasi kost, deskripsi kost, pemilik kost, biaya kost, fasilitas kost dan cara reservasi/*booking* tidak perlu bertemu langsung dengan pemilik kost. Peneliti membuat sistem menggunakan bahasa pemrograman PHP, *database MySQL* serta dirancang menggunakan pemrograman berorientasi objek. Penelitian lain dilakukan oleh Yusmaida, dkk. (2020), penelitian ini menggunakan metode *Hill Climbing* yang digunakan untuk mencari kost dengan jarak terdekat. penelitian ini menghasilkan suatu aplikasi berbasis *web*, dalam aplikasi ini terdapat petunjuk arah yang berguna untuk membantu pencari kost dengan jarak terdekat dan menyediakan informasi berkaitan dengan rumah kost bagi para pencari kost maupun pengelola rumah kost untuk mempromosikannya.

Namun disisi lain, penelitian terkait dengan prediksi harga sewa kost menggunakan algoritma *Random Forest* belum banyak dilakukan. Berikut beberapa penelitian terdahulu tentang penerapan algoritma *Random Forest* untuk memprediksi, diantaranya adalah Ardiansyah (2020), mengkaji studi kasus biaya sewa kost di Yogyakarta peneliti menggunakan *Random Forest*. Analisis *Random Forest* yang dilakukan oleh peneliti yaitu dengan membandingkan jumlah pohon yang memiliki nilai *error* terkecil. Prediksi menggunakan N_{tree} sebanyak 20 pohon percabangan, didapat nilai R^2 sebesar 65,91% dengan akurasi prediksi harga sewa kost sebesar 75,62%. Penelitian lain dilakukan oleh Susetianingtias, dkk. (2022), memprediksi penyebaran Covid-19 di Indonesia dengan membagi *dataset* menjadi *training set* dan *validation set* dengan rasio 80:20 dan mendapatkan nilai perhitungan *error* menggunakan MAE dari hasil peramalan yang dilakukan terhadap *validation set* adalah 6.4770 dengan *score* sebesar 99,8% untuk *New Cases* dan 0.2469 dengan *score* 99,3% untuk *New Deaths*, kedua uji coba menunjukkan bahwa hasil peramalan dan data aktual tidak berbeda jauh.

Penelitian serupa mengenai metode *Random Forest* dengan *Hyperparameter Tuning* yang dilakukan oleh Azhar, dkk. (2020), mengkaji prediksi pembatalan reservasi hotel menggunakan optimasi *hyperparameter* yang diaplikasikan pada algoritma *Random Forest*. Enam parameter yang digunakan diantaranya *n_estimators*, *max_features*, *max_depth*, *min_samples_split*, *min_samples_leaf* dan *bootstrap*. Tingkat akurasi yang diperoleh dalam penelitian ini sebesar 87% dari 86% yang artinya optimalisasi hiperparameter dapat meningkatkan performa model *Random Forest* tradisional. Kemudian Amalia, dkk. (2021), melakukan prediksi harga mobil bekas menggunakan algoritma *Linear Regression*, *Random Forest Regression* dan *Gradient Boost Regressor* dengan bantuan *Hyperparameter Tuning*. Didapat model dengan akurasi tinggi pada *Gradient Boost Regression* sebesar 97% setelah *tuning* dengan parameter yang ditambah *max_features* 'sqrt' dan *criteria* 'freadman_mse', sedangkan model *Random Forest Regression* mendapat akurasi sebesar 96% setelah *tuning*, dengan parameter yang ditambah *max_features* 'sqrt' dan *criteria* 'gini'.

Berdasarkan latar belakang yang telah diuraikan, penelitian terkait prediksi harga sewa kost di Kota Bandar Lampung belum pernah dilakukan peneliti lain. Dengan demikian, peneliti ingin melakukan prediksi harga sewa kost di Kota Bandar Lampung dengan variabel yang akan dilibatkan adalah harga sewa kost sebagai variabel dependen, sedangkan variabel bebasnya adalah kecamatan kost, tipe kost, waktu sewa kost dan fasilitas kostan. Peneliti menerapkan algoritma *Random Forest* dengan *Hyperparameter Tuning* dan diaplikasikan ke dalam *Web Framework Flask*.

1.2 Tujuan Penelitian

Adapun tujuan dari penelitian ini adalah :

1. Memprediksi harga sewa kost dengan algoritma *Random Forest*.
2. Mengetahui tingkat akurasi yang didapatkan dari hasil *tuning* parameter algoritma *Random Forest*.
3. Mengetahui faktor-faktor yang mempengaruhi harga sewa kost berdasarkan variabel *importance*.

1.3 Manfaat Penelitian

Manfaat yang dapat diperoleh dari penelitian ini adalah :

1. Menambah wawasan dan memperluas ilmu pengetahuan mengenai penerapan *Random Forest*.
2. Mengetahui tingkat akurasi yang didapatkan dari model *Random Forest*.
3. Memberikan gambaran kepada calon penyewa kost mengenai faktor-faktor yang harus diperhatikan dan mempengaruhi harga sewa kost.
4. Membantu calon penyewa kost dalam menentukan harga sewa kost yang sesuai dengan perekonomian dan kemampuannya.

II. TINJAUAN PUSTAKA

2.1 Kost

Kost atau indekos merupakan suatu jasa yang menyediakan kamar atau tempat untuk ditinggali dalam kurun waktu tertentu dengan sejumlah pembayaran (umumnya bulanan). Kata tersebut berasal dari ungkapan Belanda “*in de kost*”. Pengertian “*in de kost*” sebenarnya berarti makan di dalam atau bisa juga diartikan tinggal dan makan di rumah tempat menumpang tinggal (Wikipedia, 2022). Sedangkan menurut Maulana & Ginanjar (2017), kost adalah tempat tinggal sementara yang digunakan oleh beberapa kelompok masyarakat yang sengaja dibangun oleh pemiliknya untuk disewakan kepada beberapa orang dengan sistem pembayaran bulanan atau tahunan. Penyewa biasanya para mahasiswa yang datang dari luar daerah.

2.2 Data Mining

Data *Mining* adalah proses menemukan pola tersembunyi (*hidden patern*) dalam suatu bentuk informasi yang sebelumnya tidak diketahui dalam kelompok data. Data tersebut berasal dari *database*, data *werehouse*, atau media penyimpanan data lainnya. Beberapa Hal penting tentang data *mining* adalah :

1. Data *mining* adalah proses otomatis terhadap data yang ada
2. Data yang akan diolah adalah data berukuran besar

3. Tujuan dari data *mining* adalah untuk mendapatkan hubungan atau pola. Ini dapat memberikan petunjuk yang bermanfaat (Kusrini & Luthfi, 2009).

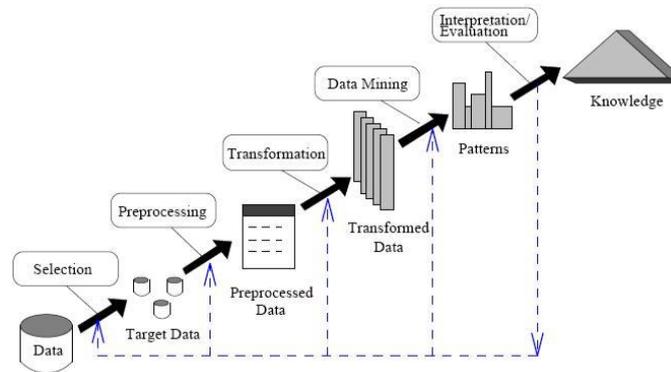
Berikut adalah pengelompokan data *mining* yang dapat dilakukan didasarkan pada tugas-tugas berikut (Mardi, 2016) :

1. *Description* (Deskripsi)
2. *Estimation* (Estimasi)
3. *Prediction* (Prediksi)
4. *Classification* (Klasifikasi)
5. *Clustering* (Pengklasteran)
6. *Association* (Asosiasi)

Data *Mining* adalah bagian dari proses *Knowledge Discovery in Database* (KDD) yang berguna untuk mengekstraksi pola atau model dari data dengan algoritma tertentu. Proses KDD adalah sebagai berikut (Agushinta & Irfan, 2008) :

1. *Data Selection*, pemilihan data dari *database* operasional harus dilakukan sebelum memulai tahap menggali informasi data di KDD. Data hasil seleksi yang digunakan dalam proses data *mining*, disimpan dalam suatu *file* terpisah dari data operasional.
2. *Preprocessing*, sebelum memulai proses data *mining* perlu melakukan data *cleaning* untuk membuang duplikasi data, memeriksa data yang tidak konsisten dan memperbaiki kesalahan pada data. proses *enrichment* juga dilakukan yaitu proses memperkaya data yang sudah ada dengan data lain yang berkaitan dan diperlukan untuk KDD seperti data eksternal.
3. *Transformation*, pengkodean adalah proses konversi data yang telah dipilih, sehingga data tersebut sesuai untuk proses data *mining*. Proses pengkodean dalam KDD adalah proses kreatif dan sangat bergantung pada jenis atau pola informasi yang akan dicari pada *database*.
4. *Data Mining*, proses menemukan pola atau informasi yang menarik dalam data terpilih dengan menggunakan teknik atau metode tertentu. Memilih metode atau algoritma yang tepat akan sangat bergantung pada tujuan dan proses KDD secara keseluruhan.

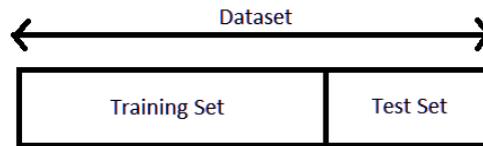
5. *Interpretation/Evaluasi*, pola informasi dari data yang dihasilkan dari proses data *mining* harus mudah dipahami oleh pihak yang berkepentingan pada saat ditampilkan. Tahap ini melakukan pemeriksaan apakah pola yang ditemukan tidak konsisten dengan fakta atau hipotesis yang sudah ada sebelumnya.



Gambar 1. Tahapan Data *Mining*

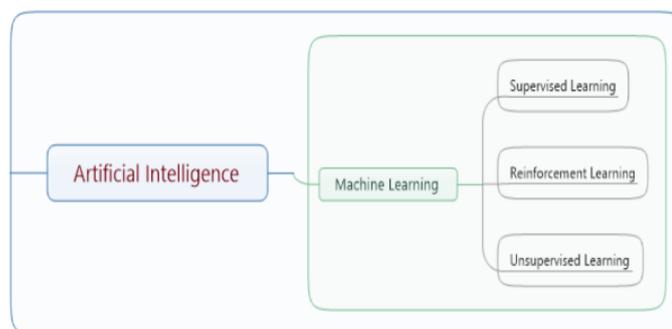
2.3 *Machine Learning*

Menurut Daqiqil (2021), *machine learning* adalah bagian dari kecerdasan buatan, hal ini memungkinkan komputer untuk mempelajari data baru tanpa pemrograman eksplisit. *Machine learning* adalah ilmu yang mempelajari tentang algoritma komputer yang dapat mengenali pola data untuk mengubah berbagai jenis data menjadi tindakan nyata dengan meminimalkan campur tangan manusia. Dengan *machine learning* memungkinkan kita untuk membuat komputer yang “belajar” dari data yang ada dan membuat keputusan sendiri tanpa diprogram ulang. Agar algoritma *machine learning* dapat melakukan proses belajar, diperlukan dua langkah yaitu penyusunan data latih (*training*) dan data uji (*testing*). Untuk jumlah data *testing* lebih sedikit dibanding data *training*, umumnya sekitar 20% dari total seluruh *dataset* (Kurniawan, 2022). Susetianingtias, dkk. (2022), melakukan percobaan pembagian data *training* dan *testing* dengan menggunakan rasio 60:40, 70:30, 80:20 dan 90:10. Rasio yang digunakan adalah rasio dengan nilai *error RMSE* terkecil.



Gambar 2. Training *Dataset* dan *Test Dataset*

Machine learning mempunyai tiga tipe yaitu *Supervised Learning*, *Unsupervised Learning* dan *Reinforcement Learning*. Bagan keterkaitan *artificial intelligence* dan *machine learning* dapat dijelaskan pada Gambar 3 (Roihan, dkk., 2019).



Gambar 3. Kerangka *Artificial Intelligence* dan *Machine Learning*

1. *Supervised Learning* adalah teknik *machine learning* yang digunakan Ketika semua data yang dipunya sudah memiliki label dan algoritmanya belajar memprediksi *output* dari *input* (Hiadayat, dkk., 2020). *Supervised learning* diklasifikasikan ke dalam masalah klasifikasi dan regresi. Masalah klasifikasi muncul ketika variabel *output* berbentuk kategori seperti merah atau biru atau penyakit dan tidak ada penyakit. Sedangkan masalah regresi muncul ketika variabel *output* adalah nilai riil seperti dollar atau berat (Brownlee, 2016).
2. *Unsupervised Learning* adalah metode yang digunakan ketika semua data yang dipunya tidak memiliki label dan algoritmanya mempelajari struktur bawaan data tersebut (Hiadayat, dkk., 2020). *Unsupervised learning* diklasifikasikan lebih lanjut ke dalam masalah *clustering* dan asosiasi (Brownlee, 2016).
3. *Reinforcement Learning*, dalam konteks *reinforcement learning* menggabungkan bagian pembelajaran dan pengujian. Secara aktif dalam

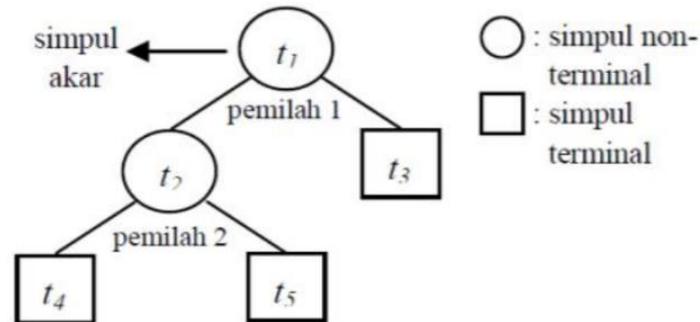
pengumpulan data dari pembelajar dengan berinteraksi ke lingkungan untuk mendapatkan tanggapan untuk setiap aktivitas pembelajar. Dengan demikian, *Reinforcement learning* bergantung pada korespondensi antara agen yang melakukan suatu kegiatan dan lingkungannya yang memberikan reaksi positif atau negatif (Gourav & Kaur, 2020).

2.4 CART (*Classification and Regression Tree*)

CART (*Classification And Regresion Tree*) adalah metode statistik *non-parametric* untuk menganalisis klasifikasi. Pertama kali diperkenalkan oleh empat ilmuwan Amerika serikat : Leo Breiman, Jerome H. Friedman, Richard A. Olshen dan Charles J. Stone pada tahun 1984 (Waluyo, dkk., 2014). CART adalah suatu algoritma dari teknik eksplorasi (pohon keputusan). Menurut Dhawangkhara & Riksakomara (2017), terdapat dua analisis pada CART yaitu *classification tree* dan *regression tree*. Jika variabelnya bersifat kategorikal maka CART menghasilkan pohon klasifikasi (*classification tree*), jika variabel dependennya bertipe kontinu atau numerik menghasilkan pohon regresi (*regression tree*).

CART menggunakan algoritma partisi rekursif biner. Biner artinya setiap kelompok data yang terkumpul dalam suatu ruang yang disebut *node* hanya dapat dibagi menjadi dua kelompok yang disebut *child nodes*. Istilah rekursi merujuk pada proses partisi biner yang dapat dilakukan secara berulang, dimana setiap *child nodes* yang diperoleh dari partisi pertama dapat dipecah ulang menjadi dua *child nodes* dilakukan sampai memenuhi kriteria tertentu. Istilah *partitioning* adalah proses klasifikasi yang dilakukan pada suatu kumpulan data dengan cara mempartisi data menjadi beberapa bagian (Brownlee, 2016). CART lebih unggul dibandingkan metode klasifikasi lainnya karena hasilnya lebih mudah diinterpretasikan, lebih akurat dan lebih cepat dihitung. Selain itu, CART dapat diaplikasikan pada kumpulan data berjumlah besar, sangat banyak dan variabel

bervariasi dengan teknik pemilahan biner (Lewis, 2000). Pembentukan pohon pada CART dapat ditampilkan pada Gambar 4 sebagai berikut.



Gambar 4. *Decision Tree* Pada CART

Decision Tree disebut juga CART (*Classification and Regression Tree*) merupakan kombinasi dari *classification tree* dan *regression tree*. CART memiliki kelemahan, CART bisa menjadi tidak stabil di *Decision Tree* karena sangat sensitif terhadap data baru. CART sangat bergantung pada jumlah sampel, jika sampel data *training* dan *testing* berubah maka hasil pohon keputusan juga berubah (Pratiwi & Zain, 2014). Dengan demikian, metode *ensemble* (gabungan) dikembangkan.

2.5 *Bootstrap Aggregating (Bagging)*

Bootstrap Aggregating (Bagging) ditemukan oleh Breiman pada tahun 1996 (Wu & Kumar, 2009). *Bagging* adalah salah satu teknik dari metode *ensemble* sederhana dan sangat kuat, yang merupakan kombinasi dari beberapa metode yang digunakan dalam model untuk meminimalkan variansi variabel independen untuk meningkatkan prediksi klasifikasi dari pohon klasifikasi tunggal (Jatmiko, dkk., 2019). Wahono & Suryana (2013), mengatakan bahwa ide dasar *Bagging* adalah memisahkan data *training* menjadi beberapa data *training* baru dengan pengambilan sampel acak dan buat model untuk data *training* baru. Hasil klasifikasi akhir diperoleh dengan *voting* masing-masing model. Ini juga membantu mengurangi varians dan menghindari *overfitting*. *Bagging* berguna

untuk meningkatkan pengklasifikasi saat mengamati kumpulan data *training* yang terdapat *noise* (Alfaro, *et al.*, 2013).

2.6 *Random Forest*

Algoritma *Random Forest* merupakan metode *ensemble*, pertama kali diusulkan oleh Breiman pada tahun 2001. Algoritma ini berasal dari kombinasi beberapa *Decision Tree* dengan distribusi yang sama untuk melatih dan memprediksi data sampel (Kuhn & Johnson, 2013). Ada dua jenis algoritma *Random Forest* yaitu regresi dan klasifikasi. Dalam *Random Forest Regression* hasilnya secara numerik atau kontinu, sedangkan hasil klasifikasi dalam kategori. *Random Forest* merupakan pengembangan lebih lanjut dari CART yang mengaplikasikan *Bootstrap Aggregating (Bagging)* dan pemilihan fitur secara acak untuk menghindari *overfitting* pada kumpulan data kecil (Breiman, 2001).

Teknik dasar *Random Forest* adalah *Decision Tree*, dimana setiap *Decision Tree* pelatihan dilakukan dengan sampel individu dan setiap fitur dipecah antara fitur antara himpunan bagian dari fitur acak pada pohon yang dipilih. Pada proses klasifikasi, individu didasarkan pada *vote* dari total suara terbanyak dalam kumpulan populasi *tree* (Saputra, dkk., 2011). Konsep *Random Forest* yaitu terdapat banyak pohon (*tree*) yang akhirnya membentuk sebuah hutan (*forest*).

Random Forest adalah salah satu cara untuk meningkatkan akurasi metode klasifikasi dengan menggabungkannya. Klasifikasi berbasis *ensemble* berfungsi paling baik jika mempunyai korelasi rendah antar *basic learner*. *Ensemble* harus membuat *basic learner* yang lemah, karena kemungkinan *learner* yang kuat memiliki korelasi tinggi dan menyebabkan *overfitting*. *Random Forest* dapat mengurangi korelasi dan mempertahankan kekokohan klasifikasi dengan mengacak proses *training* yaitu memilih sejumlah fitur dari semua fitur yang ada secara acak pada setiap melakukan *training tree*, kemudian fitur yang terpilih digunakan untuk mendapatkan percabangan pohon terbaik (Brownlee, 2016).

Tahapan Algoritma *Random Forest* yang digunakan adalah sebagai berikut :

1. Tahap pertama yaitu Inisialisasi parameter. Kemudian mengambil sampel acak dengan pengembalian dari data *training* (sampel *bootstrap* dari data *training*) sehingga membentuk kumpulan data baru berukuran N (Jatmiko, dkk., 2019).
2. Memilih variabel independen yang diambil secara acak (m_{try}) tanpa pengembalian dari semua variabel (p). Jumlah m_{try} ditentukan melalui perhitungan : Brownlee (2016)
 - Pada klasifikasi, m_{try} default yaitu \sqrt{p} dengan ukuran simpul terkecil 1.
 - Pada regresi, m_{try} default yaitu $\frac{p}{3}$ dengan ukuran simpul terkecil 5.
3. Menentukan *stopping criteria default* yaitu 1, jika didalam *subnode*/simpul anak hanya terdapat 1 sampel maka *subnode* tersebut akan berhenti melakukan *splitting* sehingga menjadi terminal *node/leaf node* dan akan dihasilkan simpul terminal sebagai hasil prediksi satu pohon CART.
4. Langkah membuat pohon pada *Random Forest* yaitu :
 - a. Setelah menentukan sampel berdasarkan hasil *bootstrap* kemudian menentukan variabel *root node* yaitu variabel independen yang terletak paling atas sebagai variabel pemisah. Variabel pemisah dipilih berdasarkan *splitting criteria*. Menurut Sartono & Syafitri (2010), kriteria pemisahan untuk variabel respon numerik, pada kasus pohon regresi antara lain adalah MSE (*Mean Square Error*), untuk pohon klasifikasi menggunakan nilai *entropy/information gain*. Variabel independen dengan nilai MSE terkecil akan menjadi variabel pemisah. Berikut adalah rumus MSE :

$$MSE_n = \frac{1}{N} \sum_{i=1}^N (y_i - \bar{y}_n)^2 \quad (2.1)$$

dimana :

- MSE_n = Nilai MSE pada pohon ke-n
- N = Total sampel pada pohon ke-n
- Y_i = Nilai sampel ke-i pada pohon ke-n
- \bar{y}_n = Nilai rata-rata sampel pohon ke-n

- b. Setelah mendapatkan atribut pemisah dengan nilai MSE terendah maka akan dibuat cabang selanjutnya dengan menentukan variabel pemisah berikutnya yang dinamakan *subnode*/simpul anak.
 - c. Hal tersebut terus dilakukan sampai mencapai *stopping criteria* yaitu minimum sampel per simpul terminal/terminal *node/leaf node* terpenuhi.
5. Mengulangi langkah 1 sampai 3 hingga diperoleh sejumlah pohon yang diinginkan. Sehingga mendapatkan k buah pohon acak.
 6. Menghitung hasil akhir dengan cara menggabungkan atau diaggregasi (*aggregate*). Misal menggunakan *majority vote* untuk kasus klasifikasi atau rata-rata untuk kasus regresi (Sartono & Syafitri, 2010). Secara sistematis nilai rata-rata seluruh prediksi pohon CART pada *Random Forest* sebagai berikut :

$$\hat{Y}_i = \frac{1}{N_{tree}} \sum_{n=1}^{N_{tree}} Y_n \quad (2.2)$$

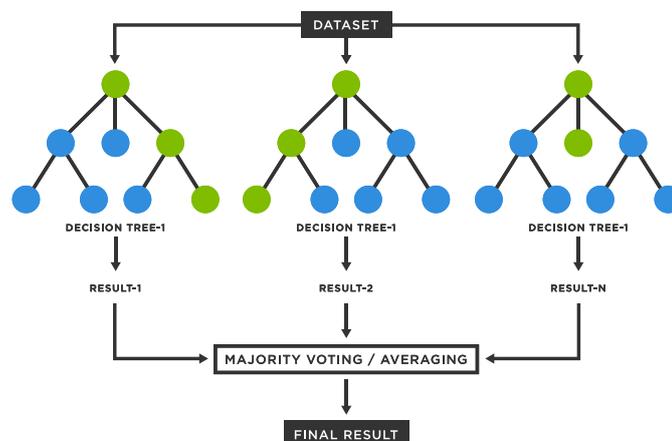
dimana :

\hat{Y}_i = Hasil prediksi akhir

N_{tree} = Total jumlah pohon pada *Random Forest*

Y_n = Hasil prediksi pohon ke-n

7. Menentukan nilai akurasi model *Random Forest* menggunakan nilai prediksi



Gambar 5. Kontruksi *Random Forest Regression*

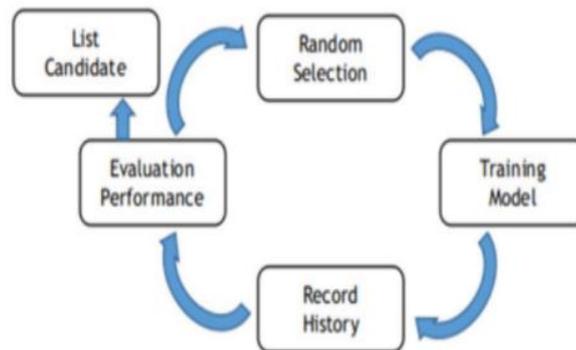
2.7 *Hyperparameter Tuning*

Sebagian besar algoritma pembelajaran memiliki parameter yang dapat disesuaikan, dan setelan parameter yang berbeda menghasilkan model dengan kinerja yang sangat berbeda. Evaluasi dan pemilihan model tidak hanya tentang memilih algoritma pembelajaran tetapi juga mengoptimalkan parameter. Proses menemukan *hyperparameter* yang tepat dikenal dengan *hyperparameter tuning* (Mubarok, dkk., 2022).

Meskipun pendekatan *Random Forest* mampu mengolah data berdimensi besar, untuk menentukan model yang optimal tidak dapat dilakukan dengan begitu saja. Sejauh ini tidak ada formulasi pasti untuk menentukan model *Random Forest* yang optimal dalam mengatasi pemodelan data berskala besar. Sementara itu, ada beberapa parameter model yang perlu ditentukan untuk mendapatkan kinerja model terbaik (Azhar, dkk., 2021). Nugraha & Sasongko (2022), mengatakan bahwa nilai *hyperparameter* tidak dapat ditentukan dari data dan selalu mempertimbangkan yang diberikan saat mendefinisikan model dengan kata lain nilai *hyperparameter* harus ditentukan dahulu sebelum menjalankan proses pembelajaran. *Hyperparameter* mempengaruhi hasil dari sebuah model. Kombinasi parameter yang dihasilkan pada proses optimalisasi dapat mewakili nilai-nilai dalam proses pelatihan model *Random Forest*. Untuk tiap metodenya kemungkinan jumlah *hyperparameter* dapat berbeda antara satu dengan lainnya. Semakin banyak parameter yang dicoba maka waktu pelatihan model semakin meningkat.

Peneliti ini menggunakan algoritma *Random Search* untuk optimasi model. algoritma tersebut memiliki kemampuan lebih baik daripada algoritma *Grid Search* untuk menentukan kombinasi parameter (Bergstra & Bengio, 2012). Algoritma *Random Search* dapat digambarkan sebagai salah satu algoritma yang menggunakan metode pencarian acak atau probabilitas (biasanya dalam bentuk *pseudo random number generator*). *Random Search* juga merupakan bagian dari optimasi *stochastik* dan optimasi global. Dengan demikian, metode ini merupakan metode pencarian langsung yang tidak membutuhkan turunan untuk menentukan

domain kontinu. Dapat dilihat pada Gambar 6 ilustrasi proses pencarian *hyperparameter* dengan *Random Search*.



Gambar 6. Ilustrasi Proses *Random Search*

Dalam praktiknya, *scikit-learn library* (sebuah *module* dari bahasa pemrograman *Python*) pada *python* menyediakan metode *RandomizedSearchCV* yang dapat diprovokasi dengan membuat rentang untuk setiap subjek pengoptimalan *hyperparameter*. Dengan menggunakan metode *RandomizedSearchCV* pada rentang yang telah ditentukan, *Random Search* dilakukan secara acak memilih kandidat *grid* kemungkinan dalam rentang tersebut, kemudian teknik *K-fold cross-validation* diterapkan pada *grid* yang dibuat. Menurut Azhar, dkk. (2021), Parameter model *Random Forest* difenisikan dan simpan dalam *grid dictionary*. Berikut adalah parameter yang disimpan pada *grid dictionary* :

Tabel 1. Parameter Pada Metode *Random Forest*

Parameter	Random Search Value	Keterangan
<i>n_estimators</i>	100, 300, 500, 700, 900, 1100	Jumlah pohon
<i>min_samples_split</i>	2, 5, 10	Jumlah sampel minimum yang dibutuhkan untuk <i>split node internal</i>
<i>min_samples_leaf</i>	1, 2, 4	Jumlah sampel minimum yang dibutuhkan disetiap simpul <i>leaf</i>
<i>max_features</i>	auto, sqrt	Jumlah fitur yang akan digunakan saat mencari <i>split</i> terbaik
<i>max_depth</i>	10, 20, 30, 40, 50, 60, 70, 80, 90, 100, 110	Kedalaman maksimum pada <i>tree</i>
<i>bootstrap</i>	<i>True, False</i>	Apakah sampel <i>bootstrap</i> saat membuat <i>tree</i> . Jika <i>false</i> , seluruh kumpulan data digunakan untuk membangun setiap <i>tree</i>

Parameter lain yang digunakan seperti 100 buah n_iter , *10-folds cross-validation* dan *squared_error criterion* digunakan untuk evaluasi kinerja model dengan 10 kali perulangan saat proses *Random Search* untuk setiap parameter.

2.8 Evaluasi Model

Performa model *machine learning* perlu dievaluasi. Untuk melakukan itu, *output* yang dihasilkan oleh model dibandingkan dengan nilai sebenarnya. Dengan menggunakan *Root Mean Square Error* (RMSE), Koefisien Determinasi (R^2) dan *Mean Absolute Percent Error* (MAPE).

1. *Root Mean Square Error* (RMSE) merupakan angka kesalahan pada hasil prediksi. Nilai RMSE yang rendah atau nilainya mendekati angka 0 menunjukkan bahwa hasil prediksi sesuai dengan data aktual. RMSE juga dapat berupa angka negatif, semakin kecil hasil RMSE semakin dekat pula prediksinya (Fachid & Triayudi, 2022). RMSE dapat dirumuskan seperti berikut :

$$RMSE = \sqrt{\frac{\sum_{t=1}^n (y_t - \hat{y}_t)^2}{n}} \quad (2.3)$$

dimana :

- n = Banyaknya data observasi
- \hat{y}_t = Nilai prediksi dari y
- y_t = Nilai aktual atau data asli

2. Koefisien Determinasi (R^2) adalah mengukur proporsi keragaman variabel dependen yang mampu dijelaskan oleh variabel independen dalam model. R^2 menunjukkan kebaikan model, semakin besar R^2 semakin baik modelnya. Rentang R^2 adalah antara 0 sampai 1. Semakin dekat R^2 ke nilai 1 maka

regressor yang terbentuk dapat memprediksi variabel Y dengan baik (Siagian & Sugiarto, 2000). R^2 dapat dirumuskan seperti berikut :

$$R^2 = \frac{\sum_{t=1}^n (y_t - \hat{y}_t)^2}{\sum_{t=1}^n (y_t - \bar{y})^2} \quad (2.4)$$

dimana :

- n = Banyaknya data observasi
- \hat{y}_t = Nilai prediksi dari y
- y_t = Nilai aktual atau data asli
- \bar{y} = Nilai rata-rata y

3. *Mean Absolute Percentage Error* (MAPE) adalah perhitungan statistika untuk menunjukkan keakuratan model prediksi. MAPE adalah rata-rata selisih persentase antara data aktual dan hasil prediksi (Fachid & Triayudi, 2022).

MAPE dapat dirumuskan seperti berikut :

$$MAPE = \frac{1}{n} \left(\sum_{t=1}^n \frac{|y_t - \hat{y}_t|}{y_t} \right) \times 100\% \quad (2.5)$$

dimana :

- n = Banyaknya data observasi
- \hat{y}_t = Nilai prediksi dari y
- y_t = Nilai aktual atau data asli

Pada penelitian ini, peneliti menggunakan evaluasi model berdasarkan nilai akurasi yang diperoleh dari persamaan sebagai berikut :

$$Akurasi = 100\% - MAPE \quad (2.6)$$

Semakin besar nilai akurasi maka menyatakan semakin akurat hasil prediksi pada model.

2.9 Variabel *Importance*

Variabel *Importance* merupakan ukuran yang mencerminkan pentingnya suatu variabel prediktor prediksi. Yang bekerja berdasarkan seberapa besar suatu variabel dapat mengurangi keakuratan model atau kesalahan model sebelum dan sesudah dilakukan permutasi (Mercadier & Lardy, 2019). Variabel *importance* hanya menghitung kepentingan variabel prediktor dalam mempengaruhi variabel responnya. Dalam *Random Forest* biasanya menggunakan dua ukuran penting yaitu *Gini Importance Index* dan *Permutation Importance Index* (PIM) (Hasan, *et al.*, 2016).

Variabel *importance* dihitung menggunakan *Gini Importance* yang diasumsikan hanya dua simpul anak (pohon biner) :

$$ni_j = w_j C_j - w_{left(j)} C_{left(j)} - w_{right(j)} C_{right(j)} \quad (2.7)$$

dimana :

- ni_j = *Importance* dari simpul j
- w_j = Jumlah sampel yang mencapai simpul j
- C_j = Nilai *impurity* dari simpul j
- $left(j)$ = Anak simpul dari *split* kiri pada simpul j
- $right(j)$ = Anak simpul dari *split* kanan pada simpul j

Ukuran kepentingan untuk setiap *feature* pada pohon keputusan dihitung sebagai berikut :

$$fi_i = \frac{\sum_{j: \text{node } j \text{ splits on feature } i} ni_j}{\sum_{k \in \text{all nodes}} ni_k} \quad (2.8)$$

dimana :

- fi_i = *Importance feature* i
- ni_j = *Importance* simpul j

Kemudian nilai dapat dinormalisasi menjadi antara 0 dan 1 yang dapat dihitung menggunakan rumus berikut :

$$normfi_i = \frac{fi_i}{\sum_{j \in all\ features} fi_j} \quad (2.9)$$

Dalam *Random Forest* ukuran kepentingan dihitung rata-rata, membagi dengan jumlah total pohon :

$$RFfi_i = \frac{\sum_{j \in all\ trees} normfi_j}{T} \quad (2.10)$$

dimana :

$RFfi_i$ = Nilai ukuran kepentingan i dari semua pohon dalam model *Random Forest*

$normfi_j$ = Normalisasi nilai ukuran kepentingan untuk i pada pohon j

T = Jumlah pohon

2.10 Web Scraping

Web Scraping adalah proses mengambil data dari internet, biasanya data tersebut diambil dari halaman *web* yang menggunakan bahasa *markup* seperti HTML atau XHTML dan dokumen tersebut dianalisis dan diambil data tertentu yang diperlukan untuk tujuan tertentu (Turland, 2010). *Web scraping* dapat melakukan semua jenis pencarian yang tidak dapat dilakukan oleh mesin pencari tradisional seperti *Google Search* (Mitchell, 2015).

Adapun Langkah-langkah melakukan *Web scraping* (Josi, dkk., 2014) :

1. *Create Scraping Template*, pemogram memeriksa dokumen HTML dari situs *web* yang akan diambil datanya untuk *tag* HTML berisi data yang akan diambil.
2. *Explore Site Navigation*, pemogram mengeksplorasi metode navigasi pada situs *web* yang akan digunakan datanya untuk direplikasi pada aplikasi *web scraper* yang akan dibangun

3. *Automate Navigation and Extraction*, berdasarkan data yang diperoleh pada langkah 1 dan 2, dibuat aplikasi *web scraper* untuk mengotomatisasi pengambilan data.
4. *Extracted Data and Package History*, data yang diperoleh pada langkah 3, kemudian disimpan dalam bentuk tabel *database*.

2.11 *Web Framework Flask*

Web framework merupakan sebuah perangkat lunak yang dirancang untuk mendukung pengembangan aplikasi *web*, *web service* dan *web* dinamis. Sedangkan *framework* adalah sekumpulan fungsi dan perintah inti yang disusun dalam berbagai aturan dan dapat dihubungkan bersama, sehingga ketika mengembangkan aplikasi dalam bentuk *website* harus mengikuti prosedur atau aturan yang ada dalam *framework* (Wardana, 2010).

Flask merupakan *web framework* yang ditulis dalam bahasa pemrograman *Python* yang diklasifikasikan sebagai jenis *microframework*. Fungsi *flask* sebagai kerangka kerja untuk aplikasi dan tampilan *web*. *Flask* memiliki dua *external libraries* yaitu *WSGI toolkit* dan *Jinja2 template engine*. *Flask* tidak memerlukan *library* tertentu dalam penggunaannya. *Flask* dapat menggunakan ekstensi untuk menambahkan fitur dan komponen yang sudah disediakan oleh pihak ketiga dan tidak terpasang secara standar pada *Flask* seperti *Form Validation*, *Upload Handling* dan *Database* (Aslam, *et al.*, 2015).

Flask dibagi menjadi dua bagian yaitu *static file* yang berisi semua kode yang diperlukan *website* seperti kode *CSS*, *javascript* dan *file* gambar. Dan *template file* yang berisi semua *template Jinja* termasuk halaman *HTML*. Saat mengembangkan aplikasi *framework* ini, *virtual environment* diperlukan untuk menampung *library* yang akan digunakan (Ningtyas & Setiyawati, 2021).

III. METODOLOGI PENELITIAN

3.1 Waktu dan Tempat Penelitian

Penelitian ini dilaksanakan pada semester ganjil tahun ajaran 2022/2023 bertempat di Jurusan Matematika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Lampung.

3.2 Data Penelitian

Data yang digunakan dalam penelitian ini merupakan data sekunder yaitu data kost di Kota Bandar Lampung yang diperoleh dari hasil *scraping* pada *website* mamikos.com yang merupakan *platform e-commerce* pencari kost nomor 1 di Indonesia, pengambilan data dilakukan pada bulan Oktober tahun 2022.

Data kost yang diperoleh dari hasil *scraping* sebanyak 920 kost di Kota Bandar Lampung dengan variabel yang digunakan dalam penelitian ini yaitu Harga Sewa, Kecamatan, Tipe Kost, AC, WiFi, Kamar Mandi Dalam, Jangka Waktu Sewa dan Akses 24 Jam. Dalam penelitian ini terdapat data kategorik yang harus diubah ke dalam bentuk skala pengukuran nominal terlebih dahulu. Adapun penjelasan setiap variabel yang digunakan ditampilkan pada tabel 2 sebagai berikut :

Tabel 2. Variabel Penelitian

No.	Variabel	Definisi Variabel	Kategorisasi
1	Harga Sewa	Harga sewa kost (Rp) yang wajib dibayarkan oleh konsumen (penyewa kost) pada pemilik kost secara rutin.	-
2	Kecamatan	Pembagian wilayah kost yang akan dihuni berdasarkan tingkat kecamatan.	0 = Bumi Waras 1 = Enggal 2 = Kedamaian 3 = Kedaton 4 = Kemiling : 20 = Way Halim
3	Tipe Kost	Variabel yang menunjukkan kost memiliki beberapa jenis yang membedakan penggunaan kamar kost.	0 = Campur 1 = Putra 2 = Putri
4	AC	Fasilitas penyejuk udara yang disediakan oleh pemilik kost.	0 = AC 1 = Tidak ada AC
5	WiFi	Fasilitas jaringan internet yang disediakan oleh pemilik kost.	0 = Tidak Ada WiFi 1 = WiFi
6	K. Mandi Dalam	Fasilitas kamar mandi yang ditawarkan oleh pemilik kost terdapat dua macam yaitu kamar mandi dalam bersifat pribadi dan kamar mandi luar bersifat umum bersama penghuni lain.	0 = K. Mandi Dalam 1 = K. Mandi Luar
7	Jangka Waktu Sewa	Lamanya penyewaan kamar kost.oleh konsumen (penyewa kost).	0 = 6 Bulan 1 = Bulanan 2 = Mingguan 3 = Tahunan
8	Akses 24 Jam	Fasilitas yang mengizinkan penghuni kost dapat keluar dan masuk kost dalam 24 jam.	0 = Akses 24 Jam 1 = Tidak Ada Akses

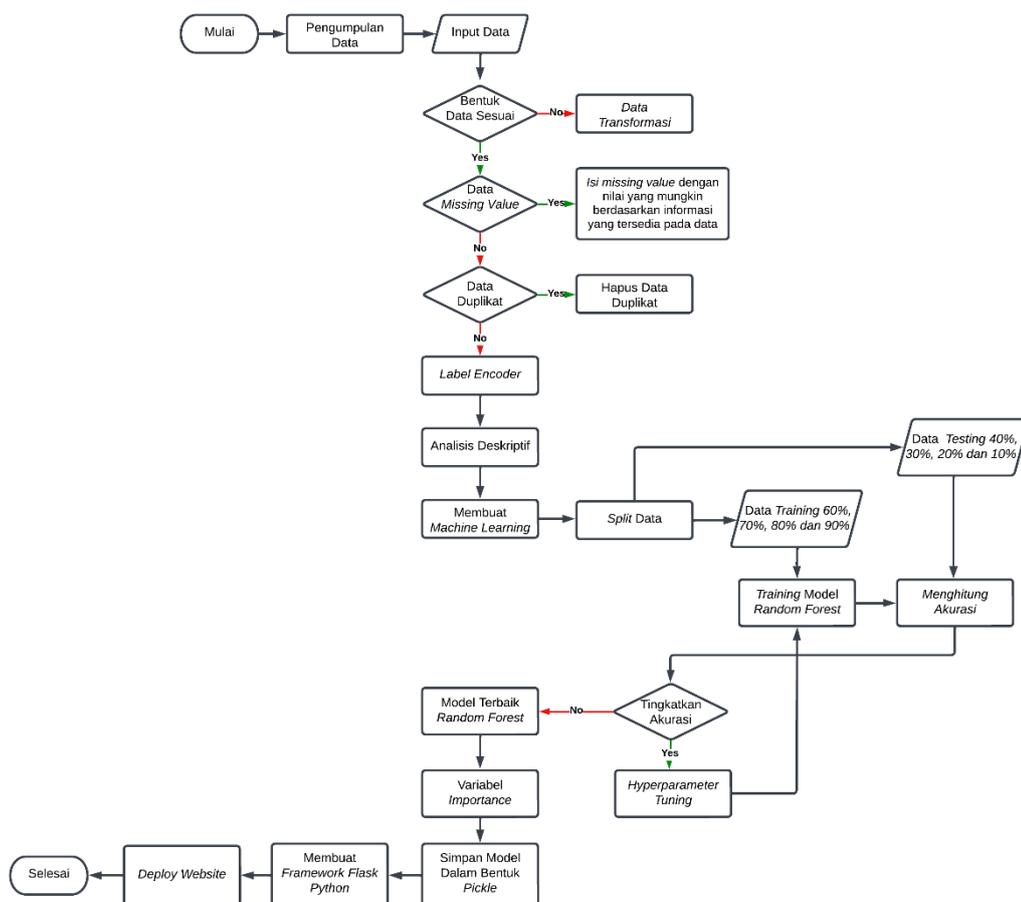
3.3 Metode Penelitian

Dalam penelitian ini akan menerapkan metode *Random Forest Regression* dengan *Hyperparameter Tuning* dengan bantuan *software python* untuk membuat model yang dapat memprediksi harga sewa kost di Kota Bandar Lampung. Model yang telah diperoleh kemudian diaplikasikan ke dalam aplikasi berbasis *website* menggunakan *web framework flask*. Berikut langkah-langkah penelitian yang dilakukan :

1. Melakukan *input* data kost di Kota Bandar Lampung yang diperoleh dari hasil *scraping* pada *website* mamikos.com ke dalam *python*.
2. Melakukan *preprocessing* data, yaitu :
 - a. Data Transformasi, melakukan transformasi bentuk yaitu menghapus kata “Rp” dan tanda titik pada nominal harga dengan menggunakan *excel*.
 - b. Data *Cleaning*, membersihkan data yang tidak diperlukan seperti tanda baca. Memastikan data tidak memiliki data hilang (*missing value*), untuk menangani *missing value* dapat dilakukan dengan mengisi nilai yang mungkin berdasarkan informasi yang tersedia pada data dan menghapus duplikasi data.
 - c. *Handling Data Categorical*, memberikan label pada data yang berbentuk kategorik dengan menggunakan *label encoder*.
3. Melakukan analisis deskriptif dan visualisasi untuk melihat gambaran umum mengenai harga sewa kost di Kota Bandar Lampung.
4. Membagi data menjadi data *training* dan *testing*. Pada penelitian ini, peneliti mencoba beberapa rasio yaitu 60:40, 70:30, 80:20 dan 90:10.
5. Melakukan analisis *Random Forest Regression* :
 - a. Menentukan banyaknya m_{try} variabel yaitu $\frac{p}{3}$ merupakan *default* dalam *Random Forest Regression* pada *python*.
 - b. Menentukan jumlah pohon (N_{Tree}) yaitu 100 yang merupakan nilai *default*.
 - c. Melatih data *training* dan melakukan evaluasi pada data *testing* dengan menghitung nilai RMSE, R^2 dan MAPE.

- d. Meningkatkan akurasi model dengan cara *tuning* parameter.
 - e. Membuat plot ukuran kepentingan (variabel *importance*).
 - f. Menyimpan model terbaik dalam bentuk *pickle*, yang digunakan untuk menyimpan dan membaca data dari *file* berformat `.pkl` yang nantinya dapat dibaca ke dalam *web application* saat proses *deployment*
6. Setelah mendapatkan model terbaik, selanjutnya membuat desain tampilan *website* prediksi harga sewa kost di Kota Bandar Lampung dan mengekspor *machine learning* dalam bentuk *pickle* dengan menggunakan *framework flask*.
 7. Setelah program berhasil, selanjutnya melakukan *hosting website* agar dapat diakses oleh semua orang.

Secara singkat proses penerapan algoritma *Random Forest* dengan *Hyperparameter Tuning* digambarkan dalam diagram alir berikut :



Gambar 7. Diagram Alir *Random Forest* dengan *Hyperparameter Tuning*

V. KESIMPULAN

Berdasarkan hasil dan pembahasan yang telah diuraikan pada Bab IV, didapatkan kesimpulan sebagai berikut :

1. Harga sewa kost yang diprediksi dengan 7 variabel independen yaitu Kecamatan, Tipe Kost, AC, WiFi, K. Mandi Dalam, Waktu Sewa dan Akses 24 Jam. Berdasarkan hasil analisis *Random Forest Regression* menggunakan rasio pembagian data 90:10 dengan parameter *default* merupakan hasil terbaik dibanding rasio lainnya. Kemudian model dilakukan *Hyperparameter Tuning* dengan parameter yang digunakan yaitu jumlah pohon yang digunakan sebanyak 700 pohon, *max_depth* sebanyak 10 percabangan, *min_samples_leaf* sebanyak empat, *min_samples_split* sebanyak 10, *max_features* "auto" dan pembuatan pohon menggunakan sampel *bootstrap* mampu meningkatkan akurasi model sebesar 0,75%
2. Tingkat akurasi hasil prediksi harga sewa kost menggunakan *Random Forest Regression* setelah dilakukan *Hyperparameter Tuning* diukur dengan nilai RMSE, R^2 , dan MAPE dari data *testing*. Menghasilkan nilai RMSE sebesar 971099, nilai R^2 sebesar 0,92 artinya variabel independen mempengaruhi variabel dependen harga sewa, nilai MAPE sebesar 19,32% dengan tingkat akurasi sebesar 80,68%.
3. Faktor yang paling mempengaruhi harga sewa kost berdasarkan variabel *importance* yaitu Waktu Sewa sebesar 0,726079, sedangkan yang memiliki pengaruh paling kecil yaitu Kamar Mandi sebesar 0,002046.
4. *Website* prediksi harga sewa kost yang dibuat menggunakan *framework flask* dan *dideploy* dengan bantuan *PythonAnywhere* dapat diakses melalui <http://prediksisewakostbdl.pythonanywhere.com/> .

DAFTAR PUSTAKA

- Agustina, I., Mulyani, Y., Septiana, T., & Mardiana. 2012. Analisis Pengembangan Model Prediksi Kesuksesan Kickstarter Menggunakan Algoritma Backpropagation dan Random Forest. *Jurnal Informatika dan Teknik Elektro Terapan*. **10**(3): 173-182.
- Alfaro, E., Gamez, M., & Garcia, N. 2013. Adabag: An R Package for Classification with Boosting and Bagging. *Journal of Statistical Software*. **54**(2): 1-35.
- Alzubi, J., Nayyar, A., & Kumar, A. 2018. Machine Learning from Theory to Algorithms: An Overview. *Journal of Physics: Conference Series*. **1142**(1): 2-15.
- Amalia., Radhi, M., Sitompul, D.R.H., Sinurat, S.H., & Indra, E. 2021. Prediksi Harga Mobil Menggunakan Algoritma Regresi dengan Hyper-Parameter Tuning. *Jurnal Sistem Informasi dan Ilmu Komputer Prima*. **4**(2): 28-32.
- Ardiansyah, F. 2020. Sistem Prediksi Harga Sewa Kost dengan Menggunakan Random Forest Analytics (Tugas Akhir). Program Studi Statistika FMIPA Universitas Islam Indonesia, Yogyakarta.
- Aslam, F.A., Mohammed, H.N., & Lokhande. 2015. Efficient Way of Web Development Using Python and Flask. *International Journal of Advanced Research in Computer Science*. **6**(2): 54-57.
- Azhar, Y., Mahesa, G.A., & Mustaqim, M.C. 2021. Prediksi Pembatalan Pemesanan Hotel Menggunakan Optimalisasi Hiperparameter pada Algoritme Random Forest. *Jurnal Teknologi dan Sistem Komputer*. **9**(1): 15-21.

- Badan Pusat Statistik Kota Bandar Lampung. 2022. *Kota Bandar Lampung Dalam Angka 2022*. Badan Pusat Statistik Kota Bandar Lampung, Lampung.
- Bergstra, J. & Bengio, Y. 2012. Random Search for Hyper-Parameter Optimization. *Journal of Machine Learning Research*. **13**: 281-305.
- Breiman, L. 2001. Random Forest. *Machine Learning*. **45**: 5-32.
- Brownlee, J. 2016. *Master Machine Learning Algorithms*. Machine Learning Mastery, San Francisco.
- Budiman, A., Wahyuni, L.S., & Bantun, S. 2019. Perancangan Sistem Informasi Pencarian dan Pemesanan Rumah Kos Berbasis Web (Studi Kasus: Kota Bandar Lampung). *Jurnal TEKNOKOMPAK*. **13**(2): 24-30.
- Dhawangkhar, M. & Riksakomara, E. 2017. Prediksi Intensitas Hujan Kota Surabaya dengan Matlab menggunakan Teknik Random Forest dan CART (Studi Kasus Kota Surabaya). *Jurnal Teknik ITS*. **6**(1): 94-99.
- Fachid, S. & Triayudi, A. 2022. Perbandingan Algoritma Regresi Linier dan Regresi Random Forest dalam Memprediksi Kasus Positif Covid-19. *Jurnal Media Informatika Budidarma*. **6**(1): 68-73.
- Gourav, K. & Kaur, A. 2020. A Study of Reinforcement Learning Applications & Its Algorithms. *International Journal of Scientific & Technology Research*. **9**(3): 4223-4228.
- Hasan, M.A.M., Nasser, M., Ahmad, S., & Molla, K.I. 2016. Feature Selection for Intrusion Detection Using Random Forest. *Journal of Information Security*. **7**: 129-140.
- Hiadayat, S.P., U, A.R.I., & Fathona, I.W. 2020. Penentuan Parameter Kinerja Bangunan dengan Metode Inverse Modeling Menggunakan Machine Learning, hlm. 1214-1220. e-Proceeding of Engineering, Bandung.
- Daqiqil, I. 2021. *Machine Learning: Teori, Studi Kasus dan Implementasi Menggunakan Python*. UR PRESS, Riau.

- Jatmiko, Y.A., Padmadisastra, S., & Chadidjah, A. 2019. Analisis Perbandingan Kinerja CART Konvensional, Bagging dan Random Forest pada Klasifikasi Objek: Hasil dari Dua Simulasi. *Media Statistika*. **12**(2): 1-12.
- Josi, A., Abdillah, L.A., & Suryayusra. 2014. Penerapan Teknik Scraping pada Mesin Pencarian Artikel Ilmiah. *Jurnal Sistem Informasi*. **5**: 159-164.
- Kuhn, M. & Johnson, K. 2013. *Applied Predictive Modeling*. Springer, New York.
- Kurniawan, D. 2022. *Pengenalan Machine Learning dengan Python*. Elex Media Komputindo, Jakarta.
- Kusrini & Luthfi, E.T. 2009. *Algoritma Data Mining*. Andi Offset, Yogyakarta.
- Lewis, R.J. 2000. *An Introduction to Classification and Regression Tree (CART) Analysis*. Annual Meeting of the Society for Academic Emergency Medicine in San Fransisco, California.
- Liashchynskyi, P. & Liashchynskyi, P. 2019. Grid Search, Random Search, Genetic Algorithm: A Big Comparison for NAS. *arXiv:1912.06059*.
- Mambang & Byna, A. 2017. Analisis Perbandingan Algoritma C.45, Random Forest dengan Chaid Decision Tree untuk Klasifikasi Tingkat Kecemasan Ibu Hamil. *Seminar Nasional Teknologi Informasi dan Multimedia*. **5**(1): 103-108.
- Mardi, Y. 2016. Data Mining : Klasifikasi Menggunakan Algoritma C4.5. *Jurnal Edik Informatika*. **2**(2): 213-219.
- Maulana, I. & Ginanjar, R. 2017. Sistem Informasi Manajemen Kost Berbasis Web. *Information System Application*. **2**(1): 11-19.
- Mercadier, M. & Lardy, J. 2019. Credit Spread Approximation and Improvement Using Random Forest Regression. *European Journal of Operational Research*. 1-37.

- Mitchell, R. 2015. *Web Scraping with Python*. Springer, New York. 2th Edition. O'reilly Media, Sebastopol.
- Mubarok, M.R., Muliadi., & Herteno, R. 2022. Hyper-parameter Tuning pada XGBOOST untuk Prediksi Keberlangsungan Hidup Pasien Gagal Jantung. *Kumpulan Jurnal Ilmu Komputer*. **9**(2): 391-401.
- Ningtyas, D.F. & Setiyawati, N. 2021. Implementasi Flask Framework pada Pembangunan Aplikasi Purchasing Approval Request. *Jurnal Janitra Informatika dan Sistem Informasi*. **1**(1): 19-34.
- Nugraha, W. & Sasongko, A. 2022. Hyperparameter Tuning pada Algoritma Klasifikasi dengan Grid Search. *Jurnal Sistem Informasi*. **11**(2): 391-401.
- Nurhayati., Busman., & Iswara, R.P. 2019. Pengembangan Algoritma Unsupervised Learning Technique pada Big Data Analysis di Media Sosial Sebagai Media Promosi Online Bagi Masyarakat. *Jurnal Teknik Informatika*. **12**(1): 79-96.
- Mustika, E. & Sukarti. 2017. Sistem Informasi Pencarian Tempat Kos Berbasis Geografis di Bandar Lampung. *Jurnal Cendikia*. **14**(1): 13-18.
- Pratiwi, F.E. & Zain, I. 2014. Klasifikasi Pengangguran Terbuka Menggunakan CART (Classification and Regression Tree) di Provinsi Sulawesi Utara. *Jurnal Sains dan Seni Pomits*. **3**(1): 54-59.
- Agushinta, D. & Irfan, M. 2008. Perancangan Aplikasi Data Mining untuk Memprediksi Permintaan Customer pada Perusahaan Persewaan Mobil, hlm. 206-213. Proceeding Seminar Ilmiah Nasional Komputer dan Sistem Intelijen, Depok.
- Roihan, A., Sunarya, P.A., & Rafika, A.S. 2019. Pemanfaatan Machine Learning dalam Berbagai Bidang: Review Paper. *Indonesian Journal on Computer and Information Technology*. **5**(1): 75-82.
- Rosmala, D., Ichwan, M., & Gandalisha, M.I 2011. Komparasi Framework MVC (Codeigniter, Dan Cakephp) Pada Aplikasi Berbasis Web. *Jurnal Informatika*. **2**(2): 22-30.

- Samhadi, R.M., Basuki, B.M., & Sugiono. 2022. Implementasi Multiple Linear Regression pada Machine Learning untuk Prediksi Harga Kost di Kota Malang. *Science Electro*. **14**(2): 1-7.
- Saputra, W.S.J., Sujatmika, A.R., & Arifin, A.Z. 2011. Seleksi Fitur Menggunakan Random Forest dan Neural Network. *Electronic Engineering Polytechnic Institute of Surabaya*. 93-97.
- Sartono, B. & Syafitri, U.D. 2010. Metode Pohon Gabungan: Solusi Pilihan Untuk Mengatasi Kelemahan Pohon Regresi dan Klasifikasi Tunggal. *Forum Statistika dan Komputasio*. **15**(2): 1-7.
- Siagian, D. & Sugiarto. 2000. *Metode Statistika Untuk Bisnis dan Ekonomi*. Gramedia Pustaka Utama, Jakarta.
- Susetianingtias, D.T., Patriya, E., & Rodiah. 2022. Model Random Forest Regression Untuk Peramalan Penyebaran Covid-19 di Indonesia. *Jurnal Manajemen Sistem Informasi dan Teknologi*. **6**(1): 32-41.
- Turland, M. 2010. *php|architect's Guide to Web Scraping with PHP*. Marco Tabini & Associates, Canada.
- Wahono, R.S. & Suryana, N. 2013. Combining Particle Swarm Optimization based Feature Selection and Bagging Technique for Software Defect Prediction. *International Journal of Software Engineering and Its Applications*. **7**(5): 153-166.
- Waluyo, A., Mukid, M.A., & Wuryandari, T. 2014. Perbandingan Klasifikasi Nasabah Kredit Menggunakan Regresi Logistik Biner dan CART (Classification and Regression Trees). *Media Statistika*. **7**(2): 95-104.
- Wardana. 2010. *Menjadi Master PHP dengan Framework Codeigniter*. Elex Media Komputindo, Jakarta.
- Widiatmoko, R. & Santosa, A. 2015. Perancangan Furniture pada Hunian Kost Pekerja di Kawasan Surabaya Timur. *Jurnal Intra*. **4**(1): 63-72.

Wikipedia. 2022. Indekos. <https://id.wikipedia.org/wiki/Indekos>. Diakses pada 5 November 2022.

Wu, X. & Kumar, V. 2009. *The Top Ten Algorithms in Data Mining*. CRC Press, Boca Raton.

Yusmaida., Neneng., & Ambarwari, A. 2020. Sistem Informasi Pencarian Kos Berbasis Web dengan Menggunakan Metode Hill Climbing. *Jurnal Teknologi dan Sistem Informasi (JTSI)*. **1**(1): 68-74.