

**SISTEM PREDIKSI HARGA SEWA KOST DENGAN
MENGUNAKAN *RANDOM FOREST ANALYTICS*
(Studi Kasus : Kost Eksklusif di Daerah Istimewa Yogyakarta)**

TUGAS AKHIR



Faisal Ardiansyah

16611085

**PROGRAM STUDI STATISTIKA
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS ISLAM INDONESIA
YOGYAKARTA
2020**

HALAMAN PERSETUJUAN PEMBIMBING

TUGAS AKHIR

Judul : Sistem Prediksi Harga Sewa Kost menggunakan
Random forest Analytics (Studi Kasus : Kost
Eksklusif di Daerah Istimewa Yogyakarta)

Nama Mahasiswa : Faisal Ardiansyah

Nomor Mahasiswa : 16611085

**TUGAS AKHIR INI TELAH DIPERIKSA DAN DISETUJUI UNTUK
DIUJIKAN**

Yogyakarta, 3 November 2020

Menyetujui,

Dosen Pembimbing 1



Dr. Raden Bagus Fajriya Hakim, S.Si., M.Si

HALAMAN PENGESAHAN

TUGAS AKHIR

Sistem Prediksi Harga Sewa Kost menggunakan *Random Forest Analytics*

(Studi Kasus : Kost Eksklusif di Daerah Istimewa Yogyakarta)

Nama Mahasiswa : Faisal Ardiansyah

NIM : 16611085

**TUGAS AKHIR INI TELAH DIUJIKAN
PADA TANGGAL: 8 OKTOBER 2020**

Nama Penguji:

Tanda Tangan

1. Dr. Raden Bagus Fajriya Hakim, S.Si., M.Si. 
2. Ayundyah Kesumawati, S.Si., M.Si. 
3. Dina Tri Utari, S.Si., M.Sc. 

Mengetahui,

Dekan Fakultas Matematika dan Ilmu Pengetahuan Alam



Prof. Riyanto, S.Pd., M.Si., Ph.D.

KATA PENGANTAR



Assalamualaikum Wr.Wb

Puji syukur penulis panjatkan atas kehadiran Allah SWT karena atas berkat, rahmat, kesehatan dan kekuatan yang diberikan oleh-Nya tugas akhir ini dapat berjalan dengan lancar. Tugas akhir ini memberikan begitu banyak pembelajaran yang kemudian dapat di kembangkan pada penelitian selanjutnya.

Keberhasilan pembuatan tugas akhir ini tentunya tidak terlepas dari pihak-pihak yang membantu, memberi semangat dan turut memberi dukungannya selama kegiatan ini berlangsung. Ucapan terima kasih ini disampaikan kepada :

1. Bapak Prof. Riyanto, S.pd., M.Si., Ph.D., selaku Dekan Fakultas Matematika dan Ilmu Pengetahuan Alam.
2. Bapak Dr. Edy Widodo, M.Si selaku Ketua Jurusan Statistika Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Islam Indonesia..
3. Ibu RB Fajriya Hakim selaku Dosen Pembimbing penelitian yang selalu sabar membimbing dan mengarahkan penulis.
4. Bapak M. Muhajir, M.Si selaku Dosen Pembimbing Akademik penulis yang sangat penulis kagumi dan memberikan ilmu yang sangat bermanfaat untuk penulis.
5. Seluruh dosen dan karyawan Program Studi Statistika Universitas Islam Indonesia yang selalu memberikan ilmu baik dalam bidang akademik maupun non akademik.
6. Orang tua penulis, Bapak Kaolan dan Ibu Tusmirah yang telah memberi dukungan dan motivasi penulis dalam menulis penelitian ini.
7. Reza Aditya Pratama selaku kakak kandung penulis sebagai donatur resmi penulis juga Adik saya Kartika Sulistyaningsari yang memberikan semangat untuk penulis.
8. Keluarga besar INVISIO yang selalu ada dan menemani saya selama berkuliah di Jurusan Statistika.

9. Teman-teman seperbimbingan, Alfa, Reni, Luna, Gifa, Lina, Yaya, Mita, Indira yang berpegangan tangan untuk berjalan beriringan dalam melakukan perjuangan bersama.
10. Keluarga Besar Artcos, Statistika UII angkatan 2016.
11. Seluruh pihak yang mendoakan penulis dalam diam yang tidak bisa disebutkan satu-persatu.

Semoga segala bentuk kegiatan yang dilakukan selama tugas akhir ini dapat bermanfaat. Penulis menyadari sepenuhnya bahwa tugas akhir ini masih jauh dari sempurna, oleh karena itu segala kritik dan saran yang sifatnya membangun selalu penulis harapkan. Semoga tugas akhir ini dapat bermanfaat bagi penulis khususnya dan bagi semua yang membutuhkan umumnya. Akhir kata, semoga Allah SWT selalu melimpahkan rahmat serta hidayah-Nya kepada kita semua, Amin amin ya robbal 'alamiin

Wassalamualaikum Wr.Wb

Yogyakarta, 8 Oktober 2020

Penulis

Faisal Ardiansyah

DAFTAR ISI

HALAMAN SAMPUL	i
ABSTRAK	xii
ABSTRACT	xiii
BAB I PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	5
1.3 Batasan Masalah	5
1.4 Tujuan Penelitian	5
1.5 Manfaat Penelitian	6
BAB II TINJAUAN PUSTAKA	7
BAB III LANDASAN TEORI	12
3.1 Pengertian Kos	12
3.2 Metode Statistika	12
3.2.1 Web Scraping	12
3.2.2 Parsehub	12
3.2.3 Data inputation	13
3.2.4 Statistika Deskriptif	13
3.2.5 Prediksi	14
3.2.6 Pohon Keputusan (Decision Tree)	14
3.2.7 Classification and Regression Trees (CART)	15
3.2.8 Random Forest	16
3.2.9 Jumlah Pohon Random Forest	18
3.2.10 Akurasi	18
3.2.11 Koefiesien Determinasi (R^2)	18
3.2.12 Feature Importance	19
3.2.13 Python	19
3.2.14 Web Development	20
3.2.15 Flask	20
3.2.16 Heroku	21

3.2.17 Label Encoder.....	21
BAB IV METODOLOGI PENELITIAN	23
4.1. Populasi dan Sampel Penelitian.....	23
4.2. Tempat dan Waktu Penelitian	23
4.3. Jenis Penelitian	23
4.4. Data dan Variabel Penelitian	23
4.5. Metode Analisis.....	25
4.7 <i>Flowchart</i>	26
BAB V HASIL DAN PEMBAHASAN.....	28
5.1 Analisis Deskriptif.....	28
5.2 Analisis <i>Random Forest</i>	32
5.3. Membangun Web Aplikasi Menggunakan <i>Flask</i> dan <i>Heroku</i>	37
BAB VI PENUTUP	40
6.1 Kesimpulan.....	40
6.2 Saran	41
DAFTAR PUSTAKA	42
LAMPIRAN.....	47

DAFTAR TABEL

Tabel 2. 1 Perbedaan dan Persamaan Tinjauan Pustaka.....	9
Tabel 3. 1 Interpretasi Koefisien	19
Tabel 4. 1 Definisi Operasional Variabel (DOV).....	24
Tabel 5. 1 <i>Data Missing Value</i>	28
Tabel 5. 2 <i>Data imputation</i>	28
Tabel 5. 3 Rata-rata Harga Sewa berdasarkan Jenis Kost dan Daerah	30
Tabel 5. 4 Pembagian Data <i>Train</i> dan <i>Test</i>	33
Tabel 5. 5 Jumlah Pohon dengan Error	33
Tabel 5. 6 Perbandingan Data Aktual dan Prediksi.....	34
Tabel 5. 7 R^2 Data, MAPE dan Akurasi.....	35

DAFTAR GAMBAR

Gambar 3. 1 <i>Decision Tree</i>	15
Gambar 3. 2 <i>Random forest</i>	17
Gambar 4. 1 <i>Flowchart</i> Penelitian.....	27
Gambar 5. 1 Tipe Koston.....	29
Gambar 5. 2 Ketersediaan Koston	29
Gambar 5. 3 Jangka Waktu Sewa berdasarkan Daerah	30
Gambar 5. 4 <i>Label Encoder</i>	32
Gambar 5. 5 <i>Label encoder</i> dengan <i>Value</i>	32
Gambar 5. 6 Perbandingan Data Aktual dan Prediksi	34
Gambar 5. 7 <i>Feature Importance</i>	36
Gambar 5. 8 <i>Save Model</i> Prediksi	36
Gambar 5. 9 <i>Home.html</i>	37
Gambar 5. 10 <i>App.py</i>	38
Gambar 5. 11 Web Aplikasi yang Sedang Aktif	38
Gambar 5. 12 Tampilan <i>Prototype</i> Prediksi Kost.....	39

DAFTAR LAMPIRAN

Lampiran 1 Data Harga Sewa Kost.....	47
Lampiran 2 Analisis Random forest Menggunakan manual	53
Lampiran 3 Analisis <i>Random Forest</i> Menggunakan <i>Python</i>	60
Lampiran 4 <i>Deploy Flask Script (home2.html)</i>	62
Lampiran 5 <i>Flask Script (app.py)</i>	65
Lampiran 6 <i>Deploy ke Heroku</i>	66



PERNYATAAN

Dengan ini saya menyatakan bahwa dalam Tugas Akhir ini tidak terdapat karya yang sebelumnya pernah diajukan untuk memperoleh gelar kesarjanaan di suatu Perguruan Tinggi dan sepanjang pengetahuan saya juga tidak terdapat karya atau pendapat yang pernah ditulis atau diterbitkan orang lain, kecuali yang diacu dalam naskah ini dan disebutkan dalam daftar pustaka.

Yogyakarta, 3 November 2020



(Faisal Ardiansyah)



SISTEM PREDIKSI HARGA SEWA KOST DENGAN MENGGUNAKAN RANDOM FOREST ANALYTICS

Oleh: Faisal Ardiansyah

Program Studi Statistika

Fakultas Matematika dan Ilmu Pengetahuan Alam

Universitas Islam Indonesia

ABSTRAK

Yogyakarta adalah kota yang memiliki daya tarik bagi pendatang dengan tingkat pertumbuhan penduduk yang tinggi. Kota ini banyak menarik masyarakat untuk datang, tidak hanya untuk berwisata, melainkan untuk melangsungkan pendidikan, bahkan untuk bekerja dan menetap sementara waktu. Pemilihan kost sebagai hunian menjadi solusi alternatif bagi para pendatang tersebut. Permintaan yang besar, memunculkan adanya jasa pengiklan hunian secara daring. Mamikos adalah salah satunya, menyajikan informasi kamar kostan, lengkap dengan fasilitas kost, harga kost, dan dekorasi kamar beserta foto desain kamar yang menggambarkan kondisi sebenarnya. Mamikos dalam websitenya menampilkan harga sewa dengan penyesuaian kriteria yang dibutuhkan seperti Daerah, WiFi, Akses 24 jam, Kamar Mandi, Kecamatan, Jenis Kost dan Jangka Waktu Sewa. Namun, belum ada fasilitas untuk pengunjung mengetahui rata-rata / harga prediksi kostan dengan kriteria yang diinginkan, sehingga perbandingan harga dilakukan secara manual. Berdasarkan permasalahan tersebut, peneliti menerapkan metode random forest yang mampu membantu pengunjung menentukan keputusan berupa harga sewa prediksi dengan variabel yang diinginkan dalam suatu prototype yang memiliki User Experience yang mudah, sederhana dan fungsional. Berdasar kalkulasi dari random forest membentuk model prediksi dari 1111 data kostan di Yogyakarta yang dibagi menjadi 80% data train dan 20% data terbentuklah 20 pohon untuk memprediksi harga sewa kostan dengan tingkat akurasi sebesar 75.62%, R-squared data testing sebesar 65.91% dan R-Squared data training sebesar 94.9%. Model yang terbentuk memiliki feature importance, hal tersebut berpengaruh pada pembentukan model prediksi dari faktor kepentingan terbesar yaitu oleh variabel Jangka Waktu Sewa sebesar 0.682 dan variabel Akses 24 jam dengan faktor kepentingan terkecil sebesar 0.0103. Selanjutnya model tersebut diintegrasikan menggunakan Flask dan Heroku sehingga dapat menjadi prototype yang dapat diakses oleh semua orang dengan fungsi melakukan prediksi harga sewa kost berdasarkan kriteria yang diinginkan.

Kata kunci : Yogyakarta, Harga sewa kostan, Random forest, Flask dan Heroku.

BOARDING HOUSE PRICE PREDICTION SYSTEM WITH RANDOM FOREST ANALYTICS

Author: Faisal Ardiansyah

Department of Statistics

Faculty of Mathematics and Natural Sciences

Islamic University of Indonesia

ABSTRACT

Yogyakarta is a city with high population growth due to its attractiveness for migrants. Yogyakarta's problems are different from other, it is because this city have a big number of students and tourist. Some of them are studying, traveling, working and staying temporarily there. They come from various place in Indonesia, need a boarding house as option of their temporary living place. At the same time, they need a proper information before they decide to stay. Some of people use online residential / boarding home advertising platform, one of it is Mamikos. Mamikos share a lot of information about boarding rooms and other complement information such as additional facilities, prices, and rooms decorations along with design photos that describe actual conditions as much as possible. People will visit its address and they can looking for proper boarding house based on their urge. Some people will consider their choise through prices that based on location and facilities. We found that Mamikos has not implemented one function that can facilitate it. It needs additional function that can predict boarding prices according to desired criteria. Based on these problems a prototype was created by applying random forest and implemented with Flask where the random forest formed a prediction model from 1.111 boarding home data in Yogyakarta City which was divided into 80% train data data and 20% data formed 7 trees to predict boarding rental prices with an accuracy of 75.62%, R-squared data testing is 65.91% and R-squared training data is 94.9%. The model formed has feature importance, this feature importance has an influence on the formation of a model for the prediction model of the greatest interest faktor, namely by the variabel rental period of 0.638912 and the Accesbility 24 hours variabel with the smallest interest factor of 0.0103. Furthermore, the model is integrated using Flask and Heroku and can be accessed by everyone to predict boarding prices based on the desired criteria

Keywords: Yogyakarta, Boarding home price, Random forest, Flask and Heroku

BAB I

PENDAHULUAN

1.1 Latar Belakang

Daerah Istimewa Yogyakarta (Jawa: Dhaérah Istiméwa Ngayogyakarta) adalah Daerah Istimewa setingkat provinsi di Indonesia yang merupakan peleburan Negara Kesultanan Yogyakarta dan Negara Kadipaten Paku Alaman. Daerah Istimewa Yogyakarta terletak di bagian selatan Pulau Jawa, dan berbatasan dengan Provinsi Jawa Tengah dan Samudera Hindia. Daerah Istimewa yang memiliki luas 3.185,80 km² ini terdiri atas satu kotamadya, dan empat kabupaten, yang terbagi lagi menjadi 78 kecamatan, dan 438 desa/kelurahan (Indonesia.go.id).

Permasalahan yang dihadapi sekarang adalah bahwa pertumbuhan penduduk di Kota Yogyakarta masih relatif tinggi karena daerah ini memiliki daya tarik tersendiri bagi pendatang. Fenomena permasalahan yang berada di Kota Yogyakarta berbeda dengan permasalahan yang lain. Hal itu karena Kota Yogyakarta merupakan kota pelajar sekaligus kota wisata yang ada di Indonesia, tidak heran jika setiap tahunnya kota Yogyakarta mengalami laju pertumbuhan yang sangat tinggi. Namun tidak semua orang yang datang di kota Yogyakarta sekedar berwisata saja, banyak pendatang dari berbagai daerah guna melangsungkan pendidikan dan menetap sementara waktu. Sehingga mengakibatkan peningkatan pertumbuhan penduduk (Yanuar, 2019).

Kebanyakan orang yang ingin tinggal di Yogyakarta untuk menuntut ilmu dimulai sejak lulus sekolah menengah pertama (*junior high scholl*), kemudian melanjutkan di sekolah menengah atas. Ada dua jenis sekolah menengah atas ini, yaitu sekolah menengah umum (SMA) dan sekolah menengah kejuruan (SMK). Harapannya agar mereka setelah lulus SMA itu dapat diterima masuk perguruan tinggi di Yogyakarta. Jumlah SMA di Provinsi Daerah Istimewa Yogyakarta ada 165 sekolah terdiri dari 69 sekolah negeri dan 96 sekolah swasta. Sementara SMK berjumlah 203 sekolah yang terdiri dari 51 sekolah negeri dan 152 sekolah swasta (Utomo, 2009)

Bukan hanya pelajar saja, pendatang yang berstatus merantau mereka tidak sedikit yang memiliki ekonomi rendah. Kost-kost an menjadi hunian favorit oleh

para perantau tersebut, karena kost-kostan memiliki harga sewa yang murah (dengan fasilitas minim), lokasi tersebar dimana-mana yang memudahkan bagi para pekerja untuk bisa memilih dekat dengan lokasi pekerjaannya, serta akses dan kebebasan yang tinggi ini menjadi nilai lebih bagi mereka untuk memilih kost-kostan untuk menjadikan sebagai tempat tinggal dan menyimpan barang-barang (Widiatmoko, 2015)

Adanya kendala terkait ekonomi masyarakat entah itu dari kalangan pelajar, mahasiswa dan pekerja menyebabkan menyewa kost sebagai solusi alternatif. Menyewa kosan, akan lebih hemat daripada harus mengontrak atau sewa apartemen. Pilih kostan yang memiliki tempat strategis dan dekat dengan kantor atau kampus, dengan begitu tidak perlu keluar ongkos banyak (Asmita, Gaya Hidup, n.d.). Salah satu situs yang dapat digunakan untuk melihat kostan adalah Mamikos yang dapat diakses di <https://www.mamikos.com/>. Mamikos menyajikan informasi kamar kostan, lengkap dengan fasilitas kost, harga kost, dan dekorasi kamar beserta foto desain kamar yang sebisa mungkin menggambarkan kondisi sebenarnya. Mamikos juga memiliki tim, Agen dan *Mamichecker*, yang saat ini mengunjungi, memastikan, dan mengecek properti kost satu per satu. Adapun kost yang sudah dicek memiliki tanda *Mamichecker*. Informasi ketersediaan kamar kost dan harga kost selalu *update* untuk memastikan info kost kami akurat dan bermanfaat untuk anak kost (MAMIKOS, n.d.).

Teknologi dalam perkembangan arus produksi, konsumsi dan distribusi informasi memegang peranan penting. Urgensi peranan teknologi dalam proses massifikasi informasi terjadi ketika hasil teknologi membantu mengubah pola komunikasi yang dibatasi oleh ruang dan waktu menjadi pola komunikasi informasi tanpa batas. Dengan demikian, pada dasarnya teknologi bersifat baik, sehingga tidak mengherankan apabila terjadi perubahan dari media massa tradisional menjadi media massa baru. Pada akhirnya media baru dalam konteks teknologi dan globalisasi mengalami perubahan yang sedemikian kompleks. Globalisasi menjadi salah satu faktor penting dalam industri dan teknologi media komunikasi (Wuryanta, 2004)

Globalisasipun juga terjadi dalam bidang teknologi informasi, membuat perusahaan dibidang apapun meluaskan usahanya dalam bentuk digital. Perusahaan perlu merancang sebuah sistem yang dapat membantu konsumen yang sudah loyal ataupun konsumen baru yang segan untuk bertanya, dengan mudah mendapatkan informasi terkait profil perusahaan ataupun produk yang ditawarkan seperti harga dan spesifikasi lainnya melalui akses internet secara *realtime*.

Perancangan *prototype* sistem dapat digunakan sebagai salah satu strategi pemasaran dalam menawarkan produk untuk memenuhi kebutuhan dan memudahkan konsumen dengan beberapa prinsip seperti produk, nilai, biaya, hubungan, pasar, dan yang lainnya (Santoso, 2014). Pembentukan *prototype* bertujuan untuk memberikan gambaran aplikasi yang dirancang sebelum proses evaluasi dan dijadikan acuan dalam sebuah aplikasi (Nugraha, 2018) .

Adapun (Bruno Klaus Afonso, 2020) mengemukakan dalam penelitian yang berjudul “*Housing Prices Prediction with a Deep Learning and Random forest Ensemble*” Pengembangan model prediksi harga rumah dapat membantu penjual rumah atau agen real estate untuk membuat keputusan berdasarkan informasi yang lebih baik berdasarkan penilaian harga rumah. Studi ini menganalisis dataset yang terdiri dari 12.223.582 iklan perumahan, yang dikumpulkan dari situs *web* Brasil dari 2015 hingga 2018. Untuk memprediksikan harga properti, Bruno menggabungkan dua arsitektur ML yang berbeda, berdasarkan *Random forest* (RF) dan *Recurrent Neural Networks* (RNN). Studi ini menunjukkan bahwa memperkaya dataset dan menggabungkan berbagai pendekatan *Machine Learning* dapat menjadi alternatif yang lebih baik untuk prediksi harga perumahan di Brasil.

Selain itu juga (Lokhande P. P., 2015) dalam penelitian berjudul “*Efficient Way Of Web Development Using Python And Flask*” mengungkapkan *Web* adalah bantuan jaringan yang paling sering digunakan yang memenuhi persyaratan semua jenis pengguna; ini memberikan solusi untuk semua jenis masalah definisi. Untuk mendesain dan mengembangkannya, terstruktur dengan baik dan dengan tampilan *web* yang bagus, kita harus memilih teknologi yang tepat. Kebutuhan teknologi untuk memuaskan portal *web* yang baik dapat dipenuhi dengan “*Python*” dan “*Flask*”. Untuk menjaga kompleksitas seperti teknologi seperti *Python Jinja Flask*

lebih berguna. Teknologi tersebut juga membantu menciptakan antarmuka yang lebih ramah pengguna untuk pengambilan data dari *www*. karena *Python* dapat digunakan untuk membuat *web* lebih kuat, cepat dan efisien dengan bantuan *Flask*.

Berdasarkan masalah diatas perancangan *prototype* sistem dapat digunakan sebagai salah satu strategi pemasaran dalam menawarkan produk untuk memenuhi kebutuhan dan memudahkan konsumen dengan beberapa prinsip seperti produk, nilai, biaya, hubungan, pasar, dan yang lainnya. Kemudian dikembangkan kedalam bentuk *web* aplikasi agar mudah diakses oleh khalayak. Hasil penelitian dapat digunakan sebagai bahan pertimbangan dalam mengatur perekonomian dan kemampuan calon konsumen dalam memenuhi kebutuhan.

Terdapat beberapa pendekatan analisis yang dapat digunakan untuk melakukan prediksi, salah satunya adalah *data mining*. *Data mining* adalah bidang yang digabung dari beberapa bidang keilmuan yang meyatukan teknik dari pembelajaran mesin, pengenalan pola, statistik, *database*, dan visualisasi untuk pengenalan permasalahan pengambilan informasi dari *database* yang besar (Larose, 2014) Salah satu metode yang terdapat dalam *data mining* adalah metode *random forest*.

Random forest merupakan salah satu metode dalam *Decision Tree*. *Decision Tree* atau pohon pengambil keputusan adalah sebuah diagram alir yang berbentuk seperti pohon yang memiliki sebuah *root node* yang digunakan untuk mengumpulkan data, Sebuah *inner node* yang berada pada *root node* yang berisi tentang pertanyaan tentang data dan sebuah *leaf node* yang digunakan untuk memecahkan masalah serta membuat keputusan. *Decision tree* mengklasifikasikan suatu sampel data yang belum diketahui kelasnya kedalam kelas – kelas yang ada. Penggunaan *decision tree* agar dapat menghindari *overfitting* pada sebuah set data saat mencapai akurasi yang maksimum (Aditya Yanuar r, 2018). Variabel dependen yang akan digunakan dalam penelitian adalah harga sewa kost, sedangkan variabel independen antara lain, jenis kostan, tipe kost, daerah kostan, fasilitas kostan.

Dengan demikian, penelitian ini dilakukan dengan judul “**Sistem Prediksi Harga Sewa Kost dengan menggunakan *Random forest Analytics* (Studi Kasus : Biaya Kost Tinggal di Yogyakarta)**”. Diharapkan melalui penulisan ini bisa

memberikan manfaat kepada pengguna saat menggunakan mesin prediksi harga sewa kost.

1.2 Rumusan Masalah

Berdasarkan uraian latar belakang di atas, maka rumusan masalah dalam penelitian ini sebagai berikut:

1. Bagaimana analisis deskriptif harga kost dan faktor-faktor yang mempengaruhinya?
2. Bagaimana pengaruh dan tingkat akurasi faktor-faktor yang mempengaruhi harga kost menggunakan metode *random forest*?
3. Bagaimana hasil rancang bangun program *website* mengenai prediksi harga kost di D.I. Yogyakarta dengan pendekatan *random forest*?

1.3 Batasan Masalah

Untuk meminimalisir perluasan dalam pembahasan penelitian, adapun batasan–batasan masalah yang diberikan sebagai berikut:

1. Data yang digunakan adalah data sekunder yang diperoleh dengan *scrapping* menggunakan *Parsehub* yang tersedia di *website* MamiKos khusus Daerah Istimewa Yogyakarta *dataset* juga berupa kost eksklusif.
2. Variabel yang digunakan antara lain harga sewa kost sebagai variabel dependen dan tipe kost, jangka waktu penyewaan, fasilitas dan daerah sebagai variabel independen.
3. Analisis deskriptif dilakukan menggunakan *Ms. Excel* 2016 dan analisis prediksi dilakukan menggunakan metode *random forest* dengan *Python*. Selain itu tampilan *prototype* kedalam *website* dilakukan menggunakan *Flask* dan *Heroku*.

1.4 Tujuan Penelitian

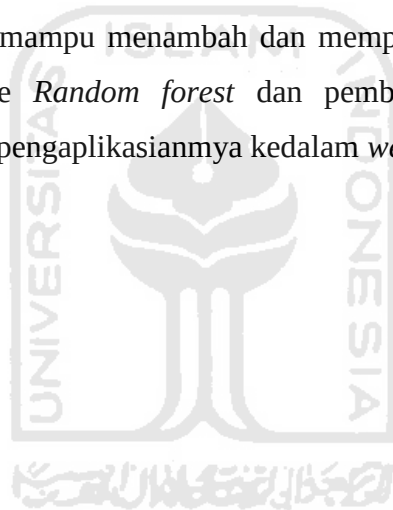
Berdasarkan latar belakang dan rumusan masalah yang telah diuraikan, maka tujuan dilakukannya penelitian ini adalah:

1. Mengetahui gambaran umum harga sewa kos dan faktor-faktor yang mempengaruhinya.

2. Mengetahui pengaruh dan tingkat akurasi faktor-faktor yang mempengaruhinya harga kost menggunakan metode *random forest*.
4. Mengetahui hasil rancang bangun program *website* mengenai prediksi harga kost di D.I. Yogyakarta dengan pendekatan *random forest*.

1.5 Manfaat Penelitian

Berdasarkan tujuan yang akan dicapai, penelitian ini diharapkan mampu memberikan manfaat seperti memberikan gambaran bagi konsumen terkait faktor-faktor yang perlu diperhatikan dan dapat mempengaruhi harga sewa kost. Kemudian dapat menjadi acuan sebagai bahan pertimbangan dalam mengatur perekonomian dan kemampuan konsumen dalam mencari kost. Selain itu, penelitian ini diharapkan mampu menambah dan memperluas ilmu pengetahuan terkait penerapan metode *Random forest* dan pembuatan sebuah *prototype* menggunakan *Flask* serta pengaplikasianmya kedalam *website*.



BAB II

TINJAUAN PUSTAKA

Penulisan yang akan dilakukan membutuhkan penulisan terdahulu sebagai acuan, tujuannya untuk mengetahui hubungan antara penulisan yang dilakukan sebelumnya dengan penulisan yang penulis lakukan saat ini serta dapat menghindari adanya duplikasi. Hal ini perlu dilakukan untuk menunjukkan bahwa penulisan yang dilakukan mempunyai arti penting sehingga dapat diketahui kontribusinya terhadap ilmu pengetahuan.

Beberapa tinjauan pustaka yang diambil terkait dengan *random forest* salah satunya Menurut pendapat (Alshamsi, 2014) mengenai penelitiannya yang berjudul “*Predicting car insurance policies using random forest*” pada penelitian ini bertujuan untuk membantu perusahaan asuransi dalam memprediksi pilihan pelanggan untuk memberikan layanan yang lebih kompetitif. *Random forest* digunakan untuk mengembangkan model klasifikasi yang dapat diterapkan dalam memprediksi polis asuransi mana yang akan dipilih oleh pelanggan. Kinerja model yang dikembangkan dibandingkan dengan beberapa teknik penambangan data seperti *ZeroR classifier*, *Simple Logistics Function*, *Decision Tree* dan *Naïve Bayes* pada dataset berisi 7 kebijakan berbeda. Hasil penelitian menunjukkan bahwa *random forest* adalah teknik yang paling tepat dengan akurasi keseluruhan 97,9%.

Implementasi *random forest* pernah dilakukan oleh (Fransiska Amalia Kurniawan, 2011) mengenai menerapkan algoritma *Random forest* dan *CART* untuk membangun model klasifikasi pada *misuse IDS. Intrusion Detection System* (IDS) memainkan peranan penting dalam menjaga *integrity*, *confidentiality* dan *availability* dari sumber daya jaringan komputer. Tujuan utama IDS adalah mengklasifikasikan aktivitas jaringan apakah termasuk aktivitas normal atau intrusi. Penulisan ini membahas algoritma *random forest* dan *CART* sudah mampu diimplementasikan untuk klasifikasi dalam sistem pendeteksi intrusi, namun berdasarkan parameter pengujian *precision*, *recall*, *F Measure*, sistem yang dibangun menggunakan *random forest* dan *CART* belum dapat dikatakan akurat

melihat nilai akurasi tertinggi yang dapat dicapai sebesar 70,72% saat jumlah *record* = 224 *records* dan banyaknya pohon = 60 pohon.

Hal ini diungkapkan juga oleh (Aji Primajaya, 2018) dalam penelitian yang berjudul “*Random forest Algorithm for Prediction of Precipitation*” memiliki tujuan menciptakan suatu model untuk memprediksi hujan, dari penelitian ini didapatkan *random forest* lebih cocok menggunakan seluruh data untuk training data dan testing data jika dilihat dari evaluasi akurasi, presisi, *recall*, *f-measure*, *kappa statistic*, *MAE*, *RMSE* dan *ROC Area*. Akurasi *random forest* dengan menggunakan teknik *10-fold cross validation* sebesar 71,09% sedangkan dengan teknik menggunakan seluruh data sebesar 99,45%.

Penelitian dengan judul “*Klasifikasi Pembiayaan Warung Mikro Menggunakan Metode Random forest dengan Teknik Sampling Kelas Imbalanced*” yang ditulis oleh (Widiastuti, 2018) penerapannya adalah mengklasifikasikan pembiayaan Warung Mikro Bank Syariah Mandiri KC Jambi menggunakan metode *Random forest*. Dalam penelitian ini menggunakan teknik sampling kelas *imbalanced*, nilai *sensitivity*, *precision*, dan *G-mean* meningkat dibandingkan menggunakan data asli. Didapatkan *Random forest* dengan teknik *oversampling* mempunyai nilai *AUC* tertinggi yaitu sebesar 66.69%.

Dari (HONG, 2020) juga melakukan penelitian dengan judul “*A House Price Valuation Based on The Random forest Approach: The Mass Appraisal of Residential Property in South Korea*”. Penelitian ini memiliki focus pada prediksi harga rumah dengan *random forest* untuk dibandingkan dengan metode yang sudah ada sebelumnya yaitu regresi linier. Data dalam penelitian ini merupakan data transaksi rumah pada periode 2006 hingga 2017 di daerah Gangnam, Korea Selatan. Prediksi dengan metode *random forest* dapat menjadi pelengkap metode regresi, karena kekurangan yang dihasilkan dari metode regresi dapat di pulihkan dengan *random forest* yang berhasil membentuk model yang lebih kompleks. Hal tersebut menunjukkan bahwa *random forest* dapat membentuk model lebih baik karena lebih tepat dalam memprediksi harga rumah dibandingkan dengan metode regresi.

Penggunaan analisis *random forest* dalam penelitian dengan judul “*How Much Is My Car Worth? A Methodology for Predicting Used Cars’ Prices Using*

Random forest”, (Pal N., 2018) mengungkapkan data kenaikan penjualan mobil bekas secara eksponensial meningkat karena Negara berkembang memilih sewa daripada membeli mobil baru. Penjual mobil terkadang memanfaatkan skenario ini dengan mendaftar harga yang tidak realistis karena permintaan. Dalam hal ini, timbul kebutuhan akan model yang dapat menetapkan harga untuk kendaraan dengan mengevaluasi fitur-fiturnya dengan mempertimbangkan harga mobil lain. Analisis *random forest* kali ini digunakan untuk memprediksi harga mobil bekas, dengan 500 Pohon Keputusan dibuat untuk melatih data didapatkan nilai akurasi pelatihan ditemukan menjadi 95,82%, dan akurasi pengujian adalah 83,63%. Model dapat memprediksi harga mobil secara akurat dengan memilih fitur yang paling berkorelasi.

Penelitian lain diungkapkan oleh (R. Delshi Howsalya Devi, 2020) makalahnya yang berjudul “*Prediction of Disease Using Random Forest Classification Algorithm*” memprediksi penyakit umum berdasarkan gejala pasien. Di dalam prediksi penyakit umum, kebiasaan hidup orang dan informasi pemeriksaan dipertimbangkan untuk prediksi yang akurat. Penggunaan algoritma *random forest* untuk mengklasifikasikan data pasien karena data medis saat ini tumbuh sangat luas dan itu perlu memproses data yang ada untuk memprediksi penyakit yang tepat berdasarkan gejala yang dijadikan sebagai variabel dependen. Hasil *random forest* yaitu bisa memprediksi risiko penyakit umum yang akurat, penggunaan *random forest* juga karena mengarah pada konsumsi waktu yang singkat dan biaya yang minimal untuk prediksi penyakit dan prediksi risiko.

Tabel 2. 1 Perbedaan dan Persamaan Tinjauan Pustaka

No.	Penulis	Judul Penelitian	Metode	Persamaan	Perbedaan
1.	Fransiska Amalia Kurniawan dkk (2011)	Analisis dan Implementasi <i>random forest</i> classification dan <i>classification and regression Tree</i> Misuse Intrusion Detection System	<i>Random forest</i> dan <i>Cart</i>	Sama-sama menggunakan <i>Random forest</i>	Studi kasus penulisan sekarang yaitu prediksi harga sewa kostan dan melakukan analisis dengan metode pembandingan yaitu <i>cart</i> .

No.	Penulis	Judul Penelitian	Metode	Persamaan	Perbedaan
2.	A. S. Alshamsi (2014)	Predicting car insurance policies using <i>random forest</i> .	<i>Random forest</i>	Sama-sama menggunakan <i>Random forest</i>	Studi kasus penulisan sekarang yaitu prediksi harga sewa kostan.
3.	Aji Primajaya, Betha Nurina (2018)	<i>Random forest</i> Algorithm for Prediction of Precipitation	<i>Random forest</i>	Sama-sama menggunakan <i>Random forest</i>	Studi kasus penulisan sekarang yaitu prediksi harga sewa kostan.
4.	Julia Widiastuti (2018)	Klasifikasi Pembiayaan Warung Mikro Menggunakan Metode <i>Random forest</i> dengan Teknik Sampling Kelas Imbalanced	<i>Random forest</i>	Sama-sama menggunakan <i>Random forest</i>	Studi kasus penulisan sekarang yaitu prediksi harga sewa kostan dan pada penelitian ini menggunakan teknik sampling kelas <i>imbalanced</i> juga lebih kearah <i>web development</i> .
5.	Jengei HONG, Heeyoul CHOI, dan Woo-sung KIM (2019)	<i>A House Price Valuation Based on The Random forest Approach: The Mass Appraisal of Residential Property in South Korea.</i>	<i>Random forest</i>	Sama-sama menggunakan <i>Random forest</i> dan kearah prediksi harga hunian	Studi kasus penulisan sekarang yaitu prediksi harga sewa kostan juga penelitian kearah <i>web development</i> . Tidak menggunakan regresi linier
6.	Pal N., Arora P., Kohli P., Sundararaman D., Palakurthy S.S. (2019)	How Much Is My Car Worth? A Methodology for Predicting Used Cars' Prices Using <i>Random forest</i>	<i>Random forest</i>	Sama-sama menggunakan <i>Random forest</i>	Studi kasus penulisan sekarang yaitu prediksi harga sewa kostan dan kearah <i>web development</i> .
7.	R. Delshi Howsalya Devi,dkk (2020)	Prediction of Disease Using <i>Random forest</i> Classification Algorithm	<i>Random forest</i>	Sama-sama menggunakan <i>Random forest</i>	Studi kasus penulisan sekarang yaitu prediksi harga sewa kostan dan kearah <i>web development</i> .

Referensi di atas dijadikan sebagai acuan dalam penulisan ini. Tabel 2.1 merupakan tabel rangkuman perbandingan dengan penulisan sebelumnya yang

berkaitan dengan fokus yang sedang penulis teliti dan terdapat beberapa kemiripan dengan penelitian yang berkaitan yaitu *random forest* juga pembeda antara penelitian ini dengan penelitian sebelumnya adalah tema dan variabel yang digunakan. Penelitian ini menggunakan data sekunder terkait harga sewa kostan yang diperoleh dengan proses *scraping* pada website www.mamikos.com. Variabel yang digunakan dalam penelitian antara lain jenis kostan, tipe kost, daerah kostan, fasilitas kostan.



BAB III

LANDASAN TEORI

3.1 Pengertian Kos

Menurut Kamus Besar Bahasa Indonesia kost berasal dari kata indekos yang artinya tinggal di rumah orang lain dengan atau tanpa makan (dengan membayar setiap bulan) (KBBI, n.d.), Sedangkan dalam (Wikipedia, n.d.) memuat indekos atau kost adalah sebuah jasa yang menawarkan sebuah kamar atau tempat untuk ditinggali dengan sejumlah pembayaran tertentu untuk setiap periode tertentu. Dalam (Hajar, 2005) berpendapat bahwa tempat kost memiliki fungsi yang sama dengan rumah, dimana fungsi kos-kosan adalah sebagai sarana tempat tinggal sementara bagi mahasiswa yang pada umumnya berasal dari luar daerah selama masa studinya.

3.2 Metode Statistika

Metode statistika merupakan prosedur atau cara yang digunakan dalam melakukan sebuah penelitian dari proses pengumpulan data hingga analisis dan akhirnya diperoleh suatu kesimpulan. Metode-metode yang digunakan dalam penelitian ini antara lain:

3.2.1 Web Scraping

Web Scraping (Turland, 2010) adalah proses pengambilan sebuah dokumen semi-terstruktur dari internet, umumnya berupa halaman-halaman *web* dalam bahasa markup seperti *HTML* atau *XHTML*, dan menganalisis dokumen tersebut untuk diambil data tertentu dari halaman tersebut untuk digunakan bagi kepentingan lain.

3.2.2 Parsehub

Parsehub adalah perangkat lunak untuk ekstraksi *web* yang mendukung pengambilan data yang rumit dari situs yang menggunakan *AJAX*, *JavaScript*, *redirect*, dan *cookie*. Teknologi *machine learning* digunakan untuk membaca dan

menganalisis dokumen. Analisis memungkinkan *ParseHub* untuk menghasilkan data yang relevan. Perangkat lunak ini tersedia sebagai Klien desktop untuk *Windows, Mac dan Linux*. Selain itu, ada Aplikasi *Web*, yang bisa digunakan di dalam *browser*. Aplikasi ini memungkinkan untuk mengekstraksi hingga 200 halaman (Draxl, 2018).

3.2.3 *Data imputation*

Imputasi data melibatkan nilai-nilai yang hilang dalam satu set data. Data yang hilang menciptakan sejumlah tantangan potensial untuk analisis statistik. Pada dasarnya, nilai-nilai yang hilang mempertanyakan validitas *dataset* untuk mewakili kasus yang diamati dan, pada akhirnya, kerangka sampling. Dari sudut pandang statistik, nilai-nilai yang hilang dapat meningkatkan kemungkinan membuat kesalahan Tipe I dan Tipe II, mengurangi kekuatan statistik, dan membatasi reliabilitas interval kepercayaan (Streiner, 2002).

Ada prosedur statistik yang tersedia bagi para peneliti untuk mengganti nilai yang hilang dengan estimasi yang masuk akal. Metode dasar, seperti substitusi rata-rata, imputasi regresi, dan imputasi *hot deck* dapat membiaskan nilai imputasi tergantung pada mekanisme ketiadaan. Metode lanjutan termasuk maksimisasi ekspektasi, kemungkinan maksimum informasi lengkap, dan imputasi ganda. Metode-metode ini menghasilkan estimasi nilai-nilai yang hilang yang lebih dapat diandalkan, terutama ketika nilai yang hilang tidak secara acak (Rosenthal, 2017).

Pada penelitian ini menggunakan data imputasi dengan mengubah nilai-nilai kosong dengan modus data dikarenakan data berupa kategorik, hal ini juga merupakan mean sampel merupakan estimasi tak bias modus populasi.

3.2.4 *Statistika Deskriptif*

Statistika deskriptif merupakan cara mendeskripsikan atau menggambarkan data yang telah terkumpul sebagaimana adanya tanpa dengan tujuan memberikan kesimpulan yang berlaku untuk umum/generalisasi. Ciri-ciri analisis data deskriptif yaitu penyajian data lebih ditekankan dalam bentuk tabel, grafik, dan ukuran-ukuran statistik seperti persentase, rata-rata, variansi, korelasi, dan angka indeks.

Selain itu, analisis ini tidak menggunakan uji signifikansi dan taraf kesalahan karena tidak ada kesalahan generalisasi (Purwoto, 2007)

Statistik deskriptif atau statistik deduktif adalah bagian dari statistik yang mempelajari cara pengumpulan data dan penyajian data sehingga mudah dipahami. Statistik deskriptif hanya berhubungan dengan hal menguraikan atau memberikan keterangan-keterangan mengenai suatu data atau keadaan atau fenomena. Dengan kata lain statistik deskriptif berfungsi menerangkan keadaan, gejala, atau persoalan (Iqbal, 2001). Penarikan kesimpulan pada statistik deskriptif (jika ada) hanya ditujukan pada kumpulan data yang ada. Didasarkan pada ruang lingkup bahasannya, statistik deskriptif mencakup:

1. Distribusi frekuensi beserta bagian-bagiannya seperti :
 - a. Grafik distribusi (histogram, poligon frekuensi, dan ogif);
 - b. Ukuran nilai pusat (rata-rata, median, modus, kuartil dan sebagainya);
 - c. Ukuran dispersi (jangkauan, simpangan rata-rata, variasi, simpangan baku, dan sebagainya);
 - d. Kemencengan dan keruncingan kurva
2. Angka indeks.
3. *Times series*/deret waktu atau berkala.
4. Korelasi dan regresi sederhana.

3.2.5 Prediksi

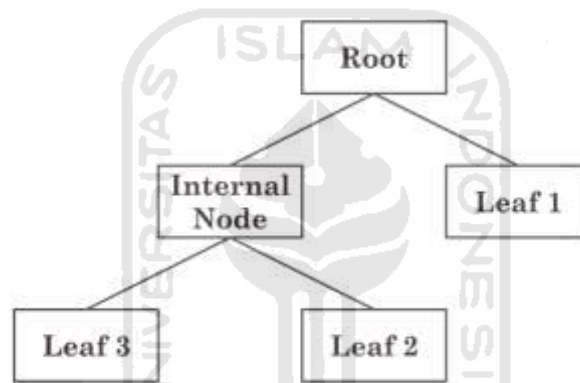
Prediksi adalah suatu proses memperkirakan secara sistematis tentang sesuatu yang paling mungkin terjadi di masa depan berdasarkan informasi masa lalu dan sekarang yang dimiliki, agar kesalahannya (selisih antara sesuatu yang terjadi dengan hasil perkiraan) dapat diperkecil. Prediksi tidak harus memberikan jawaban secara pasti kejadian yang akan terjadi, melainkan berusaha untuk mencari jawaban sedekat mungkin dari yang akan terjadi (Herdianto, 2013).

3.2.6 Pohon Keputusan (*Decision Tree*)

Decision tree learning adalah suatu metode belajar yang sangat populer dan banyak digunakan secara praktis. Metode ini merupakan metode yang berusaha menemukan fungsi-fungsi pendekatan yang bernilai diskrit dan tahan terhadap data-

data yang terdapat kesalahan (*noise data*) serta mampu mempelajari ekspresi-ekspresi *disjunctive*. *Iterative Dichotomiser 3* (ID3), *Assistant* dan *C4.5* merupakan jenis dari *decision tree learning*. Dalam membangun *decision tree learning* dibutuhkan evaluasi semua atribut yang ada menggunakan suatu ukuran statistik untuk mengukur efektifitas suatu atribut dalam mengklasifikasikan kumpulan sampel data. Dalam hal ini *information gain* adalah yang paling banyak digunakan (Suyanto, 2011).

Proses pada pohon keputusan dengan mengubah bentuk data (tabel) menjadi model pohon, mengubah model pohon menjadi *rule*, dan menyederhanakan *rule* (Basuki, 2003).



Gambar 3. 1 *Decision Tree*

Pada pohon keputusan terdapat 3 jenis *node*, yaitu:

- Rote Node*, merupakan *node* yang paling atas. Pada *node* ini tidak ada *input* dan bisa tidak mempunyai output atau mempunyai output lebih dari satu.
- Internal Node*, merupakan *node* pencabangan. Pada *node* ini hanya terdapat satu *input* dan mempunyai *output* minimal dua.
- Leaf Node* atau *Terminal Node*, merupakan *node* akhir. Pada *node* ini hanya terdapat satu *input* dan tidak mempunyai *output*.

3.2.7 *Classification and Regression Trees (CART)*

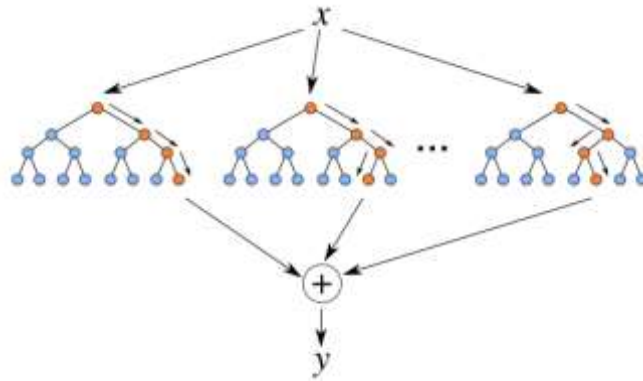
Classification and Regression Trees (CART) adalah salah satu metode atau algoritma dari teknik pohon keputusan. *CART* adalah suatu metode statistik nonparametrik yang dapat menggambarkan hubungan antara variabel respon

(variabel dependen) dengan satu atau lebih variabel prediktor (variabel independen) (Alia Hartati, 2012). Apabila variabel respon berbentuk kontinu maka metode yang digunakan adalah metode regresi pohon (*regression trees*), sedangkan apabila variabel respon memiliki skala kategorik maka metode yang digunakan adalah metode klasifikasi pohon (*classification trees*). Variabel respon dalam penelitian ini berskala kategorik, sehingga metode yang akan digunakan adalah metode klasifikasi pohon (Breiman L. F., 1984).

3.2.8 *Random Forest*

Metode *random forest* adalah pengembangan dari metode *CART* (Dewi, 2011), yaitu dengan menerapkan metode *bootstrap aggregating (bagging)* dan *random feature selection*. Pada *random forest*, banyak pohon ditumbuhkan sehingga terbentuk hutan (*forest*), kemudian analisis dilakukan pada kumpulan pohon tersebut. Pada gugus data yang terdiri atas n amatan dan p peubah penjelas, *random forest* dilakukan dengan cara (Breiman L. , 2001):

- a. Tahapan pertama adalah tahapan *bootstrap*, dimana tahapan ini melakukan penarikan contoh acak berukuran n dengan pemulihan pada gugus data..
- b. Setelah itu menggunakan contoh *bootstrap*, pohon dibangun sampai mencapai ukuran maksimum (tanpa pemangkasan). Pada setiap simpul, pemilihan pemilah dilakukan dengan memilih m peubah penjelas secara acak, dimana $m \ll p$. Pemilah terbaik dipilih dari m peubah penjelas tersebut. Tahapan ini adalah tahapan *random feature selection*.
- c. Berikutnya, ulangi langkah 1 dan 2 sebanyak k kali, sehingga terbentuk sebuah hutan yang terdiri atas k pohon.



Gambar 3. 2 *Random forest*

Dalam penyelesaian *Random forest* yaitu dengan pemilihan nilai “ k ” sampel yang diambil secara acak dengan pengembalian atau *replacement*. Dalam hal ini dengan menggunakan dataset yang sama untuk membangun *decision tree*, pada penyelesaian ini metode *CART* digunakan untuk menentukan setiap *node* pada *tree*. Perhitungan informasi *gain* dengan persamaan sebagai berikut:

$$Gain(A) = Info(D) - Info_A(D) \quad 3.1$$

Nilai *info* (D) dapat diperoleh menggunakan rumus 3.2 dan 3.3 untuk mendapatkan nilai $info_A(D)$:

$$Info(D) = - \sum_{i=1}^n p_i \log_2(p_i) \quad 3.2$$

Keterangan:

n = jumlah kelas target

p_i = proporsi kelas i terhadap partisi D

$$Info_A(D) = \sum_{j=1}^v \frac{|D_j|}{|D|} \times Info(D_j) \quad 3.3$$

Keterangan:

v = jumlah partisi

D_j = total partisi ke j

D = total baris pada semua partisi

Nilai *information gain* pada atribut dengan nilai kontinu atau numerik harus menentukan nilai pembelah (*split point*) terbaik untuk mengelompokkan nilai. *Split*

point terbaik diperoleh dari dengan cara mengurutkan data terlebih dahulu. Kemudian median atau nilai tengah setiap pasangan nilai yang saling berdekatan dianggap sebagai kemungkinan *split point* yang dapat digunakan. Apabila atribut A merupakan atribut dengan nilai kontinu maka seluruh nilai A diurutkan, kemudian menentukan nilai tengahnya sehingga kemungkinan jumlah partisi pada persamaan 3.3 adalah dua atau $v = 2$ ($j = 1$ dan 2) (Harista, 2019).

3.2.9 Jumlah Pohon *Random Forest*

Penelitian dengan judul *How Many Trees in a Random forest?* menyatakan bahwa hampir tidak ada literatur yang menunjukkan berapa banyak pohon yang harus digunakan untuk membangun sebuah hutan acak dengan hasil yang maksimal (Oshiro, 2012). Penelitian lain juga mengungkapkan banyaknya pohon yang akan dibentuk sangat berpengaruh terhadap tingkat akurasi hasil klasifikasi. Semakin banyak pohon, semakin akurat hasil klasifikasinya. Selain itu juga RF dapat menangani *input* variabel yang besar, menyeimbangkan *error* dalam *unbalanced dataset* (Fransiska Amalia Kurniawan, 2011).

3.2.10 Akurasi

Percobaan dari penulisan dievaluasi dengan pengukuran akurasi Xhemali, et al (2009). *Precision* dan *Recall* adalah dua perhitungan yang banyak digunakan untuk mengukur kinerja dari sistem/metode yang digunakan (Suwanto Sanjaya, 2015). *Precision* adalah tingkat ketepatan antara informasi yang oleh pengguna dengan jawaban yang diberikan oleh sistem. Sedangkan *Recall* adalah tingkat keberhasilan sistem dalam menemukan kembali sebuah informasi. Sedangkan di dunia lain seperti dunia statistika dikenal juga istilah *accuracy*. Akurasi di definisikan sebagai tingkat kedekatan antara nilai prediksi dengan nilai aktual (Hidayat, 2015).

3.2.11 Koefisien Determinasi (R^2)

Koefisien determinasi (R^2) digunakan untuk mengetahui seberapa besar persentase sumbangan pengaruh variabel independen secara bersamaan terhadap terhadap variabel dependen. Nilai *R square* dikatakan baik jika diatas 0.5 karena

nilai *R square* berkisar antara 0 sampai 1. Koefisien korelasi mempunyai kriteria-kriteria diantaranya sebagai berikut (Sugiyono, 2015).

Tabel 3. 1 Interpretasi Koefisien

Interval Koefisien	Tingkat Hubungan
0.80 – 1.000	Sangat Kuat
0.60 – 0.799	Kuat
0.40 – 0.599	Cukup Kuat
0.20 – 0.399	Rendah
0.00 – 0.199	Sangat Rendah

Koefisien determinasi biasanya dilambangkan dengan R^2 yang diperoleh dari hasil pembagian antara nilai *Sum of Squares Error* (SSR) dengan *Sum of Squares Total* (SST) ,dimana nilai SSR dan SST didefinisikan sebagai berikut: (Kim, 2018)

$$R^2 = \frac{\sum_{i=1}^n (\hat{y}_i - \bar{y})^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad 3.6$$

Keterangan:

y_i = variabel dependen
 $\bar{y}$ = nilai rata-rata variabel dependen
 $\hat{y}_i$ = nilai prediksi

3.2.12 Feature Importance

Feature importance antara lain untuk menghindari *overfitting* dalam model dan mengetahui tentang faktor yang mendasari pembentukan model. Pemilihan *feature importance* akan dilakukan ketika pembentukan model. Dua ukuran penting yang sering digunakan dalam *random forest* adalah *Gini Importance Index* dan *Permutation Importance Index* (PIM) (Hasan, 2015).

3.2.13 Python

Python merupakan sebuah bahasa pemrograman yang cukup terkenal yang memiliki banyak manfaat untuk mendukung pemrograman yang berorientasi objek dan dapat berjalan diberbagai macam *platform* sistem operasi, seperti *PCs*,

Macintosh, UNIX. Beberapa kelebihan dari bahasa pemrograman *Python* diantara lain (Ginting, 2020):

- a. Pengembangan program dilakukan dengan cepat dan *coding* yang lebih sedikit
- b. Mendukung multi *platform*
- c. Memiliki sistem pengelolaan memori yang otomatis
- d. *Python* bersifat *Object Oriented Programming* (OOD)

3.2.14 *Web Development*

Web development adalah sebuah kegiatan pengembangan halaman-halaman *web* yang ada di internet. Halaman *web* adalah barang utama yang ada di internet, dimana seluruh pengguna internet pasti akan mengakses suatu *web* tertentu yang ada di internet (Irsyad, 2019). *Web development* mencakup beberapa hal seperti penulisan *web*, halaman *web*, situs *web*, dan yang lainnya. Saat ini terdapat beberapa bahasa pemrograman yang bersaing kompetitif seperti *Java, Python, PHP, Ruby*, dan masih banyak lagi (Taneja, 2014).

3.2.15 *Flask*

Flask adalah *Web Application Framework* untuk mempermudah dalam pengembangan atau pembangunan *website*. *Flask* ditulis dalam bahasa pemrograman *Python* dan tersedia dalam bentuk *Library* dan dapat di-*install* melalui *Package Manager Python* (Pahlevi, 2020). Berdasarkan penjelasan tersebut, maka *Flask* memiliki beberapa keunggulan seperti (Irsyad, 2019) :

- a. Kesederhanaan *core* yang dimiliki membuat *Flask* ringan ketika dijalankan
- b. Mampu mengatasi fungsi *HTTP request* dengan sangat mudah
- c. Memiliki API baik dan juga koheren
- d. Dokumentasi yang tidak sedikit dapat terstruktur dengan baik dan dapat langsung digunakan
- e. Mudah saat proses *deploy* untuk produksi
- f. Mudah ketika diperiksa secara keseluruhan
- g. Memiliki fleksibilitas yang tinggi dan konfigurasi yang mudah disesuaikan.

3.2.16 Heroku

Heroku adalah sebuah *cloud platform* yang menjalankan bahasa pemrograman tertentu, *Heroku* mendukung bahasa pemrograman seperti *Ruby*, *Node.js*, *Python*, *Java*, *PHP*, dan lain-lain. *Heroku* termasuk ke dalam kriteria *Platform As A Service* (PaaS), sehingga jika deploy aplikasi ke *Heroku* cukup hanya dengan melakukan konfigurasi aplikasi yang ingin dideploy dan menyediakan *platform* yang memungkinkan pelanggan untuk mengembangkan, menjalankan, dan mengelola aplikasi tanpa kompleksitas membangun dan memelihara infrastruktur yang biasanya terkait dengan pengembangan dan peluncuran aplikasi (Hariyadi, 2017).

Heroku selaku PaaS, mampu mempercepat proses dalam *deploy*, *build*, *manage* dan menciptakan aplikasi / program besar dengan mengabaikan faktor infrastruktur. Infrastruktur sendiri adalah hal yang penting dan terkadang membebani proses development. Apabila terdapat gangguan internet atau server yang tidak beroperasi maksimal maka sistem tidak akan berjalan. *Heroku* dapat dikatakan sebagai perantara antara infrastruktur dan bagian penyedia kebutuhan sehingga *developer* dapat fokus dalam pembuatan program yang sesuai keinginannya tanpa terbebani *maintenance* infrastruktur untuk pemrograman. Selain itu, *Heroku* mendukung hampir seluruh bahasa *open source*. Hal ini memungkinkan penggunaanya untuk memilih dan mengkombinasikan berbagai bahasa pemrograman, apabila dapat diimplementasi oleh siapapun maka besar kemungkinan untuk berbagai *developer* berkolaborasi dalam satu *platform* ini (Rockford, 2017).

3.2.17 Label Encoder

Label Encoder adalah teknik pengkodean yang populer untuk menangani variabel kategori. Setiap label diberi bilangan bulat unik berdasarkan urutan abjad (Sethi, 2020). *Label Encoder* mengacu pada pengubahan label menjadi bentuk numerik untuk mengubahnya menjadi bentuk yang dapat dibaca mesin. Algoritma *machine learning* kemudian dapat memutuskan dengan cara yang lebih baik tentang bagaimana label tersebut harus dioperasikan. Ini adalah langkah pra-pemrosesan

yang penting untuk menciptakan suatu kumpulan data terstruktur dalam proses *supervised machine learning* (Chugh, 2019).



BAB IV

METODOLOGI PENELITIAN

4.1. Populasi dan Sampel Penelitian

Populasi dari penelitian adalah semua data kosten yang ada berupa Harga, Daerah, Tipe Kost dll yang terdapat pada website www.mamikos.com. Sampel yang digunakan adalah data kosten yang ada di wilayah Provinsi Yogyakarta. Objek penelitian adalah Harga Sewa, Daerah, Kecamatan, WiFi, Kamar Mandi, Akses 24 jam, Jangka Waktu Sewa. Penelitian ini juga menggunakan set data kost eksklusif yang diiklankan di website.

4.2. Tempat dan Waktu Penelitian

Data penelitian ini diambil melalui situs <https://www.mamikos.com/> menggunakan metode *web scrapping*. Pengambilan data dilakukan pada bulan Juli 2020.

4.3. Jenis Penelitian

Jenis Penelitian berupa analisis statistika deskriptif dan metode *random forest* untuk penentuan prediksi harga sewa kost. Algoritma model yang telah terbentuk dari analisis *random forest* disimpan dan dikembangkan menggunakan *Flask* dan *Heroku*. Hasil yang akan ditampilkan adalah *prototype* yang dikembangkan dengan *web* aplikasi menampilkan 7 *form* berupa variabel kriteria kosten seperti Harga Sewa, Daerah, Kecamatan, WiFi, Kamar Mandi, Akses 24 jam, Jangka Waktu Sewa.

4.4. Data dan Variabel Penelitian

Data merupakan data sekunder dimana peneliti mengambil dalam situs <https://www.mamikos.com/> adapun Definisi Operasional Variabel dari data berupa :

Tabel 4. 1 Definisi Operasional Variabel (DOV)

No	Nama Variabel	Kategori	DOV	Satuan/ Skala	Contoh
1	Harga Sewa	-	Nilai barang atau uang yang harus dibayarkan oleh konsumen atas kost yang disewa	Numerik	750000
2	Daerah	0 = Bantul 1 = Gunung Kidul 2 = Kota Yogyakarta 3 = Kulon Progo 4 = Sleman	Pembagian wilayah yang akan dihuni berdasarkan tingkat kabupaten/kota	Kategorik	Sleman
3	Tipe Kost	0 = Campur 1 = Putra 2 = Putri	Tipe atau jenis untuk membedakan penggunaan kamar kost untuk putra, putri, atau campur.	Kategorik	Campur
4	Kecamatan	0 = Banguntapan 1 = Danurejan 2 = Depok 3 = Gede ... 25 = Wonosari	Pembagian wilayah yang akan dihuni berdasarkan tingkat kecamatan didalam suatu daerah tertentu.	Kategorik	Wonosari
5	WiFi	0 = Tidak ada WiFi 1 = WiFi	Fasilitas jaringan internet yang disediakan pemilik kamar kost.	Kategorik	WiFi
6	K. Mandi Dalam	0 = K. Mandi Dalam 1 = K. Mandi Luar	Fasilitas kamar mandi yang ditawarkan, kamar mandi dalam untuk kost yang disediakan kamar mandi didalam kost dan bersifat pribadi sedangkan kamar mandi luar disediakan untuk umum bersama penghuni kost yang lain.	Kategorik	K. Mandi Dalam
7	Akses 24 Jam	0 = Akses 24 Jam 1 = Tidak Ada Akses	Fasilitas dimana penghuni kost bisa keluar masuk ke dalam kost dalam 24 jam.	Kategorik	Akses 24 jam
8	Jangka waktu	0 = Bulanan 1 = Harian 2 = Mingguan 3 = Tahunan	Lamanya penyewaan kamar yang ditawarkan kepada penghuni kostan	Kategorik	Mingguan

Variabel terbagi menjadi dependen dan independen. variabel berupa:

1. Variabel dependen merupakan variabel terikat yang dipengaruhi oleh variabel independen. Variabel dependen dalam penelitian ini Harga sewa dengan satuan skala numerik
2. Variabel independen merupakan variabel yang memberikan pengaruh terhadap variabel dependen. Adapun variabel independen dalam penelitian ini adalah Daerah, Tipe kost, Kecamatan, *WiFi*, Akses 24 jam, Kamar mandi, Jangka Waktu.

4.5. Metode Analisis

Software yang digunakan pada penulisan ini adalah *Microsoft Excel 2016*, dan *Anaconda3 5.3.1*. Ada beberapa metode analisis data yang digunakan pada penulisan ini, yaitu :

- a. Analisis deskriptif, digunakan untuk menggambarkan dan memetakan gambaran umum data kostan di Provinsi Yogyakarta.
- b. Analisis *random forest*, digunakan dalam membentuk model prediksi harga sewa kost beserta melihat error dan akurasi dari model.
- c. Model yang terbentuk *dideploy* dengan rancangan *web* yang telah dibentuk dengan *Notepad++* melalui *Command Prompt* yang bisa diakses melalui *localhost*.
- d. Pembuatan *system prototype* yang hanya bisa diakses melalui *localhost* *dideploy* ke *Heroku* dengan *Git*, *jinja sklearn* menggunakan *Command Prompt*.

1.6. Alat dan Cara Organisir Data

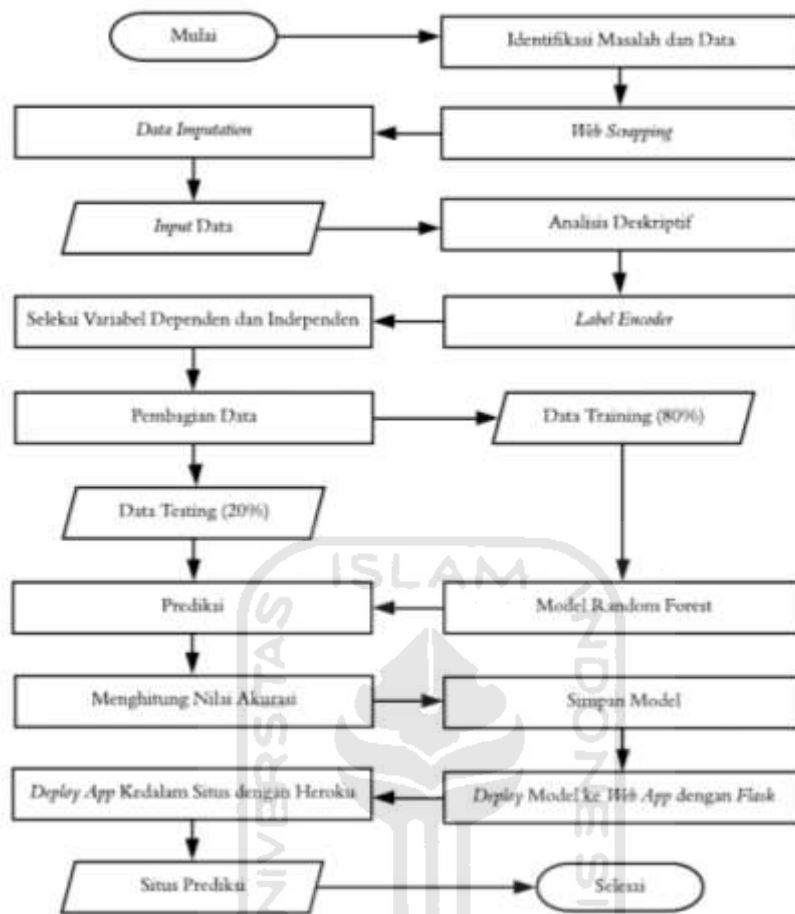
Proses analisis dalam penelitian ini menggunakan bantuan *Microsoft Excel 2016*, dan *Anaconda3 5.3.1*. Ada beberapa metode yang digunakan dalam penelitian ini, diantaranya sebagai berikut :

- a. Identifikasi masalah dan objek penelitian yang digunakan. Dilanjutkan pengambilan data menggunakan *Parsehub* ketika data didapatkan maka dilanjutkan dengan metode *imputation* untuk menghilangkan *missing value* dengan modus. Adapun dilanjutkan dengan melihat statistika deskriptif.

- b. Melakukan *input* data ke dalam *Python* serta memberikan *label encoder* terhadap data, juga menyeleksi antara variabel dependen dan independen. Data yang ada kemudian dibagi menjadi data *train* dan data *testing* untuk pembentukan model yang akan digunakan untuk pengembangan *web* aplikasi.
- c. Dari berbagai percabangan pohon *random forest* yang terbentuk dilihat tingkat kesalahan terkecil dan akurasi terbesar model yang terbaik dan disimpan dalam bentuk *pickle*.
- d. Peneliti membentuk tampilan *web* berisi *form* yang merupakan kriteria dalam memilih kost. Tampilan *web* yang terbentuk dihubungkan dengan *script app.py* yang *dideploy* menggunakan *Command Prompt* dari sini *prototype* bisa digunakan hanya melalui *localhost*.
- e. Pengembangan *prototype* yang hanya bisa diakses melalui *localhost* dihubungkan dengan *Heroku* dengan mengaktifkan fungsi *Git* dan *CLI Heroku* jika terhubung maka *prototype* bisa digunakan ketika mengklik *url* yang terbentuk.

4.7 Flowchart

Dari penjabaran poin 4.6 maka bisa dilihat dalam bentuk diagram alir sebagai berikut :



Gambar 4. 1 *Flowchart Penelitian*

BAB V

HASIL DAN PEMBAHASAN

Imputasi data pada penelitian ini digunakan untuk memberikan atau memperkirakan nilai pada missing data dengan cara mendapatkan pola dari data. Data penelitian memiliki nilai *missing value* yang persentasenya lebih dari 2% dimana *missing value* terdapat pada variabel wifi, k.mandi dalam dan akses 24 jam.

Tabel 5. 1 Data Missing Value

A	B	C	D	E	F	G	H
Daerah	Tipe_Kost	Kecamatan	WiFi	K._Mandi_Dalam	Akses_24_Jam	Jangka_Waktu	Harga_Sewa
Kota	Putri	Gondomanan	WiFi	K. Mandi Dalam		Harian	250000
Kota	Campur	Danurejan		K. Mandi Luar	Akses 24 Jam	Harian	125000
Kota	Campur	Danurejan	Tidak Ada WiFi		Akses 24 Jam	Harian	175000
Kota	Campur	Danurejan	Tidak Ada	K. Mandi Dalam		Harian	225000
Kota	Campur	Gedong Tengen		K. Mandi Dalam	Akses 24 Jam	Harian	145000
Kota	Campur	Gedong Tenger WiFi		K. Mandi Luar	Akses 24 Jam	Harian	50000
Kota	Putri	Danurejan	WiFi	K. Mandi Dalam	Tidak Ada Akses 24	Harian	100000

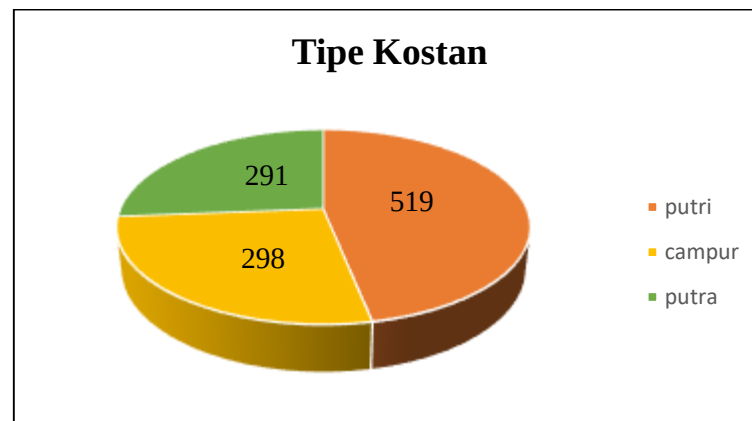
Tabel 5. 2 Data imputation.

	A	B	C	D	E	F	G	H
1	Daerah	Tipe_Kost	Kecamatan	WiFi	K._Mandi_Dalam	Akses_24_Jam	Jangka_Waktu	Harga_Sewa
2	Kota	Putri	Gondomanan	WiFi	K. Mandi Dalam	Akses 24 Jam	Harian	250000
3	Kota	Campur	Danurejan	Tidak Ada	K. Mandi Luar	Akses 24 Jam	Harian	125000
4	Kota	Campur	Danurejan	Tidak Ada	K. Mandi Dalam	Akses 24 Jam	Harian	175000
5	Kota	Campur	Danurejan	Tidak Ada	K. Mandi Dalam	Akses 24 Jam	Harian	225000
6	Kota	Campur	Gedong Tenger WiFi		K. Mandi Dalam	Akses 24 Jam	Harian	145000
7	Kota	Campur	Gedong Tenger WiFi		K. Mandi Luar	Akses 24 Jam	Harian	50000
8	Kota	Putri	Danurejan	WiFi	K. Mandi Dalam	Tidak Ada Akses 24	Harian	100000
9	Kota	Campur	Danurejan	WiFi	K. Mandi Dalam	Akses 24 Jam	Harian	165000
10	Kota	Putra	Danurejan	Tidak Ada	K. Mandi Luar	Akses 24 Jam	Harian	30000
11	Kota	Campur	Gedong Tenger WiFi		K. Mandi Luar	Akses 24 Jam	Harian	60000

Data tabel 5.1 dan 5.2 merupakan contoh *Data imputation*. *Data Imputation* yang dilakukan kali ini adalah dengan melihat modus, karena data berupa kategorik. Konsep tersebut juga dilakukan pada *missing value* yang terdapat di variabel lainnya.

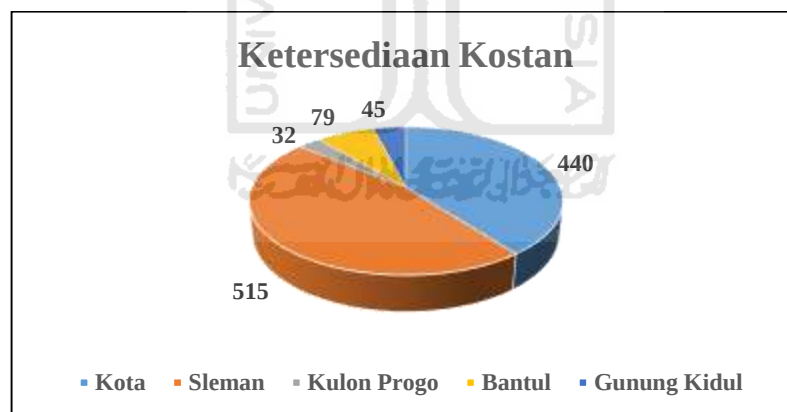
5.1 Analisis Deskriptif

Analisis deskriptif adalah statistik yang digunakan untuk menganalisa data dengan cara mendeskripsikan atau menggambarkan data sebelum dilakukan analisis. Analisis deskriptif penelitian kali ini untuk melihat gambaran variabel independen yaitu harga sewa juga melihat gambaran dari variabel dependen berupa WiFi, Tipe kostan, Daerah, Akses 24 jam dan Jangka Waktu Sewa.



Gambar 5. 1 Tipe Koston.

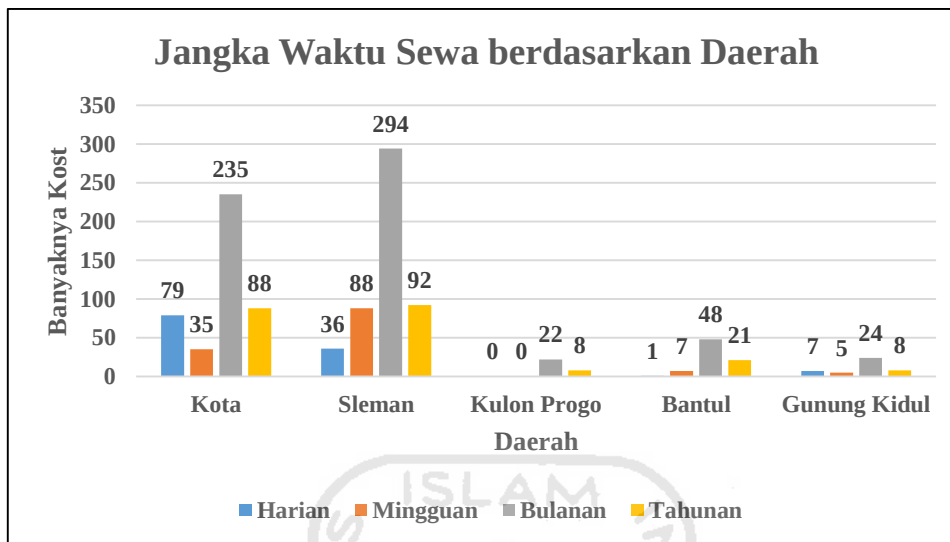
Chart pada gambar 5.1 diketahui bahwa situs website mamikos paling banyak mengiklankan tipe koston putri sebesar 46% dengan jumlah 519 koston digambarkan dengan pie chart berwarna *orange*. Pada chart tipe koston juga untuk koston campur dan koston putri jumlah yang diiklankan di *website* hampir sama yaitu dengan jumlah sebesar 291 koston putra atau sebesar 26%, tipe koston campur yang diiklankan di *website* sebesar 298 atau sebesar 27% dari keseluruhan koston yang diiklankan.



Gambar 5. 2 Ketersediaan Koston

Ketersediaan koston daerah Yogyakarta yang terhimpun dari 5 daerah yaitu Kota Yogyakarta, Sleman, Gunung Kidul, Kulon Progo dan Bantul bisa dilihat pada gambar 5.2 Kabupaten Sleman memiliki ketersediaan koston terbanyak yaitu sebesar 515 koston, diikuti oleh Kota Yogyakarta sebanyak 440 koston, pada kabupaten lainnya memiliki jumlah koston yang diiklankan pada *website* yang lebih sedikit seperti pada Kabupaten Bantul hanya 79 koston, 45 koston tersedia pada

Kabupaten Gunung Kidul dan paling sedikit yang tersedia pada *website* adalah Kabupaten Kulon Progo sebanyak 32 kosten.



Gambar 5. 3 Jangka Waktu Sewa berdasarkan Daerah

Pada gambar 5.3 untuk jangka waktu sewa harian terbanyak terdapat di daerah Kota Yogyakarta sebanyak 79 kosten sedangkan pada Kabupaten Kulon Progo tidak terdapat kost harian. Kabupaten Sleman memiliki jumlah terbanyak untuk jangka waktu sewa mingguan, bulanan, dan Tahunan dengan jumlah 88 kost mingguan, 294 kost bulanan, dan 92 kost tahunan. Kabupaten Kulon Progo hanya terdapat kosten untuk bulanan dan tahunan. Daerah Bantul terdapat 1 kost harian, 7 mingguan, 48 kost bulanan, dan 21 kost tahunan. Pada Kabupaten Gunung Kidul terdapat 7 kost harian, 5 kost mingguan, 24 kost bulanan, dan 8 tahunan.

Tabel 5. 3 Rata-rata Harga Sewa berdasarkan Jenis Kost dan Daerah

Waktu	Jenis Kost	Daerah				
		Bantul	Gunung Kidul	Kota	Kulon Progo	Sleman
Harian	Campur	-	Rp 225.000	Rp 137.149	-	Rp 295.000
	Putra	-	Rp 300.000	Rp 88.545	-	Rp 66.200
	Putri	-	Rp 23.750	Rp 170.000	-	Rp 113.846
Mingguan	Campur	Rp 810.000	-	Rp 727.400	-	Rp 1.163.590
	Putra	-	-	Rp 333.333	-	Rp 688.197
	Putri	Rp 300.000	Rp 144.000	Rp 796.875	-	Rp 604.435
Bulanan	Campur	Rp 1.433.333	Rp 672.222	Rp 2.043.813	Rp 550.000	Rp 1.873.068
	Putra	Rp 514.412	Rp 600.000	Rp 709.623	Rp 391.667	Rp 919.763
	Putri	Rp 653.478	Rp 468.750	Rp 828.472	Rp 564.474	Rp 954.540
Tahunan	Campur	Rp 21.000.000	Rp 5.433.333	Rp 14.896.556	-	Rp 18.760.000
	Putra	Rp 5.380.000	-	Rp 8.567.568	Rp 9.125.000	Rp 8.787.209
	Putri	Rp 7.517.500	Rp 4.840.000	Rp 8.557.813	Rp 5.020.000	Rp 8.669.355

Pada tabel 5.3, peneliti mengabaikan variabel WiFi, Akses 24 Jam, dan Kamar Mandi dan hanya menggunakan Jenis Kos, Jangka Waktu Sewa, Daerah dan rata-rata Harga Sewa. Peneliti menggunakan warna merah untuk mendeteksi nilai rata-rata tertinggi berdasarkan waktu sewa dan daerah dengan penjabaran sebagai berikut:

- a. Berdasarkan ketersediaan jenis kost, tidak terdapat kost harian di daerah Bantul dan Kulon Progo, untuk jenis kost mingguan tidak terdapat di Kulon Progo.
- b. Berdasarkan jangka waktu sewa :
 - Untuk jangka waktu harian rata-rata harga tertinggi untuk kost campur terdapat di Sleman, rata-rata harga tertinggi untuk kost putra terdapat di Gunung Kidul, rata-rata harga untuk kost putri tertinggi terdapat di Kota Yogyakarta.
 - Untuk jangka waktu mingguan rata-rata harga tertinggi untuk kost campur terdapat di Sleman, rata-rata harga tertinggi untuk kost putra terdapat di Sleman, rata-rata harga untuk kost putri tertinggi terdapat di Kota Yogyakarta.
 - Untuk jangka waktu bulanan rata-rata harga tertinggi untuk kost campur terdapat di Kota Yogyakarta, rata-rata harga tertinggi untuk kost putra terdapat di Sleman, rata-rata harga untuk kost putri tertinggi terdapat di Kota Yogyakarta.
 - Untuk jangka waktu Tahunan rata-rata harga tertinggi untuk kost campur terdapat di Bantul, rata-rata harga tertinggi untuk kost putra terdapat di Kulon Progo, rata-rata harga untuk kost putri tertinggi terdapat di Kota Yogyakarta.
- c. Secara keseluruhan, rata-rata harga sewa kost tanpa memperhatikan variabel WiFi, Akses 24 Jam, dan Kamar Mandi dan hanya menggunakan Jenis Kos, Jangka Waktu Sewa, Daerah dan rata-rata Harga Sewa, maka Daerah Kota Yogyakarta memiliki kecenderungan harga kost yang tinggi dimana terdapat 5 kriteria kost dengan harga rata-

rata tertinggi disusul Sleman terdapat 4 kriteria kost, dan untuk Kulon Progo, Gunung Kidul, dan Bantul hanya terdapat masing-masing 1 kriteria kost dengan harga tertinggi.

5.2 Analisis *Random Forest*

Analisis *random forest* digunakan untuk membuat sebuah model, model *random forest* didapatkan dari data sampel yang digunakan. Model yang terbentuk digunakan kembali pada data yang baru untuk melihat gambaran sebenarnya, Dalam hal ini penelitian menggunakan 8 variabel berupa; Daerah, Tipe Kost, Kecamatan, Wifi, Kamar Mandi, Akses 24 jam dan Harga Sewa. Harga sewa merupakan variabel independen dengan tipe data numerik yang akan diprediksi nilainya dari variabel lain dimana variabel tersebut berupa data yang bersifat kualitatif. Penelitian ini menggunakan data sebanyak 1111 data kost yang terhimpun di Yogyakarta.

	Daerah	Tipe_Kost	Kecamatan	WiFi	K_Mandi_Dalam	Akses_24_Jam	Jangka_Waktu
0	2	2	6	1	0	0	1
1	2	0	1	0	1	0	1
2	2	0	1	0	0	0	1
3	2	0	1	0	0	0	1
4	2	0	4	1	0	0	1

Gambar 5. 4 Label Encoder

```
#Daftar Kategori
mapping_dict={}
for col in category_col:
    dataset[col] = LabelEncoder.fit_transform(dataset[col])
    le_name_mapping = dict(zip(LabelEncoder.classes_, LabelEncoder.transform(LabelEncoder.classes_)))
    mapping_dict[col]=le_name_mapping
print(mapping_dict)
```

```
{'Daerah': {'Bantul': 0, 'Gunung Kidul': 1, 'Kota': 2, 'Kulon Progo': 3, 'Sleman': 4}, 'Tipe_Kost': {'Campur': 0, 'Putra': 1, 'Putri': 2}, 'Kecamatan': {'Banguntapan': 0, 'Danurejan': 1, 'Depok': 2, 'Gede': 3, 'Gedong Tengen': 4, 'Gondokusuman': 5, 'Gondokusuman': 6, 'Inogiri': 7, 'Jetis': 8, 'Kraton': 9, 'Mantriweron': 10, 'Mergangsan': 11, 'Mlati': 12, 'Ngaglik': 13, 'Ngemplak': 14, 'Ngemplak': 15, 'Pakualaman': 16, 'Pangasih': 17, 'Sewon': 18, 'Sleman': 19, 'Tegalrejo': 20, 'Temon': 21, 'Umbulharjo': 22, 'Umbulmartani': 23, 'Mates': 24, 'Monosari': 25}, 'WiFi': {'Tidak Ada WiFi': 0, 'WiFi': 1}, 'K_Mandi_Dalam': {'K. Mandi Dalam': 0, 'K. Mandi Luar': 1}, 'Akses_24_Jam': {'Akses 24 Jam': 0, 'Tidak Ada Akses 24 Jam': 1}, 'Jangka_Waktu': {'Bulanan': 0, 'Harian': 1, 'Mingguan': 2, 'Tahunan': 3}}
```

Gambar 5. 5 Label encoder dengan Value

Tahapan selanjutnya adalah *label encoder*, pada tahap ini prosel pelabelan pada data kategorik menjadi data dengan skala nominal. Dari gambar dapat diketahui terdapat 5 kategorik untuk variabel Daerah, 3 kategorik untuk Tipe Kost, 25 kategorik untuk Kecamatan, 2 kategorik untuk Wifi, 2 kategorik untuk Kamar mandi dalam, 2 kateorik untuk akses 24 jam, 4 kategorik untuk Jangka waktu. Pelabelan data ini dimulai dari nilai 0,1,2, dan seterusnya.

Penelitian ini dilanjutkan dengan membagi data menjadi 2 yaitu *training* dan *testing*, data *training* digunakan untuk melatih algoritma untuk membentuk sebuah model dan data *testing* digunakan untuk mengetahui algoritma yang sudah dilatih dan terbentuk sebelumnya pada data training. Model dengan performa terbaik dapat menentukan sebuah prediksi, dari sebuah model dievaluasi dari akurasi atau seberapa banyak prediksi yang dibuat yang sesuai dengan kenyataan.

Tabel 5. 4 Pembagian Data *Train* dan *Test*

Daerah	Data Training	Data Testing
Kota	352	88
Sleman	412	103
Kulon Progo	25	7
Bantul	63	16
Gunung Kidul	36	9
Jumlah	888	223

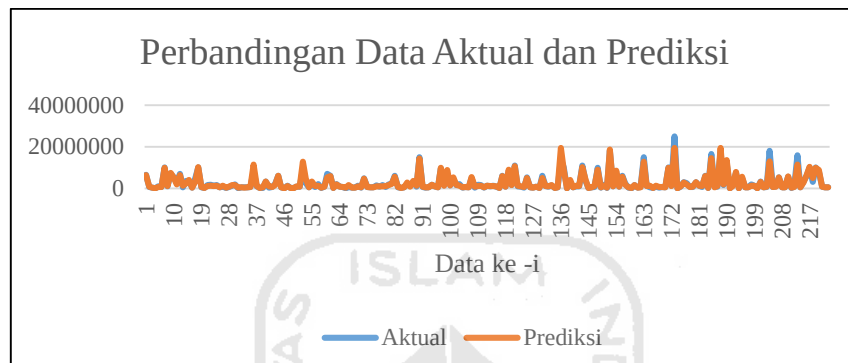
Pembagian data *training* harus lebih besar dibanding dengan *testing*, jika *testing* lebih besar maka akan terjadi overlap dan model tidak akan “belajar”. Proporsi yang digunakan kali ini 0.8:0.2 yang diambil secara acak. Dari 1111 data, terbentuk 888 data training dan 223 data testing. Dari tabel data *training* dan *testing* berdasarkan daerah maka dapat dilihat Daerah Kota Yogyakarta terbagi menjadi 352 data *training* 88 data *testing*, Daerah Sleman terbagi menjadi 412 data *training* 103 data *testing*, Daerah Kulon Progo terbagi menjadi 25 data *training* 7 data *testing*, Daerah Bantul terbagi menjadi 63 data *training* 16 data *testing* dan Daerah Gunung Kidul terbagi menjadi 36 data *training* 9 data *testing*.

Penelitian dilanjutkan dengan analisis *random forest*, sebelum pembentukan model peneliti menentukan jumlah pohon untuk mendapatkan nilai prediksi dengan nilai akurasi terbaik. Penentuan pohon sangat berpengaruh pada nilai akurasi dimana jumlah pohon menentukan nilai *error* yang terbentuk, semakin kecil nilai *error* semakin baik model yang terbentuk.

Tabel 5. 5 Jumlah Pohon dengan *Error*

Jumlah Pohon	MAPE (%)
5	26.81
7	26.57
9	25.62
10	25.49
20	24.37
50	24.41
150	25.58
250	25.91
400	25.59

Perbandingan dilakukan peneliti dengan memilih jumlah pohon yang terbentuk dan melihat nilai *error* terkecil yang dilihat dengan MAPE. Tabel 5.5 menunjukkan jumlah percabangan dengan 20 pohon memiliki *error* terkecil yaitu diangka 24.37% dibanding dengan percabangan lainnya. Dari data *training* diketahui analisis *random forest* dengan 20 pohon didapatkan model terbaik dilanjutkan dengan melihat data prediksi dan aktual dari model, didapatkan seperti;



Gambar 5. 6 Perbandingan Data Aktual dan Prediksi

Gambar 5.6 menunjukkan data *testing* prediksi dengan warna *orange* dan data aktual. Data prediksi yang terbentuk pada grafik mendekati nilai aktual pada beberapa data nilai prediksi bisa diatas ataupun dibawah nilai aktual. Dalam hal ini data bisa dilanjutkan dengan algoritma *random forest*. Ketepatan algoritma ini dapat menghasilkan model yang baik dengan akurasi yang tinggi, sehingga nilai prediksi semakin mendekati nilai aktual.

Tabel 5. 6 Perbandingan Data Aktual dan Prediksi

No	Data ke-i	Harga Aktual	Harga Prediksi
1	385	Rp 6.500.000	Rp 6.340.000
2	652	Rp 600.000	Rp 1.040.000
3	33	Rp 300.000	Rp 238.242
4	76	Rp 200.000	Rp 55.250
5	112	Rp 900.000	Rp 628.500
6	227	Rp 650.000	Rp 735.714
7	376	Rp 10.000.000	Rp 9.757.167
8	680	Rp 900.000	Rp 1.099.783
9	1102	Rp 7.000.000	Rp 7.215.000
10	385	Rp 6.500.000	Rp 6.340.000
...	...	...	...
223	852	Rp 450.000	Rp 506.702

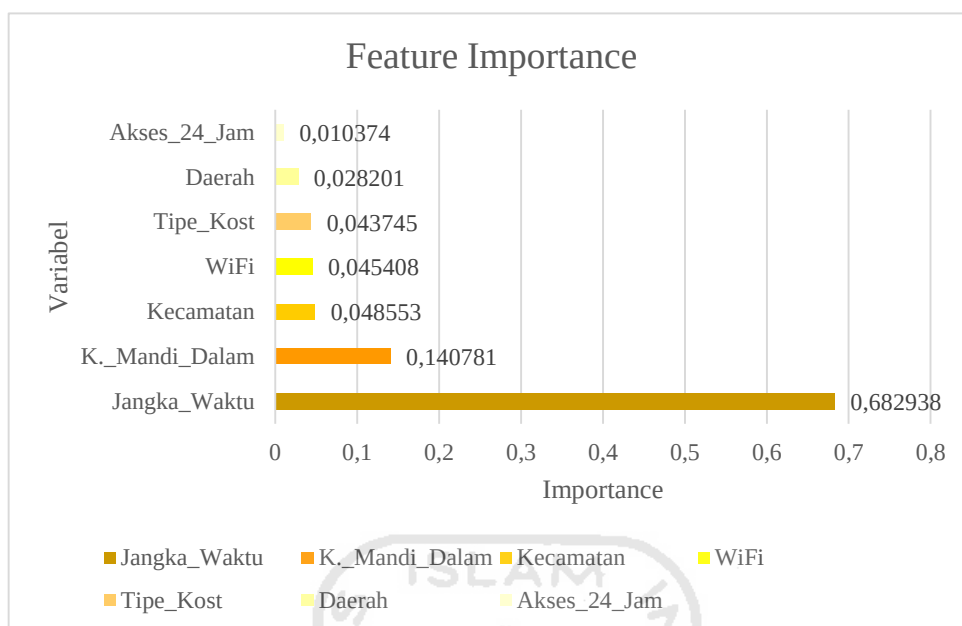
Pada tabel 5.6 adalah nilai dari data *testing* dengan jumlah 223 data yang akan dibandingkan nilainya dengan data aktual, data tersebut adalah nilai harga sewa untuk kost di Daerah Provinsi *Yogyakarta*. Dari data aktual dengan prediksi nilainya tidak berbeda jauh, contoh pada data ke 385 didapatkan harga sewa kost Rp 6.500.000 pada data prediksi didapatkan harga sewa kost Rp 6.340.000.

Tabel 5. 7 R^2 Data, MAPE dan Akurasi

R^2 Data Testing	R^2 Data Training	MAPE	Akurasi
65.91%	94.9%	24.37%	75.62%

Dengan menggunakan percabangan 7 pohon pada pembentukan model *random forest* didapatkan nilai R^2 Data Testing pada data testing sebesar 65.91%, dalam hal ini variabel independen dalam data testing mempengaruhi variabel dependen berupa harga sewa kost sebesar 65.91%, sedangkan 34.09% dipengaruhi oleh faktor lain yang tidak dimasukkan ke dalam model. Pada R^2 Data Training didapatkan nilai 94.9% yang artinya pada data *training* dengan jumlah 888 data variabel independen memberikan pengaruh sebesar 94.9% kepada variabel dependen yaitu harga sewa kost, sedangkan sisanya dipengaruhi faktor lain yang tidak dimasukkan kedalam model.

Nilai MAPE diambil dari 20 pohon percabangan yaitu 24.37% dimana MAPE adalah nilai rata-rata dari keseluruhan selisih antara nilai aktual dan prediksi. Semakin kecil nilai MAPE semakin kecil tingkat kesalahan yang terjadi, dalam hal ini maka akurasi untuk model *random forest* sebesar 75.62%.



Gambar 5. 7 Feature Importance

Feature importance memiliki fungsi memberikan seberapa sering suatu atribut digunakan dalam membangun pohon yaitu dengan menghitung *information gain*. Semakin tinggi nilai tersebut, semakin tinggi tingkat kepentingan atribut tersebut. Pada gambar dan tabel bisa diketahui bahwa variabel Jangka Waktu memberi peranan sangat besar untuk penentuan prediksi harga sewa kost yaitu sebesar 0.682, diikuti variabel Kamar mandi dengan *importance* sebesar 0.1407, pada variabel Kecamatan sebesar 0.048553, variabel WiFi sebesar 0.045, juga variabel Tipe kost 0.043, diikuti variabel Daerah sebesar 0.028, dan variabel yang memiliki pengaruh paling kecil yaitu Akses 24 jam sebesar 0.0103.

```
In [102]: #Save Model
import pickle
pickle.dump(regressor, open('model1.pkl', 'wb'))
model = pickle.load(open('model1.pkl', 'rb'))
print(model.predict([[1,2,24,1,0,1,3]]))

[1285500.]
```

Gambar 5. 8 Save Model Prediksi

Model yang telah diperoleh kemudian disimpan untuk pengembangan *prototype website*, pada *Python* disimpan dengan perintah *pickle*. Model ini juga akan tersimpan dan memudahkan untuk memprediksi suatu data karena peneliti

hanya memanggil model tanpa melakukan analisis yang sebelumnya dilakukan. Pada gambar mencoba memprediksi harga sewa kamera dengan kriteria *model.predict([[1,2,14,1,0,1,3]])* maka secara otomatis model akan membaca kriteria secara berurutan yaitu variabel Daerah, Tipe Kost, Kecamatan, WiFi, K. mandi dalam, Akses 24 jam, Jangka waktu secara berurutan seperti 1 = Gunung Kidul, 2 = Putri, 24 = Wonosari, 1 = WiFi, 0 = K. mandi dalam, 1 = Akses 24 jam, 3 = Tahunan maka prediksi harga sewa kostnya sebesar Rp 12.855.000 .

5.3. Membangun Web Aplikasi Menggunakan Flask dan Heroku

Setelah tahapan pembentukan model, model terbaik akan digunakan menjadi *prototype* aplikasi yang dapat memprediksi harga sewa kost di Yogyakarta. Prediksi didapatkan dengan model *random forest* dimana model dipengaruhi dari beberapa variabel independen seperti Daerah, Kecamatan, WiFi, Tipe kost, Akses 24 jam, K mandi dalam. Dari hal ini dibutuhkan *form* yang berisi opsi dari masing-masing variabel dependen, karena berupa *prototype form* bersifat sederhana yang nantinya bisa dikembangkan.



```

1 <!DOCTYPE html>
2 <html>
3 <head>
4 <title>Prediksi Harga</title>
5 <style type="text/css">
6 .login {
7   margin: 50px auto;
8   width: 500px;
9   height: 400px;
10  padding: 10px;
11  border: 5px solid #00008B;
12 }
13 </style>
14 </head>
15 <body style="background: #6495ED;">
16 <font face="helvetica">
17 <div class="login">
18 <center><div><font color="#00008B">Prediksi: Harga Sewa Kost</font color> <font color="#00008B"></div></center>
19 <br>
20 <!-- Main Input for Receiving Query to our ML -->
21 <center><form action="{ url_for('result') }" method="post">
22 <label for="Daerah">Daerah</label>
23 <select id="Daerah" name="Daerah">
24 <option value="0">Kota</option>
25 <option value="1">Sleman</option>
26 <option value="2">Kulon Progo</option>
27 <option value="3">Bantul</option>
28 <option value="4">Gunung Kidul</option>
29 </select>
30 <br>
31 <label for="Tipe Kost">Tipe Kost</label>
32 <select id="Tipe Kost" name="Tipe Kost">

```

Gambar 5. 9 Home.html

Jika dilihat pada gambar *form* dibentuk dengan *notepad++* dimana pembentukan *form* terbentuk dari data kategori, pada *syntax* pemberian *value* sesuai dengan *label encoder* yang telah dilakukan sebelumnya di *Python* dengan membentuk *form* untuk semua data independen.

```

1  import os
2  import numpy as np
3  import flask
4  import pickle
5  from flask import Flask, render_template, request
6
7  app = Flask(__name__, template_folder='templates')
8  @app.route('/')
9  def student():
10     return render_template("home2.html")
11 def ValuePredictor(to_predict_list):
12     to_predict = np.array(to_predict_list).reshape(1,5)
13     loaded_model = pickle.load(open("model.pkl","rb"))
14     result = loaded_model.predict(to_predict)
15     return round(result[0],2)
16 @app.route('/', methods = ['POST', 'GET'])
17 def result():
18     if request.method == 'POST':
19         to_predict_list = request.form.to_dict()
20         to_predict_list = list(to_predict_list.values())
21         to_predict_list = list(map(float, to_predict_list))
22         result = float(ValuePredictor(to_predict_list))
23         return render_template("home2.html", result_text = "Prediksi Harga Sewaan Rp. {}".format(result))
24 if __name__ == '__main__':
25     app.run(debug = True)
26

```

Gambar 5. 10 App.py

Script selanjutnya adalah *app.py*, script ini memiliki fungsi untuk menghubungkan model yang telah disimpan dalam bentuk *pickle* agar terhubung dengan *form* yang telah dibuat untuk penentuan harga sewa kost. Sistem script ini adalah ketika tampilan *web* menampilkan *form* peneliti memasukan kriteria kost kemudian program akan mengeksekusi dengan model yang terbentuk kemudian keluar nilai prediksi harga sewa kost.

Tahapan selanjutnya untuk pengembangan *web* adalah *deploy* ke *Heroku* agar semua orang bisa memakai, karena sebelum *deploy* program hanya bisa dijalankan didalam *localhost*. *Web* aplikasi yang hanya bisa dijalankan di *localhost* *deploy* dengan script *Flask* didalam *Command Prompt* dengan mengaktifkan beberapa packages seperti *venv*, *git* dan *unicorn jinja2 numpy sklearn*.

```

* Restarting with stat
* Debugger is active!
* Debugger PIN: 314-291-864
* Running on http://127.0.0.1:5000/ (Press CTRL+C to quit)

```

Gambar 5. 11 Web Aplikasi yang Sedang Aktif

Apabila telah aktif maka *web* aplikasi dapat diakses dengan alamat <http://127.0.0.1:5000/>. Hal ini hanya bisa diakses di *localhost* dilanjutkan dengan *deploy* menggunakan *Heroku*. Proses selanjutnya jika rancangan *prototype* sudah *deploy* maka rancangan dapat diakses oleh banyak orang dengan alamat <https://prediksisewakost.herokuapp.com/> saat *web* diakses maka tampilannya seperti:

Gambar 5. 12 Tampilan *Prototype* Prediksi Kost.

Dalam hal ini *prototype web* terbentuk dari 7 *form*, dimana setiap *form* memiliki pilihan yang sesuai dengan kategori dari masing-masing variabel. Cara penggunaan dari *website* ini adalah memilih kriteria kost yang sesuai mengklik masing-masing opsi dari *form* mulai dari Daerah, Tipe Kost, sampai Jangka Waktu kemudian klik *button* Prediksi maka akan keluar nilai prediksi harga sewa kost Daerah Yogyakarta. Jika harga prediksi sesuai dengan kebutuhan dari pencari kos maka akan dilanjutkan dengan mengklik “Ayok Cari Kos” dengan Mudah maka akan diteruskan menuju web mamikos.com dimana situs web tersebut menyediakan banyak kost. Mamikos dalam websitenya dapat menampilkan harga sewa dengan penyesuaian kriteria yang dibutuhkan seperti Daerah, WiFi, Akses 24 jam, Kamar Mandi, Kecamatan, Jenis Kost dan Jangka Waktu Sewa. Untuk melakukan perbandingan harga sewa, pengunjung harus mengecek satu per satu tampilan iklan kost kemudian membandingkan satu sama lain dengan patokan harga sewa yang ditentukan sendiri oleh pengunjung. Hal ini kurang efektif karena harga yang menjadi patokan perbandingan tidak terstandarisasi dan cenderung berbeda antar individu. Berdasarkan permasalahan tersebut, peneliti menerapkan metode *random forest* yang mampu membantu pengunjung menentukan keputusan berupa harga sewa prediksi dengan variabel yang diinginkan. Metode random forest diimplementasikan menjadi sebuah *prototype* dengan *Flask* yang mudah digunakan oleh pengunjung dan dapat mengkalkulasi prediksi harga di suatu daerah dengan fasilitas tertentu.

BAB VI PENUTUP

6.1 Kesimpulan

Berdasarkan penelitian yang telah dilakukan, maka dapat disimpulkan beberapa hal sebagai berikut :

1. Dari 1.111 data didapatkan Kab. Sleman memiliki jumlah kosten terbanyak sebesar 515 kosten dan paling sedikit Kab. Kulon Progo sebesar 32 kosten, dimana dari data tersebut 47% atau sebesar 519 data berupa kost putri. Dari data rata-rata harga sewa berdasarkan daerah wilayah Kota memiliki rata-rata harga sewa tertinggi untuk jangka waktu bulanan, sedangkan untuk jangka waktu harian, mingguan, dan tahunan wilayah Sleman rata-rata harga sewa lebih besar dibanding wilayah lain.
2. Model didapatkan dari analisis *random forest* dengan 20 percabangan pohon, variabel yang memberi banyak peran dalam pembentukan model adalah variabel jangka waktu dan variabel dengan pengaruh paling kecil kepada model adalah K. mandi dalam. Dari hal ini didapatkan nilai akurasi prediksi harga sewa kost sebesar 75.62%.
3. *Prototype* yang terbentuk merupakan *web* aplikasi yang dapat diakses semua orang ketika mengklik url <https://prediksisewakost.herokuapp.com/>, dengan penggunaan memilih kriteria kosten yang diinginkan dari 7 *form* yang ada berupa Daerah, Tipe Kost, Kecamatan, K. mandi dalam, WiFi, Akses 24 jam, dan Jangka waktu lalu klik button Prediksi maka akan keluar harga prediksi. Jika sesuai dengan yang diinginkan maka bisa langsung mengklik “Ayok Cari Kost dengan Mudah” maka akan langsung terhubung dengan *website* mamikos.com untuk memesan kosten.

6.2 Saran

Berdasarkan penelitian yang dilakukan, adapun saran yang diberikan oleh penulis sebagai berikut:

1. Pada penelitian selanjutnya diharapkan melakukan penambahan variabel independen seperti fasilitas yang ada didalam kosan seperti ; AC, TV, Lemari, Kasur juga mencari faktor lainnya yang berpengaruh pada pemilihan kost.
2. Penelitian ini dilakukan menggunakan fungsi *random forest regressor*. Pada penelitian selanjutnya dapat menggunakan fungsi *random forest classification* atau metode *machine learning* lainnya.



DAFTAR PUSTAKA

- Aditya Yanuar r. (2018, Juli 28). *Random Forest*. Retrieved from Menara Ilmu Machine Learning: <https://machinelearning.mipa.ugm.ac.id/2018/07/28/random-forest/>
- Aji Primajaya, B. N. (2018). Random Forest Algorithm for Prediction of Precipitation. *Indonesian Journal of Artificial Intelligence and Data Mining (IJAIDM)*, 27-31.
- Alia Hartati, I. Z. (2012). Analisis CART (Classification And Regression Trees) pada Faktor-Faktor yang Mempengaruhi Kepala Rumah Tangga di Jawa Timur Melakukan Urbanisasi. *JURNAL SAINS DAN SENI ITS* , 100-105.
- Alshamsi, A. (2014). Predicting Car Insurance Policies Using Random Forest. *0th International Conference on Innovations in Information Technology*, 128-132.
- Asmita, F. N. (n.d.). *Gaya Hidup*. Retrieved from KOINWORKS: <https://koinworks.com/blog/ngekos-ngontrak-atau-sewa-apartemen/>
- Asmita, F. N. (n.d.). *Lebih Hemat Ngekos, Ngontrak atau Sewa Apartemen?* Retrieved from <https://koinworks.com/blog/ngekos-ngontrak-atau-sewa-apartemen/>: <https://koinworks.com/blog/ngekos-ngontrak-atau-sewa-apartemen/>
- Basuki, A. &. (2003). *Modul Ajar Decision Tree*. Surabaya: ITS.
- Breiman, L. (2001). *Random Forests*. Berkeley: University of California.
- Breiman, L. F. (1984). *Classification and Regression*. New York: Chapman & Hall/CRC.
- Bruno Klaus Afonso, L. M. (2020). *Housing Price Prediction with a Deep Learning and Random Forest Ensemble*. São Paulo: University of São Paulo.
- Chugh, A. (2019, Agustus 07). *ML | Label Encoding of datasets in Python*. Retrieved from GeeksforGeeks: <https://www.geeksforgeeks.org/ml-label-encoding-of-datasets-in-python/>
- Daniel T. Larose, C. D. (2014). *Discovering Knowledge in Data: An Introduction to Data Mining*. Hoboken, New Jersey: Wiley.

- Dewi, N. K. (2011). PENERAPAN METODE RANDOM FOREST DALAM DRIVER ANALYSIS. *Forum Statistika dan Komputasi*, 35-43.
- Draxl, V. (2018). *Web Scraping Data Extraction from websites*. Austria: Technikum Wien.
- Fransiska Amalia Kurniawan, A. A. (2011). *ANALISIS DAN IMPLEMENTASI RANDOM FOREST DAN CLASSIFICATION DAN REGRESSION TREE (CART) UNTUK KLASIFIKASI PADA MISUSE INTRUSION DETECTION SYSTEM*. Bandung: Telkom University.
- Ginting, V. S. (2020). PENERAPAN ALGORITMA C4. 5 DALAM MEMPREDIKSI KETERLAMBATAN PEMBAYARAN UANG SEKOLAH MENGGUNAKAN PYTHON. *JurTI (Jurnal Teknologi Informasi)*, 1-6.
- Hajar, S. S. (2005). Faktor-faktor yang Memengaruhi Keputusan Mahasiswa dalam Memilih Rumah Kost. *e-Jurnal*, 25-31.
- Harista, R. A. (2019). *Penerapan Metode Random Forest Untuk Prediksi Win Ratio Pemain Player Unknown Battleground*. Yogyakarta: Universitas Sanata Dharma.
- Hariyadi, E. (2017, Maret 7). *Membuat Proyek Pertama Heroku*. Retrieved from Codepolitan: <https://www.codepolitan.com/membuat-proyek-pertama-heroku-58b872c6217eb>
- Hasan, M. A. (2015). Feature Selection for Intrusion Detection Using Random Forest. *Journal of Information Security*, 133-144.
- Herdianto. (2013). *Prediksi Kerusakan Motor Induksi Menggunakan Metode Jaringan Saraf Tiruan*. Medan: Universitas Sumatera Utara.
- Hidayat. (2015). Analisis Sentimen terhadap Wacana Politik pada Media Masa Online Menggunakan Algoritma Support Vector Machine dan Naive Bayes. *Jurnal Elektronik Sistem informatikaa dan Komputer (JESIK)*, 1-7.
- HONG, J. H. (2020). A House Price Valuation Bases On The Random Forest Approach: The Mass Appraisal of Residential Property In South Korea. *International Journal of Strategic Property Management*, 150-151.

- indonesia.go.id. (2017). *Daerah Istimewa Yogyakarta*. Retrieved from indonesia.go.id: <https://indonesia.go.id/province/daerah-istimewa-yogyakarta>
- Iqbal, H. (2001). *Pokok-Pokok Materi Statistik 1 (Statistik Deskriptif)*. Jakarta: PT Bumi Aksara.
- Irsyad, M. (2019, Oktober 15). *Ilmu Wajib Bagi Calon Web Developer*. Retrieved from <https://maulanairsyad.com/ilmu-wajib-bagi-calon-web-developer/>
- KBBI. (n.d.). *Pengertian Indekos*. Retrieved from Kamus Besar Bahasa Indonesia: <https://kbbi.web.id/indekos>
- Kim, H. (2018, Agustus 2018). Statistical notes for clinical researchers: simple linear regression 2 - evaluation of regression line.
- Larose, D. T. (2014). *DATA MINING*. United States: WILEY.
- Lebih Hemat Ngekos, Ngontrak atau Sewa Apartemen?* (n.d.). Retrieved from KOINWORKS: <https://koinworks.com/blog/ngekos-ngontrak-atau-sewa-apartemen/>
- Lokhande, P. P. (2015). Efficient Wa Of Web Development Using Python And Flask. *International journal of Advanced Research in Computer Science*, 54-57.
- Lokhande, P. P. (2015). Efficient Way Of Web Development Using Python And Flask. *International Journal of Advanced Research in ComputerScience* (pp. 54-57). India : Anjuman-I-Islam's Kalsekar Technical Campus.
- MAMIKOS. (n.d.). *MAMIKOS*. Retrieved from MAMIKOS: mamikos.com
- Nugraha, W. M. (2018). Penerapan Metode Prototype dalam Perancangan Sistem Informasi Perhitungan Volume dan Cost Penjualan Minuman Berbasis Website. *JUSIM*, 99.
- Oshiro, T. M. (2012). How Many Trees in a Random Forest? *International Workshop on Machine Learning and Data Mining in Pattern Recognition*, 154-158.
- Pahlevi, M. R. (2020, Maret 1). *Memasang Python 3 di Windows (dengan cara yang benar)*. Retrieved from medium:

- <https://medium.com/@elpahlevi/memasang-python-3-di-windows-dengan-cara-yang-benar-193fe15f28a6>
- Pal N., A. P. (2018). How Much Is My Car Worth? A Methodology for Predicting Used Cars' Prices Using Random Forest. *Future of Information and Communication Conference* (pp. 413-422). India: Springer.
- Purwoto, A. (2007). *Panduan Laboratorium Statistik Inferensial*. Jakarta: Gramedia Widiasarana.
- R. Delshi Howsalya Devi, P. K. (2020). PREDICTION OF DISEASES USING RANDOM FOREST CLASSIFICATION ALGORITHM. *Zeichen Journal*, 19-26.
- Riza Widiatmoko, A. S. (2015). Perancangan Furniture pada Hunian Kost. *JURNAL INTRA Vol. 4*, 63-72 .
- Rockford, M. (2017, September 12). *An Introduction to Heroku*. Retrieved from medium: <https://medium.com/@GoRadialspark/an-introduction-to-heroku-c11c6fcbffa>
- Rosenthal, S. (2017). *The International Encyclopedia of Communication Research Methods*. Amerika Serikat: John Wiley & Sons, In.
- Santoso, B. (2014). Prototype Sistem Informasi Pemasaran dengan Strategi Pipeline Management Spancop Berbasis E-Commerce. *Faktor Exacta*, 249-250.
- Sethi, A. (2020, Maret 6). *One-Hot Encoding vs. Label Encoding using Scikit-Learn*. Retrieved from Analyticvidhya: <https://www.analyticsvidhya.com/blog/2020/03/one-hot-encoding-vs-label-encoding-using-scikit-learn/>
- Streiner. (2002). The Case of the Missing Data: Methods of Dealing with Dropouts and other Research Vagaries. *The Canadian Journal of Psychiatry*, 70-77.
- Subroto, G. (2014, 9 29). HUBUNGAN PENDIDIKAN DAN EKONOMI:. *Universitas Negeri Jakarta*, pp. 391 - 400.
- Sugiyono. (2015). *Metode Penelitian Manajemen*. Bandung: Alfabeta.
- Suwanto Sanjaya, E. A. (2015). Pengelompokan Dokumen Menggunakan Winnowing Fingerprint dengan Metode K-Nearest Neighbour. *Jurnal CoreIT*, 50-56.

- Suyanto. (2011). *Artificial Intelegent (Cetakan kedua)*. Bandung: Informatika.
- Taneja, S. (2014). Python as a Tool for Web Server Application Development. *International Journal of Information, Communication and Computing Technology*, 77-78.
- Turland, M. (2010). *php| architect's Guide to Web, Introduction-Web Scraping*.
- Utomo, P. (2009). *Dinamika pelajar dan mahasiswa di sekitar kampus Yogyakarta*. Yogyakarta: Universitas Negeri Yogyakarta.
- Widiastuti, J. (2018). *KLASIFIKASI PEMBIAYAAN WARUNG MIKRO MENGGUNAKAN METODE RANDOM FOREST DENGAN TEKNIK SAMPLING KELAS IMBALANCED*. Yogyakarta: Universitas Islam Indonesia.
- Widiatmoko, R. (2015). Perancangan Furniture pada Hunian Kost. *JURNAL INTRA*, 121-131.
- Wikipedia. (n.d.). *Indekos*. Retrieved from [id.wikipedia.org: https://id.wikipedia.org/wiki/Indekos](https://id.wikipedia.org/wiki/Indekos)
- Wuryanta, A. E. (2004). *Digitalisasi masyarakat: Menilik kekuatan dan kelemahan dinamika era informasi digital dan masyarakat informasi*. Jakarta: Universitas Atma Jaya Yogyakarta.
- Yanuar, A. A. (2019). DAMPAK URBANISASI TERHADAP PENDUDUK TAHUN 2017 DI KOTA YOGYAKARTA. *Jurnal Intra*, 63-72. Retrieved from ResearchGate.

LAMPIRAN

Lampiran 1 Data Harga Sewa Kost

Daerah	Tipe Kost	Kecamatan	WiFi	K. Mandi Dalam	Akses 24 Jam	Jangka Waktu	Harga Sewa
Kota	Putri	Gondomanan	WiFi	K. Mandi Dalam	Akses 24 Jam	Harian	250.000
Kota	Campur	Danurejan	Tidak Ada WiFi	K. Mandi Luar	Akses 24 Jam	Harian	125.000
Kota	Campur	Danurejan	Tidak Ada WiFi	K. Mandi Dalam	Akses 24 Jam	Harian	175.000
Kota	Campur	Danurejan	Tidak Ada WiFi	K. Mandi Dalam	Akses 24 Jam	Harian	225.000
Kota	Campur	Gedong Tengen	WiFi	K. Mandi Dalam	Akses 24 Jam	Harian	145.000
Kota	Campur	Gedong Tengen	WiFi	K. Mandi Luar	Akses 24 Jam	Harian	50.000
Kota	Putri	Danurejan	WiFi	K. Mandi Dalam	Tidak Ada Akses 24 Jam	Harian	100.000
Kota	Campur	Danurejan	WiFi	K. Mandi Dalam	Akses 24 Jam	Harian	165.000
Kota	Putra	Danurejan	Tidak Ada WiFi	K. Mandi Luar	Akses 24 Jam	Harian	30.000
Kota	Campur	Gedong Tengen	WiFi	K. Mandi Luar	Akses 24 Jam	Harian	60.000
Kota	Campur	Danurejan	WiFi	K. Mandi Luar	Tidak Ada Akses 24 Jam	Harian	100.000
Kota	Campur	Danurejan	WiFi	K. Mandi Luar	Akses 24 Jam	Harian	50.000
Kota	Campur	Ngampilan	Tidak Ada WiFi	K. Mandi Dalam	Tidak Ada Akses 24 Jam	Harian	150.000
Kota	Campur	Gedong Tengen	WiFi	K. Mandi Luar	Tidak Ada Akses 24 Jam	Harian	60.000

Kota	Campur	Danurejan	WiFi	K. Mandi Dalam	Akses 24 Jam	Harian	135.000
Kota	Campur	Gondomanan	WiFi	K. Mandi Luar	Akses 24 Jam	Harian	50.000
Kota	Campur	Gedong Tengen	WiFi	K. Mandi Luar	Akses 24 Jam	Harian	50.000
Kota	Campur	Jetis	WiFi	K. Mandi Dalam	Akses 24 Jam	Harian	300.000
Kota	Putri	Danurejan	WiFi	K. Mandi Dalam	Tidak Ada Akses 24 Jam	Harian	250.000
Kota	Putri	Gondomanan	WiFi	K. Mandi Dalam	Akses 24 Jam	Harian	250.000
Kota	Campur	Gedong Tengen	WiFi	K. Mandi Dalam	Akses 24 Jam	Harian	145.000
Kota	Campur	Gedong Tengen	WiFi	K. Mandi Dalam	Akses 24 Jam	Harian	275.000
Kota	Campur	Gedong Tengen	WiFi	K. Mandi Luar	Akses 24 Jam	Harian	50.000
Kota	Campur	Gedong Tengen	WiFi	K. Mandi Dalam	Akses 24 Jam	Harian	190.000
Kota	Campur	Danurejan	WiFi	K. Mandi Dalam	Akses 24 Jam	Harian	165.000
Kota	Putra	Danurejan	Tidak Ada WiFi	K. Mandi Luar	Akses 24 Jam	Harian	30.000
Kota	Campur	Gedong Tengen	WiFi	K. Mandi Luar	Akses 24 Jam	Harian	60.000
Kota	Putri	Jetis	Tidak Ada WiFi	K. Mandi Luar	Akses 24 Jam	Harian	80.000
Kota	Campur	Danurejan	WiFi	K. Mandi Luar	Tidak Ada Akses 24 Jam	Harian	100.000
Kota	Campur	Gedong Tengen	WiFi	K. Mandi Luar	Akses 24 Jam	Harian	90.000
Kota	Putri	Jetis	Tidak Ada WiFi	K. Mandi Luar	Tidak Ada Akses 24 Jam	Harian	100.000

Kota	Campur	Gedong Tengen	WiFi	K. Mandi Dalam	Akses 24 Jam	Harian	120.000
Kota	Campur	Ngampilan	Tidak Ada WiFi	K. Mandi Dalam	Tidak Ada Akses 24 Jam	Harian	150.000
Kota	Campur	Gedong Tengen	WiFi	K. Mandi Dalam	Akses 24 Jam	Harian	300.000
Kota	Campur	Gedong Tengen	WiFi	K. Mandi Luar	Tidak Ada Akses 24 Jam	Harian	60.000
Kota	Campur	Danurejan	WiFi	K. Mandi Dalam	Akses 24 Jam	Harian	135.000
Kota	Campur	Gedong Tengen	WiFi	K. Mandi Dalam	Akses 24 Jam	Harian	135.000
Kota	Campur	Gedong Tengen	WiFi	K. Mandi Dalam	Akses 24 Jam	Harian	150.000
Kota	Putra	Gondomanan	Tidak Ada WiFi	K. Mandi Luar	Akses 24 Jam	Harian	75.000
Kota	Campur	Gondomanan	WiFi	K. Mandi Dalam	Akses 24 Jam	Harian	50.000
Kota	Putra	Gondomanan	Tidak Ada WiFi	K. Mandi Dalam	Tidak Ada Akses 24 Jam	Harian	125.000
Kota	Campur	Umbulharjo	WiFi	K. Mandi Dalam	Akses 24 Jam	Harian	250.000
Kota	Putra	Mergangsan	WiFi	K. Mandi Dalam	Akses 24 Jam	Harian	165.000
Kota	Putri	Gondomanan	WiFi	K. Mandi Luar	Tidak Ada Akses 24 Jam	Harian	75.000
Kota	Putra	Gondomanan	WiFi	K. Mandi Luar	Akses 24 Jam	Harian	75.000
Kota	Putra	Gondomanan	Tidak Ada WiFi	K. Mandi Luar	Tidak Ada Akses 24 Jam	Harian	100.000
Kota	Putra	Gondomanan	WiFi	K. Mandi Dalam	Akses 24 Jam	Harian	150.000

Kota	Putri	Mergangsan	Tidak Ada WiFi	K. Mandi Luar	Tidak Ada Akses 24 Jam	Harian	75.000
Kota	Putri	Mergangsan	WiFi	K. Mandi Dalam	Tidak Ada Akses 24 Jam	Harian	550.000
Kota	Putri	Mergangsan	WiFi	K. Mandi Dalam	Tidak Ada Akses 24 Jam	Harian	350.000
Kota	Putri	Mergangsan	Tidak Ada WiFi	K. Mandi Luar	Tidak Ada Akses 24 Jam	Harian	50.000
Kota	Putra		WiFi	K. Mandi Luar	Akses 24 Jam	Harian	50.000
Kota	Putri	Umbulharjo	Tidak Ada WiFi	K. Mandi Luar	Tidak Ada Akses 24 Jam	Harian	250.000
Kota	Putri	Umbulharjo	Tidak Ada WiFi	K. Mandi Dalam	Tidak Ada Akses 24 Jam	Harian	100.000
Kota	Putra		WiFi	K. Mandi Dalam	Akses 24 Jam	Harian	75.000
Kota	Campur	Mergangsan	WiFi	K. Mandi Dalam	Tidak Ada Akses 24 Jam	Harian	135.000
Kota	Campur	Mergangsan	WiFi	K. Mandi Dalam	Akses 24 Jam	Harian	165.000
Kota	Campur	Umbulharjo	WiFi	K. Mandi Dalam	Akses 24 Jam	Harian	50.999
Kota	Campur	Umbulharjo	WiFi	K. Mandi Dalam	Tidak Ada Akses 24 Jam	Harian	50.000
Kota	Campur	Mantrijeron	WiFi	K. Mandi Dalam	Akses 24 Jam	Harian	150.000
Kota	Campur	Mantrijeron	Tidak Ada WiFi	K. Mandi Dalam	Akses 24 Jam	Harian	125.000
Kota	Campur	Mergangsan	WiFi	K. Mandi Dalam	Tidak Ada Akses 24 Jam	Harian	150.000

Kota	Campur	Mantrijeron	Tidak Ada WiFi	K. Mandi Dalam	Akses 24 Jam	Harian	175.000
Kota	Campur	Mantrijeron	WiFi	K. Mandi Dalam	Tidak Ada Akses 24 Jam	Harian	150.000
Kota	Putri	Umbulharjo	Tidak Ada WiFi	K. Mandi Dalam	Tidak Ada Akses 24 Jam	Harian	75.000
Kota	Campur	Mantrijeron	WiFi	K. Mandi Dalam	Akses 24 Jam	Harian	135.000
Kota	Putri	Umbulharjo	WiFi	K. Mandi Dalam	Tidak Ada Akses 24 Jam	Harian	250.000
Kota	Campur	Mergangsan	WiFi	K. Mandi Dalam	Akses 24 Jam	Harian	135.000
Kota	Campur	Mantrijeron	WiFi	K. Mandi Dalam	Tidak Ada Akses 24 Jam	Harian	135.000
Kota	Campur	Mantrijeron	WiFi	K. Mandi Dalam	Akses 24 Jam	Harian	50.000
Kota	Campur	Mantrijeron	WiFi	K. Mandi Luar	Akses 24 Jam	Harian	75.000
Kota	Putri	Umbulharjo	WiFi	K. Mandi Dalam	Tidak Ada Akses 24 Jam	Harian	175.000
Kota	Campur	Mergangsan	WiFi	K. Mandi Dalam	Akses 24 Jam	Harian	500.000
Kota	Putri	gede	Tidak Ada WiFi	K. Mandi Luar	Akses 24 Jam	Harian	50.000
Kota	Putri	gede	WiFi	K. Mandi Dalam	Akses 24 Jam	Harian	200.000
Kota	Putra	Banguntapan	WiFi	K. Mandi Dalam	Tidak Ada Akses 24 Jam	Harian	99.000
Kota	Putri	gede	WiFi	K. Mandi Dalam	Tidak Ada Akses 24 Jam	Harian	150.000
Kota	Putri	Banguntapan	Tidak Ada WiFi	K. Mandi Luar	Tidak Ada	Harian	20.000

					Akses 24 Jam		
Kota	Putri	Danurejan	WiFi	K. Mandi Dalam	Tidak Ada Akses 24 Jam	Mingguan	1.200.000
Kota	Putri	Gondomanan	WiFi	K. Mandi Dalam	Akses 24 Jam	Mingguan	1.575.000
Kota	Campur	Danurejan	Tidak Ada WiFi	K. Mandi Luar	Akses 24 Jam	Mingguan	500.000
Kota	Campur	Danurejan	Tidak Ada WiFi	K. Mandi Dalam	Akses 24 Jam	Mingguan	900.000
Kota	Campur	Danurejan	Tidak Ada WiFi	K. Mandi Dalam	Akses 24 Jam	Mingguan	1.200.000
Kota	Campur	Gedong Tengen	WiFi	K. Mandi Dalam	Akses 24 Jam	Mingguan	1.000.000
Kota	Campur	Gedong Tengen	WiFi	K. Mandi Luar	Akses 24 Jam	Mingguan	300.000
Kota	Putri	Danurejan	WiFi	K. Mandi Dalam	Tidak Ada Akses 24 Jam	Mingguan	500.000
Kota	Campur	Danurejan	WiFi	K. Mandi Dalam	Akses 24 Jam	Mingguan	892.500
Kota	Putra	Danurejan	Tidak Ada WiFi	K. Mandi Luar	Akses 24 Jam	Mingguan	150.000
Kota	Campur	Gedong Tengen	WiFi	K. Mandi Luar	Akses 24 Jam	Mingguan	360.000
Kota	Campur	Danurejan	WiFi	K. Mandi Luar	Akses 24 Jam	Mingguan	200.000
Kota	Campur	Ngampilan	Tidak Ada WiFi	K. Mandi Dalam	Tidak Ada Akses 24 Jam	Mingguan	700.000
Kota	Campur	Danurejan	WiFi	K. Mandi Dalam	Akses 24 Jam	Mingguan	1.050.000
Kota	Campur	Danurejan	WiFi	K. Mandi Dalam	Akses 24 Jam	Mingguan	850.000
Kota	Campur	Gondomanan	WiFi	K. Mandi Luar	Akses 24 Jam	Mingguan	350.000

Kota	Putri	Danurejan	WiFi	K. Mandi Dalam	Tidak Ada Akses 24 Jam	Mingguan	1.200.000
Kota	Campur	Gedong Tengen	WiFi	K. Mandi Dalam	Akses 24 Jam	Mingguan	1.000.000
Kota	Campur	Gedong Tengen	WiFi	K. Mandi Dalam	Akses 24 Jam	Mingguan	980.000
Kota	Campur	Gedong Tengen	WiFi	K. Mandi Luar	Akses 24 Jam	Mingguan	300.000
Kota	Campur	Gedong Tengen	WiFi	K. Mandi Dalam	Akses 24 Jam	Mingguan	1.300.000

Lampiran 2 Analisis Random forest Menggunakan manual

1. Peneliti mengambil sampel

Daerah	Tipe Kost	Kecamatan	WiFi	K. Mandi Dalam	Akses 24 Jam	Jangka Waktu	Harga Sewa
2	2	11	1	0	0	3	Rp 10.000.000
2	1	22	1	0	0	3	Rp 10.000.000
4	2	2	1	0	0	3	Rp 1.600.000
4	1	2	1	0	0	0	Rp 1.600.000
4	0	2	1	0	0	0	Rp 1.800.000
1	0	25	1	0	0	0	Rp 1.400.000
3	2	24	1	0	0	0	Rp 850.000
0	2	18	1	0	0	0	Rp 850.000
0	0	18	1	0	0	2	Rp 10.000.000
2	0	1	1	0	0	2	Rp 10.000.000

2. Peneliti membagi data menjadi *train* dan *testing*.

	Daerah	Tipe Kost	Kecamatan	WiFi	K. Mandi Dalam	Akses 24 Jam	Jangka Waktu	Harga Sewa
Data Training	2	2	11	1	0	0	3	Rp 10.000.000
	2	1	22	1	0	0	3	Rp 10.000.000
	4	1	2	1	0	0	0	Rp 1.600.000
	4	0	2	1	0	1	0	Rp 1.600.000
	1	0	25	1	0	0	0	Rp 1.800.000
	0	2	18	1	0	0	0	Rp 1.400.000
	0	0	18	1	0	0	2	Rp 850.000
	2	0	1	1	0	0	2	Rp 850.000
Data Testing	4	2	2	1	0	0	3	Rp 10.000.000
	3	2	24	1	0	0	0	Rp 1.600.000

3. Menentukan *node* 1. Hitunglah nilai *entropy* dan *informasi gain* dari masing-masing variabel.

Atribut	Partisi	Total data	850.000	1.400.000	1.600.000	1.800.000	10.000.000	Entropy	Information Gain
Total		8	2	1	2	1	2	2,25	
Daerah	<=2	6	2	1	0	1	2	0	2,25
	>2	2	0	0	2	0	0	0	
Tipe Kost	0	4	2	0	1	1	0	0	2,25
	>0	4		1	1	0	2	0	
Kecamatan	<=15	4	1	0	2	0	1	0	2,25
	>15	4	1	1	0	1	1	0	
WiFi	0	8	2	1	2	1	2	2,25	2,25
	>0	0	0	0	0	0	0	0	
Kamar Mandi	0	8	2	1	2	1	2	2,25	2,25
	>0	0	0	0	0	0	0	0	
Akses	0	7	2	1	1	1	2	2,235926	0,293564
	>0	1	0	0	1	0	0	0	
Jangka Waktu	<=1	4	0	1	2	1	0	0	2,25
	>1	4	2	0	0	0	2	0	

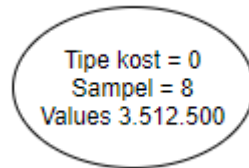
$$\begin{aligned}
 Info(Total) &= - \sum_{i=1}^n p_i \log_2(p_i) \\
 &= -\frac{2}{8} \log_2\left(\frac{2}{8}\right) - \frac{1}{8} \log_2\left(\frac{1}{8}\right) - \frac{2}{8} \log_2\left(\frac{2}{8}\right) - \frac{1}{8} \log_2\left(\frac{1}{8}\right) - \frac{2}{8} \log_2\left(\frac{2}{8}\right) \\
 &= 0,5 + 0,375 + 0,5 + 0,375 + 0,5 \\
 &= 2,25 \\
 Info_A(D) &= \sum_{j=1}^v \frac{|D_j|}{|D|} \times Info(D_j) \\
 &= \frac{1}{8} \times 0 + \frac{7}{8} \times 2,23 \\
 &= 1,95
 \end{aligned}$$

Selanjutnya adalah menghitung nilai informasi gain variabel fisik. Informasi gain tersebut akan digunakan untuk menentukan variabel yang menjadi node-node dalam pembangunan pohon.

$$\begin{aligned}
 Gain(A) &= Info(D) - Info_A(D) \\
 &= 2,25 - 1,95 \\
 &= 0,29
 \end{aligned}$$

Variabel yang digunakan sebagai *node* merupakan variabel yang memiliki nilai informasi gain tertinggi dibandingkan variabel lain. Namun, apabila nilai gain tertinggi terjadi pada beberapa variabel atau dengan kata lain beberapa variabel

memiliki nilai informasi gain yang sama, maka akan dipilih salah satu untuk dijadikan *node* dalam pohon. Seperti yang tercantum dalam tabel diatas yang menunjukkan bahwa variabel produk, merek, model, dan karet kamera memiliki informasi gain yang sama, sehingga variabel produk yang akan dipilih untuk menjadi node dalam pohon. Hasil pohon memiliki tampilan seperti gambar berikut:



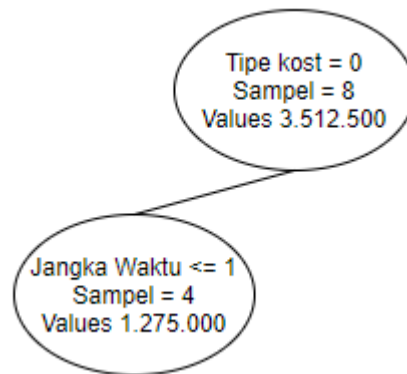
4. Menentukan *node* 2. Urutkan data berdasarkan tipe kost 0 seperti gambar dibawah ini:

Daerah	Tipe Kost	Kecamatan	WiFi	K. Mandi Dalam	Akses 24 Jam	Jangka Waktu	Harga Sewa
4	0	2	1	0	0	0	Rp 1.600.000
1	0	25	1	0	0	0	Rp 1.800.000
0	0	18	1	0	0	2	Rp 850.000
2	0	1	1	0	0	2	Rp 850.000

Selanjutnya hitung kembali nilai entropy dan informasi gain dari masing-masing variabel.

Atribut	Partisi	Total data	850.000	1.400.000	1.600.000	1.800.000	10.000.000	entropy	Information Gain
Total		4	2	-	1	1	-	1,5	
Daerah	<=2	3	2	-	0	1	-	0	1.5
	>2	1	0	-	1	0	-	0	
Tipe Kost	0	4	2	-	1	1	-	0	1.5
	>0	0	0	-	0	0	-		
Kecamatan	<=15	2	1	-	1	0	-	0	1.5
	>15	2	1	-	0	1	-		
WiFi	0	0	0	-	0	0	-	0	0
	>0	4	2	-	1	1	-	1,5	
Kamar Mandi	0	4	2	-	1	1	-	1,5	0
	>0	0	0	-	0	0	-	0	
Akses	0	3	2	-	0	1	-	0	1.5
	>0	1	0	-	1	0	-	0	
Jangka Waktu	<=1	2	0	-	1	1	-	0	1.5
	>1	2	2	-	0	0	-	0	

Hasil pohon memiliki tampilan seperti gambar berikut:



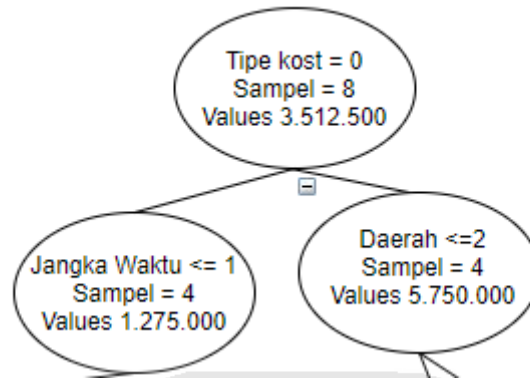
5. Menentukan *node* 3. Urutkan data berdasarkan kategori tipe kost >0, sehingga data yang diperoleh terlihat seperti berikut:

Daerah	Tipe Kost	Kecamatan	WiFi	K. Mandi Dalam	Akses 24 Jam	Jangka Waktu	Harga Sewa
2	2	11	1	0	0	3	Rp 10.000.000
2	1	22	1	0	0	3	Rp 10.000.000
4	1	2	1	0	0	0	Rp 1.600.000
0	2	18	1	0	0	0	Rp 1.400.000

Selanjutnya hitung kembali nilai entropy dan informasi gain dari masing-masing variabel.

Atribut	Partisi	Total data	850.000	1.400.000	1.600.000	1.800.000	10.000.000	entropy	Information Gain
Total		4	-	1	1	-	2	1,5	
Daerah	<=2	3	-	1	0	-	2		1,5
	>2	1	-	0	1	-	0		
Tipe Kost	0	0	-	0	0	-	0		0
	>0	4	-	1	1	-	2	1,5	
Kecamatan	<=15	2	-	0	1	-	1		0
	>15	2	-	1	0	-	1		
WiFi	0	0	-	0	0	-	0		0
	>0	4	-	1	1	-	2	1,5	
Kamar Mandi	0	4	-	1	1	-	2	1,5	0
	>0	0	-	0	0	-	0		
Akses	0	4	-	1	1	-	2	1,5	0
	>0	0	-	0	0	-	0		
Jangka Waktu	<=1	2	-	0	0	-	2		1,5
	>1	2	-	1	1	-	0		

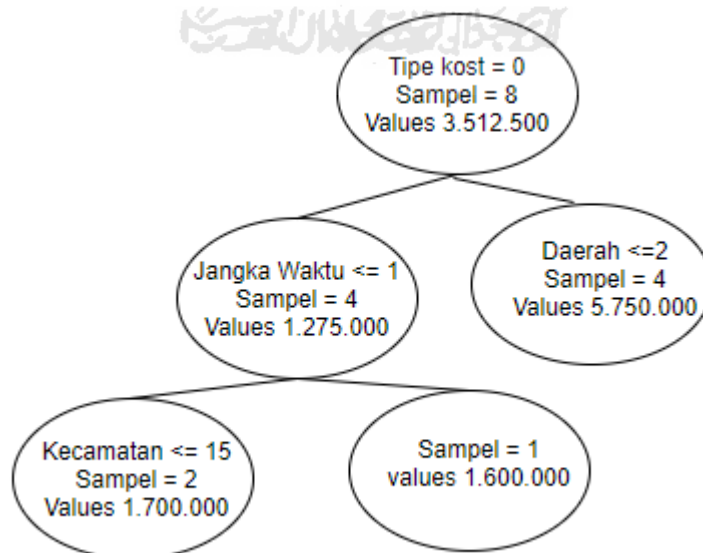
Berdasarkan informasi gain, daerah, dan jangka waktu memiliki nilai yang sama, sehingga akan dipilih salah satu variabel yang akan menjadi *node* 3 yaitu variabel daerah. Bentuk pohon yang tersusun:



6. Menentukan *node* 5. Urutkan data berdasarkan tipe kost = 0, jangka waktu <= 1, kecamatan <=15 . Data yang diperoleh sebagai berikut:

Daerah	Tipe Kost	Kecamatan	WiFi	K. Mandi Dalam	Akses 24 Jam	Jangka Waktu	Harga Sewa
4	0	2	1	0	0	0	Rp 1.600.000

Apabila data yang terbentuk hanya terdiri dari 1 data maka akan menjadi *leaf node* dengan nilai kelas yang tersisa. Bentuk pohon yang tersusun:



7. Menentukan *node* 6. Urutkan data berdasarkan tipe kost > 1 dan daerah <=2

Daerah	Tipe Kost	Kecamatan	WiFi	K. Mandi Dalam	Akses 24 Jam	Jangka Waktu	Harga Sewa
2	2	11	1	0	0	3	10.000.000
2	1	22	1	0	0	3	10.000.000
0	2	18	1	0	0	0	1.400.000

Selanjutnya hitung kembali nilai entropy dan informasi gain dari masing-masing variabel

Atribut	Partisi	Total data	850000	1400000	1600000	1800000	1000000	entropy	Informasi Gain
Total		3	-	1	-	-	2	0,918296	0,918296
Daerah	<=2	3	-	1	-	-	2	0,918296	0
	>2	0	-	0	-	-	0		
Tipe Kost	0	0	-	0	-	-	0		0,166667
	>0	2	-	1	-	-	2	0,5	
Kecamatan	<=15	1	-	0	-	-	1		0,333333
	>15	2	-	1	-	-	1	1	
WiFi	0	0	-	0	-	-	0		0
	>0	3	-	1	-	-	2	0,918296	
Kamar Mandi	0	3	-	1	-	-	2	0,918296	0
	>0	0	-	0	-	-	0		
Akses	0	3	-	1	-	-	2	0,918296	0
	>0	0	-	0	-	-	0		
Jangka Waktu	<=1	1	-	0	-	-	1		0,333333
	>1	2	-	1	-	-	1	1	

8. Menentukan *node* 7. Urutkan data berdasarkan tipe kost >0, daerah > 2. Data yang terbentuk sebagai berikut:

Daerah	Tipe Kost	Kecamatan	WiFi	K. Mandi Dalam	Akses 24 Jam	Jangka Waktu	Harga Sewa
4	1	2	1	0	0	0	1.600.000

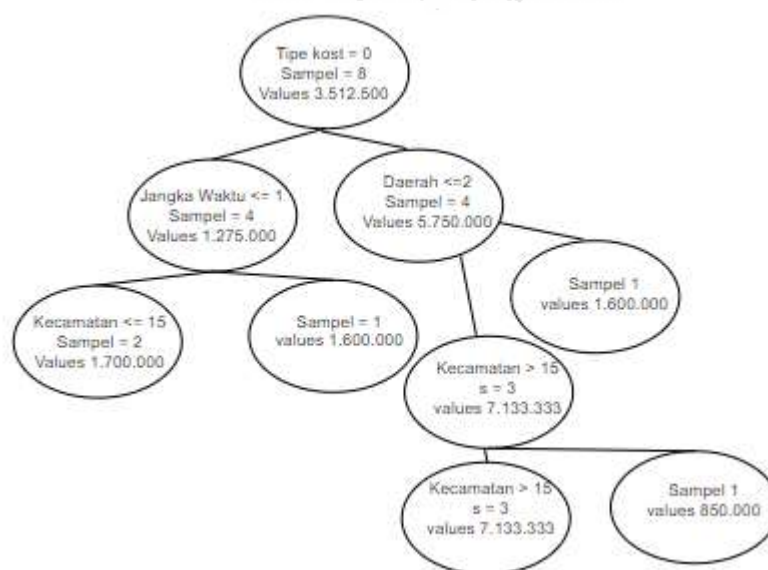
9. Menentukan *node* 8. Urutkan data berdasarkan produk >0, daerah > 2, dan kecamatan >=15 .

Daerah	Tipe Kost	Kecamatan	WiFi	K. Mandi Dalam	Akses 24 Jam	Jangka Waktu	Harga Sewa
1	0	25	1	0	0	0	1.800.000
0	0	18	1	0	0	2	850.000

Entropy yang didapat adalah

Atribut	Partisi	Total data	850.000	1.400.000	1.600.000	1.800.000	10.000.000	entropy	Information Gain
Total		2	1	-	-	1	-	1	
Daerah	≤ 2	2	2	-	-	2	-	0	1
	> 2	0	0	-	-	0	-		
Tipe Kost	0	2	1	-	-	1	-	1	0
	> 0	0	0	-	-	0	-		
Kecamatan	≤ 15	0	0	-	-	0	-		
	> 15	2	1	-	-	1	-	1	0
WiFi	0	0	0	-	-	0	-		
	> 0	2	1	-	-	1	-	1	
Kamar Mandi	0	2	1	-	-	1	-	1	0
	> 0	0	0	-	-	0	-		
Akses	0	2	1	-	-	1	-	1	0
	> 0	0	0	-	-	0	-		
Jangka Waktu	≤ 1	1	0	-	-	1	-		1
	> 1	1	1	-	-	0	-		

Pembentukan pohon 1 telah selesai dilakukan. Pembentukan pohon 2,3, dan seterusnya menggunakan cara yang sama dengan data yang telah diacak pada setiap pembentukan pohon. Setelah pohon terbentuk maka dapat dilakukan prediksi harga kamera bekas dan mengukur akurasi dari model *Random Forest* yang telah terbentuk:



Apabila jumlah pohon yang digunakan lebih dari 1, maka hasil prediksi merupakan rata-rata dari seluruh pohon yang digunakan.

Daerah	Tipe Kost	Kecamatan	WiFi	K. Mandi Dalam	Akses 24 Jam	Jangka Waktu	Harga Sewa	Prediksi
4	2	2	1	0	0	3	10.000.000	7.133.333
3	2	24	1	0	0	0	1.600.000	1.325.000

Lampiran 3 Analisis *Random Forest* Menggunakan *Python*

```
#Aktifkan Package
import pandas as pd
import numpy as np

Input Data
dataset = pd.read_csv('Data Skripsi.csv', sep=";")
dataset.head()

#Label encoder
from sklearn import preprocessing
category_col = ['Daerah', 'Tipe_Kost', 'Kecamatan', 'WiFi',
               'K. Mandi Dalam', 'Akses_24_Jam', 'Jangka Waktu']
labelEncoder = preprocessing.LabelEncoder()

#Daftar Kategorik
mapping_dict={}
for col in category_col:
    dataset[col] = labelEncoder.fit_transform(dataset[col])
    le_name_mapping = dict(zip(labelEncoder.classes_,
labelEncoder.transform(labelEncoder.classes_)))
    mapping_dict[col]=le_name_mapping
print(mapping_dict)

#Variabel Independen
X = dataset.drop(["Harga_Sewa"], axis=1)
X.head()

#Variabel Dependen
y=dataset["Harga_Sewa"]
y.head()

#DataTesting dan Training
from sklearn.model_selection import train_test_split
X_train, X_test, y_train, y_test = train_test_split(X, y,
test_size=0.2, random_state=1234)

#Proporsi Data Testing dan Training
print(X_train.shape)
print(X_test.shape)

from sklearn.ensemble import RandomForestRegressor
regressor = RandomForestRegressor(n_estimators=7,
random_state=1234)
```

```

regressor.fit(X_test, y_test)
y_pred=regressor.predict(X_test)

#Nilai Koefisiens Determinasi Data Testing
from sklearn.metrics import r2_score
r2_score(y_test, y_pred)
y_pred2=regressor.predict(X_train)
r2_score(y_train, y_pred2)

#Nilai Koefisiens Determinasi Data Training
from sklearn.metrics import r2_score
r2_score(y_test, y_pred)
y_pred2=regressor.predict(X_train)
r2_score(y_train, y_pred2)

#Perhitungan Akurasi
from sklearn import metrics
print('Mean Absolute Error:', metrics.mean_absolute_error(y_test,
y_pred))
print('Mean Squared Error:', metrics.mean_squared_error(y_test,
y_pred))
print('Root Mean Squared Error:',
np.sqrt(metrics.mean_squared_error(y_test, y_pred)))
print('MAPE :', np.mean(np.abs((y_test - y_pred) / y_test))*100)
print ('Nilai akurasi :', (1-(np.mean(np.abs((y_test - y_pred) /
y_test))))*100)

import matplotlib.pyplot as plt

#Plot Perbandingan Nilai Aktual dan Prediksi
plt.plot(y, label="Data aktual",color="yellow")
plt.plot(y_pred2, label="Data prediksi")
plt.ylabel('Harga Sewa Kost')
plt.show()

#Feature Importance
# Feature importances into a dataframe
features = list(X_train.columns)
feature_importances = pd.DataFrame({'feature': features,
'importance': regressor.feature_importances_})
feature_importances .plot(x='feature', y='importance', kind='
'barh', color="lime")
print(feature_importances)

#Save Model
import pickle
pickle.dump(regressor, open('model.pkl','wb'))
model = pickle.load(open('model.pkl','rb'))
print(model.predict([[1,2,24,1,0,1,3]]))

```

Lampiran 4 Deploy Flask Script (home2.html)

```
<!DOCTYPE html>
<html>
<head>
  <title>Prediksi Harga</title>
  <style type="text/css">
    .login {
      margin: 50px auto;
      width: 500px;
      height: 400px;
      padding: 10px;
      border: 5px solid #00008B;
    }
  </style>
</head>
<body style="background: #6495ED;">
<font face="helvetica">
  <div class="login">
    <center><h1><font color="#00008B">Prediksi Harga Sewa
Kost</font color> <font color="#00008B"></h1></center>
    <br> <br>

    <!-- Main Input For Receiving Query to our ML -->
    <center><form action="{ url_for('result') }" method="post">
      <label for="Daerah">Daerah</label>
      <select id="Daerah" name="Daerah">
        <option value="0">Kota</option>
        <option value="1">Sleman</option>
        <option value="2">Kulon Progo</option>
        <option value="3">Bantul</option>
        <option value="4">Gunung Kidul</option>
      </select>
      <br>
      <br>
      <label for="Tipe_Kost">Tipe_Kost</label>
      <select id="Tipe_Kost" name="Tipe_Kost">
        <option value="0">Putri</option>
        <option value="1">Campur</option>
        <option value="2">Putra</option>
      </select>
      <br>
      <br>
      <label for="Kecamatan">Kecamatan</label>
      <select id="Kecamatan" name="Kecamatan">
        <optgroup label="Kota">
          <option value="1">Gondomanan</option>
          <option value="2">Danurejan</option>
          <option value="3">Gedong Tengen</option>
          <option value="4">Ngampilan</option>
          <option value="5">Jetis</option>
          <option value="6">Umbulharjo</option>
          <option value="7">Mergangsan</option>
          <option value="8">Mantrijeron</option>
          <option value="9">Gede</option>
          <option value="10">Gondokusuman</option>
          <option value="11">Pakualaman</option>
```

```

<option value="12">Kraton</option>
<option value="13">Tegalrejo</option>
</optgroup>
<optgroup label="Sleman">
<option value="14">Depok</option>
<option value="15">Ngaglik</option>
<option value="16">Mlati</option>
<option value="17">Ngemplak</option>
<option value="18">Umbulmartani</option>
<option value="19">Sleman</option>
</optgroup>
<optgroup label="Gunung Kidul">
<option value="20">Wonosari</option>
</optgroup>
<optgroup label="Bantul">
<option value="21">Imogiri</option>
<option value="22">Sewon</option>
<option value="23">Banguntapan</option>
</optgroup>
<optgroup label="Kulon Progo">
<option value="24">Temon</option>
<option value="25">Wates</option>
<option value="26">Pengasih</option>
</optgroup>
</select>
<br>
<br>
<label for="WiFi">WiFi</label>
<select id="WiFi" name="WiFi">
<option value="0">WiFi</option>
<option value="1">Tidak Ada WiFi</option>
</select>
<br>
<br>
<label for="Kamar_Mandi">Kamar_Mandi</label>
<select id="Kamar_Mandi" name="Kamar_Mandi">
<option value="0">K. Mandi Dalam</option>
<option value="1">K. Mandi Luar</option>
</select>
<br>
<br>
<label for="Akses_24_Jam">Akses_24_Jam</label>
<select id="Akses_24_Jam" name="Akses_24_Jam">
<option value="0">Akses_24_Jam</option>
<option value="1">Tidak Ada Akses 24 Jam</option>
</select>
<br>
<br>
<label for="Jangka_Waktu">Jangka Waktu</label>
<select id="Jangka_Waktu" name="Jangka_Waktu">
<option value="0">Harian</option>
<option value="1">Mingguan</option>
<option value="2">Bulanan</option>
<option value="3">Tahunan</option>
</select>
<br>

```

```
<br>
<input type="submit" value="Prediksi">
</form> </center>

<br>
<center><h4>{{ result_text }}</h4></center>
</div>
<div style='text-align:center;'>
<a href="https://www.mamikos.com/">Ayok Cari Kos dengan Mudah</a>
</div>
</body>
</html>
```



Lampiran 5 *Flask Script (app.py)*

```
import os
import numpy as np
import Flask
import pickle
from Flask import Flask, render_template, request

app = Flask(__name__, template_folder='templatess')
@app.route('/')
def student():
    return render_template("home2.html")
def ValuePredictor(to_predict_list):
    to_predict = np.array(to_predict_list).reshape(1,7)
    loaded_model = pickle.load(open("model.pkl","rb"))
    result = loaded_model.predict(to_predict)
    return round(result[0],2)
@app.route('/',methods = ['POST', 'GET'])
def result():
    if request.method == 'POST':
        to_predict_list = request.form.to_dict()
        to_predict_list=list(to_predict_list.values())
        to_predict_list = list(map(float, to_predict_list))
        result = float(ValuePredictor(to_predict_list))
        return render_template("home.html",result_text = 'Prediksi
Harga Sebesar Rp. {}'.format(result))
if __name__ == '__main__':
    app.run(debug = True)
```

Lampiran 6 Deploy ke Heroku

```
C:\Users\940124\cobadeploy > pip install virtualenv
C:\Users\940124\cobadeploy > virtualenv venv
C:\Users\940124\cobadeploy > cd venv
C:\Users\940124\cobadeploy \venv>cd Scripts
C:\Users\940124\cobadeploy \venv\ Scripts> .\activate
(venv) C:\Users\940124\cobadeploy \venv > cd..
(venv) C:\Users\940124\cobadeploy > pip install Flask gunicorn
jinja2 numpy sklearn
(venv) C:\Users\940124\cobadeploy > pip freeze requirements
(venv) C:\Users\940124\cobadeploy > pip freeze > requirements.txt
```

```
#membuka cmd baru
C:\Users\940124\cobadeploy > git init
C:\Users\940124\cobadeploy > git add .
C:\Users\940124\cobadeploy > git status
C:\Users\940124\cobadeploy > git commit -m "pertama"
C:\Users\940124\cobadeploy > Heroku login
C:\Users\940124\cobadeploy > Heroku create prediksikost
C:\Users\940124\cobadeploy > git remote -v
C:\Users\940124\cobadeploy > git push Heroku master
```