

Implementasi *Machine Learning* untuk Prediksi Harga Mobil Bekas dengan Algoritma Regresi Linear Berbasis Web

Ernianti Hasibuan dan Aldian Karim

Sistem Informasi, Fakultas Ilmu Komputer dan Teknologi Informasi, Universitas Gunadarma
Jl. Margonda Raya No. 100 Pondok Cina Depok Jawa Barat
E-mail: ernianti@staff.gunadarma.ac.id, turafn@gmail.com

Abstrak

Pada saat pembeli ingin membeli mobil terdapat sesuatu masalah yang ditemui, salah satunya terkait kurangnya informasi mengenai harga mobil terutama mobil bekas. Terbatasnya informasi mengenai harga mobil bekas menyebabkan masyarakat yang ingin menjual atau membeli mobil tidak memiliki acuan terhadap harga mobil di pasaran. Informasi harga mobil sangat penting dan diperlukan untuk mengetahui prediksi kisaran harga mobil dengan mengacu pada variabel yang memang bisa diperhitungkan. Salah satu metode untuk memprediksi harga mobil adalah dengan Machine Learning (ML) yang bertujuan untuk melatih model dengan algoritma yang mampu melakukan analisis statistik dengan teknik Supervised Learning yang tepat untuk membuat prediksi. Prediksi adalah salah satu unsur statistika yang sangat penting dalam pengambilan keputusan. Algoritma Regresi Linier merupakan salah satu teknik membangun model yang digunakan untuk prediksi nilai dari data masukan yang diberikan. Pada penelitian ini algoritma Regresi Linier akan digunakan ini untuk memprediksi harga mobil bekas. Dataset yang digunakan untuk melatih model prediksi dengan algoritma Regresi Linier bersumber dari repositori dataset Kaggle untuk data harga mobil bekas. Penelitian ini melakukan prediksi berdasar model dan algoritma regresi linier untuk karakter mobil bekas yang kemudian divisualisasi hasilnya pada website. Proses prediksi menggunakan perbandingan data uji dan data latih pada proses pengujian model yang dibangun memiliki tingkat akurasi sebesar 76%.

Kata kunci : website, prediksi, linear regresi, harga mobil, dataset

Pendahuluan

Penentuan layak tidaknya harga beli atau jual mobil bekas, merupakan salah satu kendala yang dihadapi oleh masyarakat dalam pengambilan keputusan saat akan membeli atau menjual mobil atau kendaraan. Informasi yang ditemukan di internet pun terkadang sangat beragam tanpa ada kejelasan faktor yang mempengaruhi penentuan harga tersebut. Informasi mengenai harga mobil di pasaran menjadi penting dan diperlukan untuk membantu masyarakat mengetahui kisaran harga mobil dengan melihat prediksi harga mobil berdasarkan beberapa karakter dari penentuan harga mobil bekas diantaranya adalah kilometer yang sudah ditempuh (mileage), tahun produksi, model, kapasitas mesin, transmisi, termasuk pula warna meski hal ini sering juga terpengaruh pada trend warna.

Prediksi adalah suatu proses memperkirakan secara sistematis tentang sesuatu yang paling mungkin terjadi di masa depan berdasarkan infor-

masi 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 yang akan terjadi [1].

Salah satu metode komputasi untuk memprediksi harga mobil adalah dengan menggunakan Machine Learning (ML). ML merupakan metode yang membuat sebuah mesin atau komputer dapat belajar dari pengalaman atau bagaimana cara memprogram mesin untuk dapat belajar. ML membutuhkan data untuk belajar sehingga bisa juga diistilahkan dengan learn from data [2].

Metode belajar ML bisa dilakukan dengan menggunakan berbagai algoritma learning. Secara garis besar ada 3 jenis metode belajar yang digunakan yaitu Supervised Learning, Unsupervised Learning dan Reinforcement Learning.

Supervised learning merupakan metode belajar

terawasi. Program diberikan beberapa contoh data yang telah diketahui jenis / klasifikasinya sebagai bahan pembelajaran atau pelatihan [3]. Terdapat kemungkinan program akan salah dalam mengklasifikasi sebuah objek setelah dilatih. Oleh karena itu, selain menggunakan training set, juga memberikan test set. Berdasarkan hal ini akan terukur persentase keberhasilannya.

Unsupervised Learning merupakan metode belajar tidak terawasi. Metode ini menggunakan prosedur yang berusaha untuk mencari partisi dari sebuah pola. Unsupervised learning mempelajari bagaimana sebuah sistem dapat belajar untuk merepresentasikan pola input dalam cara yang menggambarkan struktur statistik dari keseluruhan pola input [3]. Berbeda dari supervised learning, metode ini tidak memiliki target output yang eksplisit sehingga biasa digunakan untuk kebutuhan pengelompokan (clustering).

Reinforcement learning adalah sebuah metode learning yang mempelajari aturan kontrol dengan cara berinteraksi dengan lingkungan yang masih asing [2]. Program akan mendapatkan hukuman jika salah dalam pengambilan keputusan dan hadiah jika benar. Pengalaman interaksi tersebut terakumulasi sehingga program dapat mengambil kesimpulan dikemudian hari menggunakan pola-pola yang telah dipelajarinya.

Pemanfaatan ML ini dapat membantu memprediksi harga mobil dengan melakukan teknik Supervised Learning yang tepat untuk membuat prediksi tersebut. Prediksi adalah salah satu unsur statistika yang sangat penting dalam pengambilan keputusan. Memprediksi dibutuhkan hubungan antar variabel atau kolom, maka algoritma yang digunakan untuk memprediksi harga mobil adalah algoritma Regresi Linier. Dikarenakan Regresi Linier dapat menghitung hubungan dan korelasi antar variabel yang dibutuhkan sehingga hasil prediksi lebih akurat. Pemanfaatan modul phyton terkait dengan memanfaatkan sklearn (scikit-learn) yang cukup mudah digunakan karena menggunakan bahasa tingkat tinggi [4].

Prediksi adalah hasil dari kegiatan memprediksi atau meramal atau memperkirakan nilai pada masa yang akan datang dengan menggunakan data masa lalu. Prediksi menunjukkan apa yang akan terjadi pada suatu keadaan tertentu dan merupakan input bagi proses perencanaan dan pengambilan keputusan. Beberapa Algoritma yang sering digunakan untuk prediksi diantaranya adalah Naïve Bayes[6] dan Regresi Linier[7].

Naive Bayes merupakan sebuah metode klasifikasi yang berakar pada teorema Bayes. Naive Bayes merupakan sebuah metode klasifikasi yang berakar pada teorema Bayes. Naive Bayes merupakan sebuah metode klasifikasi yang berakar pada teorema Bayes. Algoritma Naive Bayes kurang tepat untuk digunakan pada kasus ini karena un-

tuk membuat keputusan, diperlukan pengetahuan awal atau pengetahuan mengenai masa sebelumnya. Keberhasilannya sangat bergantung pada pengetahuan awal tersebut, banyak celah yang bisa mengurangi efektivitasnya.

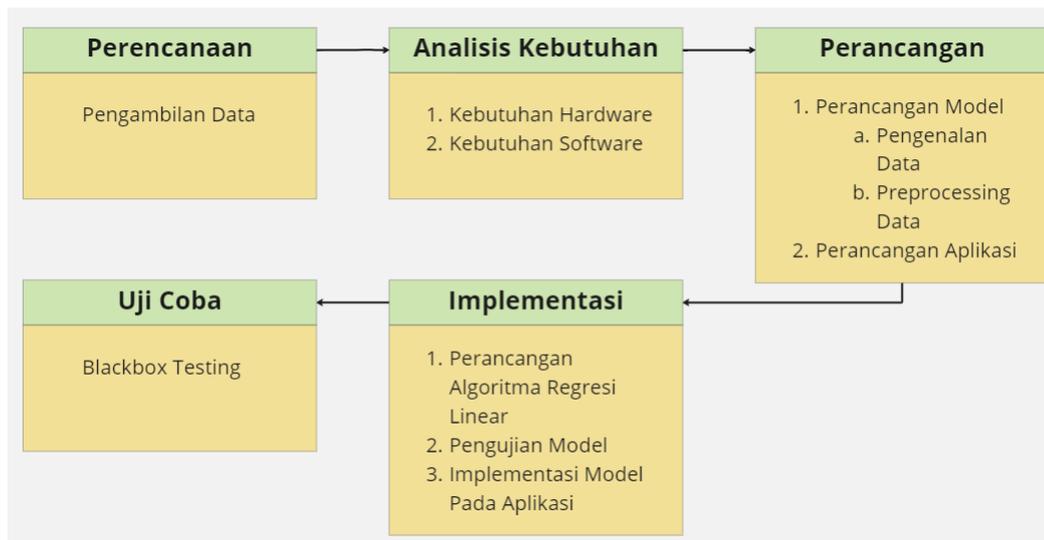
Sedangkan Regresi Linier merupakan teknik membangun model yang digunakan untuk prediksi nilai dari data masukan yang diberikan. Regresi Linier adalah ukuran statistik yang digunakan untuk menentukan kekuatan hubungan antara variabel dependen (tak bebas) dengan variabel independen (bebas). Metode utama untuk melakukan prediksi yakni membangun model regresi dengan mencari hubungan antara satu atau lebih variabel independen atau prediktor (X) dengan variabel dependen atau respons (Y) [8]. Regresi Linier memodelkan hubungan antara variabel skalar dan satu atau lebih variabel penjelas.

Langkah pertama untuk mendapatkan persamaan linear regression berganda adalah melakukan pembentukan model multiple linear regression. Nilai a , b_1 , b_2 , b_3 , dan b_4 merupakan nilai konstanta dan koefisien regresi yang dapat diperoleh dengan menggunakan perhitungan matriks determinan. Nilai a , b_1 , b_2 , b_3 , dan b_4 merupakan nilai konstanta dan koefisien regresi yang dapat diperoleh dengan menggunakan perhitungan matriks determinan [9].

Penelitian ini bertujuan untuk dapat mengetahui perkiraan harga mobil sesuai dengan kriteria yang diinginkan dengan menggunakan informasi yang berasal dari data harga sebelumnya, melalui satu layanan website.

Tahapan Penelitian

Metode yang digunakan dalam penelitian ini dilakukan melalui beberapa tahapan seperti pada Gambar 1, sebagai berikut: 1). Tahapan pengambilan dataset; untuk melatih model dengan mesin learning, dimana dataset yang digunakan data bersumber dari website penyedia data set www.kaggle.com yang berisi 10668 baris data dan 9 kolom data, 2). Tahapan data preprocessing, dimana dilakukan data cleaning dan data transformation, terhadap dataset yang didapat, 3). Tahapan modelling dan perancangan, pada tahap ini diterapkan model regresi dengan mengimplementasikan algoritma regresi linier pada dataset untuk melakukan prediksi ditahap selanjutnya dan melakukan perancangan website, 4). Tahapan implementasi, dimana pada tahap ini dirancang suatu sistem prediksi dengan memasukkan model ke dalam program dengan menggunakan bahasa pemrograman python dan framework streamlit, 5). Pengembangan Web dengan fungsi prediksi tersebut yang kemudian di uji coba, dimana pada tahap ini sistem yang dirancang akan diuji hingga menghasilkan output yang valid.



Gambar 1: Tahapan Penelitian

Perancangan

Pada tahap perencanaan yaitu merupakan tahap dalam mencari topik yang diangkat dan mengambil dataset yang akan digunakan. Pada tahap analisis kebutuhan, pada tahap ini melakukan identifikasi terkait kebutuhan perangkat keras dan perangkat lunak yang akan digunakan dalam tahap penelitian ini. Selanjutnya adalah tahap perancangan terdapat perancangan model dan perancangan aplikasi. Pada tahap perancangan model terdapat pengenalan data dan preprocessing data. Pada pengenalan data menampilkan isi dari dataset. Pada tahap preprocessing data untuk mempersiapkan data agar dapat digunakan dalam pembentukan model yaitu terdiri dari penanganan data outlier, pemilihan variabel berdasarkan tingkat korelasi, dan pemisahan data menjadi data pelatihan dan data pengujian. Setelah melakukan pemisahan data maka melakukan proses pelatihan data menerapkan algoritma regresi linear dan melakukan proses pengujian model oleh data uji untuk mendapatkan nilai akurasi. Setelah didapatkan model yang baik, maka melakukan perancangan aplikasi dan pengimplementasi model regresi yang telah dibuat pada website.

ML atau pembelajaran mesin merupakan pendekatan dalam AI yang banyak digunakan untuk menggantikan atau menirukan perilaku manusia untuk menyelesaikan masalah atau melakukan otomatisasi. Machine learning mencoba menirukan bagaimana manusia atau makhluk cerdas belajar dan menggeneralisasi, machine learning memiliki dua aplikasi utama yaitu klasifikasi dan prediksi. Ciri khas ML adalah adanya proses pelatihan, pembelajaran, atau training. Oleh karena itu, machine learning membutuhkan data untuk dipelajari yang disebut sebagai data training [1]. Adapun tahapan peramalan secara ringkas terdapat tiga tahapan yang harus dilalui dalam perancangan suatu metode peramalan, yaitu : 1) Melakukan anal-

isa pada data masa lampau. Langkah ini bertujuan untuk mendapatkan gambaran pola dari data bersangkutan, 2) Memilih metode yang akan digunakan. Terdapat bermacam-macam metode yang tersedia dengan keperluannya. Pemilihan algoritma dapat mempengaruhi hasil ramalan. Hasil ramalan diukur dengan menghitung error atau kesalahan terkecil. Oleh karena itu, tidak ada metode peramalan yang pasti baik untuk semua jenis data. 3) Proses transformasi dari data masa lampau dengan menggunakan metode yang dipilih. Apabila diperlukan maka diadakan perubahan sesuai kebutuhannya.

ML bermanfaat untuk menyelesaikan permasalahan dengan cara yang terukur. ML digunakan pada berbagai macam industri dengan memproses data yang lebih besar dan rumit dengan waktu yang lebih singkat. Pengambilan keputusan ML lebih berfokus pada pembelajaran atas input data. Dengan demikian ML digunakan karena dapat menghemat waktu dan lebih efisien saat digunakan.

Pada tahap perancangan model diawali dengan pengenalan data, dimana dataset yang digunakan bersumber dari dataset yang disediakan pada www.kaggle.com, suatu layanan website penyedia beragam dataset dengan format .csv berupa tabel. Situs ini memiliki lebih dari 6000 set data yang dapat diunduh oleh penggunaannya. Set data ini banyak dimanfaatkan dalam penelitian keilmuan di seluruh dunia untuk membuat model. Kaggle bukan hanya kumpulan dataset tetapi terdiri dari komunitas ilmuwan data terbesar.

Ketersediaan data set ini juga tentunya membutuhkan alat bantu pengelolaan data, saat ini salah satunya adalah bahasa pemrograman model skrip python yang berorientasi obyek. Beberapa fitur yang dimiliki Python adalah: 1) Memiliki kepastakaan yang luas dalam distribusi Python telah disediakan modul-modul, 2) Memiliki tata bahasa yang jernih dan mudah dipelajari, 3) Memi-

liki aturan layout kode sumber yang memudahkan pengecekan, 4) Pembacaan kembali dan penulisan ulang kode sumber, 5) Berorientasi obyek, 6) Dapat dibangun dengan bahasa Python maupun C/C++[10].

Optimasi python juga dibantu dengan tersedianya paket dasar seperti Panda yang mendukung penyediaan struktur dan fungsi data yang kaya dan khusus dirancang untuk mendukung analisis data yang kuat dan produktif [11].

Data set mobil bekas pada kaggle yang diambil pada penelitian ini ini merupakan tabel dengan jumlah 10668 data dan 9 atribut yang terdapat pada dataset yang digunakan, seperti tampak pada Gambar 2.

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 10668 entries, 0 to 10667
Data columns (total 9 columns):
#   Column          Non-Null Count  Dtype
---  ---
0   model           10668 non-null  int64
1   year            10668 non-null  int64
2   price           10668 non-null  int64
3   transmission    10668 non-null  int64
4   mileage         10668 non-null  int64
5   fuelType       10668 non-null  int64
6   tax             10668 non-null  int64
7   mpg            10668 non-null  int64
8   engineSize     10668 non-null  int64
dtypes: int64(9)
memory usage: 750.2 KB
```

Gambar 2: Tipe Data yang diambil dari repositori Kaggle.com

Data set yang diperoleh ini, kemudian akan

melalui tahapan data Preprocessing dengan menggunakan Jupyter Notebook untuk mempersiapkan data agar dapat digunakan dalam pembentukan model. Data yang didapat ini, akan melalui proses penghapusan data duplikat. Pada tahap penghapusan data duplikasi dengan tujuan agar model dapat membaca data secara berkualitas dan bukan merupakan data yang berulang. Output dari data yang sudah di duplikat akan digunakan pada tahap selanjutnya yaitu penanganan terhadap data outlier.

Data outlier adalah data yang menyimpang terlalu jauh dari rentang normal dalam suatu rangkaian data. Penghapusan data outlier akan membuat analisis data menjadi jelas dan menambah presentase akurasi data. Pada tahap ini, data yang memiliki rentang harga dibawah 25% dan diatas 75% akan dihapus karena mempengaruhi presentase akurasi data pada saat data di proses. Pemilihan variabel berdasarkan tingkat korelasi adalah tahapan berikutnya, seperti yang tampak pada Gambar 3, yang menunjukkan korelasi setiap variabel. Variabel model, tahun produksi(year), pajak(tax), kapasitas mesin(engine size) dan model transmisi(transmission) memiliki nilai tingkat akurasi yang tinggi terhadap variabel harga dependen dengan nilai 35%, 71%, 50%, 42% dan 4%. Variabel jarak tempuh(mileage), tipe bahan bakar (fuel type), rasio pemakaian bahan bakar(mpg: mile per gas), memiliki akurasi yang rendah. Dilihat pada tingkat akurasi, maka variabel yang akan digunakan pada tahap penerapan algoritma adalah model, tahun produksi(year), pajak(tax), kapasitas mesin(engine size) dan model transmisi(transmission).

	model	year	price	transmission	mileage	fuelType	tax	mpg	engineSize
model	1.00	0.05	0.35	0.01	-0.05	-0.05	0.38	-0.39	0.33
year	0.05	1.00	0.71	0.03	-0.77	0.18	0.26	-0.48	-0.06
price	0.35	0.71	1.00	0.04	-0.62	0.03	0.50	-0.75	0.42
transmission	0.01	0.03	0.04	1.00	-0.04	0.08	0.03	-0.04	-0.00
mileage	-0.05	-0.77	-0.62	-0.04	1.00	-0.29	-0.27	0.45	0.15
fuelType	-0.05	0.18	0.03	0.08	-0.29	1.00	0.10	-0.39	-0.51
tax	0.38	0.26	0.50	0.03	-0.27	0.10	1.00	-0.63	0.34
mpg	-0.39	-0.48	-0.75	-0.04	0.45	-0.39	-0.63	1.00	-0.31
engineSize	0.33	-0.06	0.42	-0.00	0.15	-0.51	0.34	-0.31	1.00

Gambar 3: Matriks Tingkat Korelasi Variabel Penentu Harga

Data ini kemudian dipisahkan menjadi data pelatihan dan data pengujian. Data training digunakan untuk melatih algoritma, sedangkan data testing dipakai untuk mengetahui performa algoritma yang sudah dilatih sebelumnya ketika menemukan data baru yang belum pernah dilihat se-

belumnya. Ini biasanya disebut dengan generalisasi. Hasil dari pelatihan tersebut bisa disebut dengan model. Pembagian data uji dan data latih coba dilakukan dengan dua percobaan dengan rasio 80%:20% dan 60%:40% dari total data yang sudah preprocessing. Pembagian data pelatihan dan

data pengujian acuan terhadap hasil nilai akurasi. Pembagian data latih dan data uji dilakukan dengan mengambil rasio 80% : 20% dari total data hasil pra-pemrosesan, sehingga berjumlah 1968 dan 7869 data karena memiliki nilai akurasi lebih besar. Data uji harus lebih ukuran rasionya dibandingkan data latih karena jika data lebih lebih besar dari dari uji, model tidak akan bisa melakukan proses uji data pada model yang digunakan. Setelah data dibagi menjadi data latih dan data uji, selanjutnya data akan digunakan pada tahap penerapan algoritma untuk melatih dan menguji data tersebut.

Penerapan Algoritma Regresi Linier

Setelah memiliki variable independen, data latih, dan data uji maka selanjutnya melakukan penerapan algoritma Regresi Linear. Regresi Linear yang digunakan dalam penelitian ini adalah Regresi Linear Berganda (multiple linear regression). Pemilihan teknik Regresi Linear Berganda karena melibatkan dari satu variable. Dalam penelitian ini variable yang digunakan adalah Tahun Mobil (year),

Kapasitas Mesin (engineSize), Model, Transmission dan Pajak (Tax) sebagai variabel yang menentukan [5]. Sehingga model regresi linear pada penelitian ini dapat dituliskan dengan persamaan (1) : (1) dimana y adalah variable response dalam hal ini adalah harga jual. Sedangkan b0 adalah intercept, b1 adalah slope untuk variable x1 yaitu model, b2 adalah slope untuk variable x2 yaitu Tahun Mobil (year), b3 adalah slope untuk variable x3 yaitu jenis transmisi(transmission), b4 adalah slope untuk variable x4 yaitu Pajak (tax) dan b5 adalah slope untuk variable x5 yaitu Kapasitas Mesin (engineSize). Nilai b0(intercept), b1 - b5 (slope) dapat dihasilkan dari perhitungan menggunakan metode least square dengan memanggil fungsi model.intercept pada library Linear Regression dari sklearn modul pada phyton. Berikut merupakan nilai intercept dan nilai slope, seperti pada Gambar 4. Sehingga persamaan regresi akan memberikan nilai seperti tampak pada Gambar 5 dan menampilkan variabel yang akan digunakan sebagai data latih dan data uji pada tahap penerapan Regresi Linear, yang kemudian diterapkan dengan perintah seperti pada Gambar 6.

```
print("Intercept : ",lig.intercept_)
print("Slope : ",lig.coef_)

Intercept : -6022573.4216657905
Slope : [ 231.84628463 2991.18641745 156.49183267 16.28700751 1037.22594122]
```

Gambar 4: Nilai Intercept dan Slope

```
x = df_clean2.drop(columns=['price', 'mileage', 'fuelType', 'mpg'])
y = df_clean2['price']
```

Gambar 5: Variabel Penerapan Regresi Linear

```
x_train, x_test, y_train, y_test = train_test_split(x,y,test_size=0.2,random_state=10)
lig=LinearRegression()
lig.fit(x_train, y_train)

LinearRegression()
```

Gambar 6: Penerapan Regresi Linear

Pengujian Model Regresi

Setelah melakukan pelatihan data menggunakan Regresi Linear maka selanjutnya akan melakukan pengujian model regresi. Pengujian model regresi menggunakan data uji dilakukan untuk mengukur performa kinerja dari model yang dinyatakan dengan nilai akurasi model. Pengujian model regresi ini dilakukan dengan data uji yang berjumlah 1968 data. Penghitungan nilai akurasi berdasarkan nilai prediksi pengujian dilakukan perhitungan R2 score dengan persamaan : $R2 = 1 - (SS_{res} / SS_{tot})$ (2)

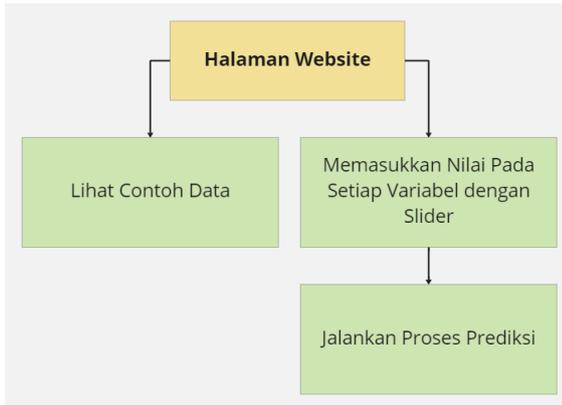
Dimana SS_{res} merupakan jumlah dari residual error yang didapat dari nilai jumlah data uji aktual

dikurangi data prediksi, kemudian dikuadratkan. dan SS_{tot} adalah jumlah error. Nilai tersebut kemudian bernilai: $R2 = 1 - (28622459939.521763 / 20347747818.16615)$ $R2 = 0.7621687114347371$

Perhitungan tersebut menentukan nilai akurasi dari model yang digunakan adalah 76, 22 % (pembulatan 2 angka).

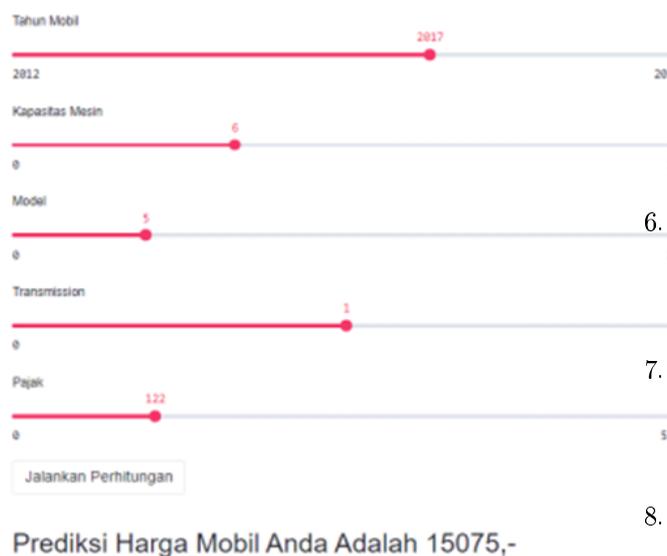
Implementasi pada Website Tahap implementasi model regresi yang dilakukan pada website dengan menggunakan pengujian black box, yang memasukkan nilai tertentu pada kolom antarmuka yang disediakan seperti tampak pada rancangan alur layanan website yang cukup sederhana dan mu-

dah dipahami seperti pada Gambar 7.



Gambar 7: Alur Layanan Website

Alur ini kemudian diimplementasi pada website dengan tampilan seperti pada Gambar 8. Antarmuka ini kemudian akan diujikan pada sistem dan hasil luarannya. Pengujian website dilakukan dengan dua tahap, yaitu pengujian antarmuka dan pengujian prediksi. Pengujian antarmuka hanya melakukan pengujian kinerja dari website, sedangkan untuk pengujian prediksi hanya untuk mengetahui hasil prediksi sesuai dengan nilai yang sudah dimasukkan oleh pengguna. Pada pengujian prediksi hasil prediksi yang diharapkan sesuai dengan nilai akurasi pada pengujian model regresi.



Gambar 8: Antarmuka Website dengan contoh Hasil Prediksi

Pengujian Antarmuka Website

Hasil perancangan website memberikan layanan pada pengguna untuk dapat berinteraksi dengan fungsi prediksi harga mobil yang dilakukan oleh website. Pengujian antarmuka dilakukan untuk menguji setiap komponen yang ada dalam aplikasi

apakah berfungsi dengan baik atau tidak. Hasil tersebut termasuk menguji apakah isian untuk setiap variable bisa difungsikan dengan baik dan memberikan hasil yang sesuai dengan fungsi perhitungannya.

Scenario Uji Black box dilakukan dengan mencoba semua View yang dirancang dan dikembangkan dalam Streamlit untuk mengetahui apakah interaksi pada web berfungsi dengan benar. Scenario uji adalah sebagai berikut:

1. Saat Website diakses, text view yang mencirikan judul layanan website akan sesuai ditampilkan.
2. Penjelasan terkait fungsi dan layanan website juga tampil dibawah bagian judul tersebut
3. Website akan memberikan layanan contoh data terkait data yang dijadikan acuan melalui aktivasi checklist, maka data ini akan tampil dengan keterangan sesuai data yang dimiliki oleh layanan
4. Website akan memberikan pilihan variabel yang mempengaruhi harga mobil bekas, tampilan akan terdiri dari 5 variabel yang dapat disesuaikan nilainya. Variabel ini terdiri dari Tahun Produksi Mobil, Kapasitas Mesin, Model, Transmisi, dan Pajak.
5. Pada tiap variabel, akan muncul slider yang bisa disesuaikan dengan nilai yang diinginkan. Slider harus dapat digerakkan sesuai dengan nilai yang diinginkan. Pada Slider Tahun Mobil, maka pengguna dapat menggerakkan slider sesuai tahun yang diinginkan, dan sistem akan mencatat sebagai variabel Tahun Produksi.
6. Pada Slider Kapasitas Mesin, maka pengguna dapat menggerakkan slider sesuai Kapasitas Mesin yang diinginkan, dan sistem akan mencatat sebagai variabel Kapasitas Mesin Mobil.
7. Pada Slider Model, maka pengguna dapat menggerakkan slider sesuai model yang diinginkan, dan sistem akan mencatat sebagai variabel Model dari Mobil.
8. Pada Slider Transmisi, maka pengguna dapat menggerakkan slider sesuai transmisi yang diinginkan, dan sistem akan mencatat sebagai variabel Transmisi dari Mobil.
9. Pada Slider Pajak, maka pengguna dapat menggerakkan slider sesuai Pajak yang dimiliki oleh mobil, dan sistem akan mencatat sebagai variabel Pajak Mobil.
10. Setelah semua Slider disesuaikan, dibagian bawah ada aktivasi untuk memulai penghitungan, maka tombol ini akan bereaksi saat di klik pengguna untuk melakukan penghitungan prediksi harga

11. Pada area text view hasil penghitungan harga akan muncul nilai harga prediksi mobil sesuai dengan hasil perhitungan harga.

Kesebelas tahapan uji ini akan dilakukan untuk menguji jalannya semua fungsi yang diberikan pada website bagi pengguna. Pengujian dilakukan dengan beberapa kali percobaan untuk tiap tahapan tersebut untuk melihat apakah ada kesalahan fungsi. Setelah dilakukan pengujian pada view, hasil dari pengujian disajikan pada Tabel 1.

Tabel 1: Hasil Pengujian Blackbox Antarmuka Website

No	Komponen/View	Fungsi	Status
1.	Text View Judul	Menampilkan text berupa judul dari website	Valid
2.	Text View Deskripsi Website	Menampilkan text deskripsi dari website	Valid
3.	Check View Sample Data	Menampilkan sampel data ketika check view dipilih	Valid
4.	Text View Pada Setiap Variabel	Menampilkan text yang mewakili nama dari setiap variabel	Valid
5.	Slider Input Tahun Mobil	Menghasilkan nilai untuk tahun mobil sesuai geseran	Valid
6.	Slider Input Kapasitas Mesin	Menghasilkan nilai untuk kapasitas mesin mobil sesuai geseran	Valid
7.	Slider Input Model Mobil	Menghasilkan pilihan model mobil	Valid
8.	Slider Input Transmisi Mobil	Menghasilkan nilai untuk transmisi mobil	Valid
9.	Slider Input Pajak Mobil	Menghasilkan nilai untuk pajak mobil	Valid
10.	Button Prediksi	Melakukan proses prediksi ketika dipili	Valid
11.	Text View Prediksi	Menghasilkan nilai hasil prediksi	Valid

Pengujian Prediksi

Tahap pertama saat pengujian prediksi adalah menentukan root mean square error (RMSE) adalah metode berbasis gradien, semakin rendah nilainya maka semakin baik prediksi yang dilakukan. RMSE akan digunakan untuk membandingkan selisih harga actual dengan harga prediksi. Nilai RMSE adalah 3813.650862 terdapat pada Gambar 9.

```
RMSE = np.sqrt(np.mean(pow(predict - y_test, 2)))
print('RMSE: %f' % RMSE)
RMSE: 3813.650862
```

Gambar 9: Pengujian dengan Root Mean Square Error (RMSE)

Pengujian prediksi dilakukan untuk mengetahui kinerja website dalam memberikan prediksi harga berdasarkan data yang dimasukan oleh pengguna. Pada Tabel 2 adalah hasil dari pengujian prediksi pada website.

Tabel 2: Pengujian Prediksi

No	Model	Year	Transmission	Tax	Engine Size	Harga Aktual	Harga Prediksi
1.	A1	2019	Semi-Auto	145	1	22950	22151
2.	A4	2017	Automatic	145	2	16800	14958
3.	A1	2017	Manual	150	1.4	12500	14599
4.	A1	2016	Manual	30	1.4	11000	9653
5.	Q3	2017	Manual	145	2	16490	17386
6.	A3	2016	Manual	150	1.4	15999	13694
7.	A1	2016	Manual	30	1.4	11000	9653
8.	A4	2017	Automatic	145	2	16800	14958
9.	A4	2018	Automatic	145	2	17200	17949
10.	A4	2018	Manual	145	1.4	15800	17509

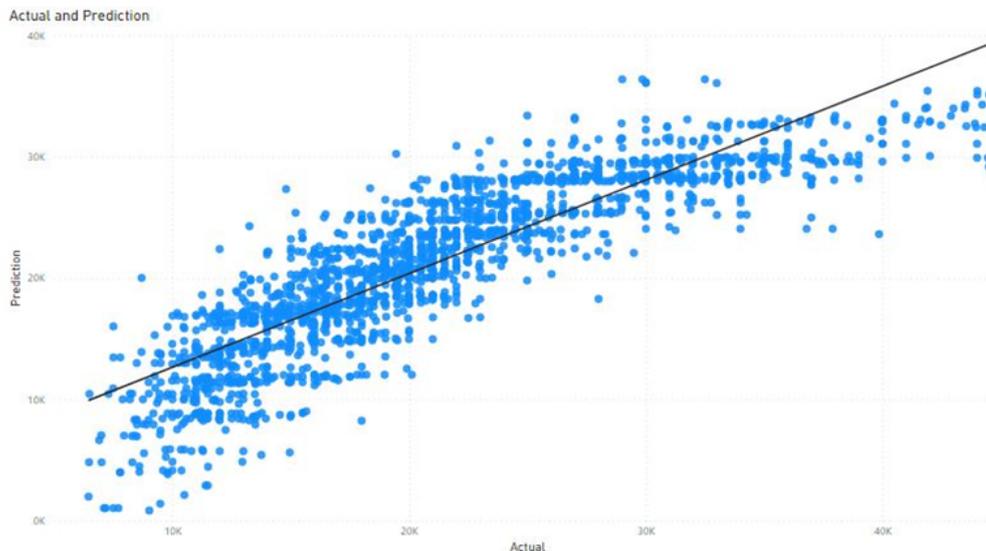
Setelah dilakukan pengujian prediksi menggunakan metode black box sebanyak 10 kali pengujian, didapatkan hasil prediksi yang relatif baik dan dapat dilihat pada grafik harga prediksi dan harga actual yang menjelaskan semakin mendekati garis linear hasil prediksi yang didapatkan semakin baik, seperti pada Gambar 10.

Analisis Hasil Prediksi

Berdasarkan hasil pengujian prediksi dengan beragam rasio data latih dan data uji seperti pada Tabel 3, rasio data uji dan data latih 80%:20% mendapatkan hasil akurasi prediksi tertinggi yaitu 76.22%. Hasil prediksi dengan RMSE pada Tabel 3 didapatkan hasil prediksi yang relatif baik pada 10 kali pengujian dan nilai RMSE yang didapatkan 3813,65. Sedangkan pada pengujian hasil layanan website menggunakan metode black box didapatkan hasil yang baik karena semua fitur yang tersedia di website dapat berjalan dengan baik.

Tabel 3: Pengujian Prediksi

No	Rasio Data	Akurasi
1.	60:40	76.03927503622086 %
2.	70:30	76.19324482069804 %
3.	80:20	76.21687114347371 %
4.	90:10	75.3895335108234 %



Gambar 10: Grafik Harga Prediksi dan Harga Actual

Penutup

Penelitian ini telah mengasihkan website yang dapat melakukan prediksi terhadap harga mobil bekas dengan mengimplementasi algoritma Linear Regression melalui pemodelan dengan machine learning. Hasil prediksi harga mobil bekas yang dilakukan telah berhasil mencapai akurasi sebesar 76.22% serta fitur yang terdapat pada website semua berjalan dengan baik, dimana hasil prediksi yang dihasilkan berdasarkan rasio data uji dan data latih yang digunakan 80%:20% menghasilkan prediksi yang relatif baik setelah dilakukan pengujian sebanyak 10 kali menggunakan metode black box dan nilai RMSE 3813,65. Pada penelitian selanjutnya dapat dilakukan perbaikan dengan memperbesar nilai akurasi keakuratan data hasil prediksi dengan cara menambahkan jumlah data dan menambahkan variable merk mobil lain serta variabel statis sebagai parameter sehingga hasil prediksi yang didapat lebih akurat

Daftar Pustaka

- [1] Abdu Ahmad, "Mengenal Artificial Intelligence, Machine Learning, Neural Network, dan Deep Learning", Yayasan Teknologi Indonesia, diakses daring pada <http://www.teknoindonesia.com>, 2017.
- [2] Ethem Alpaydin, "Introduction to Machine Learning Second Edition", MIT Press, London, 2010.
- [3] Peter Harrington, "Machine Learning in Action", Manning, New York, 2012.
- [4] Fabian Pedregosa, Gael Varoquaux and Alexandre Gramfort, "Scikit-learn: Machine Learning in Python", Journal of Machine Learning Research, Vol 12. 2825-2830, 2011.
- [5] Bagus Ardi Kurniawan dan Fitri Arianti, "Analisis Faktor yang Mempengaruhi Permintaan Riil Mobil Toyota Kijang Innova Di Kota Semarang", Skripsi pada Universitas Diponegoro. Semarang, 2013.
- [6] M. Karim dan R. M. Rahman, "Decision tree and naive bayes algorithm for classification and generation of actionable knowledge for direct marketing", Journal of Software Engineering and Applications, Vol.6 No.4, 2013.
- [7] C. R. Madhuri, G. Anuradha and M. V. Pujitha, "House price prediction using regression techniques: a comparative study", In 2019 IEEE International conference on smart structures and systems (ICSSS) (pp. 1-5), 2019.
- [8] L. Jäntschi, L. L. Pruteanu, A. C. Cozma and S. D. Bolboacă, "Inside of the linear relation between dependent and independent variables", Computational and mathematical methods in medicine, Vol 1, pp: 1-11, DOI:10.1155/2015/360752, 2015.
- [9] X. Su, X. Yan, and C. L. Tsai, "Linear regression", Wiley Interdisciplinary Reviews: Computational Statistics, 4(3), 275-294, 2012.
- [10] K. R. Srinath, "Python—the fastest growing programming language", International Research Journal of Engineering and Technology, 4(12), 354-357, 2017.
- [11] W. McKinney, "pandas: a foundational Python library for data analysis and statistics", Python for high performance and scientific computing, 14(9), 1-9, 2011.