BAB III

METODOLOGI PENELITIAN

3.1 Gambaran Umum Objek Penelitian

Objek dalam penelitian ini dilakukan untuk memprediksi harga penjualan mobil bekas, sehingga ketika Masyarakat ingin memasukkan atau menjual kendaraan mobil didalam situs *online* dapat mengetahui *range* harga yang cocok untuk kendaraan yang dimilikinya dengan keadaan kendaraannya. Dalam penelitian ini tentunya menggunakan beberapa atribut yang dibutuhkan seperti Nama kendaraan, tahun kendaraan, bahan bakar, transmisi, kapasitas kendaraan, kilometer kendaraan dan harga kendaraan. Atribut yang dipakai nantinya akan diklasifikasikan dengan menggunakan metode algoritma k-nearest neigbors sampai didapatkan informasi terbaru melalui beberapa tahap yang akan dilakukan.

Merek	Model	Kilometer	Lokasi	Transmisi	Jenis	Tahun	Warna	Kondisi	Engine Size	Harga
Toyota	Kijang Innova	68.000	Jakarta	Automatic	MPV	2019	Hitam	Bekas	2.4	Rp355.000.000
Toyota	Agya	51.000	Jawa Barat	Manual	Hatchback	2019	Kuning	Bekas	1.5	Rp81.000.000
Toyota	Avanza	32.000	Jakarta	Automatic	MPV	2019	Silver	Bekas	1.5	Rp148.000.000
Toyota	Avanza	13.000	Aceh	Automatic	MPV	2019	Hitam	Bekas	1.5	Rp145.000.000
Toyota	Agya	28.000	Banten	Manual	Hatchback	2019	Merah	Bekas	1.5	Rp109.000.000
Toyota	Voxy	88.000	Banten	Automatic	MPV	2019	Hitam	Bekas	2.0	Rp335.000.000
Toyota	Rush	68.658	Jawa Barat	Automatic	SUV	2019	Hitam	Bekas	1.5	Rp209.000.000
Toyota	Alphard	50.667	Jakarta	Automatic	MPV	2019	Hitam	Bekas	2.5	Rp885.000.000
Toyota	Fortuner	57.198	Jawa Barat	Automatic	SUV	2019	Hitam	Bekas	2.8	Rp420.000.000
Toyota	Yaris	76.405	Banten	Automatic	Hatchback	2019	Putih	Bekas	1.5	Rp197.000.000
Toyota	Avanza	28.168	Kalimantan Barat	Manual	MPV	2019	Hitam	Bekas	1.5	Rp285.000.000
Toyota	Fortuner	64.613	Jakarta	Automatic	SUV	2019	Putih	Bekas	2.8	Rp412.000.000
Toyota	Fortuner	63.434	Jakarta	Automatic	SUV	2019	Putih	Bekas	2.8	Rp414.000.000
Toyota	Fortuner	64.356	Banten	Automatic	SUV	2019	Putih	Bekas	2.8	Rp412.500.000
Toyota	Fortuner	42.352	Jakarta	Automatic	SUV	2019	Putih	Bekas	2.8	Rp418.000.000
Toyota	Fortuner	34.541	Banten	Automatic	SUV	2019	Putih	Bekas	2.8	Rp420.000.000

Gambar 3. 1 Visualisasi Dataset Mobil Bekas

Pada gambar 3.1 diatas merupakan penggunaan dataset mobil bekas yang nantinya akan digunakan dalam penelitian dimana terdapat sepuluh kolom yang berisikan:

1. Merek kendaraan = Merupakan tanda yang ditampilkan baik secara gambar, logo ataupun nama. Pada merek kendaraan ini terdapat beberapa jenis merek kendaraan yang diambil yaitu Toyota, Mitsubishi, Daihatsu, Honda dan Suzuki.
2. Model kendaraan = Merupakan tampilan yang berbeda dan dapat dibedakan berdasarkan teknologi, komponen ataupun tampilan. Pada

penelitian ini terdapat banyak sekali model yang diambil contohnya seperti Innova, Avanza, Rush dan sebagainya.

3. Kilometer = Merupakan suatu istilah yang digunakan untuk menggambarkan jarak total yang telah ditempuh oleh sebuah kendaraan sejak pertama kali digunakan. Informasi ini biasanya tercatat pada odometer
4. Lokasi = Lokasi merupakan informasi mengenai tempat atau kendaraan yang dijual itu berada.
5. Transmisi = Sistem transmisi memungkinkan kendaraan untuk bergerak maju, mundur, dan berakselerasi dengan efisien. Ada dua jenis utama transmisi: manual dan otomatis
6. Jenis = Jenis kendaraan dapat dibedakan berdasarkan desain, fungsi, dan tujuan penggunaannya. Berikut adalah beberapa jenis kendaraan yang umum seperti SUV, Sedan dan MPV.
7. Tahun = Tahun ini memberikan informasi mengenai tahun kendaraan tersebut
8. Warna = Pada bagian warn aini untuk memberikan informasi mengenai warna mobil bekas yang dijual.
9. Kondisi = Merupakan sebuah informasi untuk mengetahui kondisi dari mobil yang dijual.
10. *Engine Size* = Merujuk pada kapasitas mesin yang diukur dalam satuan volume silinder. Kapasitas mesin ini biasanya dihitung dalam desiliter (cc), liter (L), atau cubic inches (cu in). Ukuran mesin adalah salah satu faktor penting yang mempengaruhi performa, efisiensi bahan bakar, dan daya tahan suatu kendaraan.
11. Harga kendaraan = Menampilkan harga mobil yang dijual.

UNIVERSITAS
MULTIMEDIA
NUSANTARA

3.2 Metode Penelitian

Dalam penelitian ini, teknik data mining digunakan untuk memprediksi harga mobil bekas berdasarkan data yang diperoleh dari website CintaMobil. Data ini diambil di website cintamobil dikarenakan website cinta mobil memberikan informasi yang spesifik mengenai kendaraan – kendaraan yang dijualnya. Kemudian metode klasifikasi Linear Regression dan K-Nearest Neighbors (KNN) dipilih untuk menganalisis dan memahami pola dalam data yang ada. Pemilihan metode ini didasarkan pada kemampuannya dalam mengidentifikasi pola dan hubungan antara variabel yang berbeda, yang sangat penting dalam memprediksi harga yang akurat.

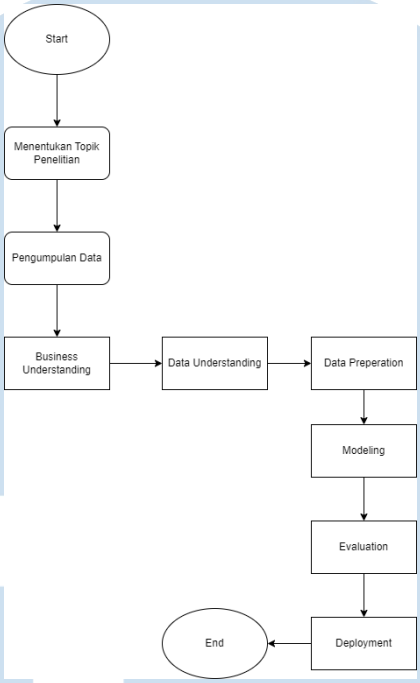
Berbagai tools dan algoritma dipilih berdasarkan kebutuhan dan tujuan penelitian. Penggunaan Jupyter Notebook, sebuah lingkungan pengembangan interaktif, memungkinkan penulisan dan pengujian kode Python secara langsung dalam dokumen yang dapat dibagikan dan diakses. Jupyter Notebook sangat berguna dalam penelitian data karena memungkinkan penampilan dan penjelasan data, analisis, dan hasil dengan cara yang interaktif dan mudah dipahami.

Selain itu, Jupyter Notebook digunakan untuk menghitung dan menampilkan nilai akurasi dari model prediksi. Nilai akurasi adalah ukuran penting yang menunjukkan seberapa baik model dapat memprediksi harga mobil bekas dengan benar. Dengan menampilkan nilai akurasi ini, kinerja model dapat dievaluasi dan penyesuaian yang diperlukan dapat dibuat untuk meningkatkan akurasi prediksi.

Dengan menggunakan teknik data mining, metode klasifikasi, dan tools seperti Jupyter Notebook, penelitian dapat dilakukan yang mendalam dan akurat tentang harga mobil bekas. Ini memungkinkan memahami faktor-faktor yang mempengaruhi harga mobil bekas dan membuat prediksi yang lebih akurat dan informatif.

UNIVERSITAS
MULTIMEDIA
NUSANTARA

3.2.1 Alur Penelitian



Gambar 3. 2 Alur Penelitian

Pada gambar 3.2 ini merupakan alur penelitian dalam proses melakukan prediksi harga penjualan mobil bekas. Berikut merupakan penjelasan alur penelitian :

1. **Menentukan Topik Penelitian**

Dalam menentukan topik penelitian ini tentunya dimulai dari mencari topik yang menarik untuk dibahas dan tentunya adanya masalah yang dialami sebuah perusahaan dan akan dibahas. Dalam topik ini nantinya akan dilakukan prediksi sehingga dapat membantu dalam mengatasi permasalahan yang ada.

2. **Pengumpulan Data**

Pengumpulan data yang dilakukan dengan cara *scraping* adalah metode yang efektif untuk mendapatkan data yang diperlukan dalam memprediksi. Dalam proses pengambilan data ini, penting untuk mempertimbangkan fitur-fitur yang nantinya bisa atau dapat digunakan

dalam melakukan pengolahan data pada penelitian ini. *Tools* yang digunakan dalam melakukan pengambilan data dengan menggunakan octoparse. *Scrapping* data yang dilakukan ini mengambil jumlah data kurang lebih sebanyak 1000 data dikarenakan membutuhkan waktu yang cukup lama ketika didalam bentuk data excel .

Pada pengumpulan data ini, fitur-fitur yang dipertimbangkan dalam proses pengambilan data melalui scraping harus memenuhi beberapa kriteria. Pertama, fitur harus relevan dengan tujuan penelitian. Misalnya, jika tujuan penelitian adalah untuk memprediksi harga kendaraan, fitur seperti tahun kendaraan, merek, dan model kendaraan mungkin sangat relevan. Kedua, fitur harus dapat diambil secara otomatis dan dalam jumlah yang cukup untuk analisis. Ketiga, fitur harus dapat diolah dan dianalisis dengan alat dan teknik yang ada.

3. *Business Understanding*

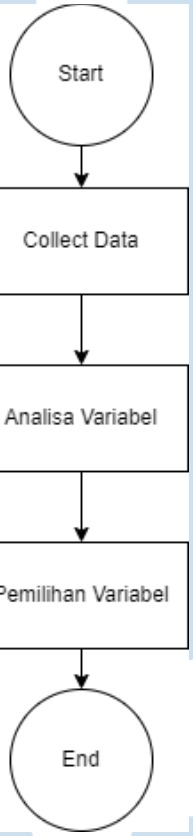
Pada penelitian ini bertujuan untuk melakukan analisa pada penjualan mobil bekas dan juga mengumpulkan data sehingga dapat memprediksi harga penjualan mobil bekas. Tahapan ini diperlukan dalam mempersiapkan seluruh proses penelitian seperti *tools* dan algoritma. Topik utama dalam penelitian ini adalah melakukan prediksi harga mobil bekas dengan menggunakan algortima Linear Regression dan KNN.

4. *Data Understanding*

Data understanding merupakan proses pertama dalam siklus analisis data yang memungkinkan untuk memahami data yang dimiliki. Proses ini sangat penting karena menentukan apa yang dapat dilakukan dengan data tersebut dan bagaimana data tersebut dapat digunakan untuk mencapai tujuan bisnis atau penelitian.

Data understanding melibatkan pengamatan, eksplorasi, dan interpretasi data untuk memahami struktur, pola, dan hubungan antara variabel. Ini mencakup pengidentifikasian dan pemahaman tentang variabel-

variabel yang ada dalam dataset, termasuk jenis data, distribusi, dan hubungan antara variabel. Proses ini juga melibatkan identifikasi dan pemahaman tentang kualitas data, termasuk ketidakakuratan, ketidaklengkapan, dan bias, yang dapat mempengaruhi hasil analisis.

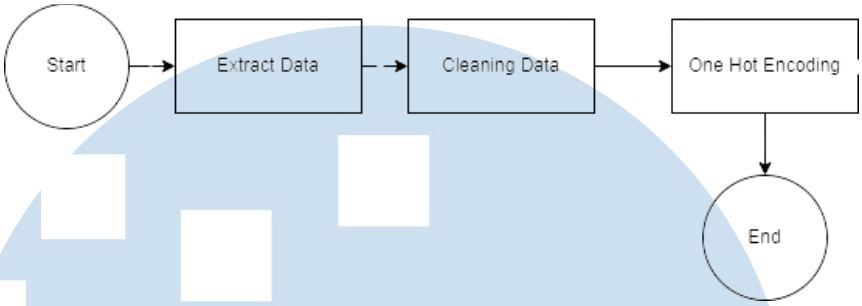


Gambar 3. 3 Diagram data understanding

Pada gambar 3.3 ini menjelaskan tahap awal mendapatkan data dengan dilakukannya scraping pada website cintamobil, setelah data didapatkan kemudian melakukan Analisa variable serta pemilihan variabel yang akan digunakan pada saat pengolah data nantinya.

UNIVERSITAS
MULTIMEDIA
NUSANTARA

5. *Data Preperation*



Gambar 3. 4 Diagram Preperation

Pada fase *Data Preparation*, proses dimulai dengan melakukan ekstraksi data, diikuti oleh pemilihan data dan fitur yang diperlukan. Selanjutnya, data tersebut diproses lebih lanjut melalui pre-processing, yang mencakup penghapusan outlier, penghilangan informasi yang tidak relevan, dan transformasi data. Untuk data yang akan diubah menjadi kolom daerah, dilakukan *one-hot encoding* untuk mengklasifikasikan daerah pada setiap data. Hal ini dilakukan karena dalam pemodelan regresi, data kategorikal tidak dapat langsung digunakan dalam proses *fitting* atau prediksi. Dengan one-hot encoding, semua data diubah menjadi kolom yang akan memberikan identifikasi. Hasil dari fase Data Preparation ini menghasilkan total data sebesar 1005 dengan 11 kolom.

6. *Modeling*

Pada tahapan Modeling, model yang digunakan untuk analisis data atau prediksi memanfaatkan teknik dan metode pengolahan data yang telah dipilih sebelumnya. Tujuan dari tahapan ini adalah untuk menghasilkan output yang sesuai dengan tujuan analisis atau prediksi yang diinginkan. Modeling adalah proses di mana algoritma atau teknik statistik dipilih dan diterapkan pada data untuk mencapai tujuan tertentu. Ini bisa mencakup berbagai jenis model, seperti model regresi untuk prediksi numerik, model klasifikasi untuk pengklasifikasian, atau model clustering untuk pengelompokan data. Pemilihan model yang tepat sangat penting, karena model yang salah dapat menghasilkan interpretasi yang salah atau tidak akurat dari data.

Table 3. 1 Perbandingan Algoritma [26][27]

Algoritma	Kelebihan	Kekurangan
K-Nearest Neighbor [26]	- Sederhana dan mudah dipahami - Dapat digunakan dalam proses klasifikasi maupun regresi - mudah diterapkan	- Waktu pemrosesan yang lama jika datasetnya sangat besar. - Kurang cocok untuk dimensi tinggi
Linear Regression [27]	- Mudah untuk diimplementasikan - Mudah untuk dilakukan evaluasi dan memiliki beragam metode atau metriks	- Sensitif terhadap outliers. - Sering menghasilkan model prediksi yang overfitting, artinya terlalu bagus dan tidak nyata.

Dari table 3.1 diatas kesimpulan mengapa menggunakan kedua algoritma tersebut adalah KNN dapat digunakan untuk regresi, di mana tujuannya adalah memprediksi hasil kontinu (seperti harga mobil bekas) berdasarkan fitur-fitur yang mirip dengan observasi lainnya. Proses implementasi KNN untuk regresi melibatkan penghitungan jarak Euclidean antara observasi yang ingin diprediksi dengan setiap observasi dalam dataset pelatihan. Observasi dengan k tetangga terdekat yang paling mirip digunakan untuk memperkirakan hasil prediksi sedangkan Algoritma K-Nearest Neighbors dapat digunakan untuk memprediksi harga jual mobil bekas dengan memanfaatkan informasi dari mobil bekas lain yang memiliki fitur yang serupa.

UNIVERSITAS MULTIMEDIA NUSANTARA

7. **Evaluation**

Tahap evaluasi menguji kinerja model siap pakai yang dipilih. Parameter terpenting yang digunakan dalam penelitian ini adalah R squared dan root mean square error (RMSE). R-squared memberikan gambaran seberapa baik model menjelaskan variasi data, dan RMSE memberikan ukuran kesalahan prediksi model.

Dengan adanya hasil atribut seperti R-squared, RMSE, dan tingkat kepentingan fitur, pengguna dapat memperoleh gambaran yang jelas tentang performa model yang mereka gunakan. Hal ini memungkinkan dalam menilai seberapa baik model menjelaskan data, seberapa akurat prediksi model, sehingga nantinya informasi ini, dapat membuat keputusan yang lebih baik tentang cara menggunakan model dalam konteks dan cara meningkatkan model untuk mencapai hasil yang lebih baik.

8. **Deployment**

Pada tahap ini, melakukan penerapan data mining dengan menggunakan dua metode algoritma yang berbeda, yaitu Linear Regression dan K-Nearest Neighbor. Proses ini melibatkan pengolahan data yang kompleks, dimana setiap data dianalisis dan diproses secara rinci untuk menemukan pola atau tren yang mungkin tidak terlihat secara langsung. Linear Regression digunakan untuk mencari hubungan antara variabel independen dan dependen, sementara K-Nearest Neighbor digunakan untuk mengklasifikasikan data berdasarkan kesamaan dengan data lain dalam dataset. Setelah data diproses melalui algoritma ini, kita mendapatkan output yang berisi informasi yang berharga. Informasi ini dapat digunakan untuk membuat keputusan yang lebih baik, memprediksi tren masa depan, atau menemukan solusi untuk masalah yang ada.

3.3 Teknik Pengumpulan Data

Penelitian ini bertujuan untuk menganalisis tren pasar mobil di Indonesia, dengan fokus pada kendaraan yang dijual di website Cintamobil. Proses mengumpulkan data yang diperlukan menggunakan teknik web scraping, yang memerlukan waktu antara

lima sampai enam jam untuk menyelesaikan proses pengambilan data. Proses scraping ini melibatkan pengumpulan berbagai jenis informasi tentang kendaraan, termasuk Nama Kendaraan, Bahan Bakar, Transmisi, Penjual, Kilometer, Kapasitas Mesin, Lokasi, Harga, *Engine Size*, dan Tahun. Data yang diambil mencakup periode dari tahun 2019 hingga tahun 2023, dengan total 1006 data yang berhasil dikumpulkan.

Setelah data diperoleh, kami melakukan proses pembersihan dan pengolahan data untuk memastikan keakuratan dan kualitas data. Proses ini melibatkan penghapusan data yang tidak relevan atau duplikat, serta pengisian nilai yang hilang atau tidak lengkap. Setelah data diproses, kami menggunakan berbagai metode analisis data untuk mengeksplorasi tren pasar, memahami preferensi konsumen, dan mengidentifikasi faktor-faktor yang mempengaruhi harga kendaraan.

Hasil dari penelitian ini diharapkan dapat memberikan wawasan yang berharga bagi pembuat keputusan di industri otomotif, baik itu produsen, dealer, atau investor. Informasi yang dikumpulkan dapat digunakan untuk merancang strategi pemasaran yang lebih efektif, memprediksi tren pasar, dan mengembangkan produk yang lebih sesuai dengan kebutuhan dan preferensi konsumen."

Kalimat ini menambahkan detail tentang tujuan penelitian, metode yang digunakan, dan bagaimana data tersebut dapat digunakan untuk mencapai tujuan penelitian.

3.4 Teknik Analisis Data

Dalam penelitian ini, teknik analisis data yang dipilih adalah teknik deskriptif, yang merupakan pendekatan yang digunakan untuk menggambarkan dan memahami karakteristik data yang telah dikumpulkan. Setelah proses scraping data untuk mendapatkan dataset yang akan diteliti, langkah selanjutnya yang dilakukan adalah penghapusan data yang terduplikat atau data ganda. Hal ini sangat penting untuk memastikan bahwa setiap entri dalam dataset adalah unik dan tidak mengandung duplikasi, yang dapat mempengaruhi akurasi dan keandalan analisis data yang akan dilakukan. Teknik analisis data deskriptif ini tidak hanya bertujuan untuk menggambarkan data yang telah diambil dan hasil dari pengumpulan data tersebut,

tetapi juga untuk memastikan bahwa data yang digunakan dalam penelitian ini adalah data yang bersih dan bebas dari duplikasi atau nilai null. Teknik analisis data deskriptif ini menggambarkan bahwa data yang telah diambil dan dilakukan pengumpulan hasil dari data tanpa adanya membuat kesimpulan [28].

Table 3. 2 Perbandingan Program

Perbandingan	Python	R
Sifat	*Open source*	*Open source*
Kelebihan	- Terdapat fungsi selain untuk melakukan analisis data. - Sintaks singkat mudah dipahami - Memiliki banyak library - Kecepatan - Powerfull untuk komputasi matematis	- Tools untuk membuat grafik. - Dibuat untuk kepentingan analisis data - Banyak fitur visualisasi data.
Kekurangan	- Jika program besar eksekusi lambat - *Memory intensive task*	- Tidak ada kemampuan untuk menganalisa multivariable - Pemrosesan yang dilakukan lebih lama

Pada pengolahan data, pemilihan bahasa pemrograman Python dengan menggunakan software Jupyter menjadi pilihan yang strategis, karena kelebihan dari penggunaan phyton ini serta python memiliki kemampuan untuk mengintegrasikan dengan berbagai sistem dan database, memudahkan pengolahan data yang kompleks.

Python dan Jupyter Notebook memungkinkan untuk bekerja dengan cepat dan efisien dalam pengolahan data. Dengan kemampuan untuk menulis kode yang efisien

dan menggunakan library yang kuat untuk pengolahan data, menyelesaikan tugas pengolahan data dengan lebih cepat dan akurat. Selain itu, kemampuan Python untuk diintegrasikan dengan berbagai sistem dan database memungkinkan analisis untuk mengakses dan memanipulasi data yang berada di berbagai sumber, meningkatkan fleksibilitas dan efisiensi dalam pengolahan data.

UMN
UNIVERSITAS
MULTIMEDIA
NUSANTARA