

BAB II

LANDASAN TEORI

2.1 Penelitian Terdahulu

Table 2. 1 Penelitian Terdahulu

No	Peneliti	Judul Penelitian	Metode Penelitian	Hasil Penelitian
1	Mohammad Aji Saputral, Martanto, Umi Hayati (2024)	ESTIMASI HARGA MOBIL BEKAS TOYOTA YARIS MENGGUNAKAN ALGORITMA REGRESI LINIER JOURNAL MAHASISWA TEKNIK INFORMATIKA Vol. 8 No. 2 (2024): JATI Vol. 8 No. 2, ISSN : 2598-828X	Dalam penelitian ini Knowledge Discovery in Databases (KDD) menjadi metode utama eksplorasi dan analisis data. Tahapan KDD meliputi seleksi data, pengolahan data, transformasi data, penggalian data dan evaluasi,.	Berdasarkan penelitian ini menggunakan model regresi linear untuk prediksi harga mobil bekas dan Evaluasi model menghasilkan nilai MSE sebesar 886,503, RMSE 941,543
2	Ernianti Hasibuan dan Aldian Karim	Implementasi Machine Learning untuk Prediksi Harga Mobil Bekas dengan Algoritma Regresi Linear Berbasis Web Jurnal Ilmiah KOMPUTASI, Volume 21 No : 4, Desember 2022, p- ISSN 1412-9434/e- ISSN 2549-7227	Metode untuk memprediksi harga mobil adalah dengan Machine Learning (ML)	Metode algoritma yang digunakan pada penelitian ini yaitu Regresi linear dan hasil dari pengujian dan nilai RMSE yang didapatkan 3813,65.
3	Tiara Lailatul Nikmah, Risma Maulidy, Syafei, Rini	Prediction Of Used Car Prices Using K-Nearest Neighbour, Random Forest And Adaptive Boosting Algorithm	Metode yang dilakukan dalam penelitian ini melalui beberapa tahapan, yaitu Pemahaman Bisnis,	Kebaruan dari penelitian ini adalah peningkatan akurasi model prediksi dari model tunggal.

No	Peneliti	Judul Penelitian	Metode Penelitian	Hasil Penelitian
	Muzayana, Asharinnisa Salsabila	Jurnal Int. Conf. Optim. Comput. Appl, Volume1, Issue1, Pages 17-22 (2022)	Pemahaman Data, Persiapan Data, Pemodelan dan Evaluasi	Hasil dari penelitian ini adalah algoritma Nilai RMSE sebesar 1163897.489566
4	Pandey, A., Rastogi, V., & Singh, S. (2020)	Car's Selling Price Prediction using Random Forest Machine Learning Algorithm. Vol 1, SSRN – 3702236. (Social Science Research Network)	Menggunakan dua algoritma yaitu Random Forest dan Extra Tress Regressor.	Model ini didasarkan pada algoritma pembelajaran mesin dan mencoba memprediksi harga jual mobil bekas berdasarkan dataset yang disediakan di Kaggle. Untuk memprediksi dataset ini, menggunakan dua algoritma machine learning yaitu Random Forest dan Extra Tress Regressor.
5	Muhammad Fikri Sofyan, Mufid Nilmada	SISTEM PAKAR PREDIKSI HARGA MOBIL BEKAS MENGGUNAKAN DECISION TREE BERBASIS WEB. UG JURNAL VOL.17 Edisi 04 April 2023	Metode penelitian yang digunakan dalam penelitian ini adalah dengan mengumpulkan data harga mobil bekas dari berbagai sumber online seperti website jual beli mobil bekas dan forum mobil..	Hasil dari penelitian ini adalah sebuah sistem pakar yang dapat membantu pengguna dalam memprediksi harga mobil bekas.

No	Peneliti	Judul Penelitian	Metode Penelitian	Hasil Penelitian
6	Reynaldi, Wahyu Syafrizal, M. Faris Al Hakim (2021)	Analisis Perbandingan Akurasi Metode Fuzzy Tsukamoto dan Fuzzy Sugeno Dalam Prediksi Penentuan Harga Mobil Bekas. Vol 44, No 2, Indonesian Journal of Mathematics and Natural Sciences.	Analisis perbandingan harga mobil Toyota Avanza yang digunakan dalam penelitian ini dilakukan dengan menggunakan metode Tsukamoto dan Sugeno dengan variabel yang sama. Sampel datanya menggunakan data penjualan mobil bekas di UD. Dito Motor, Banyumanik, Semarang.	Hasil penelitian menunjukkan bahwa metode fuzzy Tsukamoto mempunyai tingkat <i>error</i> sebesar 8% dalam memprediksi harga Toyota Avanza bekas.
7	Bambang Kriswantara (2022)	Machine Learning Used Car Price Prediction with Random Forest Regressor Model Vol 6 No 1 (2022): JISICOM: June 2022	Penelitian ini menggunakan eksperimen komparatif. Pendekatan melibatkan pengujian algoritme regresi dan membandingkan nya dengan beberapa model untuk mendapatkan perkiraan harga mobil bekas yang paling akurat. desain atau desain metodologi yang digunakan dalam penelitian;	Hasil Evaluasi Model Random Forest: Setelah menguji lima model untuk memprediksi harga mobil, Random Forest memiliki tingkat kesalahan MAE dan RMSE.
8	Evi Dewi Sri	Estimasi Harga Jual Mobil Bekas	Metode estimasi harga jual mobil	Membuat estimasi harga jual mobil

No	Peneliti	Judul Penelitian	Metode Penelitian	Hasil Penelitian
	Mulyani, Firham Mulady , Dendi Ramadhan, Ari Ariyanto no, Dikri Ramdani ,Robi Wahyundana , M. Gilang	Menggunakan Metode Regresi Linier Berganda Vol 9 No 1 (2020): e-Jurnal JUSITI	bekas adalah dengan metode Regresi Linier Berganda.	bekas menggunakan metoderegresi linier berganda ini berdasarkan variabel merk, harga beli, tahun keluar
9	Dea Miftahul Huda , Giftera Dwilestari , Ade Rizki Rinaldi, Iin. (2024).	Prediksi Harga Mobil Bekas Menggunakan Algoritma Regresi Linear Berganda. Jurnal Informatika dan Rekayasa Perangkat Lunak, e-ISSN 2685-5518, Vol. 6, No. 1, Maret 2024, Hal. 150-157.	Metode penelitian yang digunakan pada penelitian ini terdapat beberapa tahap yaitu: Data Selection, Data Cleaning, Data transformation, Data Mining dan Evaluation	Hasil nilai evaluasi yang di dapat dari perhitungan prediksi menggunakan RMSE root_mean_squared_error: 1637.49.
10	Cevi Herdian , Ahya Kamila , Felliks Feiters Tampinongkol , Agung Stefanus Kembau , I Gusti Agung Musa Budidar ma (2024)	ONE-HOT ENCODING FEATURE ENGINEERING UNTUK LABEL-BASED DATA STUDI KASUS PREDIKSI HARGA MOBILBEKAS. Jurnal Informatika Dan Teknologi Informasi, 9(1), 10–16. ISSN 2527-5240	Tujuan utama dari penelitian ini adalah untuk menerapkan metode One-Hot Encoding ke dalam prediksi harga mobil bekas dengan tujuan mendapatkan model yang lebih akurat.	Hasil penelitian yang dilakukan menunjukkan bahwa tingkat <i>error</i> yang didapatkan sebesar RMSE Validation: 5987.462748081144 .

Berdasarkan table penelitian diatas dapat dilihat:

Penelitian [9]. Research gap dari jurnal tersebut adalah bahwa penelitian sebelumnya telah fokus pada menggunakan metode regresi linier untuk memprediksi harga mobil bekas, namun penelitian ini menambah nilai dengan menggunakan algoritma regresi linier berganda untuk mengidentifikasi pengaruh atribut berdasarkan spesifikasi mobil Toyota Yaris bekas. Hal ini memberikan kontribusi baru dalam pengembangan model estimasi harga mobil bekas dengan pendekatan yang lebih komprehensif dan detail penelitian ini juga menghasilkan nilai RMSE sebesar 941,543.

Penelitian [6]. Menjelaskan mengenai Machine Learning untuk Prediksi Harga Mobil Bekas dengan Algoritma Regresi Linear Berbasis Web, selain itu juga penelitian ini akan menampilkan hasil RMSE yang didapat menggunakan algoritma yang digunakan sebesar 3813,65.

Penelitian [10]. menjelaskan prediksi harga mobil bekas dengan beberapa algoritma. Dalam memprediksi, pada penelitian ini mencari nilai RMSE pada KNN data mobil dan hasil yang didapat pada KNN adalah 1163897.489566. Selain itu juga metode yang dilakukan dalam penelitian ini melalui beberapa tahapan, yaitu Pemahaman Bisnis, Pemahaman Data, Pemahaman Bisnis, Persiapan Data, Pemodelan dan Evaluasi

Penelitian [4]. Dalam penelitian ini didasarkan pada algoritma pembelajaran mesin dan kami mencoba memprediksi harga jual mobil bekas berdasarkan dataset yang disediakan di Kaggle. Untuk memprediksi dataset ini, kami menggunakan dua algoritma machine learning yaitu Random Forest dan Extra Tress Regressor. Prediksi model ini selanjutnya dibandingkan dengan dataset uji yang dibuat dengan mengambil nilai acak dari dataset asli dan evaluasi prediksi dievaluasi lebih lanjut dengan menggunakan metode yang berbeda. Setelah evaluasi lengkap dari model prediksi, kita dapat menyimpulkan bahwa akurasi model ini sangat tinggi dan Random Forest dan Extra Tree Regression adalah salah satu algoritma terbaik untuk masalah

regresi. Kedua algoritma ini sangat akurat dan cepat dalam prediksi terlepas dari ukuran dataset.

Penelitian [3]. Metode penelitian yang digunakan pada penelitian ini adalah pengumpulan data harga mobil bekas dari berbagai sumber online. Data yang diperoleh kemudian diolah menggunakan metode preprocessing data untuk membersihkan dan mempersiapkan data setelah data sudah ada kemudian diolah dan dijadikan GUI.

Penelitian [5]. Melakukan Perbandingan Akurasi Metode Fuzzy Tsukamoto dan Fuzzy Sugeno Dalam Prediksi Penentuan Harga Mobil Bekas, penelitian ini nantinya akan mencari nilai *error* dari data yang didapat dan diolah.

Penelitian [11]. Machine learning dengan permodelan yang tetap mampu memberikan prediksi regresi terhadap harga mobil bekas. Dataset yang digunakan melalui unduh pada kaggle. Selain itu juga penelitian ini mencari model yang tepat sesuai data dan bisa memberikan prediksi harga jual mobil bekas dengan nilai error terbaik.

Penelitian [2]. Penelitian ini akan membuat estimasi harga jual mobil bekas menggunakan metode regresi linier berganda ini berdasarkan variabel merk, harga beli, tahun keluar, dan kondisi mobil ini telah mampu menghasilkan nilai pada perkiraan harga mobil bekas secara pengujian dari hasil estimasi dengan hasil yang real

Penelitian [12]. Pada penelitian ini akan mencari nilai RMSE dari data mobil bekas yang didapatkan menggunakan algoritma Linear Regresion, sebelum data ini diolah tentunya akan melakukan data dalam pemrosesan datanya.

Penelitian [13]. Pengukuran ini memberikan gambaran tentang seberapa baik model Machine Learning bekerja. Kinerja model yang lebih baik ditunjukkan oleh nilai yang lebih rendah untuk MAE, RMSE, dan MSE. RMSE pada penelitian ini mendapatkan hasil 5987.462748081144.

Kesimpulan dari sepuluh jurnal penelitian terdahulu ini yaitu pada penelitian yang akan dilakukan nantinya akan mencari nilai MAE, MSE dan RMSE dari data

yang didapatkan melalui *scraping data* pada *website* cintamobil, kemudian nantinya sebelum dihitung tingkat *error*, data yang didapat perlu dilakukan beberapa tahap seperti pembersihan data dan pemilihan data apa saja yang akan di pakai. Kemudian mengolah untuk menghasilkan tingkat nilai *error* dan tahap terakhir membuat GUI.

2.2 Teori yang digunakan

2.2.1 Prediksi

Prediksi merupakan proses mengenai perkiraan dan sesuatu yang akan terjadi dimasa depan berdasarkan informasi dan data yang telah ada pada saat ini [14]. Dalam melakukan prediksi ini tentunya memiliki tujuan untuk mendapatkan informasi yang akan terjadi di masa mendatang. prediksi ini sangat luas dan dapat diterapkan di berbagai bidang, mulai dari ilmu cuaca untuk memprediksi cuaca yang akan datang, hingga bidang olahraga yang digunakan untuk memprediksi hasil pertandingan atau kinerja seorang atlet. Prediksi ini seringkali dilakukan dalam rentang waktu yang beragam, mulai dari harian, mingguan, bulanan, hingga tahunan, tergantung pada kompleksitas dan tujuan analisis.

Proses prediksi melibatkan beberapa langkah penting yang harus dilakukan secara berurutan. Pertama, menentukan dan mengumpulkan data yang relevan dan berkualitas tinggi. Data ini harus mencakup variabel yang berpotensi mempengaruhi hasil prediksi yang diinginkan. Selanjutnya, melakukan analisis data tersebut untuk mengidentifikasi pola, tren, dan hubungan antara variabel yang berbeda. Analisis ini dapat melibatkan teknik statistik, *machine learning*, atau metode lainnya, tergantung pada jenis data dan tujuan prediksi.

Setelah pola dan hubungan antara variabel telah diidentifikasi, langkah selanjutnya adalah melakukan pengujian model prediksi untuk memastikan keakuratannya. Pengujian ini melibatkan pembagian data menjadi set pelatihan dan set pengujian, serta evaluasi model menggunakan metrik keakuratan yang sesuai. Tujuan dari pengujian ini adalah untuk memastikan bahwa model prediksi dapat dengan andal memprediksi hasil atau peristiwa yang diinginkan, dan tidak hanya mengandalkan pola yang telah terbukti dalam data historis. Dengan mengikuti langkah-langkah tersebut, prediksi menjadi proses yang sistematis dan

berdasarkan data, yang dapat membantu dalam pengambilan keputusan dan perencanaan strategis dalam berbagai bidang. Prediksi memungkinkan untuk memahami dan membantu dalam membuat proyeksi yang informasi dan berdasarkan data, bukan hanya intuisi atau asumsi".

2.2.2 Showroom

Pengertian *showroom* adalah tempat untuk memamerkan atau biasa didefinisikan sebagai tempat untuk memamerkan mobil merk umum tertentu, padahal *showroom* ini adalah tempat yang menyediakan jasa seperti otomotif, sparepart dan fasilitas pendukung lainnya yang tentunya dapat berfungsi untuk meningkatkan pemasaran [15].

Dalam konteks global, *showroom* menjadi pusat pertemuan antara produsen kendaraan dan konsumen, di mana konsumen dapat melihat, merasakan, dan memahami kualitas dan fitur kendaraan sebelum membuat keputusan pembelian. *Showroom* juga menyediakan informasi terkait kendaraan, termasuk spesifikasi, harga, dan promosi, serta menawarkan layanan purna jual seperti perbaikan dan pemeliharaan.

Showroom dapat dibagi menjadi dua jenis utama, yaitu showroom kendaraan motor dan showroom kendaraan mobil. *Showroom* kendaraan motor biasanya menampilkan berbagai jenis motor, termasuk motor sport, motor bebek, dan motor lainnya, sementara *showroom* kendaraan mobil menampilkan berbagai model mobil, dari kendaraan baru hingga kendaraan bekas.

Kategori *showroom* kendaraan baru dan showroom kendaraan bekas menunjukkan perbedaan dalam strategi penjualan dan target pasar. *Showroom* kendaraan baru biasanya menargetkan konsumen yang mencari kendaraan dengan kualitas terbaru dan fitur terkini, sementara *showroom* kendaraan bekas menargetkan konsumen yang mencari kendaraan dengan harga yang lebih terjangkau dan memiliki nilai *resell* yang tinggi.

2.2.3 Mobil Bekas

Kualitas produk mobil bekas sangat mempengaruhi minat konsumen untuk membeli. Kualitas produk mobil bekas antara lain dapat diukur dari tahun

perakitan, kilometer, kondisi fisik interior dan eksterior, kondisi mesin, kondisi ban, warna serta atribut-atribut lainnya yang membuat harga mobil tersebut dapat berubah ubah. Kualitas produk sebuah mobil bekas menjadi tolak ukur konsumen dalam menilai kelayakan mobil untuk dibeli. Mobil bekas saat ini sudah banyak diminati oleh para masyarakat untuk memulai bisnis mereka yang tentunya memerlukan kendaraan pribadi, selain itu ada juga orang – orang yang membeli mobil bekas untuk menyalurkan bakat dan hobinya dalam merestorasi mobil bekas yang bisa dibilang kondisi sudah tidak terawatt dan dimodifikasi menjadi mobil yang sangat keren.

2.3 Algoritma

2.3.1 K-Nearest Neighbor

$$Distance = \sqrt{\sum_{i=1}^n (x_{training}^i - x_{testing})^2}$$

Rumus 2. 1 K-Nearest Neighbor

Keterangan variable:

$x_{training}^i$ = Data *train* ke- *i*

$x_{testing}$ = Data *testing*

$i=1$ = Baris ke- dari table

n = Jumlah data *train*

K-Nearest Neighbors merupakan metode algoritma yang sering sekali digunakan dan sangat populer dalam penyelesaian sebuah kasus klasifikasi dan peramalan karena hasilnya yang cukup akurat. Selain hasilnya penggunaan KNN ini memiliki beberapa kelebihan yaitu mudah dimengerti, mudah diterapkan dan sederhana. Namun dari semua kelebihan tentu saja KNN ini memiliki kelemahan yaitu adanya bobot yang sama pada setiap atribut sehingga jika ada atribut yang tidak relevan maka hasilnya akan sama dengan atribut yang relevan. Selain itu penggunaan KNN adalah metode yang menggunakan *supervised*. Penggunaan

klasifikasi KNN ini bertujuan untuk mengklasifikasi objek baru berdasarkan atribut dan training [16][17].

2.3.2 Linear Regression

$$y = a + bx$$

Rumus 2. 2 Linear Regression

Dimana:

Y = variabel Dependent

x = variabel Independent

a = konstanta

b = jumlah respon yang dihasilkan oleh prediktor

Linear Regression adalah model statistik yang digunakan untuk memprediksi hubungan antara dua variabel dengan mengasumsikan adanya hubungan linier antara variabel independen dan dependen. Sederhananya, regresi linier melibatkan variabel independen dan variabel dependen. Model ini berupaya menemukan garis lurus yang paling sesuai yang mencerminkan hubungan antara variabel-variabel ini, memperkirakan garis lurus yang meminimalkan jumlah selisih kuadrat antara nilai prediksi dan nilai *actual* [18]. Dalam model ini nantinya akan dicari RMSE. Metode RMSE digunakan untuk menghitung kesalahan prediksi dengan harapan dapat mengurangi kesalahan sebanyak mungkin dengan menggabungkan semua informasi yang sangat terkait [19].

Terdapat dua jenis utama regresi linear: Regresi linear sederhana: mencakup variabel bebas dan variabel terikat. Garis lurus terbaik yang dihasilkan model ini mencerminkan hubungan linier antara variabel-variabel tersebut. Garis ini diperkirakan menggunakan kuadrat terkecil, yang berupaya menemukan garis yang paling dekat dengan titik data. Regresi Linear Berganda: Berisi beberapa variabel bebas dan satu variabel terikat. Model ini berupaya mencari garis lurus yang paling sesuai yang mencerminkan hubungan linear antara variabel terikat dan masing-masing variabel bebas. Regresi linear digunakan untuk menganalisis

dan memprediksi tren data di berbagai bidang seperti ekonomi, keuangan, dan ilmu sosial.

2.4 Tools dan Teknik

2.4.1 Python

Python adalah bahasa pemrograman yang didirikan oleh Guido van Rossum pada tahun 1991 dan pertama kali dirilis pada tahun 1994. Python dikenal dengan sintaksisnya yang sederhana dan elegan, memungkinkan pengembang untuk menulis kode yang lebih pendek dan lebih mudah dibaca dibandingkan bahasa pemrograman lainnya. Python juga dikenal sebagai bahasa pemrograman yang mudah dipelajari sehingga memungkinkan pengembang baru memahami dan menggunakan bahasa tersebut dengan cepat.

Python adalah bahasa pemrograman berorientasi objek. Artinya, semua data dan fungsi di Python adalah objek. Python mendukung berbagai paradigma pemrograman, termasuk pemrograman fungsional, pemrograman objek, dan pemrograman imperatif. Python juga mendukung pemrograman multiparadigma, memungkinkan pengembang untuk menggunakan banyak paradigma dalam satu program. Python memiliki perpustakaan standar yang sangat lengkap yang mencakup berbagai modul untuk berbagai tugas seperti pemrosesan data, pengembangan web, pengembangan GUI, dll [20].

Pustaka standar Python memungkinkan pengembang mengembangkan aplikasi kompleks dengan cepat dan efisien. Python juga mendukung pengembangan aplikasi web menggunakan berbagai *framework* seperti Django dan Flask. *Framework* ini mengurangi jumlah kode yang perlu ditulis, sehingga memudahkan pengembang untuk membuat aplikasi web yang kompleks.

Python juga memiliki komunitas besar dan aktif yang menyediakan berbagai sumber daya untuk belajar dan mendapatkan dukungan seperti Dokumentasi, Tutorial, dan Forum Diskusi. Komunitas juga memainkan peran penting dalam pengembangan dan peningkatan Python, dengan banyak pengembang yang berkontribusi terhadap pengembangan bahasa dan perpustakaan standar. Python juga dikenal karena penggunaannya dalam berbagai bidang, termasuk

pengembangan web, pengembangan aplikasi *desktop*, pengembangan game, analisis data, dan bahkan bidang ilmiah seperti pembelajaran mesin dan kecerdasan buatan. Secara keseluruhan, Python adalah bahasa pemrograman yang kuat dan fleksibel dengan dukungan luas untuk berbagai tugas pemrograman dan pengembangan aplikasi. Python juga memiliki komunitas besar dan aktif yang membantu pengembang belajar dan mengatasi tantangan.

2.4.2 Jupyter

Jupyter adalah perangkat lunak sumber terbuka untuk membuat perangkat lunak dan layanan standar komputasi interaktif dari berbagai bahasa pemrograman [21]. Aplikasi utama Jupyter adalah Jupyter Notebook. Aplikasi ini adalah aplikasi format dokumen terbuka yang memungkinkan pengguna membuat dan berbagi program yang dapat dikombinasikan dengan teks penjelasan, simbol persamaan, visualisasi interaktif, gambar, dan banyak lagi. Pada tahun 2014, Jupyter muncul sebagai proyek induknya. Python digunakan untuk komputasi ilmiah, dan bahasa pemrograman Python telah berpindah ke bidang ilmu data dan pembelajaran mesin. Aplikasi lain untuk Jupyter termasuk JupyterLab, nbconvert, Jupyter Widgets, Jupyter Hub, Binder, nbviewer, dan banyak lagi.

2.4.3 Scraping

Scraping adalah suatu proses yang melibatkan ekstraksi data dari situs web atau aplikasi web lainnya. Proses ini biasanya dilakukan dengan menggunakan bahasa pemrograman seperti Python. Dengan menggunakan Python, pengguna dapat mengakses halaman web, menavigasi struktur HTML, dan mengekstrak data yang dibutuhkan. *Scraping* memiliki berbagai aplikasi dan tujuan, termasuk analisis data, pengumpulan data untuk proyek pribadi, dan pengambilan data yang diperlukan untuk aplikasi lain. Proses ini memungkinkan pengguna untuk mengumpulkan data secara efisien dan efektif, yang kemudian dapat digunakan untuk berbagai keperluan, seperti pengembangan model prediktif, analisis sentimen, atau pengembangan algoritma pembelajaran mesin. *Scraping* juga memainkan peran penting dalam dunia akademis dan penelitian, di mana data yang diperoleh dapat digunakan untuk mendukung penelitian dan

pengembangan ilmu pengetahuan. Scraping dapat digunakan untuk berbagai tujuan, termasuk analisis data, pengumpulan data untuk proyek pribadi, dan pengambilan data yang diperlukan untuk aplikasi lain [22].

2.4.4 Octoparse

Octoparse merupakan alat gratis yang digunakan untuk web scraping dan pengumpulan data. Terdapat produk dari Octopus Data Inc., yang berbasis di Walnut, CA. Alat ini memungkinkan pengguna untuk mengekstrak data dari situs web secara otomatis dan efisien. Octoparse mendukung berbagai jenis situs web dan memiliki fitur yang memungkinkan pengguna untuk menyesuaikan proses pengumpulan data sesuai kebutuhan mereka. Selain itu, Octoparse juga menawarkan berbagai fitur lain seperti pengumpulan data dari halaman yang berisi data dinamis, pengumpulan data dari situs web yang memerlukan login, dan banyak lagi. Alat ini sangat berguna bagi pengguna yang ingin mengumpulkan data dari internet untuk analisis atau tujuan lainnya tanpa harus melakukan manual [23].

2.4.5 Streamlit

Streamlit adalah aplikasi gratis dan pengguna tidak memerlukan pengetahuan pengembangan front end tingkat lanjut untuk menggunakannya. Streamlit dapat berjalan di editor Anaconda serta seri Python 3.7 ke atas, tetapi tidak mendukung *editor* Notebook Jupyter sehingga diperlukan konversi ke Pycharm atau Visual Editor kode. Layar beranda pada aplikasi Streamlit dapat dipisahkan menjadi dua bagian, yaitu tombol pemilihan menu dan tampilan grafik visual. Ini memaksa perpustakaan NumPy dan Pandas untuk menampilkan grafik. Tombol-tombol tersebut berfungsi untuk memilih kumpulan data dari kategori negara, jenis hewan, arsitektur dan lapisan tersembunyi, pengoptimal, serta opsi untuk zaman dan prakiraan selama beberapa tahun ke depan [24].

2.4.6 Cross-Industry Standard Process for Data Mining (CRISP-DM)

CRISP-DM, singkatan dari Cross-Industry Standard Process for Data Mining, adalah sebuah kerangka kerja yang populer digunakan untuk proyek- proyek data mining dan analitik [25]. Ia memberikan pendekatan terstruktur dalam mengarahkan

tahapan-tahapan proyek, dimulai dari pemahaman masalah bisnis hingga penerapan model akhir. Terdapat enam tahap dalam CRISP-DM [25]:

1. **Pemahaman Bisnis:** pada tahap ini adalah memahami tujuan dan kebutuhan proyek dari sudut pandang bisnis.
2. **Pemahaman Data:** pada tahap ini melibatkan pengumpulan dan eksplorasi data yang akan digunakan untuk analisis. Data dikaji dari berbagai sumber, dan evaluasi kualitasnya dilakukan. Pemahaman awal tentang struktur, konten, dan hubungan data diperoleh.
3. **Persiapan Data:** Setelah pemahaman, langkah selanjutnya adalah mempersiapkan data untuk pemodelan. Pada tahap ini akan dilakukan pembersihan data, transformasi ke format yang sesuai, dan seleksi variable yang relevan.
4. **Pemodelan:** Pada tahap ini, berbagai teknik pemodelan diterapkan pada data yang telah dipersiapkan untuk membangun dan mengevaluasi model-model prediktif atau deskriptif.
5. **Evaluasi:** Model yang dipilih dievaluasi untuk memastikan kesesuaian dengan tujuan bisnis dan kinerja yang baik pada data yang tidak terlihat sebelumnya. Penilaian model menggunakan teknik evaluasi yang sesuai.
6. **Implementasi:** Setelah evaluasi model, model yang terpilih diimplementasikan dalam lingkungan operasional. Ini melibatkan integrasi model dalam proses bisnis, pemantauan performa, dan dukungan berkelanjutan.

U M W N
U N I V E R S I T A S
M U L T I M E D I A
N U S A N T A R A