

## BAB II

### LANDASAN TEORI

#### 2.1 Penelitian Terdahulu

Tabel 2. 1 Tabel Penelitian Terdahulu Mengenai Algoritma Machine Learning

No	Nama Penelitian	Nama Jurnal	Nama Penulis	Hasil Penelitian
1	<i>House Price Prediction using Machine Learning Algorithms</i>	<i>International Journal on Recent and Innovation Trends in Computing and Communication</i> Vol. 11 No.9 (2024)	Angulakshmi M, Deepa M, Mala Serene I, Thilagavathi M, Aarthi P.	Penelitian ini menghasilkan <i>Random Forest</i> yang memiliki akurasi tertinggi, yaitu 89%, dalam memprediksi harga rumah, lebih baik dibandingkan dengan algoritma lainnya seperti <i>Decision Tree</i> . Hasil ini menunjukkan bahwa <i>Random Forest</i> adalah metode yang cocok untuk <i>dataset</i> penelitian ini[14].
2	<i>Comparative Analysis of Linear Regression, Random Forest Regression and Gradient Boosted Trees Regression Method for House Price Prediction</i>	<i>Journal of Applied Computer Science and Technology (JACOST)</i> Vol. 4 No. 1 (2023)	Evita Fitri	Penelitian ini membandingkan tiga algoritma untuk memprediksi harga rumah: <i>Linear Regression, Random Forest Regression, dan Gradient Boosted Trees Regression</i> . Dengan <i>dataset</i> 2011 record, <i>Random Forest</i> memberikan akurasi terbaik 81.5% dan RMSE 0.440, lebih unggul dibandingkan <i>Linear Regression</i> (RMSE 0.515) dan <i>Gradient Boosted Trees</i> (RMSE 0.508)

No	Nama Penelitian	Nama Jurnal	Nama Penulis	Hasil Penelitian
3	<i>House Price Prediction Using Web Scraping and Machine Learning with Linear Regression Algorithm</i>	Jurnal Teknik Informatika dan Sistem Informasi Vol. 8 No. 1 (2021)	Andi Saiful, Septi Andryana, Aris Gunaryati.	Penelitian ini menggunakan <i>web scraping</i> dan <i>Linear Regression</i> untuk memprediksi harga rumah, dengan akurasi prediksi mencapai 88%. Model ini menggunakan variabel seperti luas lahan, luas bangunan, jumlah kamar tidur, kamar mandi, dan tempat parkir. Meskipun akurasi baik, nilai RMSE tercatat cukup tinggi, yaitu Rp 259,171.9 [17].
4	<i>Predicting Property Prices with Machine Learning Algorithms</i>	<i>Journal of Property Research</i> Vol. 38 No. 1 (2021)	Winky K.O. Ho, Bo-Sin Tang & Siu Wai Wong.	Penelitian ini menggunakan tiga algoritma <i>machine learning: Support Vector Machine, Random Forest, dan Gradient Boosting Machine</i> untuk memprediksi harga properti dengan data 40.000 transaksi rumah di Hong Kong. Hasilnya menunjukkan bahwa <i>Random Forest</i> dan <i>Gradient Boosting Machine</i> memberikan akurasi prediksi yang lebih baik dibandingkan <i>Support Vector Machine</i> , dengan tingkat kesalahan yang lebih rendah pada MSE, RMSE, dan MAPE[12].

No	Nama Penelitian	Nama Jurnal	Nama Penulis	Hasil Penelitian
5	<i>Research on Prediction and Analysis of Real Estate Market Based on the Multiple Linear Regression Model</i>	<i>Hindawi Scientific Programming</i> Vol. 2 No. 2 (2023)	Guangjie Liu	Penelitian ini menunjukkan penggunaan algoritma <i>Multiple Linear Regression</i> cukup efektif untuk memprediksi harga properti di China dalam kurun waktu 10 tahun terakhir. Hasil error yang didapat dari <i>Multiple Linear Regression</i> adalah kurang dari 8%[15].
6	<i>House Price Prediction using Random Forest Machine Learning Technique</i>	<i>Procedia Computer Science</i> Vol. 181 No. 2019 (2021)	Abigail Bola Adetunjia, Oluwatobi Noah Akande, Funmilola Alaba Ajala, Ololade Oyewo, Yetunde Faith Akande, Gbenle Oluwadara.	Penelitian ini menggunakan teknik <i>machine learning Random Forest</i> untuk memprediksi harga rumah dengan <i>dataset Boston Housing</i> dari <i>UCI Machine Learning Repository</i> . Model ini menggunakan 14 fitur, termasuk ukuran rumah, jumlah kamar, dan tingkat keamanan di lingkungan sekitar. Hasilnya menunjukkan bahwa model <i>Random Forest</i> dapat memprediksi harga rumah dengan selisih yang dapat diterima, yaitu $\pm 5$ dari harga aktual dan akurasi prediksi mencapai 90%. Ini menunjukkan potensi <i>Random Forest</i> sebagai alat yang efektif untuk prediksi harga rumah[18].

No	Nama Penelitian	Nama Jurnal	Nama Penulis	Hasil Penelitian
7.	<i>A Comparative Study of House Price Prediction Using Linear Regression and Random Forest Models</i>	<i>Highlights in Science, Engineering and Technology</i> Vol. 107 No. 27 (2024)	Yahan Fu	Penelitian ini membandingkan kinerja model <i>Linear Regression</i> dan <i>Random Forest</i> dalam memprediksi harga rumah. Hasilnya menunjukkan bahwa model <i>Random Forest</i> memberikan akurasi yang lebih baik, dengan RMSE 148,428.13 dan $R^2$ 0.85, dibandingkan dengan <i>Linear Regression</i> yang memiliki RMSE 214,472.76 dan $R^2$ 0.70[19].
8.	<i>Gradient Boosting–Based Machine Learning Methods in Real Estate Market Forecasting</i>	<i>Advances in Intelligent Systems Research</i> Vol. 174 No. 2 (2021)	Nikita Fedorov, Yulia Petrichenko.	Penelitian ini membandingkan empat model prediksi harga rumah: <i>Linear Regression</i> , <i>CatBoostRegressor</i> , <i>XGBRegressor</i> , dan <i>AdaBoostRegressor</i> . Hasilnya menunjukkan bahwa <i>CatBoostRegressor</i> memberikan akurasi terbaik, dengan nilai $R^2$ sebesar 0.9004 untuk Kota Chelyabinsk dan 0.8605 untuk Kota Khanty-Mansiysk. Model <i>XGBRegressor</i> dan <i>AdaBoostRegressor</i> juga menunjukkan kinerja yang baik, namun sedikit lebih rendah dibandingkan <i>CatBoostRegressor</i> .

No	Nama Penelitian	Nama Jurnal	Nama Penulis	Hasil Penelitian
				Sebaliknya, model <i>Linear Regression</i> memberikan hasil yang paling buruk, terutama untuk Kota Khanty-Mansiysk[20].
9.	<i>Comparison of Extreme Gradient Boosting and Random Forest Algorithm Performance for Predicting House Prices in Jabodetabek</i>	Jurnal Ilmu Komputer dan Sistem Informasi Vol. 8 No. 1 (2024)	Dhiwa Aqsha	Penelitian ini membandingkan kinerja algoritma <i>Extreme Gradient Boosting (XGBoost)</i> dan <i>Random Forest</i> dalam memprediksi harga rumah di Jabodetabek. Hasilnya menunjukkan bahwa model <i>Random Forest</i> memiliki performa terbaik dengan MAE 95,200,513.25, MSE 1.47e+19, dan R <sup>2</sup> 0.77, sedangkan <i>XGBoost</i> memiliki MAE 121,836,703.27, MSE 3.03e+19, dan R <sup>2</sup> 0.52. Model <i>Random Forest</i> lebih akurat dalam memprediksi harga rumah dengan kesalahan yang lebih kecil dibandingkan <i>XGBoost</i> [16].
10.	<i>Implementation of Machine Learning for House Price Prediction Using the Random Forest Algorithm</i>	Jurnal Ilmu Komputer dan Sistem Informasi Vol. 8 No. 1 (2024)	Nicholas Hadi, Jason Benedict.	Penelitian ini membandingkan tiga algoritma untuk memprediksi harga rumah: <i>Random Forest</i> , <i>Decision Tree</i> , dan <i>Polynomial Regression</i> . Hasilnya menunjukkan bahwa model <i>Random Forest</i> dengan 20 Trees memberikan akurasi terbaik, mencapai

No	Nama Penelitian	Nama Jurnal	Nama Penulis	Hasil Penelitian
				86,54% ( $R^2$ ) dengan RMSE sebesar 144,913.73. Dibandingkan dengan <i>Decision Tree</i> yang memperoleh akurasi 76,39% dan <i>Polynomial Regression</i> dengan akurasi 78,13%, <i>Random Forest</i> terbukti lebih akurat dalam memprediksi harga rumah.

Penelitian mengenai prediksi harga rumah dengan menggunakan algoritma *machine learning* menunjukkan bahwa *Random Forest* sering kali memberikan hasil yang paling akurat dibandingkan dengan algoritma lainnya. Dalam beberapa penelitian, seperti yang dilakukan oleh “*House Price Prediction using Machine Learning Algorithms*”, *Random Forest* mencapai akurasi hingga 89%, sementara model lainnya seperti *Decision Tree* memiliki akurasi yang lebih rendah[14]. Di sisi lain, penelitian yang membandingkan *Linear Regression*, *Random Forest Regression*, dan *Gradient Boosted Trees Regression* yang berjudul “Analisis Perbandingan Metode Regresi Linier, *Random Forest Regression* dan *Gradient Boosted Trees Regression Method* untuk Prediksi Harga Rumah” menunjukkan bahwa *Random Forest* unggul dengan akurasi 81.5% dan RMSE 0.440. Berbeda dengan penggunaan *Linear Regression* yang mencatatkan RMSE yang lebih tinggi[13]. Pada penelitian “Prediksi Harga Rumah Menggunakan Web Scrapping Dan Machine Learning Dengan Algoritma Linear Regression” penggunaan web scrapping dengan *Linear Regression* untuk memprediksi harga rumah menghasilkan akurasi sebesar 88%, meskipun nilai RMSE cukup tinggi[17]. Sebuah studi yang berjudul “*Predicting Property Prices with Machine Learning Algorithms*” yang menggunakan data transaksi rumah di Hong Kong menunjukkan bahwa *Random Forest* dan *Gradient Boosting Machine* memiliki akurasi yang lebih baik

dibandingkan *Support Vector Machine*, dengan error yang lebih rendah di MSE, RMSE, dan MAPE[12]. Selain itu, penelitian tentang Multiple Linear Regression di China yang berjudul “*Research on Prediction and Analysis of Real Estate Market Based on the Multiple Linear Regression Model*” menunjukkan efektivitas model tersebut dengan error kurang dari 8%. Penelitian lain tentang dataset Boston Housing yang berjudul “*House Price Prediction using Random Forest Machine Learning Technique*” mengindikasikan bahwa *Random Forest* dapat memprediksi harga rumah dengan akurasi mencapai 90% dan selisih harga aktual yang dapat diterima. Sementara itu, studi perbandingan Linear Regression dan Random Forest yang berjudul “*A Comparative Study of House Price Prediction Using Linear Regression and Random Forest Models*” menunjukkan bahwa Random Forest memberikan kinerja yang lebih baik, dengan RMSE lebih rendah dan  $R^2$  lebih tinggi[19]. *CatBoostRegressor*, *XGBRegressor*, dan *AdaBoostRegressor* juga diuji pada penelitian “*Gradient Boosting-Based Machine Learning Methods in Real Estate Market Forecasting*” dengan *dataset* harga rumah dan ditemukan bahwa *CatBoostRegressor* memberikan akurasi terbaik, terutama di dua kota Rusia, dengan nilai  $R^2$  mencapai 0.9004 dan 0.8605[20]. Penelitian lain berjudul “Perbandingan Kinerja Algoritma Extreme Gradient Boosting dan Random Forest untuk Prediksi Harga Rumah di Jabodetabek” membandingkan *XGBoost* dan *Random Forest* di Jabodetabek menunjukkan bahwa *Random Forest* lebih unggul dengan MAE 95,200,513.25 dan  $R^2$  0.77, sementara *XGBoost* memiliki  $R^2$  yang lebih rendah (0.52)[16]. Terakhir, penelitian yang berjudul “Implementasi Machine Learning untuk Prediksi Harga Rumah Menggunakan Algoritma Random Forest” membandingkan *Random Forest*, *Decision Tree*, dan *Polynomial Regression* menegaskan bahwa *Random Forest* memiliki akurasi terbaik, mencapai 86.54%, jauh lebih unggul dibandingkan dengan *Decision Tree* (76.39%) dan *Polynomial Regression* (78.13%)[21]. Secara keseluruhan, *Random Forest* muncul sebagai metode yang lebih konsisten dan akurat dalam memprediksi harga rumah dibandingkan dengan metode lainnya di sebagian besar penelitian yang dilakukan.

## 2.2 Objek Penelitian

### 2.1.1 Rumah

Rumah adalah tempat untuk beristirahat, berlindung dari cuaca, dan menjalani aktivitas sehari-hari. Rumah biasanya terdiri dari berbagai ruangan dan fasilitas yang dirancang untuk memenuhi kebutuhan dasar manusia seperti tidur, makan, mandi, dan beraktivitas[22].

Selain menjadi tempat tinggal, rumah juga memiliki arti simbolis dan emosional yang dalam di kehidupan manusia. Rumah dianggap sebagai tempat yang nyaman, aman, dan penuh kenangan. Rumah merupakan tempat di mana individu dan keluarga merasa terlindungi dan dapat mengekspresikan identitas mereka melalui dekorasi dan perabotan rumah[23].

Rumah dapat menjadi investasi penting dalam kehidupan seseorang dan dapat mencerminkan status sosial, budaya, dan ekonomi individu atau keluarga. Selain itu, konsep rumah memiliki makna yang berbeda dalam berbagai budaya dan tradisi, dan masyarakat di seluruh dunia melihatnya dengan cara yang berbeda[24].

### 2.1.2 Lamudi

Lamudi merupakan platform yang menyediakan layanan dalam bidang properti, khususnya dalam pemasaran dan pencarian properti. Pengguna platform Lamudi dapat mencari dan menyewa atau membeli properti, seperti rumah, apartemen, dan kantor. Lamudi didirikan pada tahun 2013 dan berfokus pada menyediakan layanan *online* seperti properti. Lamudi juga menjadi bagian dari Rocket Internet group, yang juga dikenal sebagai salah satu inkubator bisnis online terkemuka di dunia. Lamudi juga merupakan bagian dari EMPG (Emerging Markets Property Group), tujuannya adalah mengembangkan industri properti di Indonesia.

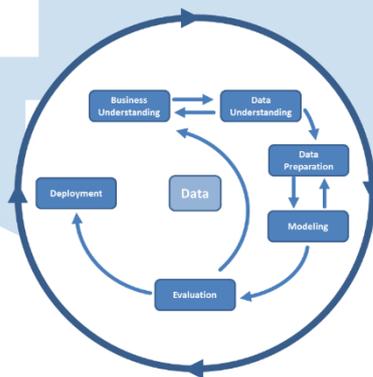
Pengguna Lamudi dapat mencari properti sesuai dengan kriteria tertentu, seperti lokasi, harga, ukuran, dan jenis properti. Mereka juga dapat melihat deskripsi, foto, dan detail terkait properti yang diiklankan. *Platform* ini

membantu orang-orang yang mencari tempat tinggal atau investasi properti dengan memudahkan pencarian dan perbandingan opsi yang tersedia.

## 2.3 Teori tentang Framework / Algoritma yang digunakan

### 2.2.1 CRISP-DM

CRISP-DM, juga dikenal sebagai *Cross-Industry Standard Process for Data Mining*, adalah suatu kerangka kerja atau metodologi dalam bidang ilmu *data mining* yang digunakan untuk mengelola proyek analisis data dengan efektif. Metodologi ini terdiri dari beberapa langkah yang membantu organisasi atau profesional data menemukan, memahami, dan mengeksploitasi pola dan informasi penting yang terkandung dalam data[25]. CRISP-DM terdiri dari enam tahap, yaitu:



Gambar 2. 1 Siklus CRISP-DM

Sumber: [26]

1. **Business Understanding**, Tahap ini melakukan pemahaman terhadap tujuan masalah dari mengidentifikasi masalah yang ingin dipecahkan dengan analisis data[26].
2. **Data Understanding**, Pada tahap ini, data yang berkaitan dengan penelitian ini akan dianalisis lebih lanjut. Pada tahap ini dilakukannya pengumpulan data, menjelajahi struktur dan karakteristiknya, serta mengevaluasi kualitas data[26].

3. **Data Preparation**, Pada tahap ini, data yang sudah dikumpulkan akan dilakukan proses *cleaning* data hingga mendapatkan data yang sempurna atau data yang sudah siap dipakai dalam penelitian ini[26].
4. **Modeling**, Pada tahap ini mengaplikasikan berbagai model data untuk mengidentifikasi pola, hubungan, tren dalam data. Fase ini termasuk penggunaan algoritma *machine learning* atau statistik untuk membangun model prediksi atau klasifikasi[26].
5. **Evaluation**, Pada tahap ini, akan dilakukan evaluasi terhadap performa model yang telah dibuat pada tahap modeling sebelumnya. Apabila model yang telah dibuat sebelumnya memberikan performa yang memuaskan, maka dapat lanjut ke tahap terakhir dalam CRISP DM. Tetapi, jika performa model kurang memuaskan, dapat dilakukan optimasi model kembali[26].
6. **Deployment**, Merupakan tahap terakhir dalam CRISP-DM. Pada tahap ini, model yang memiliki performa terbaik dapat di deploy, menyesuaikan dengan masalah yang sudah didefinisikan di tahap awal[26].

### 2.2.2 Multiple Linear Regression

*Multiple Linear Regression* adalah bagian dari teknik *Linear Regression* yang digunakan untuk memodelkan hubungan antara satu variabel dependen (Y) dan lebih dari satu variabel independen (X)[15]. Teknik ini berawal pada abad ke-19, saat para ilmuwan seperti Sir Francis Galton menganalisis hubungan antara tinggi badan orang tua dan anaknya pada tahun 1886[27]. Seiring berjalannya waktu, Adrien-Marie Legendre dan Carl Friedrich Gauss mengembangkan metode ini dengan menemukan metode *least squares* untuk mencari garis linear terbaik yang dapat menggambarkan hubungan antar variabel

dalam data. Sejak saat itu, regresi linear telah berkembang menjadi alat penting dalam analisis statistik, ekonometri, dan ilmu sosial[28].

Dalam aplikasinya, *Multiple Linear Regression* digunakan untuk memprediksi nilai dari variabel dependen berdasarkan beberapa variabel independen. Misalnya, untuk memprediksi harga rumah, kita bisa menggunakan variabel seperti luas tanah, jumlah kamar tidur, dan lokasi sebagai variabel independen untuk memprediksi harga rumah yang menjadi variabel dependen. Model ini kini banyak digunakan dalam berbagai bidang, seperti analisis tren pasar, prediksi harga saham, dan prediksi ekonomi, berkat kemajuan dalam komputasi dan perangkat lunak statistik yang memudahkan analisis dan interpretasi data[29].

Rumus dari *Multiple Linear Regression* dapat dilihat pada rumus 2.1 di bawah. Sumber : [15]

$$Y = \alpha + \beta_1 X_1 + \dots + \beta_n X_n + e \quad 2.1$$

Di mana:

$Y$  : Variabel dependen

$X$  : Variabel independen

$\alpha$  : Konstanta (Intercept)

$\beta$  : Koefisien determinasi (Slope)

$e$  : error

### 2.2.3 Gradient Boosting

*Gradient boosting* adalah teknik *ensemble learning* dari *Decision Tree* yang telah menjadi salah satu alat paling kuat dalam dunia *machine learning* dan *data science*. Pada tahun 1980-an, penelitian tentang *Decision Trees* semakin berkembang, terutama dengan salah satu metode yang disebut "*boosting*".

*Boosting* adalah metode yang menghasilkan model yang lebih kuat dengan menggabungkan model prediktif yang lemah (model yang tidak layak dalam memprediktif)[30].

Dengan perkembangan teknologi dan komputasi yang semakin canggih, serta minat yang terus tumbuh dalam *machine learning* dan *data science*, *gradient boosting* telah menjadi salah satu teknik yang paling efektif dan umum digunakan dalam pengolahan data dan prediksi.

Alur dari *Gradient Boosting* dapat dilihat dibawah:

1. Menggunakan rumus untuk melakukan inisialisasi prediksi dengan *Decision Tree* sederhana. Sumber :[31]

$$F_0(x) = \underset{y}{\operatorname{argmin}} \sum_{i=1}^n L(y_i, y) \quad 2.2$$

2. Mencari nilai prediksi dengan melakukan kalkulasi residual dengan formula.

$$\hat{y}_{im} = - \left[ \frac{\partial \Psi(y_i F(x_i))}{\partial F(x_i)} \right]_{f(x) = f_{m-1}(x)} \quad 2.3$$

3. Membuat *Decision Tree* tambahan yang memprediksi residual dari semua variabel independen.
4. Memperbarui prediksi dengan menghitung prediksi baru dengan learning rate menggunakan formula seperti

$$F_m(x) = F_{m-1}(x) + v \sum_{j=1}^{jm} y_{jm} 1(x \in R_{jm}) \quad 2.4$$

5. Mengulangi tahap ke 2 sampai 4 sesuai dengan jumlah iterasi atau *trees*.

*Gradient Booting* menjalankan *fitting* untuk mengurangi pada model tambahan dari *training* data (*square error*). *Lost Function* akan mengukur berapa banyaknya data yang menyimpang dari nilai aslinya.

#### 2.2.4 *Random Forest*

*Random Forest* adalah algoritma *ensemble learning* yang menggunakan *Decision Tree* untuk meningkatkan akurasi prediksi. Dikembangkan oleh Leo Breiman pada awal 2000-an, metode ini menjadi sangat populer karena kemampuannya untuk menangani data besar, menangani *overfitting*, dan memberikan hasil yang sangat akurat. Dalam *Random Forest*, setiap *Decision Tree* dilatih pada sampel acak dari dataset yang berbeda, dengan proses yang disebut *bootstrap sampling*. Setiap *Tree* menghasilkan prediksi, dan hasil akhir diambil dengan melakukan *voting* untuk klasifikasi atau rata-rata untuk regresi[32].

Cara kerja *Random Forest* terbagi menjadi dua fase utama: fase pertama melibatkan pembuatan sejumlah  $N$  *Decision Tree* secara acak, sedangkan fase kedua menghasilkan prediksi berdasarkan hasil dari  $N$  *Decision Tree* tersebut[14]. Rumus untuk prediksi *Random Forest* adalah sebagai berikut:

$$\text{Prediksi} = \frac{1}{K} \sum_{k=1}^K w_k h_k(x) \quad 2.5$$

Di mana:

$K$  : adalah jumlah *Decision Tree*

$w_k$  : bobot dari *Decision Tree* ke- $k$ ,

$h_k(x)$  : prediksi yang dihasilkan oleh *Decision Tree* ke- $k$  untuk input  $x$ .

### 2.2.5 Cross Validation

*Cross-validation* adalah sebuah teknik dalam *machine learning* yang digunakan untuk menilai hasil performa sebuah model. Tujuan dari penggunaan teknik *Cross-validation* yaitu untuk pengukuran data yang belum pernah dilihat sebelumnya sehingga dapat digeneralisasi dengan baik kedepannya [33].

*Cross-validation* membantu mengurangi risiko *overfitting* karena model diuji pada subset data yang tidak digunakan selama *training*. Hal ini memberikan pengukuran data yang belum pernah dilihat sebelumnya sehingga dapat digeneralisasi dengan baik kedepannya.

### 2.2.6 Metrik Evaluasi

Dalam masalah regresi, metrik evaluasi yang digunakan berbeda dengan yang digunakan pada klasifikasi. Metrik yang sering digunakan dalam regresi adalah *R-Squared* ( $R^2$ ), *Mean Absolute Percentage Error* (MAPE), dan *Root Mean Squared Error* (RMSE)[34]. Namun, dalam penelitian ini, hanya metrik  $R^2$  dan RMSE yang akan digunakan. Berikut adalah penjelasan mengenai kedua metrik tersebut.

#### a. *R-Squared* ( $R^2$ )

$R^2$  atau yang bisa disebut dengan *R-squared* adalah nilai yang menampilkan pengaruh variabel independen terhadap variabel independen. Nilai  $R^2$  berkisar antara 0 sampai dengan 1. Rumus dari  $R^2$  dapat dilihat pada rumus 2.6 di bawah: Sumber: [34]

$$R^2 = \frac{\sum_{t=1}^n (y_t - \hat{y})^2}{\sum_{t=1}^n (y_t - \bar{y})^2} \quad 2.6$$

Di mana:

$y_t$  : data yang diuji ke-t ( $t=1, \dots, n$ ),

$\hat{y}$  : prediksi dari respons ke-t ( $t=1, \dots, n$ ),

$\bar{y}$  : rata-rata,

$n$  : banyaknya data yang diuji.

Selanjutnya, tabel 2.1 di bawah menunjukkan interpretasi dari nilai  $R^2$ .

Tabel 2. 2 Interpretasi Nilai  $R^2$

Koefisien	Nilai
1 – 0,8	Sangat Kuat
0,79 – 0,6	Kuat
0,59 – 0,4	Cukup Kuat
0,39 – 0,2	Lemah
0,19 – 0	Sangat Lemah

Sumber: [34]

#### b. *Root Mean Square Error (RMSE)*

*Root Mean Square Error (RMSE)* adalah suatu metrik evaluasi dalam machine learning yang digunakan untuk mengukur seberapa besar deviasi atau kesalahan antara nilai yang diobservasi dan nilai yang diprediksi oleh model. RMSE memberikan pengukuran rata-rata dari selisih absolut antara nilai prediksi dan nilai yang sebenarnya, dan nilai RMSE yang lebih rendah menunjukkan bahwa model memiliki kesalahan prediksi yang lebih kecil[35]. Rumus pada RMSE dapat dilihat di rumus 2.7: Sumber: [35]

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (y_i - t_j)^2} \quad 2.7$$

Di mana:

$y_i$  : adalah nilai yang diprediksi oleh model untuk data ke- $i$ ,

$t_j$  : adalah nilai yang sebenarnya untuk data ke- $i$ ,

$N$  : banyaknya data yang diuji

## 2.4 Teori tentang Tools / Software yang digunakan

### 2.3.1 Python

*Python* merupakan bahasa pemrograman komputer yang sudah banyak digunakan dalam pengembangan *software*, *data analysis*, pemrograman web, dan yang lainnya. Dikembangkan pertama kali pada awal tahun 1990 oleh Guido van Rossum, *Python* dikenal dengan sintaksis yang bersifat mudah dibaca dan mudah dipahami, yang membuatnya menjadi salah satu bahasa pemrograman yang sangat populer di kalangan pemula dan pengembang berpengalaman[36].

Python ini memiliki sejumlah pustaka (*libraries*) dan kerangka kerja (*frameworks*) yang kuat, seperti NumPy, Pandas, Matplotlib untuk analisis data, Django dan Flask untuk pengembangan web, dan TensorFlow serta PyTorch untuk kecerdasan buatan (*artificial intelligence*).

### 2.3.2 Jupyter

*Jupyter* adalah sebuah *software open-source* yang menyediakan lingkungan komputasi interaktif untuk pengembangan di berbagai bahasa pemrograman. *Jupyter* sering digunakan untuk analisis data, visualisasi, dan pengembangan prototipe dalam domain ilmu data dan pengembangan perangkat lunak.

*Jupyter* mendukung banyak bahasa pemrograman, tetapi bahasa pemrograman yang paling umum digunakan di dalam notebook adalah Python. Dengan fleksibilitas dan fungsionalitasnya, *Jupyter* telah menjadi salah satu alat yang sangat berharga dalam dunia ilmu data dan pengembangan perangkat lunak[37].

### 2.3.3 Selenium

*Selenium* adalah sebuah perangkat lunak yang digunakan untuk mengotomatisasi pengujian perangkat lunak berbasis web. *Selenium* menyediakan berbagai *tools* dan *libraries* untuk mengotomatisasi tindakan yang biasanya dilakukan oleh *user* dalam menginteraksi dengan aplikasi web. *Selenium* juga sering digunakan dalam skenario pengujian otomatis, pemantauan situs web, dan otomatisasi tugas-tugas yang melibatkan browser web.

*Selenium* sangat penting dalam dunia pengujian perangkat lunak karena memungkinkan otomatisasi pengujian aplikasi web dengan lebih mudah dan cepat dibandingkan dengan pengujian manual. Selain itu, *Selenium* juga digunakan dalam pengembangan *web scraping* (*data mining* dari situs web) dan untuk tujuan otomatisasi berbagai tugas berbasis web[38].

### 2.3.4 Flask

Flask adalah kerangka kerja (*framework*) mikro untuk pengembangan aplikasi web berbasis Python. Disebut "mikro" karena memiliki desain yang ringan dan sederhana, fokus pada kebutuhan dasar untuk membangun aplikasi web tanpa banyak komponen yang sudah tertanam. Flask memberikan dasar yang kuat untuk membangun aplikasi web dengan cepat dan efisien.

Flask cocok untuk pengembangan proyek-proyek kecil hingga menengah, prototipe, dan aplikasi web yang membutuhkan kontrol yang lebih terhadap komponen-komponennya[39].

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