

BAB II

LANDASAN TEORI

2.1 Penelitian Terdahulu

Tabel 2.1 di bawah merupakan tabel penelitian terdahulu yang mengandung judul jurnal, nama jurnal, nama penulis, tahun publikasi, hasil penelitian, dan kesimpulan yang ditemukan mengenai hubungan antara penelitian terdahulu dan penelitian ini

Tabel 2.1 Penelitian Terdahulu

No	Judul Jurnal	Nama Jurnal	Penulis / Tahun	Hasil	Kesimpulan
1.	Analyzing driving factors of land values in urban scale based on <i>big data</i> and non-linear <i>machine learning</i> techniques [6]	Land Use Policy	Jun Maa, Jack C.P. Chenga , Feifeng Jiangb , Weiwei Chena , Jingcheng Zhangc, / 2020	Penelitian ini menggunakan konsep <i>big data</i> untuk mengeksplorasi fitur-fitur suatu daerah terhadap nilai dan perkembangan suatu daerah. Berdasarkan penelitian.	Penelitian ini berfungsi sebagai acuan dalam penggunaan <i>big data</i> dalam proses memahami faktor pendorong yang dapat mempengaruhi nilai properti
2.	Machine Learning based Predicting House Prices using Regression Techniques [11]	Proceedings of the Second International Conference on Innovative Mechanisms for Industry Applications	Manasa J, Narahari N S, Radha Gupta / 2020	Penelitian ini membandingkan 5 model regresi untuk mencari model dengan akurasi prediksi harga rumah terbaik. Berdasarkan penelitian SVR dan <i>XGBoost</i> merupakan 2 algoritma dengan akurasi terbaik dengan nilai r^2 66% dan 79% pada dataset <i>testing</i>	Penelitian ini dijadikan acuan dalam cara kerja implementasi SVM dan <i>XGBoost</i> , selain itu penelitian ini juga memberikan contoh metrik evaluasi yang dapat digunakan untuk membandingkan model regresi

No	Judul Jurnal	Nama Jurnal	Penulis / Tahun	Hasil	Kesimpulan
3.	Land Price Forecasting Research by Macro and Micro Factors and <i>Real estate</i> Market Utilization Plan Research by Landscape Factors: Big Data Analysis Approach [5]	Land Price Forecasting with Big Data Analysis approach	Sang-Hyang Lee , Jae-Hwan Kim and Jun-Ho Huh / 2021	Ketertarikan pelanggan terhadap pasar <i>real estate</i> dipengaruhi oleh beberapa faktor seperti akses transportasi, stabilitas ekonomi, dan saham. Berdasarkan penelitian, dari sisi ekonomi makro, fluktuasi pemegang valuta asing mempengaruhi harga <i>real estate</i> .	Penelitian ini mendukung teori terkait faktor-faktor yang mempengaruhi nilai properti secara makro dan mikro
4.	Predictive analytics using Big Data for the <i>real estate</i> market during the COVID-19 pandemic[12]	Predictive analytics on <i>real estate</i> market	Andrius Grybauskas , Vaida Pilinkienė and Alina Stundzienė / 2021	Dalam periode 4 bulan dari Mei hingga Agustus terdapat 17,2% jumlah apartemen yang mengalami penurunan harga sewa dan 10,7% harga penjualan. Perubahan harga apartemen cenderung mengecil pada setiap awal bulan. Adanya korelasi antara perubahan harga apartemen dengan dampak umum COVID-19	Penelitian ini menjadi acuan dalam penggunaan predictive analytics dalam memahami perubahan nilai properti di masa yang akan datang akibat faktor eksternal atau makro
5.	The Research Development of Hedonic Price Model-Based <i>Real estate</i> Appraisal in the Era of Big Data [13]	Land Price Appraisal in the era of Big Data	Cankun Wei , Meichen Fu , Li Wang , Hanbing Yang , Feng Tang and Yuqing Xiong / 2022	Penggunaan Big Data dinilai membantu proses <i>real estate appraisal</i> . Temuan utama penelitian terletak pada peran Big Data dalam penggunaan <i>machine learning</i>	Penelitian ini mendukung tujuan penelitian dalam pembuatan sistem prediksi menggunakan model <i>machine learning</i> untuk prediksi harga rumah

No	Judul Jurnal	Nama Jurnal	Penulis / Tahun	Hasil	Kesimpulan
				untuk meningkatkan akurasi <i>Home Price Model</i> (HPM)	
6	An Efficient Parameter Adaptive <i>Support Vector Regression</i> Using K-Means Clustering and Chaotic Slime Mould Algorithm [14]	IEEE ACCESS	Ziyi Chen, Wenbai Lu	Kombinasi algoritma <i>clustering</i> K-Means dan <i>Support Vector Regression</i> untuk meningkatkan performa algoritma SVR baik dari segi kecepatan maupun stabilitas	Penelitian ini dijadikan pedoman dalam mengimplementasi kombinasi algoritma <i>clustering</i> K-Means dan algoritma prediksi SVR
7	Streamflow forecasting using extreme gradient boosting model coupled with Gaussian mixture model [15]	Journal of Hydrology	Lingling Nia, Dong Wanga, Jianfeng Wua, Yuankun Wanga, Yuwei Taoa, Jianyun Zhang, Jiufu Liu	Kombinasi algoritma <i>clustering</i> dan <i>XGBoost</i> untuk meningkatkan performa forecasting	Penelitian ini dijadikan acuan dalam implementasi kombinasi <i>clustering</i> terhadap algoritma <i>XGBoost</i>
8	Prediction Analysis Sales for Corporate Services Telecommunications Company using GradientBoost Algorithm [16]	International Conference on Broadband Communications, Wireless Sensors and Powering	Oryza Wisesa, Andi Andriansyah, Osamah Ibrahim Khalaf	Implementasi <i>Gradient Boosting</i> untuk prediksi sales perusahaan telekomunikasi yang dievaluasi menggunakan metrik MAD, MSE, RMSE, dan MAPE	Penelitian ini dijadikan pedoman dalam menentukan metrik evaluasi yang akan digunakan untuk algoritma <i>Gradient Boost</i>

No	Judul Jurnal	Nama Jurnal	Penulis / Tahun	Hasil	Kesimpulan
9	Prediction of House Price Using <i>XGBoost</i> Regression Algorithm [17]	Turkish Journal of Computer and Mathematics Education	J.Avanijaa, Gurram Sunithab, K.Reddy Madhavic, Padmavathi Korad,R.Hites h SaiVittale	Akurasi dari prediksi harga rumah menggunakan <i>XGBoost</i> menunjukkan <i>error rate</i> yang rendah yaitu 4.4% ketika menggunakan rasio <i>splitting data</i> 80/20 untuk <i>training</i> dan <i>testing</i>	Penelitian ini dijadikan acuan kapabilitas <i>XGBoost</i> sebagai algoritma prediksi harga rumah dan pedoman dalam menentukan rasio <i>splitting data</i>
10	House Price Prediction With Statistical Analysis in Support Vector Machine Learning for Regression Estimation [18]	Curiosity: Interdisciplinary Journal of Research and Innovation	Cesar Vasquez, Vinodh Chellamuthu	Implementasi algoritma SVR untuk prediksi harga rumah dengan membandingkan 5 jenis <i>hyper-dimensional kernel</i> , <i>crossvalidation</i> , dan metrik evaluasi. Hasil setiap penggunaan <i>kernel</i> menunjukkan hasil yang baik adalah <i>kernel polynomial</i>	Penelitian ini dijadikan pedoman dalam mengatur parameter SVR untuk meningkatkan akurasi dan penggunaan metrik evaluasi untuk SVR

Berdasarkan tabel 2.1, penelitian terdahulu mendukung pembuatan sistem prediksi dapat membantu dalam meningkatkan akurasi pengecekan harga properti sesuai dengan hasil penelitian [6]. Sistem prediksi akan mempertimbangkan faktor-faktor yang mempengaruhi nilai properti seperti pada penelitian [5]. Penelitian ini akan membandingkan performa model SVR, *Gradient Boost*, dan *XGBoost* untuk prediksi harga rumah berdasarkan model yang digunakan oleh penelitian [12][11][13][17]. Model prediksi yang terbaik akan dikombinasikan dengan algoritma *clustering* seperti pada penelitian [14] [15] untuk meningkatkan akurasi prediksi serta memberikan gambaran kondisi pasar berdasarkan spesifikasi rumah.

Pemilihan metrik evaluasi untuk pencarian model prediksi dengan performa terbaik akan didasarkan pada penelitian [16][18]. Perbedaan antara penelitian terdahulu dengan penelitian ini adalah belum ditemukan penelitian yang membahas prediksi harga rumah di wilayah Bandung dan belum ditemukan juga adanya penelitian yang membandingkan performa model prediksi antara SVR, *XGBoost*, dan *Gradient Boost* yang dikombinasikan dengan *clustering* untuk memprediksi harga rumah di wilayah Bandung.

2.2 Teori Penelitian

2.2.1 Analisis Pasar *Real estate*

Real estate merupakan industri yang terbagi menjadi 5 kategori utama yaitu properti perumahan, properti komersial, lahan tanah, properti industri, dan fasilitas khusus [2]. Untuk mengetahui kondisi pasar industri *real estate* diperlukan adanya *market analysis*. Pada industri *real estate* proses analisis pasar disebut juga dengan *real estate market analysis* [5]. *Real estate market analysis* berfungsi untuk memberikan data-data penting terkait kebijakan dan proses administrasi yang ada, dimana nilai dari suatu lahan atau properti akan bergantung banyaknya permintaan dan penawaran, sama seperti pasar-pasar pada umumnya. Namun nilai *real estate* tidak hanya ditentukan oleh *supply* dan *demand*, selain itu terdapat banyak faktor-faktor mikro maupun makro seperti kebijakan pemerintah, lingkungan sosial dan ekonomi daerah tersebut, dan demografi penduduk, lokasi properti, fasilitas properti, dan infrastruktur yang tersedia di sekitar daerah Analisis terhadap faktor yang ada di daerah tujuan penting dilakukan untuk mendapatkan *insight* baru [5].

Analisis data *market real estate* dapat dilakukan menggunakan *Big Data Analytics*, menggunakan algoritma *machine learning* sebagai alat prediksi. *Big Data* merupakan data berjumlah besar baik terstruktur maupun tidak terstruktur yang terus tercipta dengan cepat dari berbagai

sumber, untuk menangani data dalam jumlah besar secara *real-time* diperlukan adanya prosesor, *software*, dan algoritma yang kuat. *Big Data Analytics* adalah suatu ilmu yang mempelajari teknik dan metode analisis terhadap Big Data. Tahapan dalam mengimplementasi *Big Data Analytics* dapat dimulai melalui pendekatan *data mining* yaitu CRISP-DM [11].

2.2.2 Property Technology

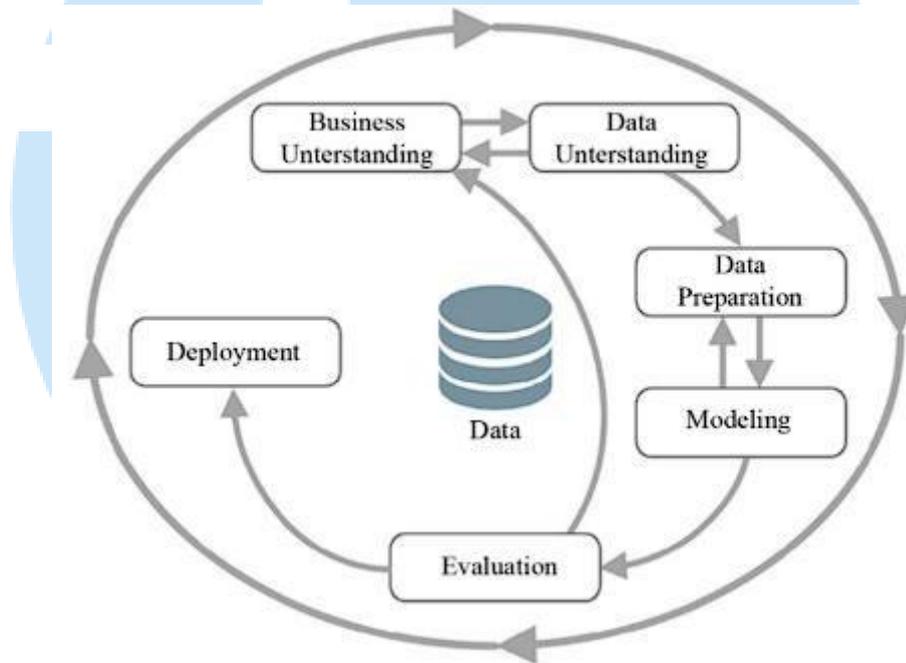
PropTech, singkatan dari *Property Technology*, merujuk pada penggunaan teknologi dalam industri properti untuk meningkatkan efisiensi, produktivitas, dan pengalaman pengguna. PropTech mencakup berbagai teknologi seperti platform digital untuk penyewaan atau penjualan properti, aplikasi untuk manajemen properti, pemantauan IoT (*Internet of Things*) untuk pemeliharaan dan keamanan properti, serta solusi keuangan digital seperti *crowdfunding* untuk investasi properti. PropTech telah mengubah cara industri properti beroperasi, memungkinkan proses transaksi yang lebih cepat dan efisien, pemeliharaan properti yang lebih baik melalui pemantauan otomatis, dan pengalaman pengguna yang lebih baik melalui platform digital yang interaktif [19].

Penerapan PropTech telah memberikan dampak yang signifikan pada berbagai aspek industri properti, mulai dari pengembang, agen real estat, hingga investor. Hal ini juga membuka peluang baru dalam inovasi dan kreativitas, dengan perusahaan-perusahaan PropTech terus mengembangkan solusi baru untuk mengatasi tantangan dalam industri properti. Dengan perkembangan teknologi yang terus berlanjut, diharapkan PropTech akan terus menjadi salah satu kekuatan utama dalam transformasi industri properti menuju masa depan yang lebih efisien dan terhubung secara digital [19].

2.3 Framework dan Algoritma

2.3.1 CRISP-DM

CRISP-DM merupakan pedoman umum yang digunakan oleh *user* untuk proses *data mining*. CRISP-DM memiliki beberapa tahap yang perlu dilakukan sebelum membuat model prediksi yang tepat.



Gambar 2.1 Framework CRISP-DM
Sumber: Jurnal Sumber

Gambar 2.1 menggambarkan alur kerja proses CRISP-DM. Tahap *Business Understanding* merupakan salah satu tahap yang vital, pada tahap ini dibutuhkan adanya pemahaman terhadap permasalahan yang sedang dihadapi oleh suatu bisnis agar dapat menentukan secara jelas tujuan *Big Data Analytics* dilakukan, menentukan batasan dari masalah, mencari tahu sumber dari data yang diperlukan, serta menentukan model *machine learning* yang akan digunakan. Tahap *Data Understanding* berfungsi untuk membangun fondasi dari proses analisis yang akan dilakukan, pada tahap ini diperlukan adanya proses pengumpulan dan penentuan data-data yang dapat digunakan untuk proses analisis. Pada tahap *Data Preparation* akan dilakukan pembersihan, normalisasi, standarisasi, dan integrasi data, proses yang dilakukan pada tahap ini berguna untuk mempersiapkan data yang ada dengan membersihkan data

redundan, *null*, penyeragaman data, dan integrasi berbagai sumber data menjadi satu dataset. *Data Modeling* berfungsi untuk menentukan jenis model statistik dari data apakah data tersebut bersifat prediktif atau deskriptif, sehingga dapat diketahui struktur dari data yang ada dan dapat menentukan model *machine learning* yang tepat untuk digunakan. Pada tahap *Machine learning Methods*, akan ditentukan model algoritma yang paling tepat digunakan untuk mencari solusi dari permasalahan yang ada [20].

2.3.2 *Support Vector Regression*

Support Vector Regression (SVR) merupakan algoritma *supervised* yang digunakan untuk memprediksi nilai diskrit. Dalam suatu SVR terdapat garis lurus yang dibutuhkan untuk memuat data yang disebut juga dengan *hyperplane*. SVR cocok digunakan untuk perancangan model prediksi nilai properti *real estate* karena kemampuannya dalam memetakan input atribut linear atau non-linear pada *high dimensional feature space* [8]. Tujuan dari algoritma ini adalah untuk menemukan *hyperplane* dalam sebuah ruang dimensional n yang dengan jelas mengklasifikasikan poin data. Beberapa *Hyperparameters* dalam vital untuk menjalankan SVR adalah *Hyperplane*, *Kernel*, *Boundary Lines* [8]. *Hyperplane* adalah batasan keputusan yang digunakan untuk memprediksi output prediksi secara terus-menerus, dimana data poin yang berada di sekitar *hyperplane* tersebut disebut juga dengan *Support Vectors* yang menjadi patokan untuk merencanakan garis prediksi pada algoritma. *Kernel* merupakan himpunan fungsi matematika yang mengambil data sebagai input dan mengubahnya menjadi bentuk yang dibutuhkan, parameter tersebut digunakan untuk menemukan *hyperplane* pada ruang dimensional yang lebih tinggi. *Boundary Lines* merupakan dua garis yang berada di sekitar *hyperplane* berdasarkan jarak epsilon yang digunakan untuk membuat margin antara poin data [21]. Berikut adalah rumus umum *Support Vector Regression* Linear [22].

$$f(x) = (w, x) + b \text{ with } w \in X, b \in R \text{ [1]}$$

Rumus 2.1 Rumus Support Vector Regression Linear

Keterangan:

- $f(x)$ adalah fungsi regresi yang dicari
- w adalah vector *weight*
- x adalah vektor fitur input
- b adalah bias

Berikut adalah rumus umum *Support Vector Regression* Non-Linear yang menggunakan *kernel*.

$$f(x) = \sum_i^{l_{sv}} (a_i - \alpha_i^*) k(x_i, x) + b \text{ [2]}$$

Rumus 2.2 Rumus Support Vector Regression Non-Linear

Keterangan:

- $f(x)$ adalah fungsi regresi yang dicari
- l_{sv} adalah jumlah data yang dilatih
- a_i dan α_i^* koefisien *Lagrange* yang diperoleh dari penyelesaian masalah optimasi.
- $k(x_i, x)$ fungsi *kernel* yang digunakan untuk memetakan data ke ruang fitur yang lebih tinggi dimensinya
- b adalah bias

2.3.3 Gradient Boosting

Gradient Boosting merupakan algoritma *machine learning* yang menggabungkan beberapa model prediksi untuk membuat satu model prediksi yang kuat. Algoritma tersebut berfungsi dengan menambahkan prediktor kepada suatu *ensemble* secara konsekutif untuk memperbaiki kekurangan dari model prediksi yang sebelumnya. Proses tersebut memanfaatkan turunan gradien untuk meminimalkan kekurangan dan

mengoptimalkan model prediksi. Karakteristik dari *Gradient Boosting* adalah fleksibilitasnya untuk proses prediksi maupun klasifikasi. Berikut adalah rumus dari *Gradient Boost* [23].

$$F_0(x) = \operatorname{arg}_y \min \sum_{i=1}^n L(y_i, \gamma) \quad [3]$$

Rumus 2.3 Rumus Gradient Boost

Keterangan:

- $F_0(x)$ adalah model awal *ensemble Gradient Boosting* yang biasanya mewakili nilai konstan
- $\operatorname{arg}_y \min$ adalah notasi matematika yang berarti "nilai argumen yang meminimalkan" dimana y adalah nilai yang dicari
- $\sum_{i=1}^n L(y_i, \gamma)$ adalah penjumlahan atas semua n titik data pelatihan dan evaluasi perbedaan nilai *loss*

2.3.4 *XGBoost*

XGBoost (*Extreme Gradient Boosting*) adalah algoritma *machine learning* yang dirancang untuk kecepatan dan kinerja. *XGBoost* bekerja dengan membangun serangkaian *decision tree* secara berurutan, di mana setiap pohon berusaha memperbaiki kesalahan yang dibuat oleh pendahulunya, sehingga meningkatkan akurasi model secara bertahap. Perbedaan utama antara *XGBoost* dan *gradient boosting* tradisional terletak pada optimalisasi dan fitur-fiturnya. *XGBoost* menawarkan peningkatan kecepatan dan efisiensi melalui algoritma paralel dan pemangkasan pohon yang lebih agresif, serta mengatasi *overfitting* dengan teknik regularisasi yang lebih kuat, sedangkan *gradient boosting* tradisional mungkin tidak seefisien dalam penanganan skala besar data atau mungkin tidak memiliki fitur regularisasi bawaan yang sama kuatnya. Berikut adalah rumus dari *XGBoost* [24].

$$\hat{y}_i = \sum_{t=1}^T f_t(x_i), f_t \in F \quad [4]$$

Rumus 2.4 Rumus Extreme Gradient Boost

Keterangan:

- \hat{y}_i adalah nilai yang diprediksikan.

- $f_t(x_i)$ adalah mewakili nilai prediksi dari pohon residual ke-t untuk residual ke-t dari x_i
- f_t adalah jumlah residual pada ronde ke-t
- F adalah ruang fungsi dari pohon residual

2.3.5 K-means

K-means *clustering* adalah algoritma pembelajaran mesin tanpa supervisi yang populer digunakan untuk membagi dataset menjadi K *cluster* yang berbeda dan tidak tumpang tindih [14]. Algoritma ini bekerja dengan cara secara iteratif mengalokasikan titik data ke *cluster centroid* terdekat dan kemudian memperbarui *centroid* berdasarkan rata-rata dari titik yang dialokasikan ke setiap *cluster*. Proses ini berlanjut sampai *centroid* tidak lagi berubah secara signifikan atau mencapai jumlah iterasi yang ditentukan. Salah satu kelebihan utama K-means adalah kesederhanaan dan efisiensi komputasinya, menjadikannya cocok untuk dataset besar.

$$\operatorname{argmin}_{c_i} \|x_i - c_i\|^2 \quad [5]$$

Rumus 2.5 Rumus K-Means

Keterangan:

- x_i adalah *data point*
- c_i adalah *centroid*
- $\|\cdot\|^2$ adalah kuadrat jarak Euclidean

2.4 Tools

2.4.1 Python

Python adalah bahasa pemrograman *open-source* yang dapat digunakan untuk berbagai kebutuhan pembuatan atau pengembangan aplikasi web, internet, *software* di bidang edukasi, sains, dan sebagainya. Bahasa pemrograman tersebut memiliki berbagai kapabilitas yang didukung oleh koleksi *package* yang banyak. Dalam bidang analisis data

beberapa *package* yang umum digunakan adalah *pandas*, *statsmodel*, *scikit-learn*, *numPy*, dan *sciPy*. *Pandas* berfungsi untuk memberikan kapabilitas pemrosesan dan manipulasi data yang, *statsmodel* berfungsi untuk pemodelan dan pengujian data, *scikit-learn* memberikan kapabilitas klasifikasi dan *machine learning*, *numPy* berfungsi untuk operasi numerik, dan *scipy* berfungsi untuk proses saintifik lainnya.

2.4.2 React

React adalah sebuah *framework* JavaScript yang populer digunakan dalam pengembangan aplikasi web modern. *Framework* ini dikembangkan oleh Facebook dan menjadi salah satu pilihan utama bagi pengembang karena kemampuannya dalam membangun antarmuka pengguna yang dinamis dan responsif. React menggunakan konsep komponen, di mana setiap bagian dari antarmuka pengguna dibuat sebagai komponen terpisah yang dapat diatur, dipelihara, dan digunakan kembali.

React juga memiliki fitur seperti virtual DOM (*Document Object Model*), yang memungkinkan pembaruan antarmuka pengguna dilakukan dengan cepat dan efisien tanpa harus melakukan *render* ulang seluruh halaman. Hal ini meningkatkan performa aplikasi secara keseluruhan. Dalam penelitian ini pembuatan *website* akan menggunakan TypeScript dibandingkan Javascript. TypeScript membuat kode lebih mudah dipahami dan dipelihara dengan memberikan informasi yang jelas tentang jenis data yang digunakan dalam komponen React. TypeScript juga mendukung fitur-fitur modern seperti *async/await* dan deklarasi tipe yang kuat, sehingga meningkatkan keamanan dan keterbacaan kode secara keseluruhan.

2.4.3 Flask

Flask adalah *framework* web yang ringan dan mudah digunakan untuk pengembangan aplikasi web di Python. Dengan pendekatan

yang minimalis dan fleksibel, Flask memberikan kontrol yang lebih besar kepada pengembang untuk merancang dan membangun aplikasi sesuai kebutuhan mereka. Kerangka kerja ini menawarkan berbagai fitur yang berguna, seperti *routing* URL, dukungan untuk *template* Jinja2, dan integrasi dengan berbagai ekstensi yang memperluas fungsionalitasnya, seperti Flask-SQLAlchemy untuk pengelolaan database, Flask-Login untuk otentikasi pengguna, dan Flask-RESTful untuk pembuatan API RESTful. Flask juga mendukung pengembangan aplikasi web skala kecil hingga menengah dengan dukungan yang luas dari komunitas Python.

Penggunaan Flask dalam pengembangan aplikasi web memberikan kelebihan dalam hal skalabilitas, kecepatan pengembangan, dan kemudahan pemeliharaan. Dengan desain yang sederhana namun kuat, Flask memungkinkan pengembang untuk fokus pada logika bisnis aplikasi tanpa terjebak dalam kerumitan teknis yang tidak perlu. Hal ini membuat Flask menjadi pilihan yang populer bagi pengembang yang ingin membangun aplikasi web yang efisien dan mudah diatur menggunakan bahasa pemrograman Python.

