

BAB III

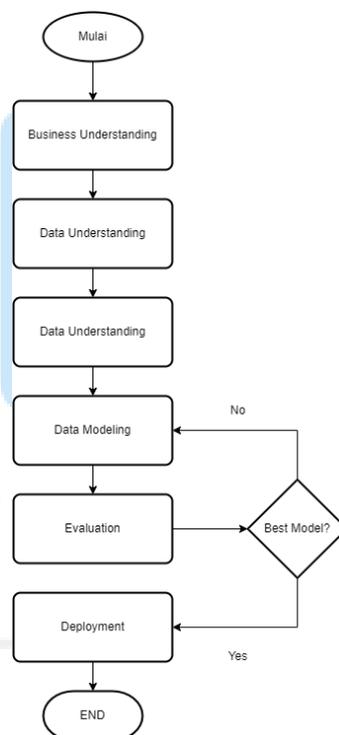
METODOLOGI PENELITIAN

3.1 Objek Penelitian

Objek dari penelitian ini adalah properti perumahan di Bandung yang mencakup daerah Kota Bandung, Kabupaten Bandung, dan Kabupaten Bandung Barat. Wilayah tersebut merupakan bagian dari wilayah Bandung Raya dengan total 77 kecamatan seperti Bojongsoang, Soreang, Ngamprah, Cileunyi, Babakan Cipayar, dan sebagainya. Data properti perumahan akan diambil melalui *website* jual beli properti perumahan yaitu Lamudi. Data dari *website* Lamudi yang akan diambil merupakan data listing rumah dari *range* harga Rp. 500.000.000 hingga Rp 10.000.000.000.

3.2 Alur Penelitian

Gambar 3.1 berikut adalah alur penelitian yang digunakan dalam penelitian ini.



Gambar 3.1 Alur Penelitian

Gambar 3.1 merupakan alur penelitian yang akan dilalui oleh penelitian ini. Alur penelitian didasarkan oleh penggunaan *framework* CRISP-DM yang akan digunakan. Penelitian dimulai dari tahap identifikasi masalah, tahap ini merupakan proses pemahaman tujuan dan kebutuhan dari penelitian. Tahap pengambilan data dilakukan dengan metode *web scrapping* menggunakan *library* Selenium. Tahap pengolahan data dilakukan untuk membersihkan data dari *missing value*, standarisasi format data, filterisasi, dan *One Hot Encoding* pada data kategorikal. Tahap pengembangan model dilakukan untuk setiap kombinasi algoritma regresi dan *clustering* yang digunakan dalam penelitian. Tahap Evaluasi model dilakukan untuk menentukan model prediksi terbaik yang akan diimplementasikan pada *website* sistem prediksi. Tahap pembentukan *website* dilakukan menggunakan *framework* React untuk *front-end* dan *framework* Flask untuk *back-end*.

3.3 Metode Penelitian

Berdasarkan tujuan dari penelitian yaitu untuk memprediksi harga rumah di Bandung, maka metodologi yang akan diimplementasi pada penelitian ini adalah metode pemahaman data. Berikut adalah beberapa contoh metodologi untuk proses data mining yaitu CRISP-DM, KDD, dan SEMMA.

Tabel 3.1 Perbandingan Metodologi Penelitian

Aspek	CRISP-DM	KDD	SEMMA
Tahapan	<ol style="list-style-type: none"> 1. Business Understanding 2. Data Understanding 3. Data Preparation 4. Modeling 5. Evaluation 6. Deployment 	<ol style="list-style-type: none"> 1. Selection 2. Preprocessing 3. Transformation 4. Data Mining 5. Interpretation/Evaluation 	<ol style="list-style-type: none"> 1. Sample 2. Explore 3. Modify 4. Model 5. Assessment

Aspek	CRISP-DM	KDD	SEMMA
Kelebihan	<ul style="list-style-type: none"> • CRISP-DM mencakup panduan langkah demi langkah yang komprehensif, mulai dari pemahaman bisnis hingga penerapan model. • Dapat diadaptasi untuk berbagai industri dan jenis proyek data mining • CRISP-DM merupakan metodologi yang paling banyak diadopsi dalam data mining dan ilmu data. 	<ul style="list-style-type: none"> • KDD fokus pada ekstraksi pengetahuan yang dapat diambil tindakan dari data yang besar. • Memberikan perhatian khusus pada pra-pemrosesan dan transformasi data untuk memastikan kualitas data yang tinggi. 	<ul style="list-style-type: none"> • SEMMA menawarkan pendekatan yang linear dan terstruktur dengan baik, membuatnya mudah untuk diikuti. • Metodologi ini menekankan pada pemilihan model dan evaluasi, yang cocok untuk analisis statistik dan pembelajaran mesin
Kekurangan	<ul style="list-style-type: none"> • Siklus iteratifnya mungkin memerlukan waktu dan upaya lebih untuk kembali ke tahap sebelumnya berdasarkan temuan baru • Tidak mencakup proses kolaborasi dalam proses <i>big data</i> 	<ul style="list-style-type: none"> • Proses KDD bisa menjadi sangat kompleks dan memakan waktu, terutama untuk dataset yang besar dan rumit. • KDD kurang menekankan pada tahap deployment model ke lingkungan produksi. 	<ul style="list-style-type: none"> • Pendekatan langkah demi langkah yang linear mungkin tidak mencerminkan sifat iteratif dari banyak proyek data science • SEMMA dikembangkan oleh SAS dan paling sering digunakan dalam ekosistem perangkat lunak SAS, yang mungkin membatasi bagi pengguna yang lebih memilih alat lain

Penelitian ini menggunakan *framework* CRISP-DM untuk karena keunggulannya yaitu fleksibel terhadap teknologi yang digunakan dan mencakup tahapan-tahapan yang sesuai dengan proses penelitian. *Framework* KDD tidak digunakan karena kurang menekankan bagian *deployment*,

sedangkan *framework* SEMMA tidak digunakan karena tidak fleksibel dalam penggunaan teknologi. Berikut adalah tahap-tahap dari *framework* CRISP-DM yang diimplementasikan dalam penelitian ini.

3.3.1 Business Understanding

Dalam tahap “Business Understanding” kebutuhan dan tujuan dari proses penelitian akan dibahas [20]. Hasil analisis ini akan didasarkan oleh pandangan terhadap permasalahan yang ada dan menghasilkan formula dari penyelesaian masalah untuk merancang strategi analisis yang tepat. Latar belakang penelitian ini membahas perkembangan pasar properti residensial, dimana terdapat kesempatan bagi pembeli rumah untuk mempertimbangkan pembelian rumah karena adanya penurunan indeks harga properti residensial dan peningkatan penjualan properti residensial berdasarkan laporan survei Bank Indonesia. Untuk mengetahui harga pasar berdasarkan spesifikasi rumah, diperlukan suatu sistem yang dapat membantu memprediksi harga rumah serta memberikan gambaran jumlah pesaing dengan spesifikasi rumah yang sama atau jumlah rumah alternatif yang dapat dibeli dengan spesifikasi yang serupa. Permasalahan data yang perlu diatasi adalah perlunya proses pencarian data yang sesuai untuk menggambarkan objek penelitian.

3.3.2 Data Understanding

Dalam tahap “Data Understanding” akan dikumpulkan data yang dibutuhkan untuk diteliti lebih detail lagi dengan tujuan mendapatkan wawasan mengenai kualitas dan informasi tersembunyi yang mungkin terdapat dalam data [20]. Pada penelitian ini, diperlukan adanya proses pengumpulan data properti di Bandung yang mengandung data spesifikasi rumah, fasilitas rumah, dan data terkait lokasi, seleksi dataset rumah dijual di Bandung, dan pembelajaran dataset yang didapatkan untuk memahami korelasi setiap variabel yang ada terhadap perubahan nilai properti.

3.4 Teknik Pengumpulan Data

Penelitian ini menggunakan data primer dengan pendekatan kuantitatif. Dataset yang digunakan adalah “Data properti rumah di Bandung” dan diperoleh melalui lamudi.com dengan menggunakan teknik *web scrapping* dengan *library* Selenium di Python. Pemilihan Lamudi.com untuk sumber data didasarkan pada struktur halaman *website* yang memudahkan pengambilan spesifikasi rumah melalui *element* HTML khusus yang disediakan oleh Lamudi dibandingkan *website* PropTech lainnya.

3.5 Teknik Analisis Data

3.5.1 Data Preparation

Dalam tahap “Data Preparation” Peneliti akan melakukan beberapa proses preparasi data menggunakan Python seperti *data cleaning*, *standardization*, *filtering*, dan implementasi *One Hot Encoding* [20]. Data cleaning adalah proses mengidentifikasi, mengoreksi, dan menghapus data yang tidak lengkap, tidak akurat, atau tidak relevan dari dataset. Standardisasi adalah langkah untuk mengubah data menjadi format yang seragam atau standar, seperti mengubah semua satuan ukuran ke dalam satu sistem satuan. Filtering adalah proses memilih subset data yang memenuhi kriteria tertentu, seperti mengambil hanya data yang relevan untuk analisis tertentu. Implementasi One Hot Encoding adalah teknik yang digunakan dalam pemrosesan data untuk mengubah data kategorikal menjadi bentuk numerik, di mana setiap nilai kategori diwakili sebagai vektor biner yang hanya memiliki nilai 0 atau 1. Setelah proses-proses preparasi data dilakukan, akan dilakukan pembagian data untuk proses pelatihan dan validasi data.

3.5.2 Data Modeling

Dalam tahap “Data Modeling” akan dibuat alur kerja *data mining* untuk memproses data menggunakan algoritma yang tepat[11]. Proses

clustering menggunakan K-means akan dilakukan terlebih dahulu sebelum proses *modeling* algoritma prediksi. Penggunaan algoritma K-means untuk *clustering* didasarkan pada penelitian terdahulu yang membuktikan metode kombinasi antara algoritma prediksi dan *clustering* K-means [14]. Hasil dari *clustering* data akan diteruskan untuk perancangan model prediksi menggunakan algoritma *Support Vector Regression*, *Gradient Boosting*, dan *XGBoost* pada setiap *cluster*. Pemilihan model regresi didasarkan pada keunggulan setiap algoritma regresi yang digunakan pada penelitian terdahulu [11] [18] [16]. Rasio data splitting yang akan digunakan adalah 80/20 masing-masing untuk testing dan training. Rasio tersebut ditentukan berdasarkan penelitian terdahulu terkait prediksi nilai properti yang pada umumnya menggunakan rasio 80/20 untuk data berjumlah besar [13].

3.5.3 Evaluation

Dalam tahap “Evaluation” akan diuji implementasi model yang telah dibuat dalam skenario praktikal dengan *real data* dan output dari pengujian akan dievaluasi berdasarkan metrik evaluasi antara SVR, *Gradient Boosting*, dan *XGBoost*. *Sensitivity Analysis* juga akan dilakukan pada ketiga model regresi untuk mengetahui fitur-fitur yang paling berpengaruh pada model regresi dan menjawab salah satu rumusan masalah penelitian. Model dengan performa terbaik akan digunakan dalam tahap *deployment* fitur prediksi harga rumah.

3.5.4 Deployment

Tahap *deployment* merupakan pengembangan *website* fitur prediksi dengan menggunakan React sebagai *framework front-end* dan Flask sebagai *framework back-end*. Pada *website* tersebut hanya akan ada satu algoritma prediksi yang diimplementasikan yaitu algoritma dengan performa terbaik. Fitur-fitur pada *website* akan dibuat berdasarkan *requirement* berikut.

Tabel 3.2 Software Requirement

website	Fitur		
	Landing Page	Specification Form	Map
Matheus Campos Housing Price	✗	✓	✓
Raul House Price Prediction	✓	✓	✗
Bangalore House Price Predictor	✓	✓	✗

Tabel 3.4 merupakan daftar fitur dari *website* prediksi harga rumah lainnya baik dari jurnal penelitian maupun *website* aktif. Berdasarkan tabel tersebut fitur yang paling sering muncul adalah *landing page* dan *specification form* oleh karena itu *website* yang akan dibuat pada penelitian ini akan mengandung fitur *landing page* yang menampilkan informasi umum terkait *website* dan *specification form* untuk memberikan *user* kapabilitas memprediksi harga rumah berdasarkan input spesifikasi rumah.