

# **DOKUMEN PROYEK**

## **12S3205 - PENAMBANGAN DATA**

### **Development of a Predictive Regression Model for House Prices Using Ensemble Stacking Techniques**

#### **Disusun Oleh:**

<b>12S22015</b>	<b>Angelina Nadeak</b>
<b>12S22029</b>	<b>Jeremy Samosir</b>
<b>12S22038</b>	<b>Ade Siahaan</b>
<b>12S22052</b>	<b>Rosari Simanjuntak</b>



**PROGRAM STUDI SARJANA SISTEM INFORMASI**

**FAKULTAS INFORMATIKA DAN TEKNIK ELEKTRO (FITE)  
INSTITUT TEKNOLOGI DEL  
TAHUN 2024/2025**

## DAFTAR ISI

<b>BAB I: PENDAHULUAN</b>	<b>4</b>
1.1 Latar Belakang	4
1.2 Rumusan Masalah	4
1.3 Tujuan Project	4
1.4 Manfaat Project	4
<b>BAB II: TINJAUAN PUSTAKA</b>	<b>4</b>
2.1 Machine Learning	4
2.2 Regressi	4
2.3 Ensemble Learning (Stacking)	4
2.4 Studi terkait	4
<b>BAB III: METODOLOGI PENELITIAN</b>	<b>4</b>
3.1 Desain Penelitian / Alur Proyek	4
3.2 Sumber Data	4
3.3 Teknik Pengolahan Data	4
3.4 Pemilihan Model	4
3.5 Evaluasi Model	4
<b>BAB IV: HASIL DAN PEMBAHASAN</b>	<b>5</b>
4.1 Visualisasi data	5
4.2 hasil model	5
4.3 evaluasi performa	5
<b>BAB V: KESIMPULAN DAN SARAN</b>	<b>5</b>
5.1 Ringkasan hasil	5
5.2 Kelebihan/kekurangan model	5

# **BAB I PENDAHULUAN**

## **1.1 Latar Belakang**

Perkembangan teknologi kecerdasan buatan dan pembelajaran mesin (machine learning) telah memberikan kontribusi signifikan dalam berbagai bidang, termasuk industri properti. Salah satu aplikasi penting dari teknologi ini adalah dalam memprediksi harga rumah secara akurat berdasarkan karakteristik atau fitur-fitur rumah yang tersedia. Informasi ini sangat berharga bagi agen properti, pembeli rumah, maupun pengembang dalam mengambil keputusan yang tepat.

Permasalahan yang umum dihadapi dalam memprediksi harga rumah adalah banyaknya variabel yang saling mempengaruhi, mulai dari ukuran rumah, jumlah kamar, lokasi, hingga kualitas material bangunan. Oleh karena itu, dibutuhkan pendekatan pemodelan yang mampu menangkap hubungan kompleks setiap fitur tersebut.

Salah satu pendekatan yang menjanjikan adalah ensemble learning, khususnya metode stacking. Teknik ini menggabungkan prediksi dari beberapa model dasar (base learners) ke dalam satu model tingkat atas (meta-learner) untuk meningkatkan akurasi dan generalisasi model. Dalam penelitian ini, base models yang digunakan adalah XGBoost, LightGBM, dan CatBoost yang merupakan tiga algoritma gradient boosting telah terbukti unggul dalam berbagai kompetisi dan aplikasi dunia nyata.

## **1.2 Rumusan Masalah**

Rumusan masalah dalam penelitian ini adalah:

1. Bagaimana membangun model regresi untuk memprediksi harga rumah menggunakan teknik ensemble stacking?
2. Sejauh mana kombinasi base models (XGBoost, LightGBM, dan CatBoost) dapat meningkatkan akurasi prediksi harga rumah?

## **1.3 Tujuan Project**

Tujuan dari project ini adalah untuk mengembangkan sebuah model regresi prediktif yang mampu memperkirakan harga rumah secara akurat dengan menggunakan pendekatan ensemble stacking. Dalam pendekatan ini, beberapa model dasar (base models) seperti XGBoost, LightGBM, dan CatBoost digunakan untuk membangun prediksi awal. Hasil prediksi dari model-model dasar tersebut kemudian digabungkan dan dilatih kembali menggunakan sebuah model tingkat atas (meta-learner) guna menghasilkan prediksi akhir yang lebih akurat dan stabil. Dengan memanfaatkan keunggulan masing-masing algoritma boosting serta pendekatan stacking, penelitian ini bertujuan untuk meningkatkan performa model dalam memprediksi harga rumah, khususnya dengan meminimalkan kesalahan yang diukur melalui metrik Root Mean Squared Error (RMSE).

## **1.4 Manfaat Project**

Penelitian ini diharapkan dapat memberikan pendekatan alternatif dalam pemodelan prediksi harga rumah yang lebih akurat dan andal melalui penerapan teknik ensemble stacking. Dengan menggabungkan kekuatan dari beberapa algoritma boosting seperti XGBoost, LightGBM, dan CatBoost, model yang dihasilkan mampu menangkap pola kompleks dalam data dan menghasilkan prediksi yang lebih stabil. Selain itu, hasil dari penelitian ini juga diharapkan dapat menjadi referensi bagi praktisi dan peneliti dalam menerapkan teknik ensemble stacking pada permasalahan regresi lainnya, khususnya yang melibatkan data dengan karakteristik dan kompleksitas serupa.

## **BAB II: TINJAUAN PUSTAKA**

### **2.1 Machine Learning**

### **2.2 Regresi**

### **2.3 Ensemble Learning (Stacking)**

### **2.4 Studi terkait**

## **BAB III: METODOLOGI PENELITIAN**

### **3.1 Desain Penelitian / Alur Proyek**

### **3.2 Sumber Data**

### **3.3 Teknik Pengolahan Data**

### **3.4 Pemilihan Model**

### **3.5 Evaluasi Model**

## **BAB IV: HASIL DAN PEMBAHASAN**

### **4.1 Visualisasi data**

### **4.2 hasil model**

### **4.3 evaluasi performa**

## **BAB V: KESIMPULAN DAN SARAN**

### **5.1 Ringkasan hasil**

### **5.2 Kelebihan/kekurangan model**