



CALIFORNIA HOUSING PRICE PREDICTION

Presented by :
Bonifasius Sinurat
JCDS 3004 - 009



Outline

Chapter 1 - Data & Business Understanding

Chapter 2 - Preprocessing

Chapter 3 - Modeling & Interpretation

Chapter 4 - Business Calculations

Chapter 5 - Conclusion & Recommendations





CHAPTER 1

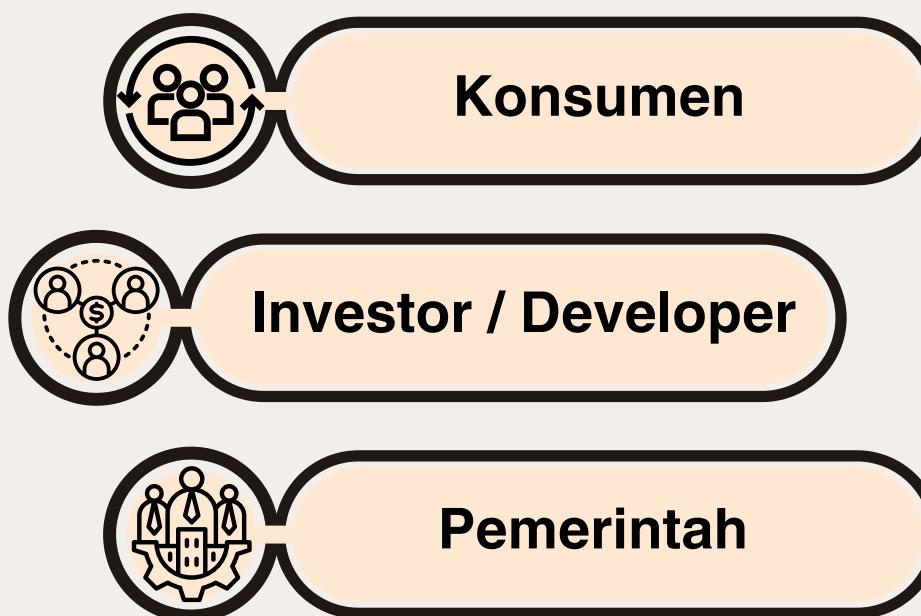
DATA & BUSINESS UNDERSTANDING

CONTEXT



Indikator utama ekonomi & kesejahteraan sosial

Aktor-aktor yang berperan:



Variasi harga dipengaruhi oleh :



CALIFORNIA HOUSING PRICE DATASET

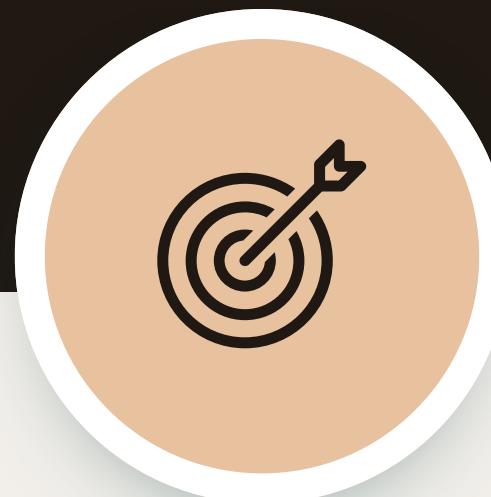
- Data Sensus California Tahun 1990 (hunian per distrik)
- 14448 *entries*
- 137 *missing values* & 0 data duplikat
- 9 kolom numerik, 1 kolom kategori

Source: Pace & Barry, 1997



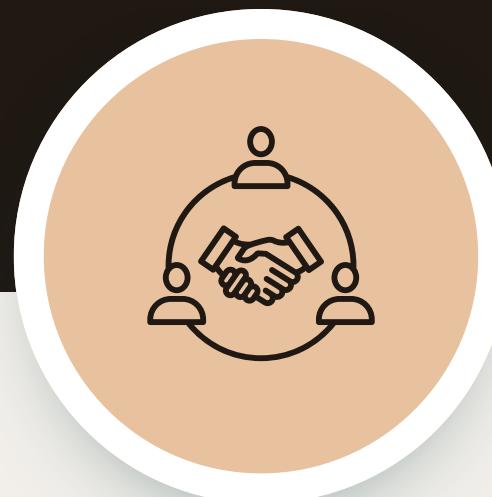
Problem Statement

Bagaimana cara **memprediksi berbagai tipe harga rumah secara wajar** untuk dapat **menetapkan harga yang lebih kompetitif?**



Goals

- **Mengidentifikasi 3 faktor utama** yang berpengaruh terhadap harga rumah.
- Toleransi tingkat akurasi prediksi model **maksimal 15%** dari harga aktual.



To Whom It May Concern

★ **Chief Marketing Officer**

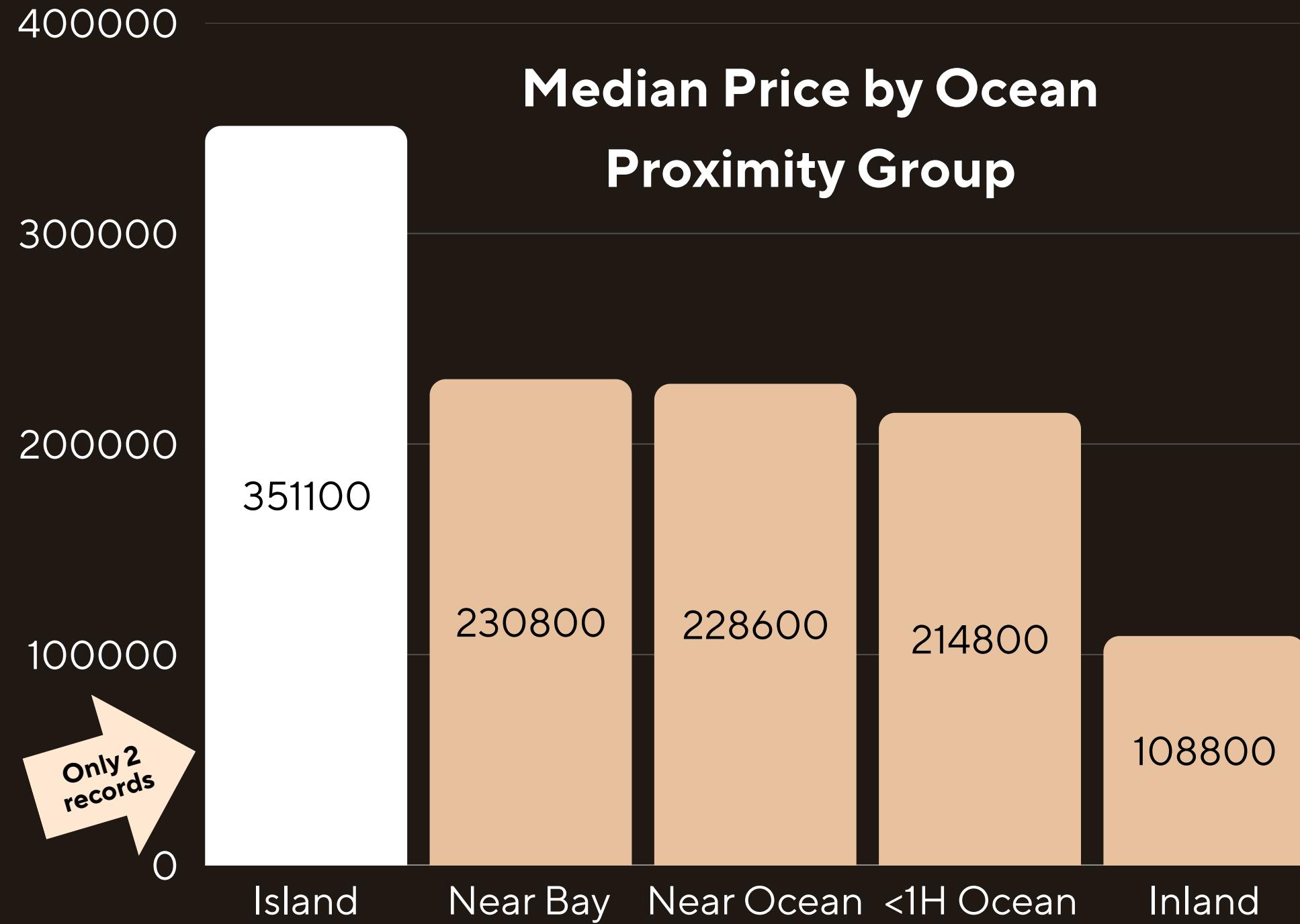
★ **Sales Director**



CHAPTER 2

PREPROCESSING

UNIQUE FINDINGS



House Value	Count
500001.0	678
500000.0	18
483300.0	6
...	...

Terdapat **censoring** pada data sensus pada saat pengumpulan data dilakukan.

GENERATING NEW FEATURES

roomsPerHouseholds

Rasio total kamar terhadap rumah tangga

bedroomsPerRoom

Proporsi kamar tidur terhadap total kamar

popPerHouseholds

Rasio total populasi terhadap rumah tangga

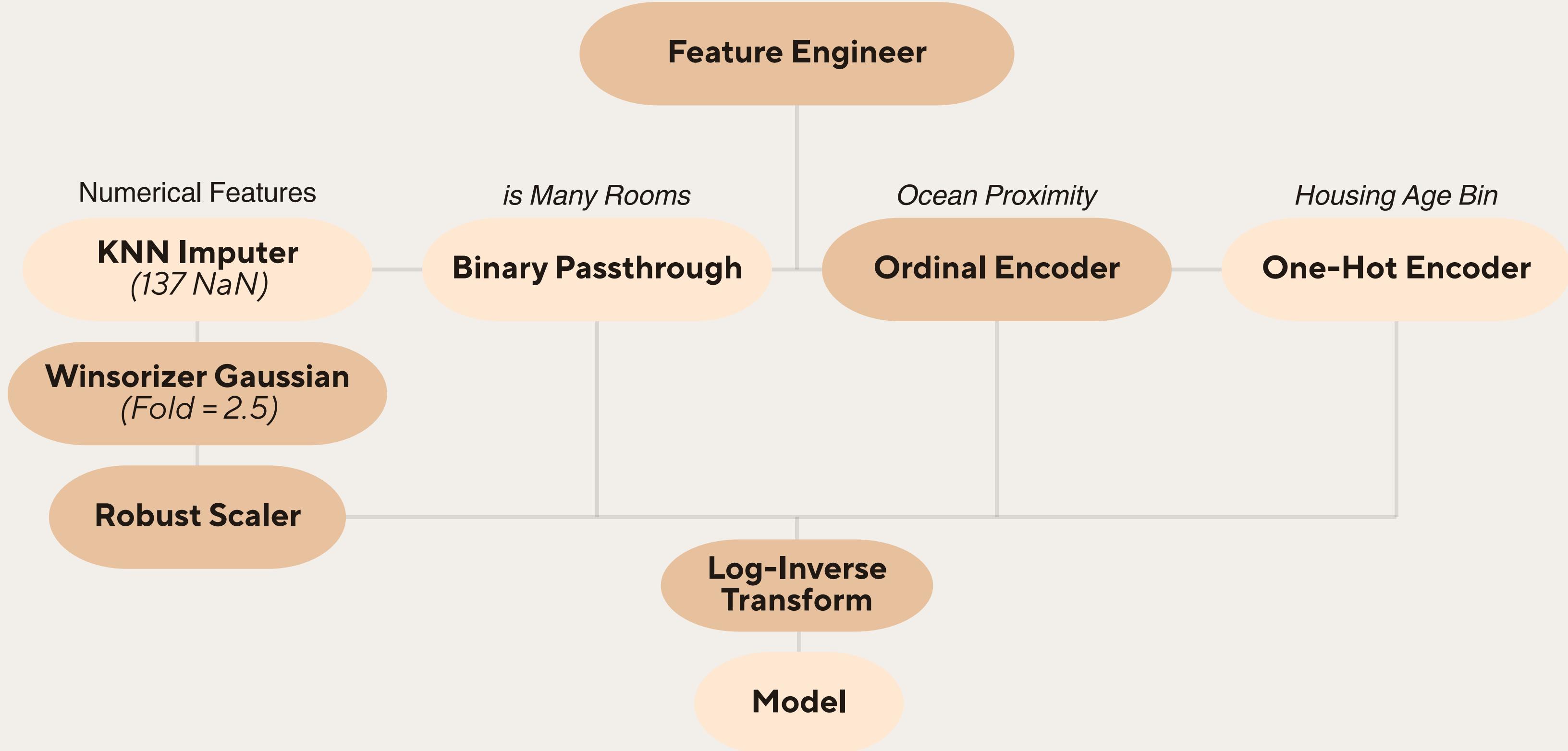
isManyRooms

Fitur biner yang menandai rumah-rumah dengan jumlah kamar sangat banyak (di atas quantile ke-95)

housingAgeBin

Fitur kategorikal yang mengelompokkan usia rumah menjadi kelompok (*young, middle, old*) berdasarkan hasil binning

PIPELINE





CHAPTER 3

MODELING & INTERPRETATION

MODEL SELECTION



Linear Regression 01

Ridge Regression 02

Lasso Regression 03

K-Nearest Neighbors 04

Decision Tree 05

Random Forest 06

XGBoost 07

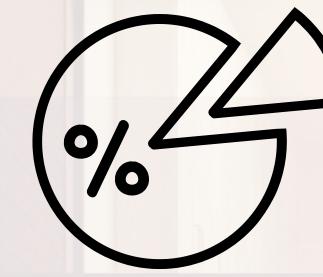
CatBoost 08

BEST MODEL : CATBOOST



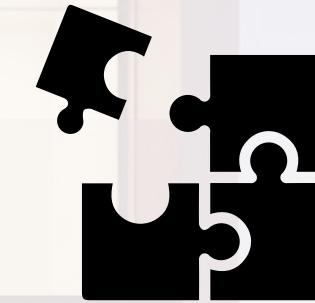
MAE : \$ 27,341

Rata-rata besaran selisih dolar (\$) nilai prediksi dengan nilai aktual.



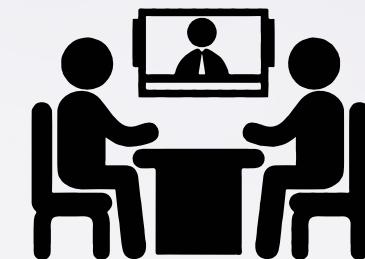
MAPE : 14.90%

Rata-rata persentase selisih nilai prediksi dengan nilai aktual.



RMSLE : 0.2111

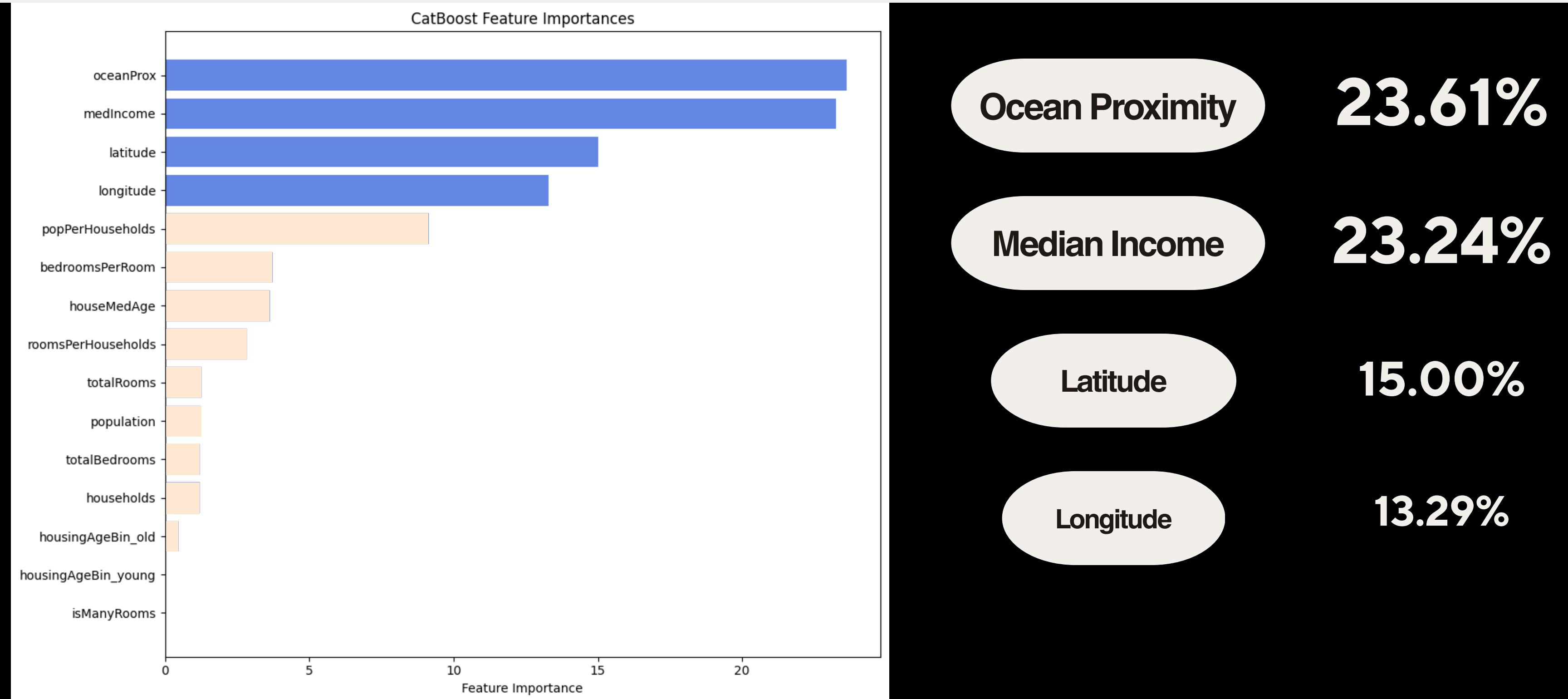
Mengukur selisih error relatif sebagai perbandingan performa setiap model terhadap dataset yang ada



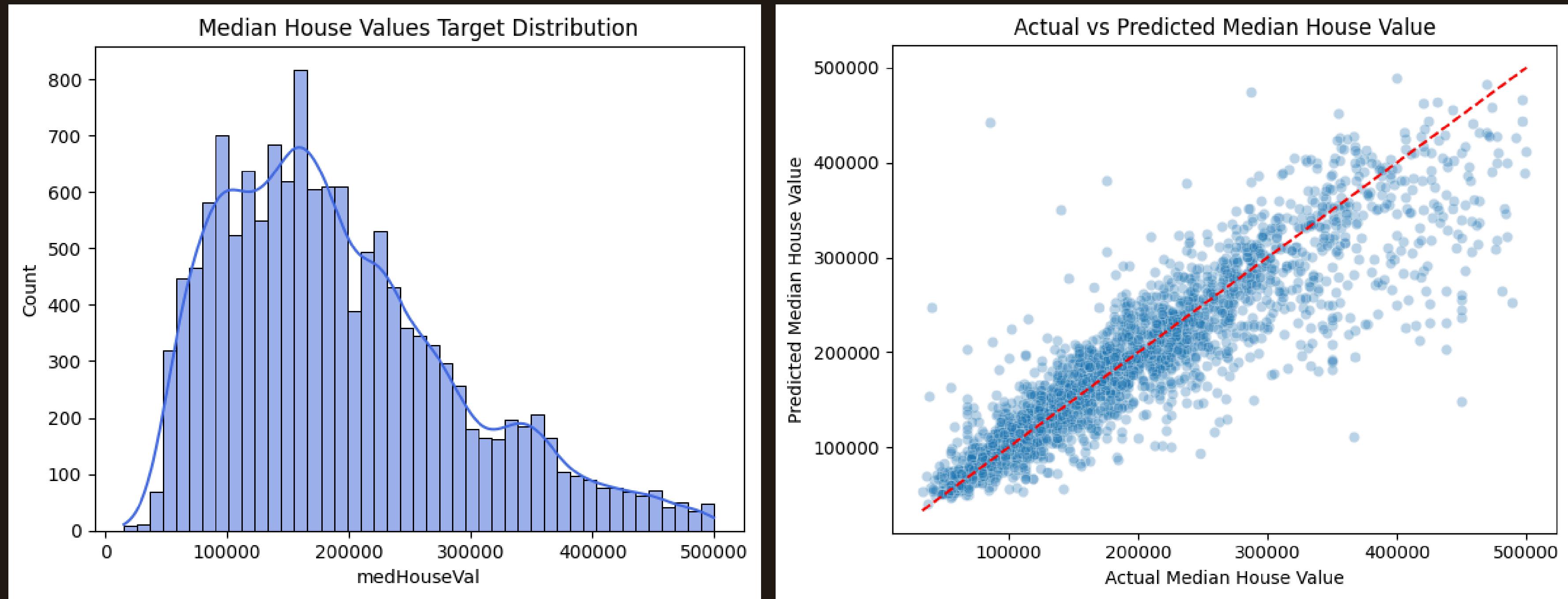
Action to Take

- Acuan rekomendasi harga
- Monitoring harga per wilayah

FEATURE IMPORTANCES



ACTUAL VS PREDICTED VALUES



MODEL LIMITATION



Error yang Meningkat Seiring dengan Harga

Penting untuk menganalisa margin error yang terjadi pada rumah mewah dan menggunakan model sebagai alat pendukung, bukan sebagai penentu harga mutlak.



Keterbatasan pada Kualitas Data

Dibutuhkannya evaluasi untuk hasil prediksi pengembangan baru, tipe rumah mewah, atau wilayah geografis yang tidak tercakup dalam data latih model.



Ketidakhadiran Fitur Geospasial Analisis

Untuk properti yang benar-benar baru, dengan karakteristik unik atau berada di lokasi baru, menghasilkan prediksi dengan akurasi yang rendah.

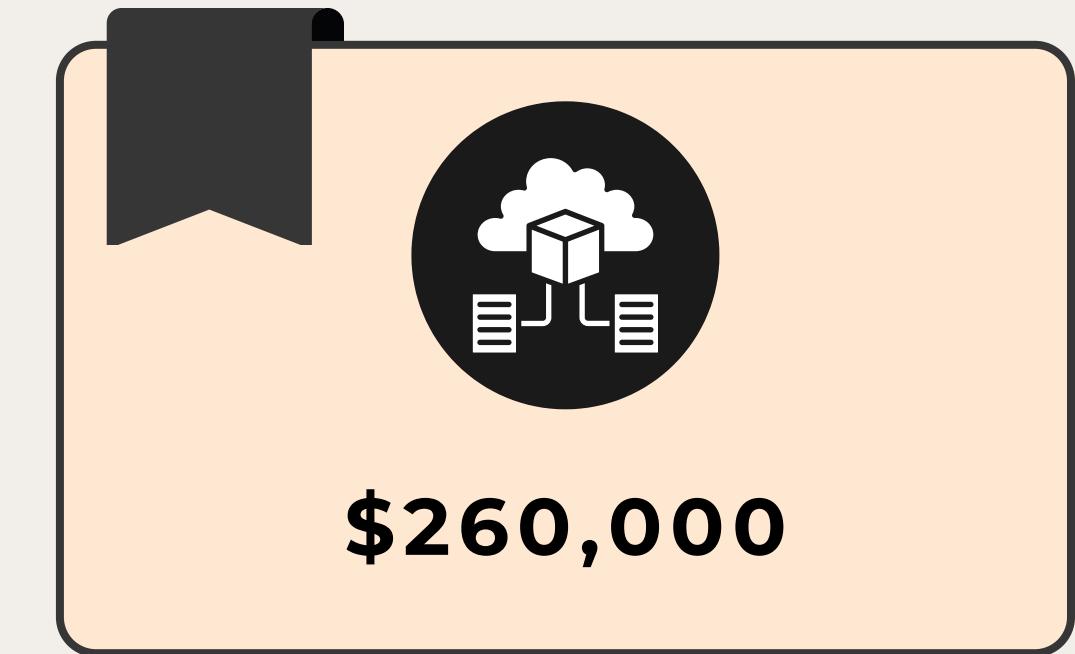
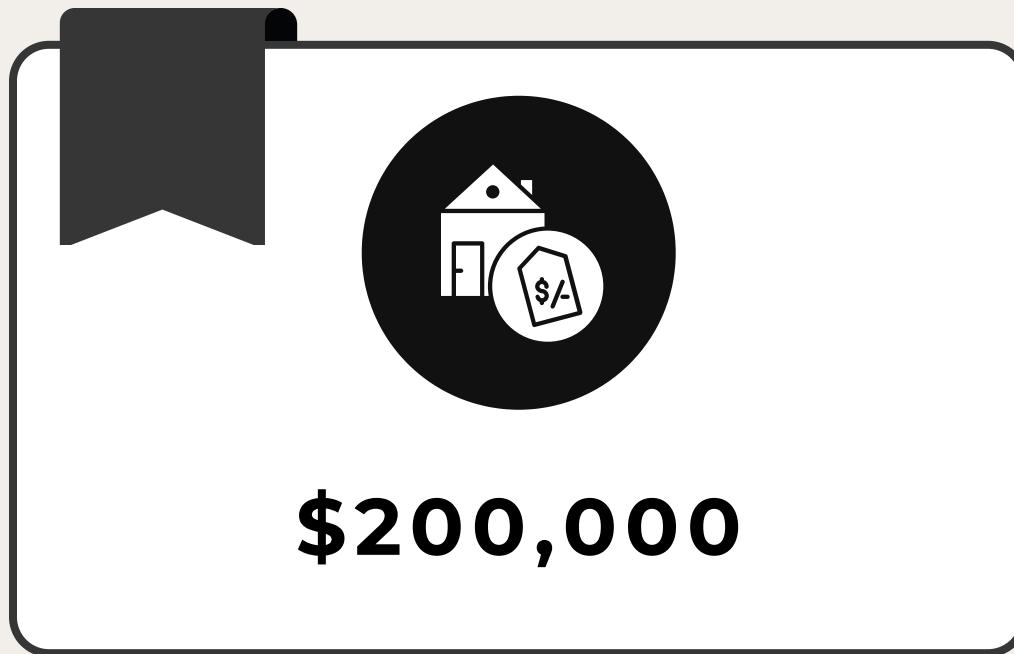




CHAPTER 4

BUSINESS CALCULATIONS

RETURN ON INVESTMENT (ROI)



Total Investment = \$236,000

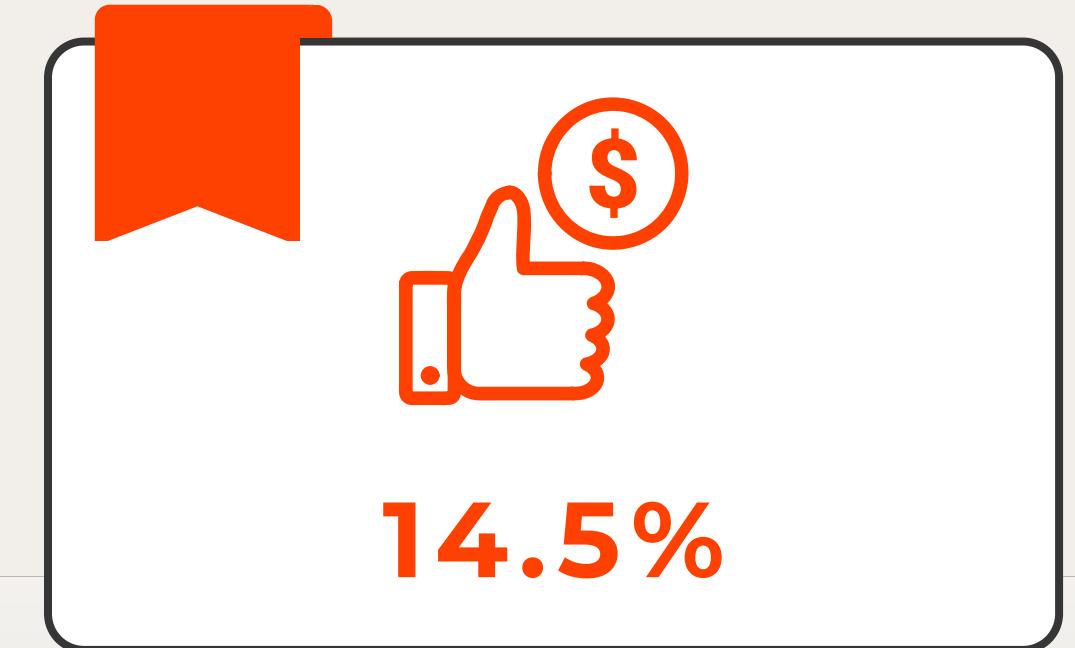
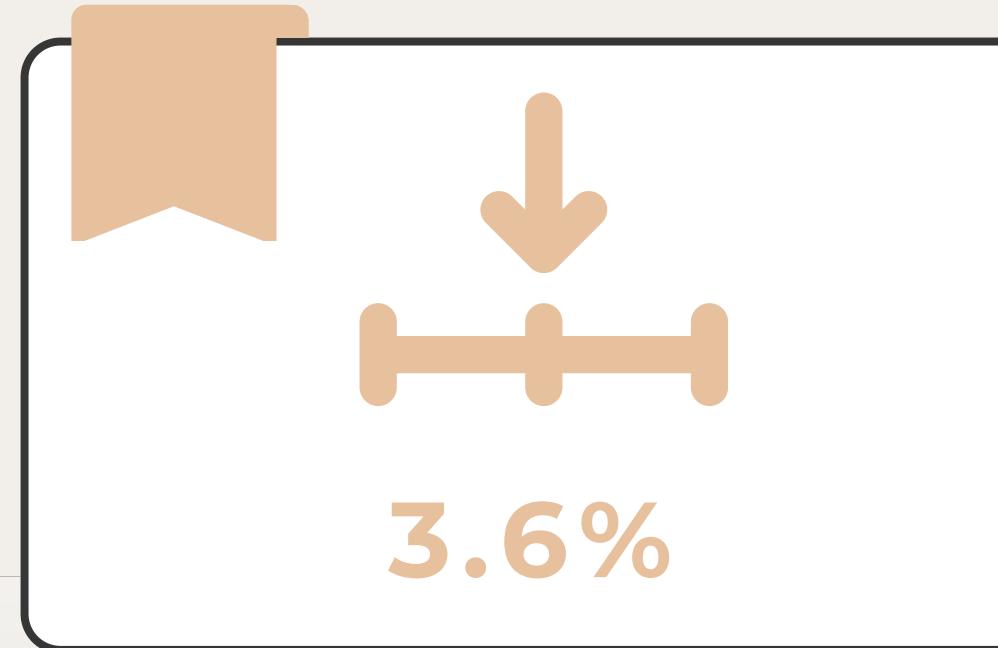
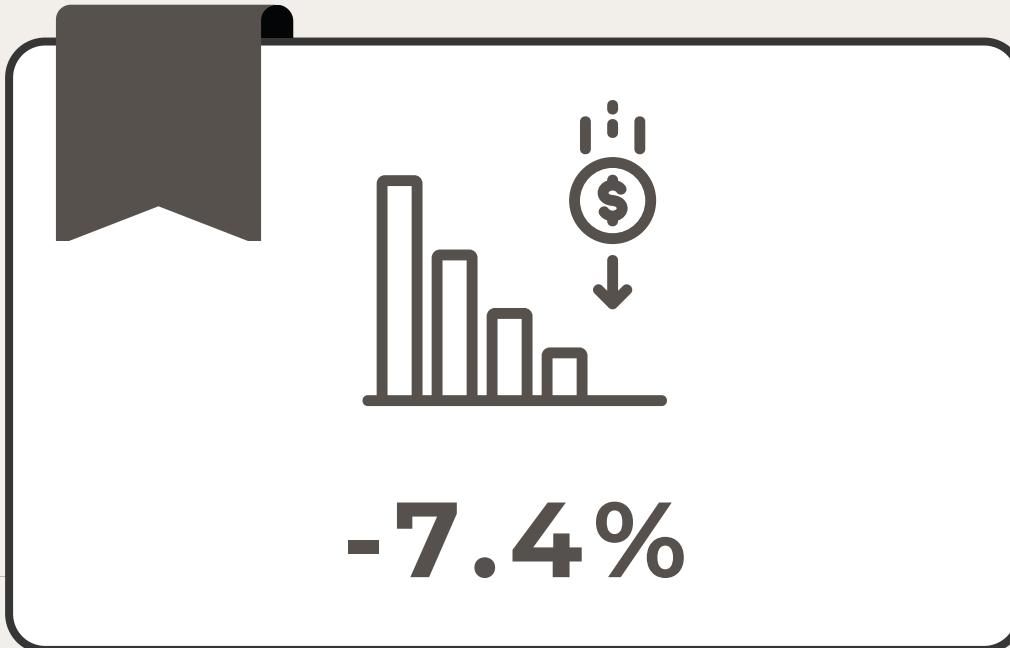
Selling Fee 6%

ROI Target $\geq 10\%$



ROI NET IN 6 MONTHS

AFTER 6% SELLING FEE



Sale Net = \$218,699

Sale Net = \$244,400

Sale Net = \$270,101

- Prediksi \$260k belum tembus target ROI 10%
- *Action to Take: No Deal Below \$276,170*

PRICE NEGOTIATION

MAE $\approx \$27,341 \rightarrow$ *Model Prediction*
\$260,000

Komisi
6% $\times \$27,341$
 $\approx \$1,640$
per transaksi

\$232,659 — \$287,341

PELUANG KONVERSI LEBIH CEPAT

FOKUS MENINGKATKAN MARGIN

- *Listing di sekitar harga prediksi model atau sedikit di atasnya untuk mempercepat penjualan sekaligus menjaga margin.*
- *Memakai MAE sebagai confidence band operasional untuk jangkar negosiasi.*

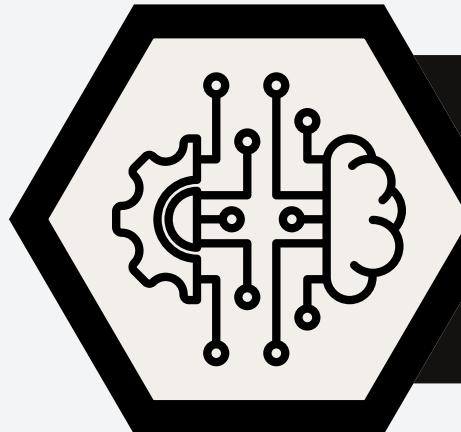


CHAPTER 5

CONCLUSION & RECOMMENDATIONS

CONCLUSION

MACHINE LEARNING PERSPECTIVE



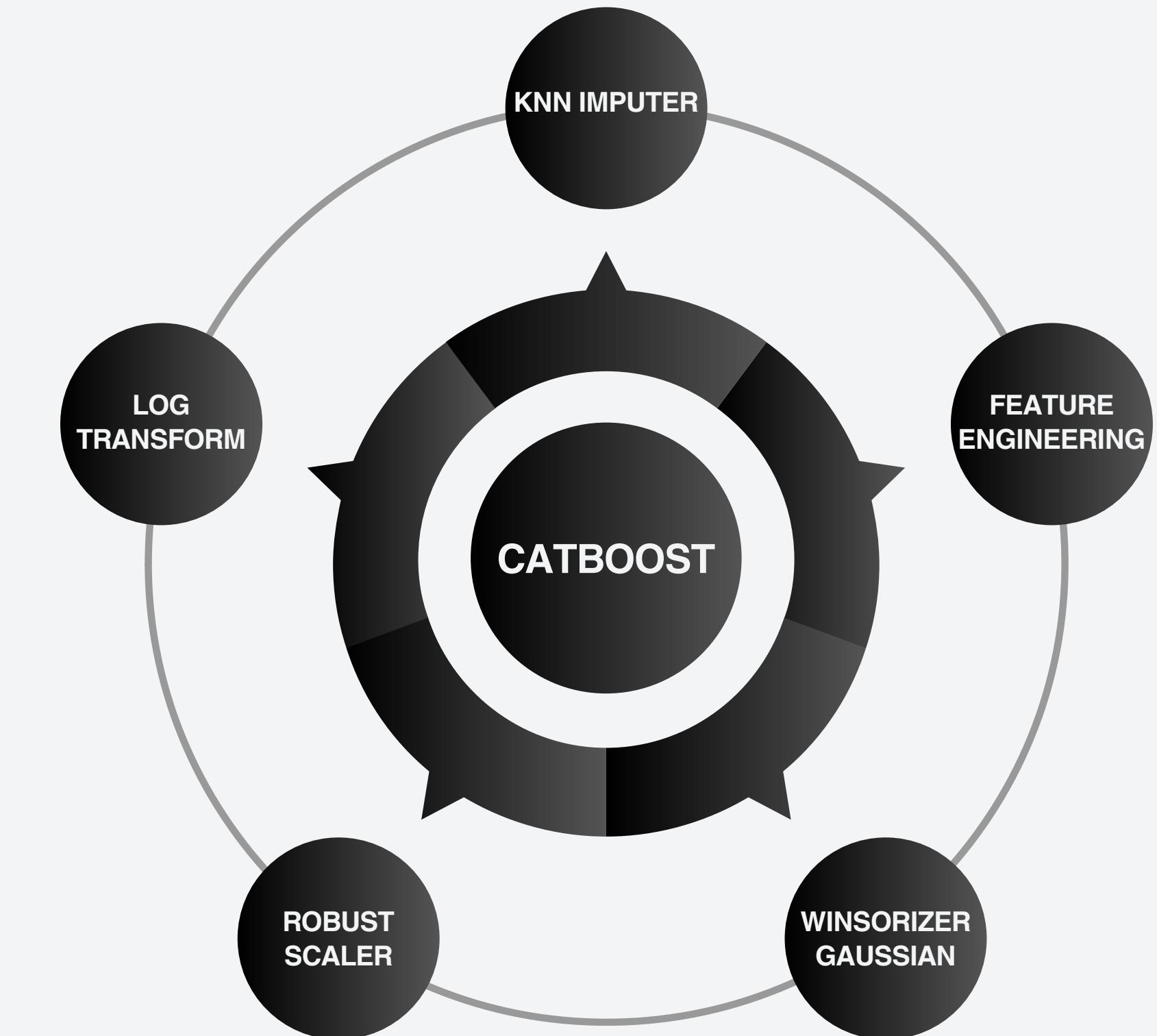
CatBoost secara konsisten memberikan akurasi dan stabilitas tertinggi di seluruh metrik evaluasi, dengan nilai error yang paling rendah.



Performa prediksi model yang didorong oleh fitur-fitur yang berhubungan dengan lokasi geografis dan karakteristik ekonomi lingkungan.



Limitasi berada pada rumah mewah akibat ketidakseimbangan data dan heteroskedastisitas alami dalam harga pasar properti.



CONCLUSION

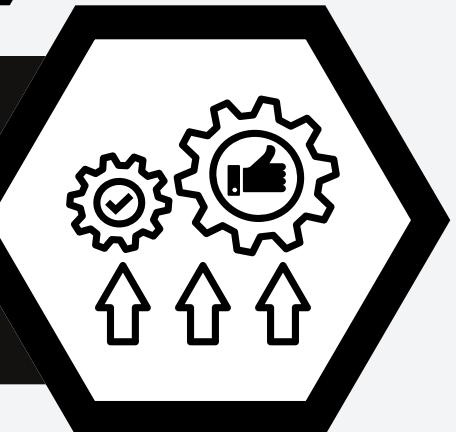
BUSINESS PERSPECTIVE

Model dapat memprediksi properti pada segmen mainstream atau menengah dengan baik.



Untuk properti bernilai tinggi atau unik, hasil prediksi model sebaiknya dianggap sebagai panduan negosiasi harga.

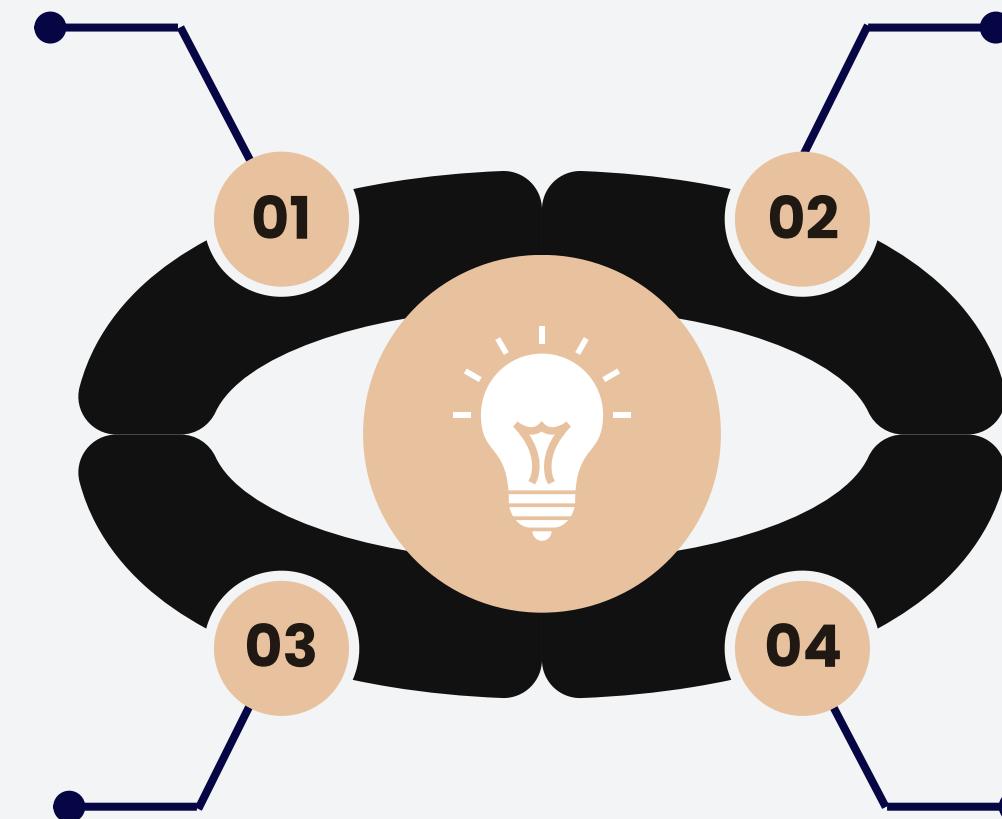
Integrasi model ML sebagai upaya dalam pengambilan keputusan bisnis yang lebih cepat dan objektif dalam pricing, akuisisi, perencanaan renovasi, dan mitigasi resiko



RECOMMENDATIONS

MACHINE LEARNING & DATA SCIENCE TEAM

Monitoring model secara berkala seperti **retraining & revalidation seiring data baru dimasukkan**. Untuk dapat menangani **perubahan pasar** atau **penambahan fitur baru**.



Memperluas cakupan data untuk segmentasi rumah mewah atau kawasan baru agar dapat mengurangi error prediksi pada area kritis ini.

Mengembangkan **feature engineering dengan berfokus pada geospasial** karena model yang ada **menitikberatkan splitting pada fitur-fitur lokasi seperti Ocean Proximity, Latitude, serta Longitude**.

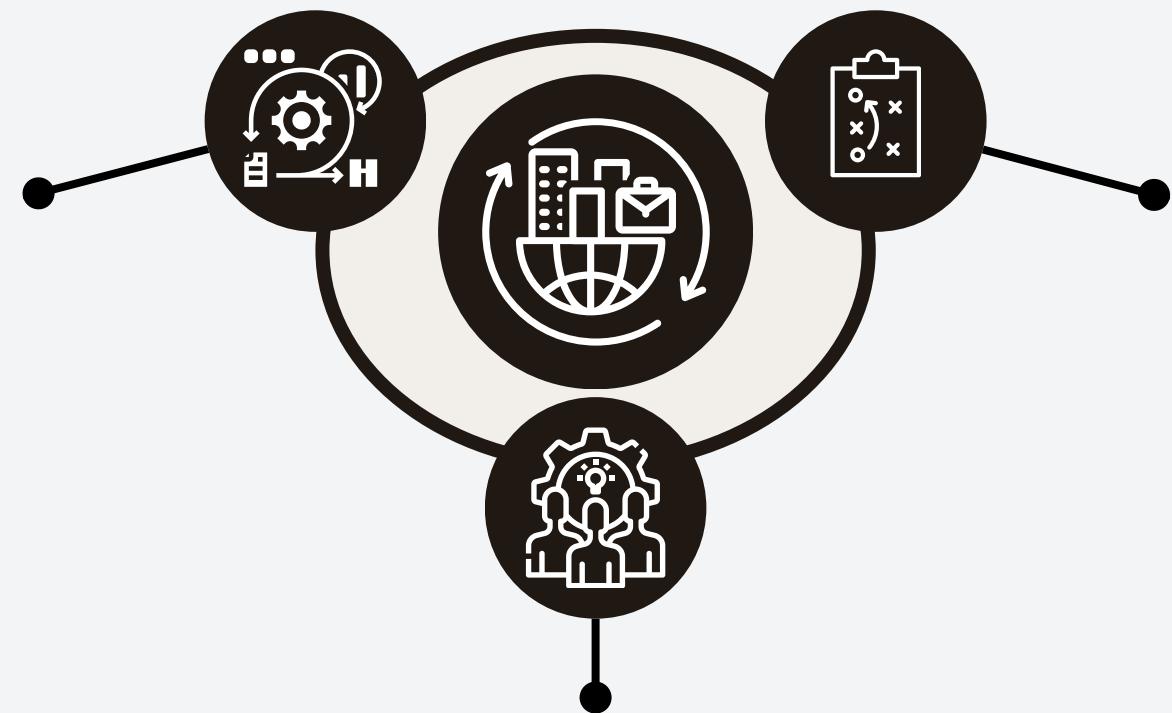
Konsistensi fokus dengan model yang dapat didukung **teknik Explainable AI seperti SHAP Value** untuk **menjaga transparansi** dan **membangun kepercayaan stakeholder**.

RECOMMENDATIONS

STAKEHOLDERS & DECISION MAKERS

Implementasi di Pasar Utama:

Gunakan model CatBoost sebagai mesin utama untuk valuasi properti, rekomendasi harga, dan analisis pasar di segmen mainstream.



Untuk Properti Mewah:

Selalu padukan hasil prediksi model dengan analisis manusia dan riset pasar untuk properti bernilai tinggi atau tidak umum.

Penggunaan yang Transparan:

Pastikan semua keputusan berbasis model tetap adil, transparan, dan sesuai dengan regulasi serta standar etika dalam industri real estate.



THANK YOU

