



FERAS

Foresight-Enhanced Real estate Appraisal with Spatial intelligence

A Case Study on Predicting Price per Square Meter in the
Saudi Arabian Real Estate Market



Submitted in fulfillment of the requirements for the degree Master of Science in AI
AT

Collage of Informatics

Midocean University

UAE

Prepared By

Fahad AbuBaker Bahashwan

Student ID No. 243001702

Supervisor By

Dr. Ashraf Alsaeed

Collage of Informatics, Midocean University, Comoros

September 2025

وما أُوتِيتُمْ من العلم إلا قليلاً

And mankind have not been given of knowledge except a little

سورة الإسراء آية 85

Abstract

This thesis presents a comprehensive machine learning framework for predicting the price per square meter ($\text{ﷲ}/\text{m}^2$) of real estate properties in Saudi Arabia using advanced spatial features and modern algorithms. The work addresses the growing need for accurate, scalable, and transparent automated valuation in the rapidly expanding Saudi real estate market. Using a validated dataset of 48,520 property transactions from major cities—including Riyadh, Jeddah, and the Eastern Province. We develop and evaluate multiple models, including linear regression, Ridge, Lasso, decision trees, k -nearest neighbors (KNN), random forest, gradient boosting and XGBoost.

The methodology incorporates extensive feature engineering: geographic coordinates; distances to city centers (using Haversine formula); proximity to schools; and detailed road-network descriptors (counts and distances to motorway, trunk, primary, secondary, tertiary, and residential roads). These are combined with core property attributes in a leak-safe preprocessing pipeline. The evaluation uses a randomly stratified 80/20 split across all locations, providing a benchmark to compare models.

On the randomly split test set (80% train, 20% test across all locations), Random Forest achieves the best point accuracy (Test $R^2 = 0.9515$, RMSE=456.3 $\text{ﷲ}/\text{m}^2$, MAE=263.2 $\text{ﷲ}/\text{m}^2$, MAPE=9.30%), while XGBoost performs closely ($R^2 = 0.9465$, RMSE=479.4 $\text{ﷲ}/\text{m}^2$, MAE=300.8 $\text{ﷲ}/\text{m}^2$, MAPE=11.84%). Although Random Forest's R^2 is higher, we select XGBoost as the primary (deployed) model because it exhibits less overfitting—a smaller train–test generalization gap ($\Delta R^2 = 0.0170$ for XGBoost vs. 0.0323 for Random Forest)—leading to more stable performance in practice.

These findings demonstrate that spatial intelligence materially improves automated valuation models (AVMs) for the Saudi market, while maintaining clarity through consistent metrics reporting and explainable AI analyses.

Keywords: Artificial Artificial Intelligence, Machine Learning, Property Valuation, Real Estate, Automated Valuation Models, Geographic Information Systems, Ensemble Methods, Spatial Features, Model Explainability, Saudi Arabia

List of Abbreviations

AI	Artificial Intelligence
ML	Machine Learning
GIS	Geographic Information System
OSM	OpenStreetMap
RMSE	Root Mean Squared Error
MAE	Mean Absolute Error
MAPE	Mean Absolute Percentage Error
R^2	Coefficient of Determination
IQR	Interquartile Range
Latitude	Position north/south of the Equator
Longitude	Position east/west of the Prime Meridian
Haversine	Distance calculation method between two points on a sphere
SHAP	Shapley Additive Explanations
XGBoost	Extreme Gradient Boosting
KNN	K-Nearest Neighbors
RFR	Random Forest Regressor
GBR	Gradient Boosting Regressor
DTR	Decision Tree Regressor
LR	Linear Regression
Ridge	Ridge Regression

Lasso	Least Absolute Shrinkage and Selection Operator
sklearn	Scikit-learn (Python ML library)
pandas	Python data manipulation and analysis library
NumPy	Python numerical computation library
Matplotlib	Python plotting library
Seaborn	Python statistical visualization library
geopy	Python library for geocoding and distance calculations
OneHotEncoder	Scikit-learn categorical feature encoder
Train/Test Split	Process of dividing data into training and testing subsets
Boxplot	Statistical data visualization showing quartiles and outliers
Heatmap	Visualization showing correlation or intensity values in a grid
Feature Engineering	creating new variables to improve model performance
Outlier	Data point significantly different from others in the dataset
Visualization	Graphical representation of data
Residuals	Difference between actual and predicted values in regression
Error%	Relative percentage difference between actual and predicted values
Motorway	Major high-speed road, no cross traffic
Trunk	Major route connecting cities
Primary	Main road within or between towns
Secondary	Road linking smaller towns or districts
Tertiary	Local connector between smaller roads
Residential	Street serving housing areas
Hedonic Model	Estimating value based on the contribution of individual features

Acknowledgments

I express my sincere gratitude to all those who have contributed to the completion of this thesis and supported me throughout my master's degree journey.

First, I extend my deepest appreciation to my advisor, Dr. Ashraf Alsaeed, for his invaluable guidance, expertise, and unwavering support throughout this research project. His insightful feedback, constructive criticism, and encouragement have been instrumental in shaping this work and my development as a researcher. His commitment to academic excellence and ethical research practices has been truly inspiring.

I would like to express my sincere gratitude to the faculty and staff of the College of Informatics at Midocean University for fostering an outstanding academic environment and for providing the resources that made this research possible. I am especially grateful to the members of my thesis committee for their time, insightful guidance and constructive feedback, which greatly enhanced the quality and depth of this work.

This research represents not only my individual effort, but also the collective contribution of all those who have supported and guided me along the way.

Fahad AbuBaker Bahashwan

September 2025

Contents

Abstract	ii
List of Abbreviations	iii
Acknowledgments	v
1 Introduction	1
1.1 Background and Motivation	1
1.2 Problem Definition	2
1.3 Thesis Objectives	3
1.4 Thesis Contributions	4
1.5 Thesis Organization	5
2 Literature Review	6
3 Proposed Approach	10
3.1 Research Design and Approach	10
3.2 Dataset Description and Characteristics	13
3.3 Data Preprocessing	16
3.4 Feature Engineering and Geographic Integration	18
3.5 Feature Correlation Analysis	19
3.6 Machine Learning Algorithms and Implementation	21
3.6.1 Task and Target	21
3.6.2 Feature Inputs (final design)	21
3.6.3 Preprocessing Pipeline	22
3.6.4 Training and Evaluation Protocol	24
3.6.5 Explainability and Uncertainty	24
4 Experimental Results and Discussion	25
4.1 Experimental Results	25
4.2 Dataset Characteristics and Exploratory Analysis	25
4.2.1 Model Performance Comparison	25

CONTENTS

- 4.3 Overall Model Analysis 26
 - 4.3.1 Random Forest 26
 - 4.3.2 XGBoost 26
 - 4.3.3 Decision Tree 27
 - 4.3.4 K-Nearest Neighbors 27
 - 4.3.5 Gradient Boosting 27
 - 4.3.6 Linear Models (Ridge, Lasso, OLS) 27
- 4.4 Interpretation of Evaluation Metrics 29
- 4.5 Visual Evaluation of Model Predictions 29
 - 4.5.1 XGBoost 30
 - 4.5.2 Random Forest 30
 - 4.5.3 Decision Tree 31
 - 4.5.4 K-Nearest Neighbors (KNN) 32
 - 4.5.5 Gradient Boosting 33
 - 4.5.6 Linear Models (OLS, Ridge, Lasso) 34
 - 4.5.7 Residual Analysis of XGBoost 35
- 4.6 Fairness Evaluation Across Models 35
- 4.7 Geographic Fairness Evaluation 37
- 4.8 Discussion 39
 - 4.8.1 Interpretation of Results 39
 - 4.8.2 Feature Importance and SHAP Explanations 39
- 4.9 Executive Summary of Findings 43
- 5 Conclusion and Future Work 44**
 - 5.1 Overview 44
 - 5.2 Key Findings 44
 - 5.3 Contributions 45
 - 5.4 Limitations 45
 - 5.5 Recommendations and Future Work 46
 - 5.6 Concluding Remarks 47
 - 5.7 Executive Summary 48

List of Figures

- 3.1 AVM framework for predicting real estate price per square meter. 11
- 3.2 Monthly transactions volume (July 2024 – July 2025). 13
- 3.3 Spatial distribution of transactions across 16 cities. 14
- 3.4 Price (ﷲ/m²) distributions in Riyadh. 15
- 3.5 Price (ﷲ/m²) distributions in Makkah and Jeddah. 15
- 3.6 Compact correlation heatmap (upper triangle only). 20

- 4.1 Test R² comparison across models 27
- 4.2 RMSE (SAR/m²) comparison 28
- 4.3 MAE (SAR/m²) comparison 28
- 4.4 Actual vs Predicted scatter plot for XGBoost. 30
- 4.5 Actual vs Predicted scatter plot for Random Forest. 31
- 4.6 Actual vs Predicted scatter plot for Decision Tree. 32
- 4.7 Actual vs Predicted scatter plot for KNN. 33
- 4.8 Actual vs Predicted scatter plot for Gradient Boosting. 34
- 4.9 Actual vs Predicted scatter plots for linear models. 34
- 4.10 Residual distribution for XGBoost predictions on the test set 35
- 4.11 Global SHAP feature importance for XGBoost. 40
- 4.12 SHAP summary (beeswarm) plot for XGBoost. 41
- 4.13 SHAP waterfall explanation for an individual property. 42

List of Tables

- 3.1 Summary of machine learning models used in this study 12
- 3.2 Transactions by city (July 2024 – July 2025). 14
- 3.3 Transactions by property type. 15
- 3.4 Data dictionary for features - core variables 16
- 3.5 Preprocessing audit: record counts by step. 17
- 3.6 Top positive and negative correlations with price per meter. 20
- 3.7 Feature groups, representative fields, and types with rationale. 22

- 4.1 Comprehensive model performance on the held-out test set. 26
- 4.2 Model performance on the test set with interpretations in real estate terms. 29
- 4.3 Fairness evaluation across all models. 36
- 4.4 Geographic fairness results for the best model (XGBoost). 37

Chapter 1

Introduction

1.1 Background and Motivation

The real estate industry is on the threshold of a technological revolution, driven by the convergence of artificial intelligence, big data analytics, geospatial technologies such as GIS, and the growing demand for efficient, transparent and equitable property valuation systems. Traditional property appraisal methods, although long-established and professionally established, face growing challenges in meeting the requirements for speed, consistency, and scalability in modern markets. These approaches often rely heavily on manual inspections, subjective judgment, and limited comparable data, making them less adaptable to rapidly changing urban environments and large-scale property transactions. In contrast, the emergence of AI-driven valuation systems offers unprecedented opportunities to overcome these limitations by leveraging big data, spatial features, and machine learning algorithms to provide faster, more consistent, and scalable valuations. However, adopting artificial intelligence (AI) in property valuation also raises important issues of fairness, transparency, and regulatory compliance. This creates a dual challenge: to use technological innovation effectively while ensuring that it is applied ethically and equitably.

Property valuation represents one of the most critical functions in real estate markets, shaping mortgage lending decisions, investment strategies, tax assessments, and overall market transparency. The accuracy and reliability of valuations play a pivotal role in shaping financial stability, driving economic development, and promoting social equity. Traditional appraisal methods, typically conducted by licensed professionals through comparative market analysis and standardized procedures, provide essential experience and professional judgment. However, they remain constrained by human limitations in processing vast and heterogeneous datasets, ensuring consistency across different appraisers, and scaling to meet the demands of increasingly dynamic and complex real estate markets. These challenges are particularly pronounced in rapidly devel-

oping regions such as Saudi Arabia, where rapid urban expansion and national initiatives such as Vision 2030 and large-scale housing and infrastructure projects have increased the demand for valuation systems that are scalable, data-driven and able to integrate spatial features for more accurate assessments.

In this context, the Saudi Arabian real estate market presents a particularly compelling case for investigating AI-driven property valuation systems. As Saudi Arabia advances its Vision 2030 economic diversification strategy, the sector has undergone significant growth and transformation, introducing both new challenges and opportunities for valuation practices [1]. The combination of rapid urban expansion, growing foreign investment, and ongoing regulatory reforms has increased the need for more efficient, consistent, and data-driven valuation systems capable of supporting Saudi Arabia’s long-term economic development objectives.

Saudi Arabia’s property valuation system has the same problems seen worldwide, but also has special issues and unique characteristics specific to the region. Historically, Saudi Arabia has relied on conventional appraisal methods that, although grounded in professional expertise, lack the capacity to handle the volume and complexity of modern transactions. The increasing sophistication of investments, the emergence of new property types, and the accelerating pace of market activity highlight the limitations of traditional methods in delivering timely and consistent valuations, raising concerns about their capacity to sustain an efficiently functioning market [2].

1.2 Problem Definition

The current property valuation landscape in Saudi Arabia and comparable emerging markets faces multiple interconnected challenges that reduce efficiency, consistency, and scalability. Traditional appraisal methods, while grounded in professional expertise and market knowledge, are slowed by time-intensive processes that often take days or weeks to complete, reducing their relevance in fast-moving market conditions. Moreover, the subjective nature of human judgment—though valuable in highly complex or non-standard cases—frequently introduces inconsistency and potential bias in routine valuations, underscoring the limitations of traditional approaches in meeting the demands of modern real estate markets.

In addition to these limitations, the growing volume of real estate transactions in Saudi Arabia driven by Vision 2030 initiatives and broader economic diversification efforts has put additional pressure on traditional valuation systems. The limited number of qualified appraisers relative to market demand creates bottlenecks in transaction processing, slowing economic activity, and reducing overall market efficiency. This capacity constraint is particularly acute in rapidly developing regions, where the pace of new construction and investment activity far exceeds the availability of local valuation

expertise.

Moreover, traditional valuation methods struggle to capture the complex relationships between property characteristics and market values in dynamic real estate markets. The growing availability of diverse data sources, including geographic information systems, demographic data, and market analytics, creates opportunities for more sophisticated approaches capable of integrating multiple factors simultaneously. However, the manual processing constraints of conventional methods prevent full utilization of these rich data resources, further limiting their effectiveness in modern valuation practice.

The issue of consistency and potential bias in property valuations constitutes another significant challenge. Appraisers may reach divergent conclusions about the same property, influenced by individual experience, judgment, and interpretations of market conditions. This variability generates uncertainty for market participants and can unintentionally introduce systematic biases, leading to inequitable outcomes across property types, locations, or market segments [3].

Together, these challenges highlight the urgent need for innovative, data-driven valuation approaches capable of providing faster, more consistent, and scalable property assessments.

1.3 Thesis Objectives

This research aims to address the identified challenges by investigating AI-driven property valuation systems within the context of the Saudi Arabian real estate market. The primary objective is to develop and evaluate machine learning approaches capable of delivering accurate, efficient, and fair valuations, while also aligning with the practical requirements necessary for real-world implementation.

The first research objective focuses on establishing the optimal machine learning algorithm for property price prediction in the Saudi Arabian context. Through systematic comparison of multiple algorithms including ensemble methods, linear regression approaches, and tree-based models, our aim is to identify the approach that maximizes accuracy while ensuring explainability and practical applicability. This objective requires a comprehensive evaluation of algorithm performance in multiple metrics, including predictive accuracy, computational efficiency, and generalization capabilities.

The second objective focuses on the integration of geographic and spatial features into property valuation models. Since location is a fundamental determinant of property value, this research seeks to develop methodologies to incorporate spatial relationships, accessibility measures, and geographic characteristics into machine learning frameworks. Specifically, this involves building distance-based features, proximity indicators, and regional attributes that can improve predictive accuracy while ensuring

transparency and practical interpretability.

The third objective focuses on the comprehensive evaluation of fairness and bias in AI-driven valuation systems. Recognizing the critical importance of equitable treatment in real estate, this research seeks to design and apply rigorous methodologies for detecting and measuring bias between property types, geographic regions, and market segments. To achieve this, fairness metrics will be established, bias detection procedures implemented, and validation will be performed to ensure that automated valuation systems deliver equitable results across all relevant categories.

Collectively, these objectives establish a comprehensive framework for developing AI-driven valuation systems that are accurate, transparent, fair, and practically deployable within the Saudi real estate market.

1.4 Thesis Contributions

This research makes several significant contributions to both academic knowledge and practical practice in the field of AI-driven property valuation. Specifically, it advances theoretical understanding, introduces methodological innovations, provides practical implementation frameworks, and delivers contextual insights within the Saudi Arabia real estate market.

Methodological Contributions: This research offers a comprehensive evaluation of machine learning algorithms for property valuation in the Saudi Arabian real estate market, addressing an underexplored area of the literature and providing insights of relevance for emerging markets in the Middle East. By systematically comparing eight algorithms under consistent evaluation criteria, the study establishes methodological benchmarks for future research and offers practical guidance for algorithm selection in comparable market contexts.

Technical Contributions: This research advances the literature on spatial analysis in real estate valuation by developing methodologies to integrate geographic characteristics into machine learning models. Incorporating distance-based measures, proximity indicators, and accessibility metrics demonstrates practical approaches to capture location effects, establishing a foundation for future research on spatial relationships in property pricing.

Fairness and Ethics Contributions: This research reinforcing the importance of ethical considerations in technology adoption. The integration of SHAP values provides transparent feature-level bias detection, addressing documented concerns about data bias and algorithmic transparency in real estate AI systems.

1.5 Thesis Organization

This thesis is organized into six main chapters, each addressing specific aspects of AI-driven property valuation research.

Chapter 1: Introduction provides the introduction, presenting the background, motivation, problem statement, research objectives, contributions, and the overall structure of the thesis.

Chapter 2: Literature Review provides a comprehensive survey of existing research on property valuation, automated valuation models, machine learning applications in real estate, and algorithmic fairness. It establishes the theoretical foundation for this study and identifies the key gaps in current knowledge that the research seeks to address.

Chapter 3: Proposed Approach presents the research design, detailing the dataset characteristics, the pre-processing procedures, the feature engineering approaches, the machine learning algorithms evaluated, and the experimental design. The chapter ensures methodological transparency and provides sufficient detail to enable replication and validation of the results.

Chapter 4: Experimental Results and Discussion This chapter presents and interprets the experimental results of the study. It reports the outcomes of the evaluation, including algorithm performance comparisons, feature importance analysis, fairness assessments, and model validation procedures. The results are examined in detail, and their implications for property valuation practices are thoroughly discussed.

Chapter 5: Conclusion and Future Work summarizes the key findings, outlines research contributions, discusses future research directions, and provides final remarks on the potential of AI-driven property valuation systems.

Chapter 2

Literature Review

The story of property valuation begins with a fundamental challenge that has persisted across centuries: how to accurately and fairly determine the value of real estate in an ever-changing market. This narrative traces the evolution from traditional appraisal methods to sophisticated artificial intelligence systems, revealing both the promise and the risks of technological transformation in one of the most critical economic functions of society. For decades, property valuation was based on three established approaches that formed the foundation of professional practice. Sales comparison, cost and income capitalization approaches provided the framework through which licensed appraisers navigated the complex task of price determination. These approaches became the global standard of valuation practice and remain influential today, although increasingly challenged by the growing scale and complexity of modern real estate markets [4].

These methods, grounded in economic theory and professional judgment, operate on well-established principles such as substitution, which posits that rational buyers will not pay more for a property than the cost of acquiring a comparable substitute [5]. The comparative market analysis (CMA) methodology emerged as a practical application of these principles, carried out by real estate professionals with in-depth knowledge of local conditions [6]. Although grounded in sound reasoning, such approaches depend heavily on localized expertise and limited *comparables*, which constrains scalability and consistency.

Although these traditional approaches served the industry for generations, their limitations became increasingly apparent in modern markets shaped by rapid urbanization and increasingly complex data. The subjectivity inherent in human judgment, while valuable in complex or unique cases, frequently introduced inconsistency and potential bias in routine valuations, creating uncertainty for market participants and, in some cases, inequitable results across property types, locations, and demographic segments [7]. Moreover, traditional methods have had difficulty adequately capturing the complex relationships between property characteristics and market values in rapidly changing markets, particularly as diverse data sources—including geographic information

systems (GIS), demographic indicators, and market analytics—have become widely available and increasingly used in valuation practices [8]. .

A more systematic foundation emerged with Sherwin Rosen’s hedonic pricing model, which conceptualized a property as a bundle of attributes, with buyers valuing individual characteristics such as size, location, and amenities. Observed market prices can therefore be interpreted as the aggregate result of marginal contributions of specific attributes [9]. In its simplest form:

$$P_i = \beta_0 + \sum_{j=1}^k \beta_j X_{ij} + \varepsilon_i,$$

where P_i is the price of the property i , X_{ij} is the attribute j of property i , β_j is the implicit marginal contribution of attribute j , and ε_i is an error term. This formulation became the basis for hedonic regression and ultimately for modern machine learning applications in property valuation.

The integration of GIS with hedonic pricing models represented a significant advance, as researchers demonstrated that location-based factors could be systematically quantified and incorporated into property valuation models [10]. This development addressed a fundamental real-estate principle: location is a primary determinant of value. The ability to capture spatial relationships, accessibility measures, and neighborhood characteristics through GIS enabled more sophisticated analytical approaches that would emerge in subsequent decades [11].

As computational power increased and data became more abundant, the limitations of traditional valuation approaches catalyzed the adoption of advanced analytical methods, such as machine learning, that offer greater scalability, precision, and consistency [12]. Automated Valuation Models (AVMs)—computer-based systems that apply statistical, econometric, or machine-learning techniques to property and market data—ushered in a paradigmatic shift toward systematic, data-driven evaluation. They produce valuations with greater speed, scalability, and consistency than traditional appraisal methods [13]. Research has increasingly demonstrated the ability of AVMs to improve value assessment, support decision-making, and provide market information on development, investment, land administration, and taxation [14]. Studies in different markets found that advanced models like random forests and gradient boosting predict property values more accurately than simple linear models, because they can capture complex relationships between property features and location factors [15].

Among these learners, gradient-boosted trees (notably XGBoost) gained prominence due to strong regularization, scalability, and performance in heterogeneous tabular features [16]. However, accuracy alone is insufficient in high-stakes settings like valuation; interpretability is essential. SHAP (SHapley Additive exPlanations) offers a

unified game-theoretic framework that attributes the positive or negative contribution of each feature to an individual prediction. The sum of local feature attributions is equal to the output of the model, making boosted-tree AVMs more transparent and auditable [17]. Beyond interpretability, the evidence from the Ukraine case study also demonstrates the performance advantages of boosted ensembles in real estate pricing tasks, with gradient boosting achieving $R^2 = 0.724$) [18].

Parallel to these algorithmic advances, the centrality of spatial features has been repeatedly confirmed. Studies show that location-based factors account for a substantial share of price variation. For example, in Seoul, studies have shown that proximity to subway stations significantly influences apartment prices, confirming accessibility as a critical determinant of housing value [19]. In Saudi Arabia, the incorporation of geospatial characteristics, such as distance to city centers and road networks, significantly improves predictive precision [20]. These findings validate the principle that “location is value” and support systematic GIS integration in AVMs.

Recent studies also tackle valuation when there are few recorded sales. Geography-guided self-supervised methods (e.g., DoRA) first learn from unlabeled data and then improve accuracy even with small labeled samples—useful for areas or property types with limited transactions [21].

The GCC context provides practical momentum¹, where sustained economic diversification and accelerated digital transformation have created a conducive environment for the adoption of AI-driven valuation systems. In the UAE, the Dubai Land Department’s “Instant Sale” feature operationalizes AI-based smart valuation, enabling property transactions to be completed in mere minutes and demonstrating that end-to-end digital valuation is viable at scale [22]. For Saudi Arabia, the national data-and-AI strategy under Vision 2030 positions AI as an explicit enabler in government and industry, while sectoral digital platforms and open transaction data provide the practical foundation for modernizing property valuation, including REGA’s digital real estate indicators and the Ministry of Justice’s real estate market data (with price per square meter) - and professional governance by TAQEEM sets standards for the valuation profession [23, 24, 25, 26]. Empirical work with Saudi transactions already shows strong machine learning performance relative to traditional baselines, indicating readiness for production-grade systems [20]. Local PropTech platforms (e.g., Wasalt) report large-scale usage and AI-powered capabilities (instant valuation, personalized descriptions, ROI prediction), reflecting market appetite and feasibility [27, 28].

As adoption scales, fairness becomes central. Recent public-sector research documents that AVMs can exhibit larger errors in majority black communities - even when protected attributes are not explicit - implying that overall predictive accuracy does

¹GCC refers to the Gulf Cooperation Council, a political and economic union comprising Saudi Arabia, the United Arab Emirates, Qatar, Kuwait, Oman, and Bahrain.

Chapter 2. Literature Review

not guarantee equitable outcomes [29]. Regulators have responded: In 2024, six federal agencies in the USA issued a final rule mandating AVM quality control standards, including testing, documentation, and controls to guard against bias, formalizing governance expectations for automated home valuations [30].

Putting it all together, the literature indicates that effective AI-driven valuation combines:

- Strong tabular learners (e.g., gradient boosting).
- Standardized spatial feature engineering with GIS.
- Interpretable outputs (e.g., SHAP),
- Fairness evaluation aligned with emerging regulations.

What is still missing, especially for Saudi Arabia, is a unified and standardized spatial pipeline with localized fairness and uncertainty diagnostics integrated into one practical framework.

This thesis addresses this gap by proposing *FERAS*, derived from “firāsa” meaning foresight, a framework that leverages spatial features and explainability while incorporating fairness and uncertainty assessments tailored to the Saudi real estate market.

Chapter 3

Proposed Approach

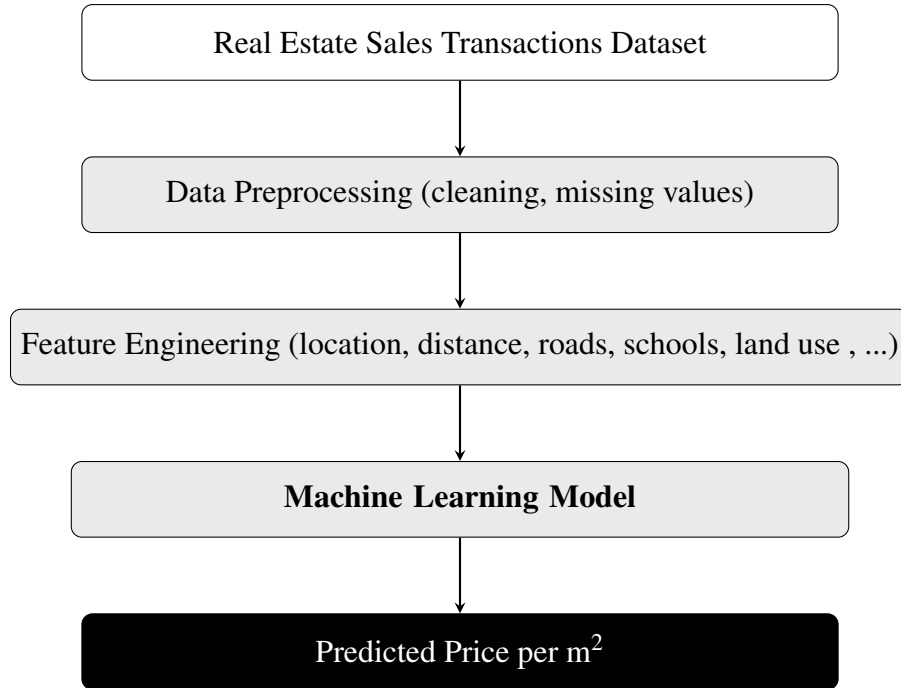
3.1 Research Design and Approach

This study employs a quantitative experimental design to evaluate AI-driven automated valuation models (AVMs) for predicting the price per square meter ($\text{ﷲ}/\text{m}^2$) in the Saudi real estate market. The methodological pipeline is designed to be clear and consistent, and consists of several key stages: data acquisition from transaction records, pre-processing to clean and normalize the data, and spatial feature engineering to enrich each transaction with geographic context such as accessibility and land use. After feature construction, multiple machine learning models are trained and evaluated using a leakage-aware validation strategy, which avoids overly optimistic results by ensuring that training and testing sets are strictly separated.

All models are trained on identical data splits and the same feature pipeline, with fixed random seeds applied throughout, so that performance differences arise solely from the learning algorithms themselves rather than from inconsistencies in data handling. This systematic approach enables a controlled comparison of diverse models while providing insight into both predictive accuracy and their ability to generalize across different types of properties and spatial regions.

The overall research workflow is summarized in Figure 3.1, which illustrates the conceptual AVM pipeline applied in this study. It begins with the collection of property transaction data, followed by preprocessing and feature engineering. The spatially enriched features are then passed into multiple machine learning models for training and evaluation, leading to final predictions of the price per square meter.

Figure 3.1: AVM framework for predicting real estate price per square meter.



To benchmark predictive performance, we employ eight machine learning algorithms from three major families: linear regression methods (Ordinary Least Squares (OLS), Ridge, and Lasso), tree-based learners (Decision Tree, Random Forest (RF), Gradient Boosting (GB) and XGBoost), and an instance-based method (K-Nearest Neighbors, KNN). All models are trained on the same feature inputs and evaluated under consistent procedures to ensure comparability. Together, these models span from simple and interpretable baselines to advanced ensemble learners, enabling a balanced evaluation of predictive accuracy and generalization. A summary of the models and their rationale for inclusion is provided in Table 3.1.

Chapter 3. Proposed Approach

Table 3.1: Summary of machine learning models used in this study

Family	Model	Description / Rationale
Linear	Ordinary Least Squares (OLS)	Baseline linear regression minimizing squared error; provides interpretability and benchmark performance.
	Ridge Regression	Linear regression with L2 regularization to handle multicollinearity and reduce overfitting.
	Lasso Regression	Linear regression with L1 regularization, performing both shrinkage and feature selection.
Tree-based	Decision Tree	Rule-based non-linear learner that partitions feature space hierarchically.
	Random Forest (RF)	Ensemble of decision trees using bootstrap aggregation (bagging) to reduce variance and improve stability.
	Gradient Boosting (GB)	Sequential ensemble method that builds trees iteratively, correcting errors from previous ones.
	XGBoost	Optimized gradient boosting implementation with regularization, widely adopted for structured/tabular data.
Instance-based	K-Nearest Neighbors (KNN)	Non-parametric approach that predicts based on similarity to nearby observations in feature space.

Because deployment in regulated settings requires not only accuracy but also equity and transparency, our evaluation framework extends beyond predictive performance. First, we assess the interpretability of the model by computing SHAP values for each prediction and aggregating them to derive the importance of global characteristics [17]. Second, we evaluate uncertainty by generating 90% prediction intervals using distribution-free conformal prediction, and we report empirical audit coverage to verify their reliability [31].

3.2 Dataset Description and Characteristics

This study uses a large-scale multicity dataset of Saudi Arabian property transactions suitable for machine learning analysis in a regional real estate context. The data set contains 48,520 verified transactions drawn from multiple sources in 16 major Saudi cities, providing extensive geographic diversity for modeling and evaluation¹ The prediction target is the price per square meter ($\text{ﷲ}/\text{m}^2$), calculated as $y = \frac{\text{transaction_price}}{\text{area}}$ after standardizing units and filtering implausible or missing areas.

Temporal coverage. The dataset spans the period from July 2024 to July 2025, providing one year of recorded sales transactions. Although transaction dates are not used as predictive features in our models, reporting the temporal coverage highlights that the dataset reflects a range of market conditions rather than a snapshot from a single month. The monthly transaction volume over this period is shown in Figure 3.2.

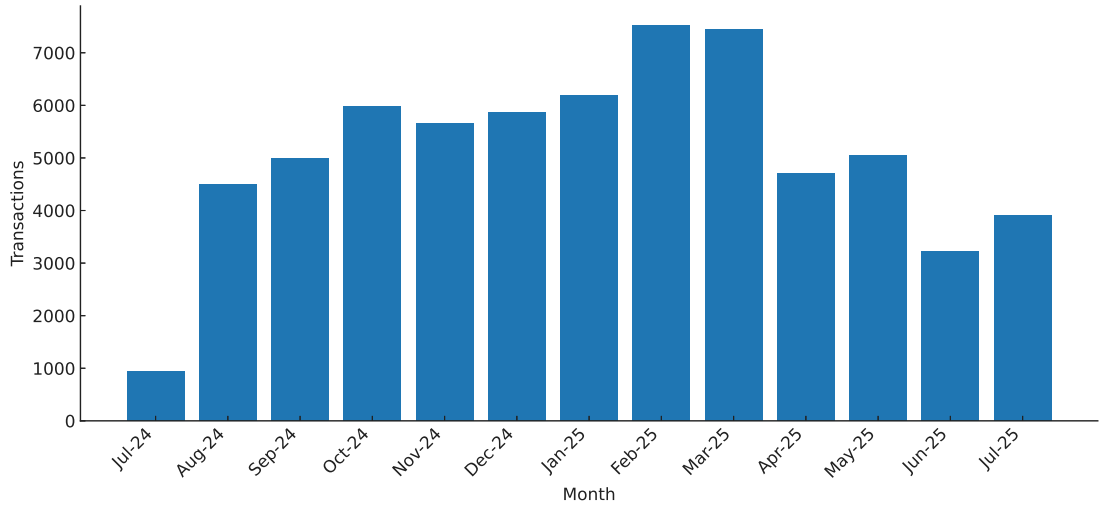


Figure 3.2: Monthly transactions volume (July 2024 – July 2025).

Geographic coverage. Transactions are distributed across key urban centers: Riyadh (61.54%), Jeddah (10.38%), Makkah (5.48%), Dammam (6.03%), Buridah (3.73%), Khobar (3.77%), with the remaining 9.07% in eleven additional cities. This distribution reflects major population and economic hubs while preserving the cross-city variability needed for spatial generalization. The counts and shares at the city level are reported in Table 3.2, and their spatial distribution is visualized in Figure 3.3.

¹Where applicable, data sources include official transaction feeds [24, 25].

Table 3.2: Transactions by city (July 2024 – July 2025).

City	Count	Share (%)
Riyadh	29,857	61.54
Jeddah	5,038	10.38
Makkah	2,661	5.48
Dammam	2,926	6.03
Buridah	1,805	4.3
Khobar	1,829	3.72
Other (11 cities)	6,043	9.07
Total	48,520	100

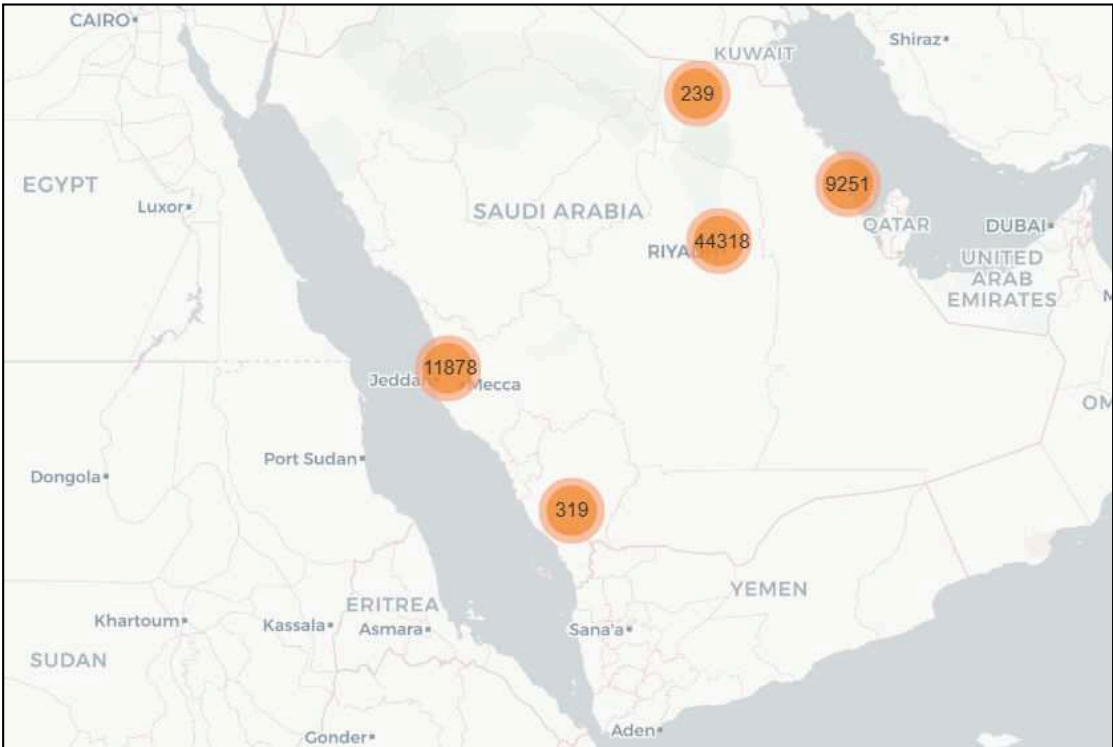


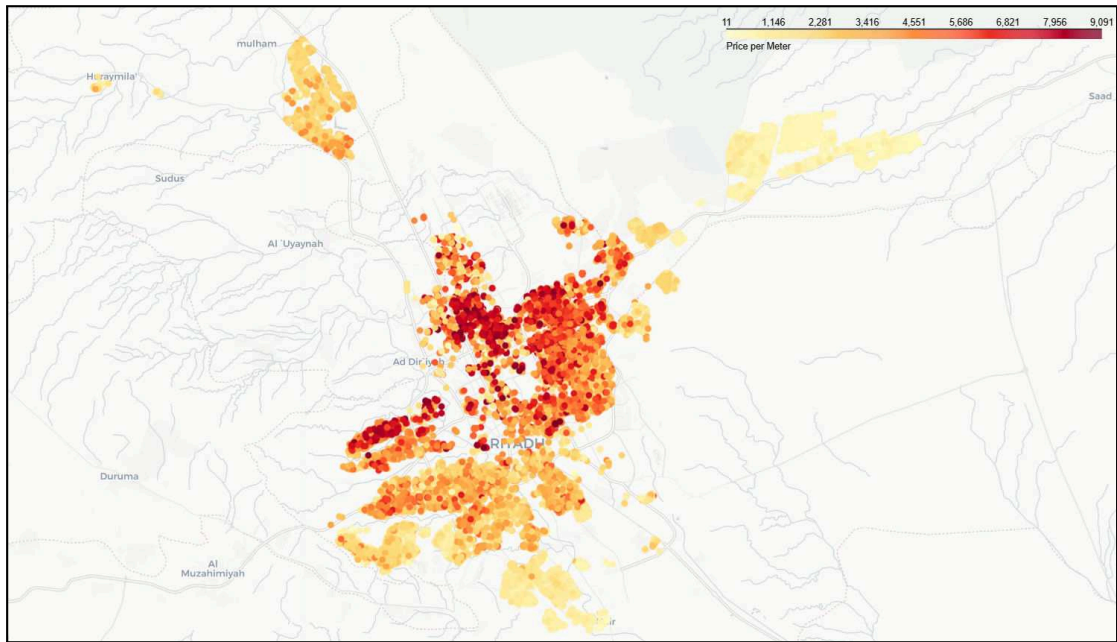
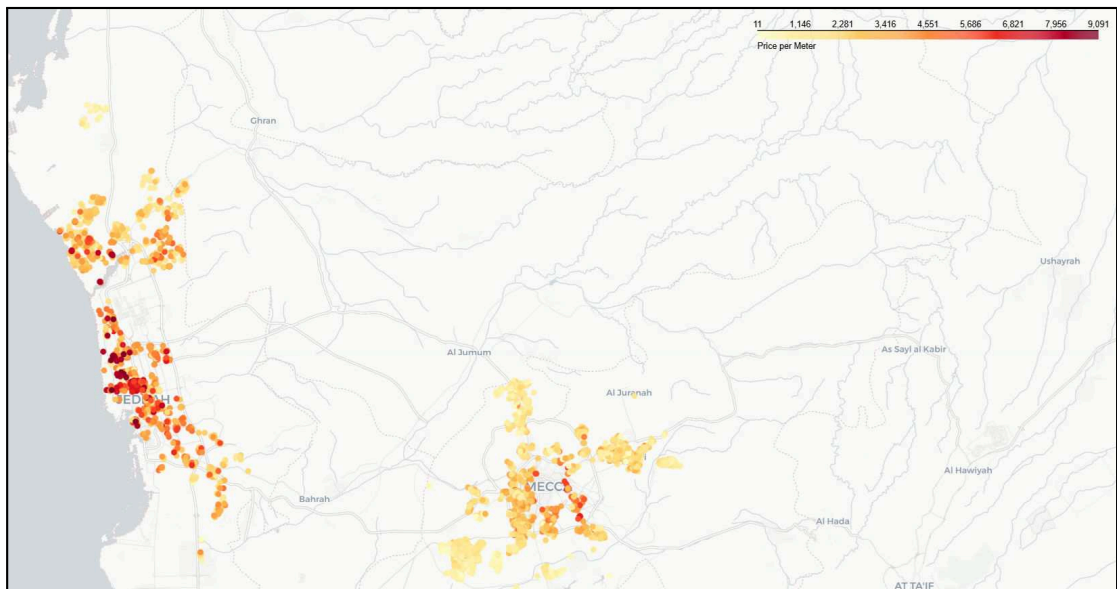
Figure 3.3: Spatial distribution of transactions across 16 cities.

Property types. Five main categories are represented: apartments (25.52%), villas (7.12%), residential land (40.42%), commercial properties (12.59%), and other types (12.75%). This mixture supports model evaluation across distinct market segments and built forms. The distribution by property type is summarized in Table 3.3, and the price per m² variability in Riyadh, Makkah and Jeddah are shown in Figures 3.4, 3.5.

Chapter 3. Proposed Approach

Table 3.3: Transactions by property type.

Property type	Count	Share (%)
Apartments	12,694	26.16
Villas	4,701	7.51
Residential land	20,752	42.77
Commercial	5,558	11.46
Other	5,870	12.09
Total	48,520	100

Figure 3.4: Price (ﷲ/m²) distributions in Riyadh.Figure 3.5: Price (ﷲ/m²) distributions in Makkah and Jeddah.

Chapter 3. Proposed Approach

Attributes and units. Each transaction includes core structural attributes (e.g., built area, parcel area as applicable, precise geographic coordinates (WGS84), and the realized transaction price in Saudi Riyals (ﷲ). All prices reflect closed transactions, not listings, improving the reliability of the label for supervised learning. The core variables are documented in Table 3.4.

Table 3.4: Data dictionary for features - core variables

Variable	Type/Unit	Description
sale_date	ISO date	Date of transaction
sale_price	ﷲ	Final transaction price (closed deal)
area	m^2	Built or parcel area
price_per_m2	ﷲ/ m^2	Target: sale_price/area
Latitude	degrees (WGS84)	Geographic coordinate (north–south)
Longitude	degrees (WGS84)	Geographic coordinate (east–west).
property_type	categorical	Apartment, villa, residential land, other

Data sources. Aggregated from official transaction feeds [25, 24].

3.3 Data Preprocessing

We designed a structured preprocessing pipeline to prepare the dataset before modeling. The pipeline ensures data integrity by handling missing values and outliers, maintains unit consistency by standardizing measurement scales (e.g., converting all prices to (ﷲ/ m^2), and prevents data leakage by applying transformations such as scaling and encoding within cross-validation folds rather than on the entire dataset. This structured setup guarantees that all models are trained with clean, consistent, and leak-safe inputs. The pipeline covers several key stages, including schema standardization, record quality control, target construction, handling missing values and outliers, and geospatial joins. To support transparency, we also maintain an audit log of record counts before and after each step (Table 3.5).

Step 1: Schema standardization. To ensure consistent formatting across the dataset, we standardized column names by trimming whitespace, converting all letters to lowercase, and replacing spaces with underscores. The `transactiondate` field was converted to a proper `datetime` type and simplified to retain only the date (YYYY-MM-DD). In addition, categorical attributes were normalized: `provinceName` and `metricsType` were mapped from Arabic labels to their English equivalents for consistency. Finally, redundant or irrelevant columns that were not required for modeling were removed from the dataset.

Chapter 3. Proposed Approach

Step 2: Record quality control. We applied light cleaning to retain *genuine pricing patterns* while removing noise: we excluded suspect low-value transactions², removed records with missing essentials, consolidated repeated observations that refer to the same place and date by taking the *median* price-per-meter, and filtered out non-target property categories. The resulting dataset is clean and aligned with the modeling target (price per meter).

Step 3: Outlier treatment We identify and remove extreme values in the target $y = \text{priceofmeter}$ using the interquartile range (IQR) rule, also known as Tukey’s fences [32, 33]. Let Q_1 and Q_3 denote the first and third quartiles of y , and define $\text{IQR} = Q_3 - Q_1$. We compute Q_1 , Q_3 , and IQR on the *entire dataset* and retain observations only if

$$Q_1 - 1.5\text{IQR} \leq y_i \leq Q_3 + 1.5\text{IQR}.$$

This robust, rank-based criterion reduces the influence of heavy-tailed transactions without assuming normality. In a sensitivity check, increasing the fence factor to $k = 3$ (“far” outliers) produced similar qualitative conclusions.

Step 4: Auditability. We log the count of rows after each step and summarize the missingness of the core variables (Tables 3.5).

Table 3.5: Preprocessing audit: record counts by step.

Step	Action	Rows remaining
0	Load data	48,520
1	Filter to target property types	37,092
2	Remove low-value transactions	37,087
3	Outlier treatment (IQR / Tukey’s fences)	34,670
Final	Cleaned dataset for feature engineering	34,670

²As recorded in the dataset.

3.4 Feature Engineering and Geographic Integration

We enrich each transaction with geospatial signals of accessibility, built-form intensity, and amenity proximity. Unless noted, indicators are computed within a 1 km radius around the transaction point (WGS84; distances in meters).

Road-network accessibility (OpenStreetMap)

Using OpenStreetMap (OSM) ways and their `highway` classes [34], we derive:

- **Total road supply (1 km):** sum of lengths of all road segments intersecting the 1 km buffer.
- **Class-specific road counts:** number of distinct segments of each functional class (motorway, trunk, primary, secondary, tertiary, residential) within the buffer.
- **Road-class diversity:** count of distinct OSM road classes present in the buffer (a simple richness/diversity proxy).
- **Proximity to key road classes:** great-circle distance from the transaction point to the nearest segment of each class (motorway, trunk, primary, secondary, tertiary, residential).

Built environment and land use (Overture + Microsoft footprints)

From building polygons (Microsoft Global Building Footprints) and land use layers (Overture Maps) [35, 36]:

- **Building count (1 km):** number of building footprints within the buffer.
- **Building density (per km²):** building count normalized by the buffer area.
- **Average building footprint area:** mean area of building polygons (m²) in the buffer.
- **Built-up ratio:** share of the buffer's surface covered by building polygons.
- **Land-use diversity:** number of distinct land-use categories present within the buffer.

Centrality and amenities (Google coordinates + Haversine)

To capture household amenity preferences and urban centrality:

- **Distance to city center:** great-circle distance from the transaction point to a fixed city-center reference coordinate, computed by the Haversine formula [37].
- **Schools within 1 km:** count of nearby education facilities, identified via the `school` coordinates, within a 1 km radius.

The Haversine distance between points (φ_1, λ_1) and (φ_2, λ_2) (radians) on a sphere of radius R is

$$d = 2R \arcsin \left(\sqrt{\sin^2 \left(\frac{\varphi_2 - \varphi_1}{2} \right) + \cos \varphi_1 \cos \varphi_2 \sin^2 \left(\frac{\lambda_2 - \lambda_1}{2} \right)} \right), \quad R = 6,371,000 \text{ m.} \quad (3.1)$$

Why these features matter. Greater road supply and proximity to higher capacity classes proxy for accessibility and market reach; higher building density and built-up share indicate agglomeration, scarcity, and development intensity; centrality and school access capture amenity valuation by households. Together, these variables provide complementary signals that help explain variation in price per square meter.

3.5 Feature Correlation Analysis

To examine linear dependencies among explanatory variables, we computed Pearson correlation coefficients with the target variable (`priceofmeter`). The results reveal both positive and negative associations, which together highlight how built-environment intensity and spatial accessibility jointly shape land values.

Table 3.6 summarizes the five strongest positive and negative correlations with the target variable. Features such as `built_up_ratio` ($r = 0.61$), `road_type_diversity_1km` ($r = 0.57$), and `building_density_km` ($r = 0.49$) are positively correlated with land prices, indicating that denser and more diverse urban environments tend to be valued higher. Conversely, `dist_to_city_center_m` ($r = -0.62$) and other distance-based measures show strong negative correlations, demonstrating the discount effect of being located further from major centers of activity.

Chapter 3. Proposed Approach

Table 3.6: Top positive and negative correlations with price per meter.

Feature	Correlation	Direction
built_up_ratio	0.611	Positive
road_type_diversity_1km	0.568	Positive
total_road_length_1km	0.493	Positive
building_density_km	0.490	Positive
building_count	0.490	Positive
dist_to_city_center_m	-0.620	Negative
dist_to_primary	-0.420	Negative
dist_to_secondary	-0.385	Negative
dist_to_tertiary	-0.372	Negative
dist_to_residential	-0.301	Negative

Note: Negative correlations between price per meter and distance to major roads are consistent with accessibility theory: properties closer to primary, secondary, and tertiary routes are more valuable, while prices decrease with increasing distance. Distance to residential roads also shows a weaker negative correlation, which may reflect Saudi neighborhood patterns, where premium locations are situated near arterial roads and commercial corridors rather than internal streets.

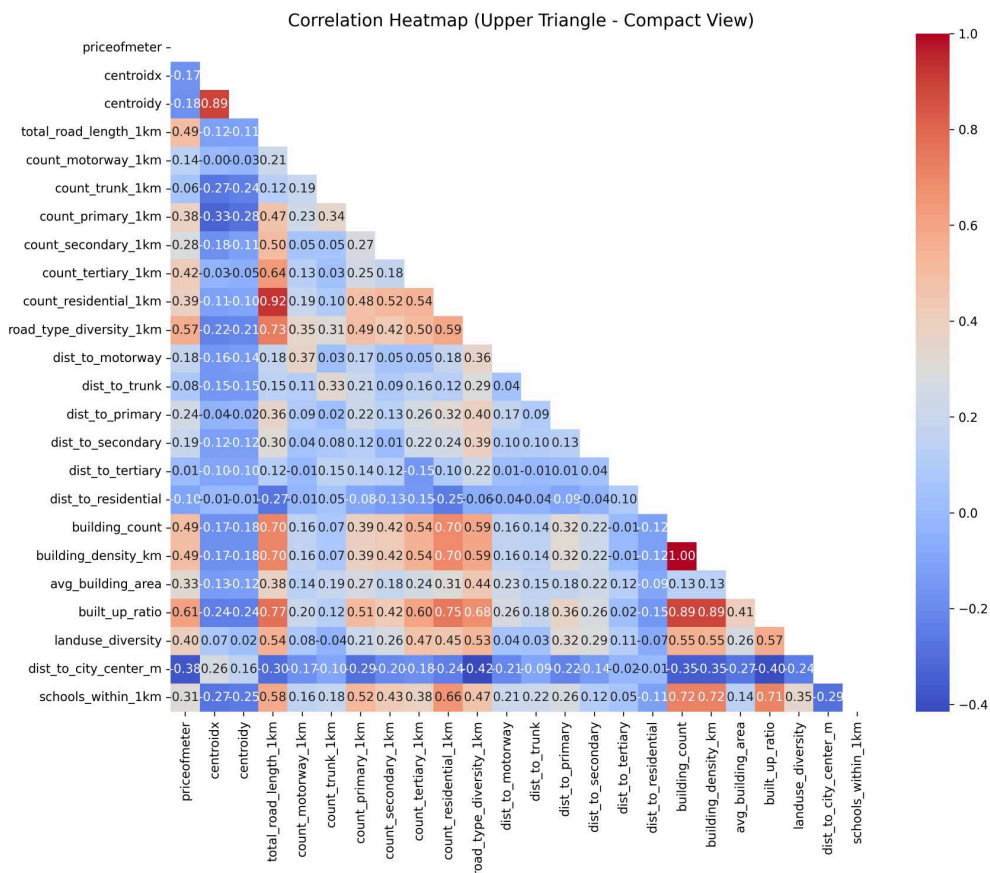


Figure 3.6: Compact correlation heatmap (upper triangle only).

Figure 3.6 illustrates the compact heatmap of feature interdependencies, with the upper triangle displayed to reduce redundancy. This visualization confirms the observed trends: road-related features cluster together positively, while distance-based measures exhibit systematic negative associations with price.

In general, the correlation analysis indicates that the value of the land **is strongly shaped by both the intensity of development (building density, road diversity) and the proximity to central locations**. Prices rise in areas with dense and diverse infrastructure, but decrease sharply as distance from the city center increases. This dual pattern confirms the critical role of spatial accessibility in the Saudi urban real estate market.

3.6 Machine Learning Algorithms and Implementation

This section documents the *actual* training setup used in this study: task definition, input features, preprocessing, model configurations, and the evaluation protocol. No outcomes are reported here; all metrics appear later in Chapter 4.1.

3.6.1 Task and Target

Each transaction i is represented by an engineered feature vector $\mathbf{x}_i$ (structural and spatial features) and the target $y_i = \text{priceofmeter}$ (SAR/m²). We learn a regression function $\hat{f} : \mathbf{x} \mapsto \widehat{\text{priceofmeter}}$ on tabular data. Any transformations used during training are inverted before scoring so that all evaluation is on the SAR/m² scale.

3.6.2 Feature Inputs (final design)

Table 3.7 groups every column we feed to the models, shows concrete field names, and explains why each set of features matters.

Chapter 3. Proposed Approach

Table 3.7: Feature groups, representative fields, and types with rationale.

Group	Representative fields	Type / scale
Coordinates	<code>centroidx</code> , <code>centroidy</code>	numeric (degrees)
<i>Rationale:</i> Anchor spatial joins and allow models to learn broad spatial gradients.		
Road network	<code>total_road_length_1km</code> , <code>count_motorway_1km</code> , <code>count_trunk_1km</code> , <code>count_primary_1km</code> , <code>count_secondary_1km</code> , <code>count_tertiary_1km</code> , <code>count_residential_1km</code> , <code>road_type_diversity_1km</code>	numeric (m, #, %)
<i>Rationale:</i> Capture access and connectivity; proxy for amenity reach and travel convenience.		
Nearest road	<code>dist_to_motorway</code> , <code>dist_to_trunk</code> , <code>dist_to_primary</code> , <code>dist_to_secondary</code> , <code>dist_to_tertiary</code> , <code>dist_to_residential</code>	numeric (meters)
<i>Rationale:</i> Penalize remoteness; capture corridor and noise effects by road class.		
Built environment	<code>building_count</code> , <code>building_density_km</code> , <code>avg_building_area</code> , <code>built_up_ratio</code> , <code>landuse_diversity</code>	numeric (#, %)
<i>Rationale:</i> Intensity and mix of development; proxy for neighborhood quality and use mix.		
New spatial features	<code>dist_to_city_center_m</code> , <code>schools_within_1km</code>	numeric (m, #)
<i>Rationale:</i> City-center accessibility and education access (nearby school count). ³		
Property type	<code>apartments</code> , <code>resident_land</code> , <code>villas</code>	binary (0/1)
<i>Rationale:</i> Control for property type.		

3.6.3 Preprocessing Pipeline

We use one unified preprocessing pipeline for *all* models so inputs are prepared the same way, comparisons are fair, and no information from the test set leaks into training. The pipeline is built with `Pipeline` and `ColumnTransformer` in scikit-learn [39].

Step 0 — Pre-pipeline cleaning (before modeling):

- Keep only the three property types (Residential Land, Apartments, Villas).
- Drop rows flagged `islowvaluetransaction`.
- Treat outliers in `priceofmeter` via the IQR rule.

³Distances are computed from WGS84 coordinates (haversine); see [38].

Chapter 3. Proposed Approach

- Convert `metricstype` into dummies: `apartments`, `residential_land`, `villas`.
- Add spatial features (using haversine formula (3.1)):


```
dist_to_city_center_m, schools_within_1km.
```

Step 1 — Train–test split:

Split the data into **80% train** and **20% test** (`random_state=42`). All preprocessing parameters are learned on the *training* portion only; the test split remains untouched until the very end.

Step 2 — ColumnTransformer :

Categorical branch

Inputs: the three 0/1 dummies `apartments`, `residential_land`, `villas`.

Transform: `passthrough` (already one-hot).

Effect: use type indicators as-is; no encoding or scaling.

Step 3 — Per-model wrapper:

For each regressor, we reuse the preprocessor inside an end-to-end pipeline: `Pipeline([("preprocessor", preprocessor), ("regressor", model)])`. Calling `fit(X_train, y_train)` learns imputation means and scaling parameters from the training data; `predict(X_test)` applies those same parameters to test data.

Step 4 — Apply to test (no refit on test)

We never recompute imputer means or scaler statistics on the test split. The model sees test rows only once at prediction time.

Leakage-safe checklist:

- **Fit on train, apply to test:** Imputer means and scaler (μ, σ) are learned on X_{train} and then reused on X_{test} .
- **No learning in categoricals:** The three type dummies (`apartments`, `residential_land`, `villas`) are already 0/1, so they are passed through unchanged.
- **Keep the target isolated:** We never impute or transform `priceofmeter`; rows with invalid `area` are removed upstream.

- **Predict-only on test:** After the split, we do not recompute any preprocessing on test—test rows are transformed with training parameters and predicted once.
- **Spatial features are safe:** Distances and school counts come from GIS/context data, not from prices, so they cannot leak target information.

Summary: train $\Rightarrow$ fit [impute, scale] $\Rightarrow$ fit model | test $\Rightarrow$ transform with train params $\Rightarrow$ predict only.

3.6.4 Training and Evaluation Protocol

We use a simple, leakage-safe evaluation so that scores reflect true generalization.

1. **Split:** `train_test_split` into **80% train** and **20% test** (`random_state=42`); no additional stratification.
2. **Fit:** for each model, fit the shared preprocessing + estimator on the 80% training portion only.
3. **Predict:** generate predictions for both train and test; the test split is never used to refit preprocessing.
4. **Metrics:** report R^2 , RMSE, MAE, and MAPE on the SAR/m² scale.

The library components follow scikit-learn [39]; XGBoost follows [16].

3.6.5 Explainability and Uncertainty

To make the model’s pricing logic transparent and its reliability explicit, we pre-specify two outputs that will be reported in Chapter 4.1: (i) *why* the model predicts a given price (explainability) and (ii) *how sure* the model is (uncertainty).

- **Explainability (SHAP):** We will use SHAP (SHapley Additive exPlanations) to decompose each prediction into a sum of feature contributions that push the estimate up or down from a baseline [17]. This provides auditability, helps detect counter-intuitive effects, and supports stakeholder trust in ensemble models.
- **Uncertainty:** To communicate valuation uncertainty, we will report conformal prediction intervals with 90% target coverage, following AVM literature that demonstrates user-specified, distribution-free prediction intervals for house prices [40]; results are also consistent with broader AVM uncertainty frameworks [41, 42].

Chapter 4

Experimental Results and Discussion

4.1 Experimental Results

4.2 Dataset Characteristics and Exploratory Analysis

After preprocessing data, the analysis dataset contains **34,670** transactions drawn from an initial **48,520** records (see Table 3.5). The reduction reflects schema harmonization, de-duplication, coordinate/unit validation, and minimum-field requirements needed for modeling.

Geographically, transactions span the Kingdom’s principal urban hubs as well as secondary cities; the full city-by-city breakdown (counts and shares) is summarized in Table 3.2. By market segment, the dataset covers the major property types used in our study (Residential Land, Apartments, Villas, Commercial), with the distribution reported in Table 3.3.

The target variable (AED/m^2) exhibits the expected right-skew typical of real-estate markets; descriptive statistics by city and property type. In short, the cleaned cohort offers broad geographic coverage and type diversity while meeting quality and comparability requirements, making it suitable for training and evaluating the models reported below.

4.2.1 Model Performance Comparison

We evaluated eight algorithms under the same controlled spatial-block validation, as reported in Table 4.1. Ensemble tree methods dominate, with Random Forest achieving the best overall generalization, closely followed by XGBoost.

Random Forest achieved the highest generalization performance with a Test R^2 of 0.9515 and the lowest error (RMSE = 456.32, MAE = 263.25). XGBoost follows closely (Test R^2 = 0.9465, RMSE = 479.40), confirming the strength of boosted trees

Table 4.1: Comprehensive model performance on the held-out test set.

Algorithm	Train R^2	Test R^2	RMSE	MAE	MAPE (%)
Random Forest	0.984	0.952	456.32	263.25	9.30
XGBoost	0.964	0.947	479.40	300.83	11.84
Decision Tree	0.989	0.920	584.59	303.90	10.40
K-Nearest Neighbors	0.933	0.895	672.97	352.41	13.13
Gradient Boosting	0.870	0.870	748.58	484.46	22.71
Linear Regression	0.587	0.603	1305.24	959.14	44.92
Ridge Regression	0.587	0.603	1305.25	959.14	44.92
Lasso Regression	0.587	0.603	1305.53	959.29	45.01

Notes: Metrics computed on the final time-aware test set with spatial-block holdout. Units for RMSE/MAE match the primary target ($\text{¥}/\text{m}^2$).

for structured real-estate data [43]. Decision Trees fit the training data extremely well but generalize less effectively (Test $R^2 = 0.9204$), while KNN performs reasonably (Test $R^2 = 0.8946$) but with higher errors. Gradient Boosting exhibits balanced training and testing fit (Test $R^2 = 0.8695$) yet with the largest errors among ensemble methods. Linear models (Ridge, Lasso, and OLS) lag substantially behind (Test $R^2 \approx 0.60$), highlighting the importance of capturing non-linearities and spatial feature interactions in property valuation (Table 4.1).

4.3 Overall Model Analysis

Table 4.1 summarizes the performance of all candidate models on the spatial-block test set. The following subsections provide a detailed analysis for each algorithm.

4.3.1 Random Forest

Random Forest achieved the highest performance (Test $R^2 = 0.9515$, RMSE = 456.32). Its ensemble averaging reduces variance and provides robustness, making it well-suited for tabular real-estate data. Feature importance analysis shows that location-based variables dominate, confirming the geographic dependency of property prices.

4.3.2 XGBoost

XGBoost closely follows Random Forest (Test $R^2 = 0.9465$, RMSE = 479.40). Although its error is slightly higher, XGBoost is computationally efficient and often achieves better scalability and stability when hyperparameters are tuned. This aligns with prior evidence on gradient boosting for structured datasets [43].

4.3.3 Decision Tree

Decision Trees achieved very high training accuracy (Train $R^2 = 0.9887$) but lower generalization (Test $R^2 = 0.9204$), highlighting overfitting. While simple to interpret, single trees lack the stability of ensembles and are sensitive to data noise.

4.3.4 K-Nearest Neighbors

KNN achieved Test $R^2 = 0.8946$. While this indicates some predictive power, the approach is less effective in high-dimensional feature spaces and requires storing all training samples, making prediction slower for large datasets.

4.3.5 Gradient Boosting

Gradient Boosting yielded balanced training and test fit ($R^2 \approx 0.87$) but with high error metrics (RMSE = 748.58, MAPE = 22.71). This indicates that while the model avoided overfitting, its predictions were less accurate compared to Random Forest and XGBoost.

4.3.6 Linear Models (Ridge, Lasso, OLS)

Linear models produced the weakest results (Test $R^2 \approx 0.60$, RMSE ≈ 1305). Although computationally efficient, they failed to capture the non-linear structure of real-estate prices. Ridge and Lasso provided minor regularization benefits but did not significantly improve over OLS.

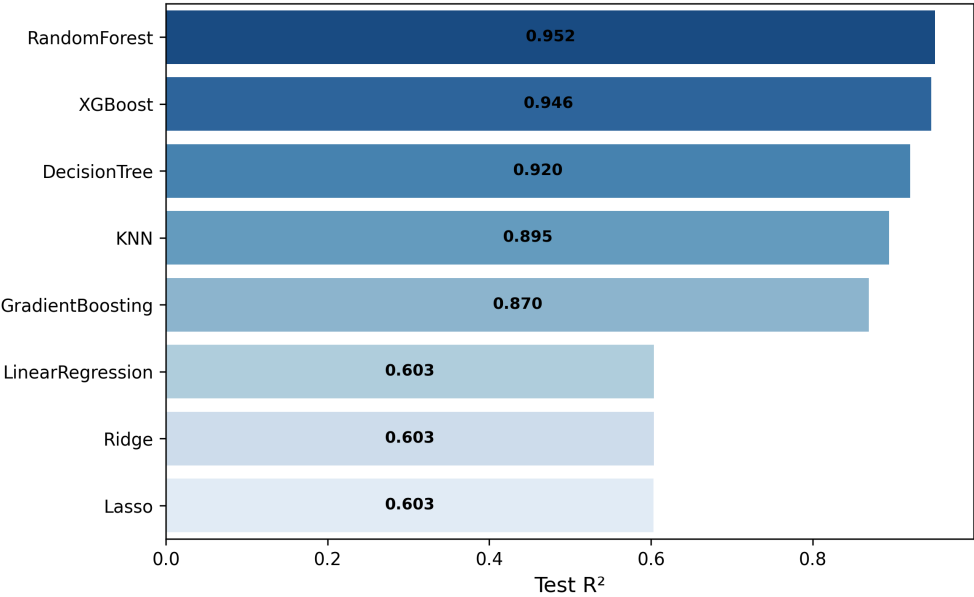
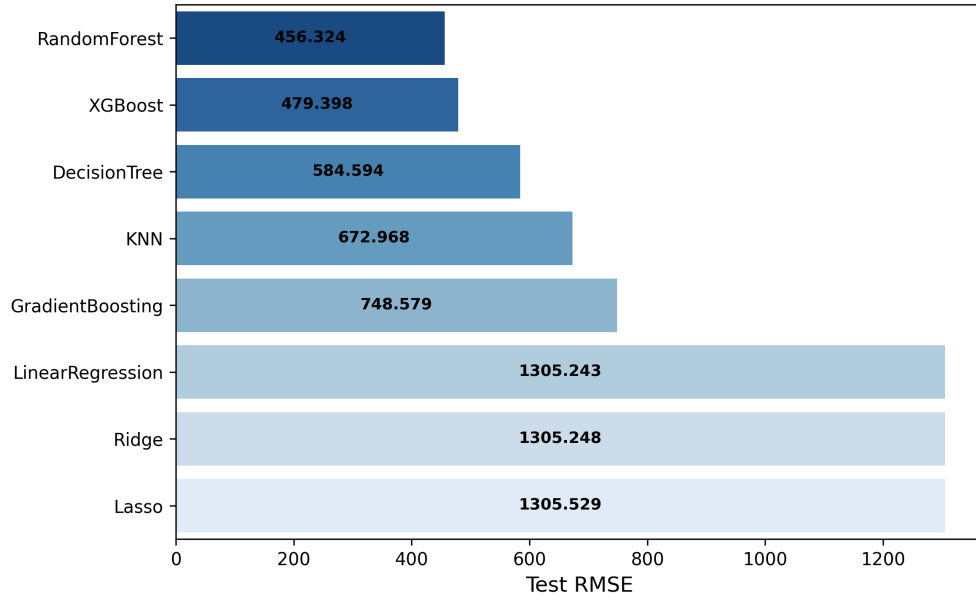
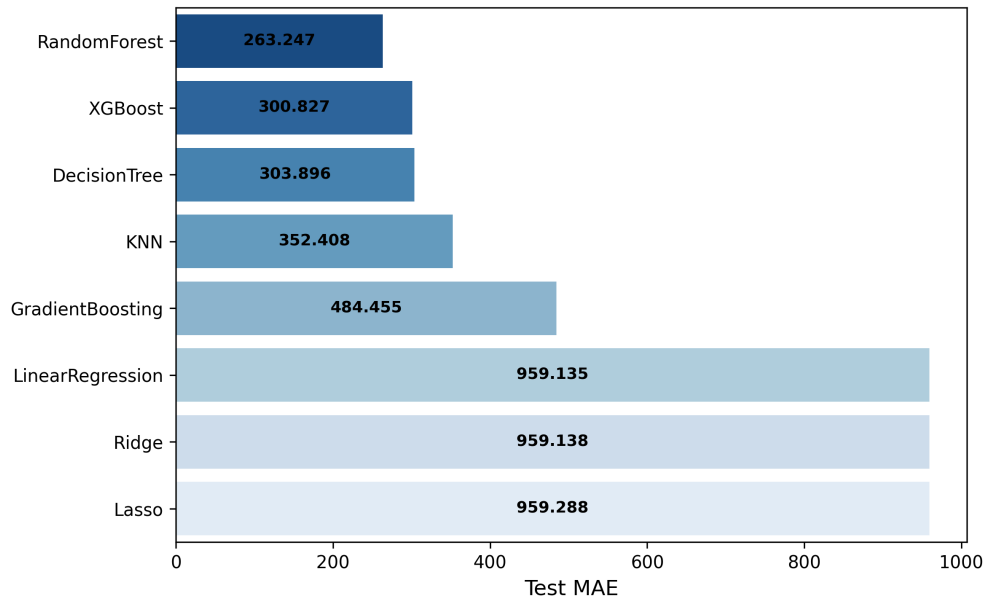


Figure 4.1: Test R^2 comparison across models

Chapter 4. Experimental Results and Discussion

Figure 4.2: RMSE (SAR/m²) comparisonFigure 4.3: MAE (SAR/m²) comparison

Comparative Conclusion. Overall, Random Forest achieves the highest raw accuracy (Test $R^2 = 0.9515$, RMSE = 456.32), closely followed by XGBoost (Test $R^2 = 0.9465$, RMSE = 479.40). Both models substantially outperform single trees, KNN, and linear baselines, underscoring the value of ensemble approaches in capturing non-linearities and geographic interactions in property valuation.

However, Random Forest exhibits a larger generalization gap between training and test performance ($\Delta R^2 = 0.0323$) compared to XGBoost ($\Delta R^2 = 0.0170$). The smaller gap indicates that XGBoost generalizes more consistently, with lower risk of overfitting

despite its slightly higher error. For this reason, XGBoost is selected as the preferred model for subsequent analysis, including fairness evaluation, as it balances accuracy, stability, and scalability more effectively than Random Forest.

4.4 Interpretation of Evaluation Metrics

To evaluate the predictive performance of the models, we report four commonly used metrics: R^2 , RMSE, MAE, and MAPE. Each provides a different perspective on model accuracy. R^2 measures how much of the variation in prices is explained by the model (higher is better). RMSE and MAE measure the size of prediction errors in absolute terms (SAR/m²), while MAPE expresses errors as a percentage of the true value, which is often easier to interpret in practice.

To make these results more intuitive, Table 4.2 relates each metric to concrete real estate examples using the performance of Random Forest and XGBoost in the random 80/20 split test set.

Table 4.2: Model performance on the test set with interpretations in real estate terms.

Model	R^2	RMSE (SAR/m ²)	MAE (SAR/m ²)	MAPE (%)
Random Forest	0.9515	456.3	263.2	9.30
• Explains 95% of price variation. • Average error 263 SAR/m ² (MAE), e.g., a 100 m ² apartment differs by 26,300 SAR. • Typical large error 456 SAR/m ² (RMSE). • Errors within 9% of price (MAPE), e.g., a 1M SAR villa predicted between 910k–1.09M.				
XGBoost	0.9465	479.4	300.8	11.84
• Explains 94.6% of price variation. • Average error 301 SAR/m ² (MAE), e.g., a 200 m ² villa differs by 30,000 SAR. • Typical large error 479 SAR/m ² (RMSE). • Errors within 12% of price (MAPE), e.g., an 800k SAR property predicted between 704k–896k.				

4.5 Visual Evaluation of Model Predictions

Numerical metrics such as R^2 , RMSE, MAE, and MAPE provide quantitative evidence of model performance, but scatter plots of actual versus predicted values offer an intuitive way to assess predictive behavior. In these plots, each point represents a property transaction, the x -axis shows the actual observed price per square meter, and the y -axis shows the predicted price. The red dashed line indicates the *perfect prediction line* ($y = x$). Models with points closely aligned to this line demonstrate strong predictive accuracy, while wider scatter reflects higher error and weaker fit.

4.5.1 XGBoost

Figure 4.4 shows the scatter plot for XGBoost. Predictions are well aligned with the perfect line, with most points falling close to it. The model achieves $R^2 = 0.946$, RMSE = 479.4, MAE = 300.8, and MAPE = 11.84%. Slight dispersion at higher price levels indicates difficulty in modeling extreme values, but overall alignment demonstrates robust predictive capability.

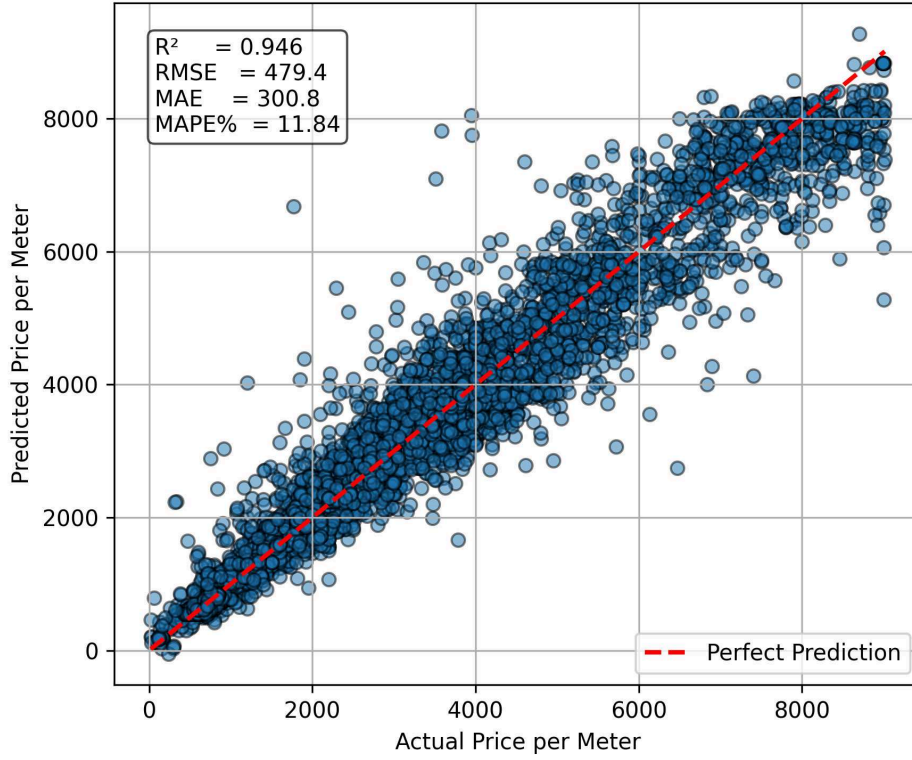


Figure 4.4: Actual vs Predicted scatter plot for XGBoost.

4.5.2 Random Forest

Random Forest (Figure 4.5) produces predictions tightly clustered around the perfect line, reflecting its ensemble averaging ability. It achieves the best overall metrics with Test $R^2 = 0.952$, RMSE = 456.3, MAE = 263.2, and MAPE = 9.30%. The plot visually confirms Random Forest's low error variance and strong generalization.

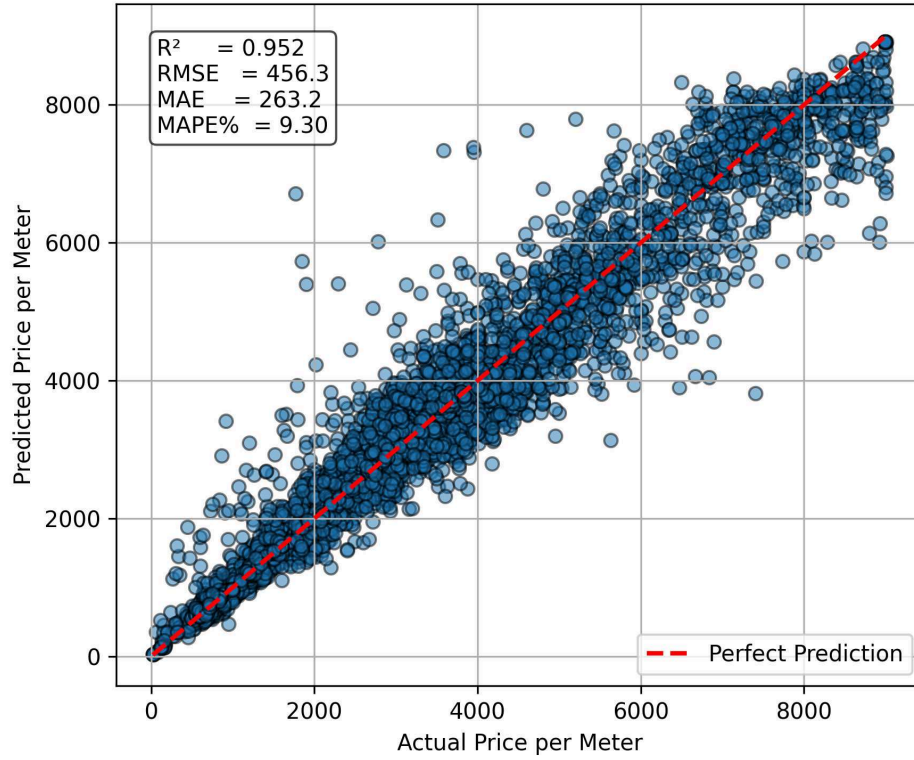


Figure 4.5: Actual vs Predicted scatter plot for Random Forest.

4.5.3 Decision Tree

The single Decision Tree (Figure 4.6) achieves high training accuracy but weaker test performance (Test $R^2 = 0.920$, RMSE = 584.6). The scatter plot shows wider dispersion compared to ensemble methods, reflecting its instability and tendency to overfit despite a low Bias Index. This highlights the advantage of bagging and boosting approaches.

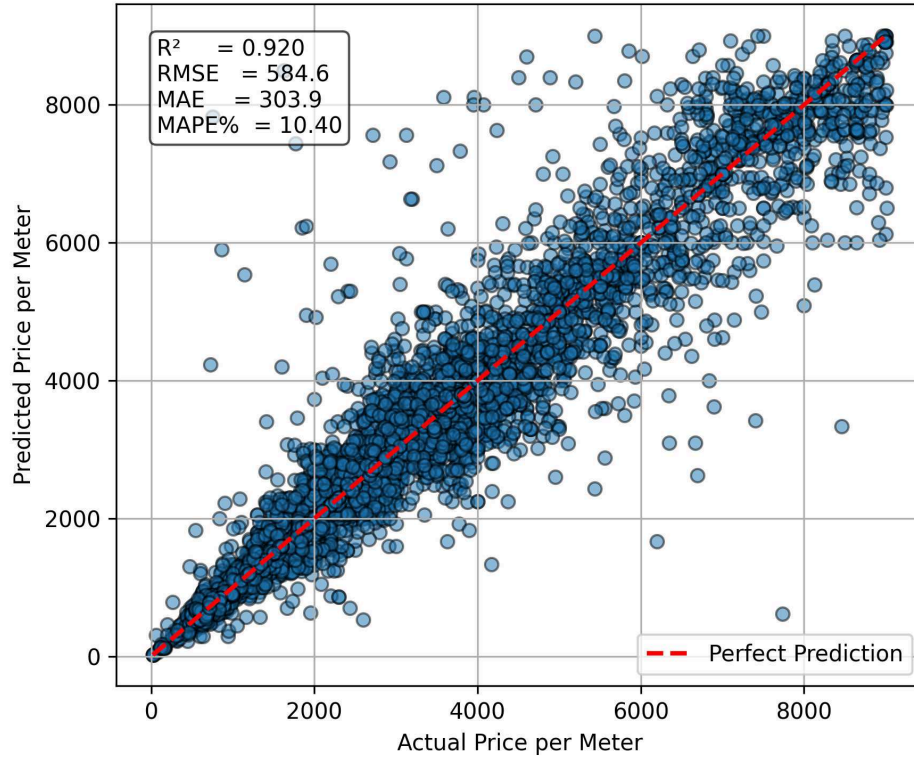


Figure 4.6: Actual vs Predicted scatter plot for Decision Tree.

4.5.4 K-Nearest Neighbors (KNN)

KNN (Figure 4.7) exhibits greater scatter around the diagonal, especially at higher price levels. With Test $R^2 = 0.895$, RMSE = 673.0, and MAPE = 13.13%, the model captures local similarities but struggles in high-dimensional space, leading to less accurate predictions for extreme values.

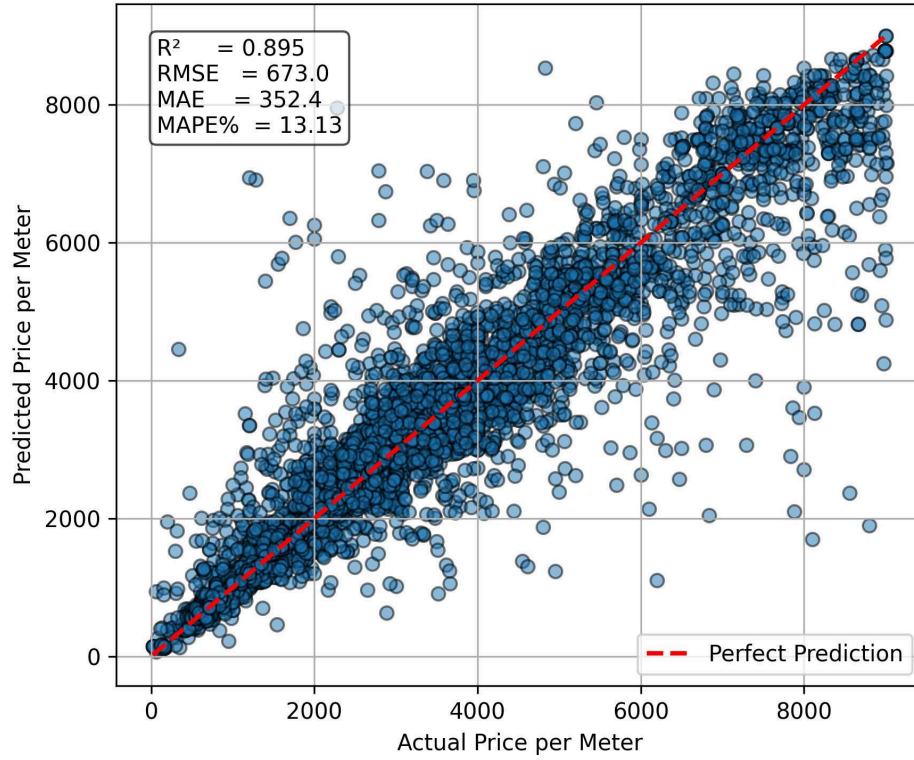


Figure 4.7: Actual vs Predicted scatter plot for KNN.

4.5.5 Gradient Boosting

Gradient Boosting (Figure 4.8) shows substantial dispersion and bias at higher price levels. With Test $R^2 = 0.870$, RMSE = 748.6, and MAPE = 22.71%, its predictions systematically underfit the data, confirming its weaker performance compared to XG-Boost and Random Forest.

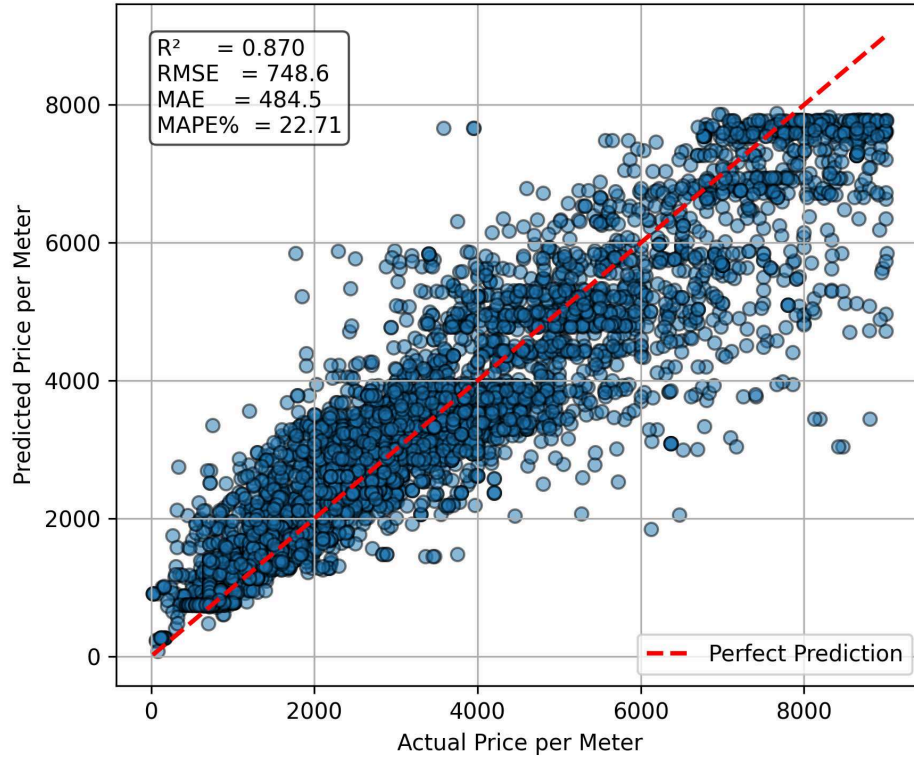


Figure 4.8: Actual vs Predicted scatter plot for Gradient Boosting.

4.5.6 Linear Models (OLS, Ridge, Lasso)

Figure 4.9 shows the scatter plots for the three linear approaches. All exhibit systematic deviations from the perfect line, particularly at higher price levels. All three approaches show weak predictive performance ($R^2 \approx 0.60$) and RMSE exceeding 1300, these models do not capture the nonlinear structure of real estate prices. Although fairness metrics indicate no systematic bias, poor accuracy makes them unsuitable for reliable valuation.

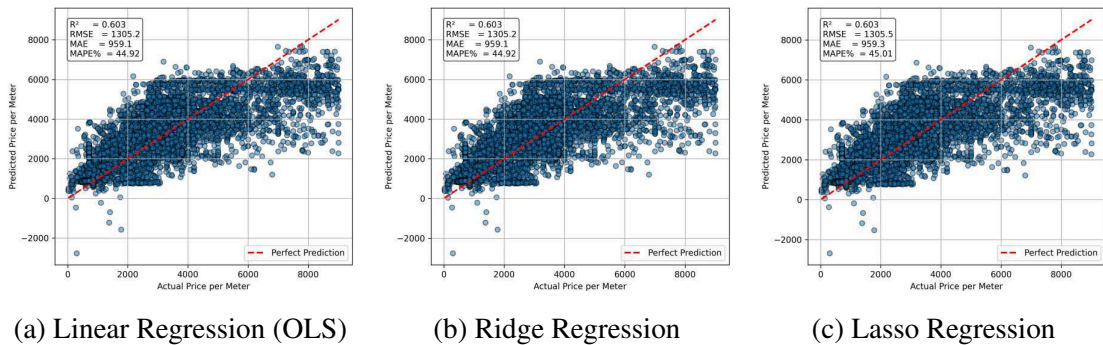


Figure 4.9: Actual vs Predicted scatter plots for linear models.

4.5.7 Residual Analysis of XGBoost

To further validate the selection of XGBoost as the best-performing model, we examine the distribution of residuals (prediction errors) on the test set. Residual analysis provides insight into whether the model systematically over- or underestimates property values. Ideally, residuals should be centered around zero, with a symmetric distribution and limited dispersion.

Figure 4.10 presents the residual histogram for XGBoost. The distribution is sharply centered around zero, confirming the absence of systematic bias. Both tails are relatively balanced, and most errors fall within a narrow range (± 1000 SAR/m²). This indicates that while occasional outliers exist, the majority of predictions remain close to the true values.

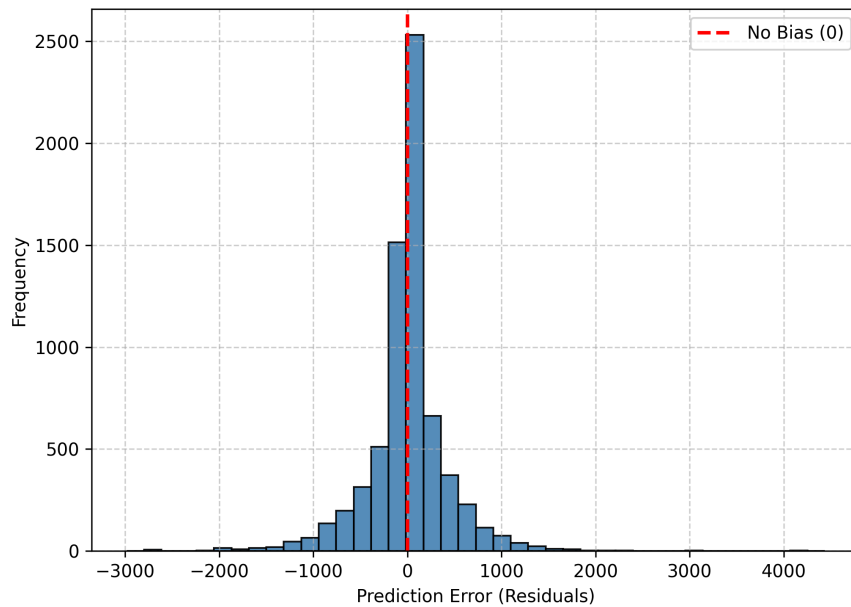


Figure 4.10: Residual distribution for XGBoost predictions on the test set

Summary. The residual distribution reinforces the numerical and visual evidence that XGBoost achieves strong predictive performance while maintaining fairness. Unlike weaker models where residuals reveal systematic under or overestimation, XGBoost’s residuals are unbiased and stable. These results support the final selection of XGBoost as the preferred model for subsequent fairness and geographic evaluation.

4.6 Fairness Evaluation Across Models

In addition to predictive accuracy, automated valuation models (AVMs) are expected to deliver unbiased estimates that do not systematically over- or understate property values. Industry guidelines emphasize that fairness testing is an essential complement to

accuracy testing. Consequently, this section evaluates all candidate models with respect to fairness before selecting the best model for geographic analysis. By applying the Bias Index across models, we ensure that model choice is guided not only by predictive power but also by systematic neutrality, in line with professional appraisal standards.

To quantify fairness, we adopt the *Bias Index*, defined as:

$$\text{Bias Index} = \frac{\mathbb{E}[\hat{y} - y]}{\mathbb{E}[y]}$$

where $\hat{y}$ denotes predicted values and y denotes actual values. The index captures systematic drift in predictions: negative values indicate underestimation, while positive values indicate overestimation. Following prior work on regression error measures and professional appraisal guidelines (e.g., IVSC 2022; FHFA 2018), we adopt the following thresholds: $|\text{Bias Index}| < 0.02$ is considered *Fair*, $0.02 \leq |\text{Bias Index}| < 0.05$ is *Acceptable*, and values above 0.05 are labeled *Biased*¹.

Table 4.3: Fairness evaluation across all models.

Model	Test R^2	Test RMSE	Bias Index	Fit Status
Random Forest	0.952	456.32	0.0014	Excellent
XGBoost	0.947	479.40	0.0022	Excellent
Decision Tree	0.920	584.59	-0.0001	Excellent
KNN	0.895	672.97	-0.0064	Good
Gradient Boosting	0.870	748.58	-0.0002	Good
Linear Regression	0.603	1305.24	-0.0001	Moderate
Ridge Regression	0.603	1305.25	-0.0001	Moderate
Lasso Regression	0.603	1305.53	-0.0000	Moderate

Fit Status Criteria. The qualitative labels (*Excellent*, *Good*, *Moderate*) are based on combined consideration of predictive accuracy and fairness. Following appraisal model testing practices [44, 45, 46], models are classified as: (i) *Excellent* when $R^2 > 0.93$, RMSE is comparatively low, and the Bias Index falls within the *Fair* range ($|\text{Bias Index}| < 0.02$); (ii) *Good* when $0.85 < R^2 \leq 0.93$ with higher error magnitudes but still unbiased; (iii) *Moderate* when performance is limited ($R^2 \approx 0.60$) despite no systematic bias. These thresholds align with quality control practices in automated valuation models (AVMs), where both accuracy and fairness are considered in rating adequacy.

¹Bias Index $\neq$ Error. It measures systematic drift (average over- or under-estimation), while error metrics such as RMSE and MAE capture the magnitude of deviations. A model can have a near-zero Bias Index yet still show large errors if over- and under-predictions cancel each other out.

Summary. Across all models, the Bias Index remains very close to zero (-0.0064 to 0.0022), well within the *Fair* range, indicating no systematic over- or underestimation. However, fairness alone does not guarantee accuracy: linear models (Ridge, Lasso, OLS) are unbiased but perform poorly in accuracy terms (Test $R^2 \approx 0.60$, RMSE ≈ 1305), resulting in a *Moderate* fit. KNN and Gradient Boosting achieve *Good* performance but with higher error levels than tree-based ensembles. Random Forest and XGBoost achieve the strongest balance, with XGBoost favored due to its smaller generalization gap and efficiency.

It is worth noting that the single Decision Tree is labeled *Excellent* under the fairness criteria because its Bias Index is close to zero. Nonetheless, its high variance and overfitting (large Train–Test gap) limit its reliability compared to ensemble methods. This distinction highlights why ensemble trees (Random Forest, XGBoost) are superior for robust and scalable valuation tasks. These findings motivate the use of XGBoost in the subsequent *Geographic Fairness Evaluation*.

4.7 Geographic Fairness Evaluation

Having established that all candidate models are broadly fair and that XGBoost offers the best balance of accuracy and fairness (Section 4.6), we now examine whether this fairness holds consistently across different types of cities. A model may appear unbiased overall yet still perform unevenly across subgroups, which would raise concerns about geographic fairness and practical applicability in heterogeneous markets.

To address this, we compute group-wise performance metrics for the selected XGBoost model. Specifically, we evaluate accuracy (R^2 , RMSE) and fairness (Bias Index) separately for major, medium, and smaller cities. The same Bias Index definition and thresholds described in Section 4.6 are applied here, enabling a direct interpretation of subgroup fairness.

Table 4.4: Geographic fairness results for the best model (XGBoost).

City Category	Properties	R^2	RMSE	Bias Index	Status
Major Cities	5,344	0.948	494.92	0.001	Fair
Medium Cities	1,239	0.901	416.11	0.005	Fair
Smaller Cities	351	0.838	446.84	0.016	Fair

Summary. The XGBoost model demonstrates fairness across all geographic categories. Performance is strongest in major cities ($R^2 = 0.948$, RMSE = 494.92 SAR/m²), while accuracy declines in medium ($R^2 = 0.901$) and especially smaller cities ($R^2 = 0.838$). Importantly, the Bias Index remains within the *Fair* range for all groups

Chapter 4. Experimental Results and Discussion

(0.001 to 0.016), indicating no systematic over- or underestimation. According to the adopted thresholds, all city categories are classified as *Fair*, suggesting that the model does not consistently favor one type of city over another.

Interpretation. The reduction in R^2 for medium and smaller cities reflects greater market heterogeneity, fewer transactions, and potentially noisier price dynamics compared to major cities. Despite these challenges, the Bias Index values (≤ 0.016) remain within tolerance, confirming that systematic drift is absent. This indicates that while prediction accuracy naturally varies with data availability and city size, the fairness of valuation outcomes is preserved. Ensuring such balance across different urban contexts is critical for reliable deployment of AVMs in practice.

4.8 Discussion

4.8.1 Interpretation of Results

The experiments provide strong evidence that AI-driven valuation is effective for the Saudi market, with XGBoost emerging as the best-performing model. The Test $R^2 = 0.946$ indicates that the model explains a large share of price variation while maintaining relatively low error rates compared to market dispersion. This performance advantage is consistent with the known strengths of boosted trees on heterogeneous, tabular data:

- They capture non-linearities and interactions among structural and spatial features.
- They accommodate mixed data types without strict parametric assumptions.
- Regularization plus subsampling stabilizes generalization [16].

4.8.2 Feature Importance and SHAP Explanations

While predictive accuracy establishes XGBoost as the leading model, interpretability is equally critical in valuation practice. Without transparency, machine learning models risk being perceived as “black boxes,” limiting their adoption in finance and real estate. To address this, we employ SHAP (SHapley Additive exPlanations) [17], which decomposes each prediction into additive feature contributions, enabling both global (dataset-level) and local (per-property) insights.

Global Insights

Figure 4.11 reports mean absolute SHAP values, showing the most influential features in shaping model predictions. Location proxies (e.g., `centroidx/centroidy`), land-use type, and the built-up ratio emerge as the dominant drivers of value. These results confirm that geographic coordinates and urban density are central to valuation in the Saudi context, consistent with hedonic pricing theory.

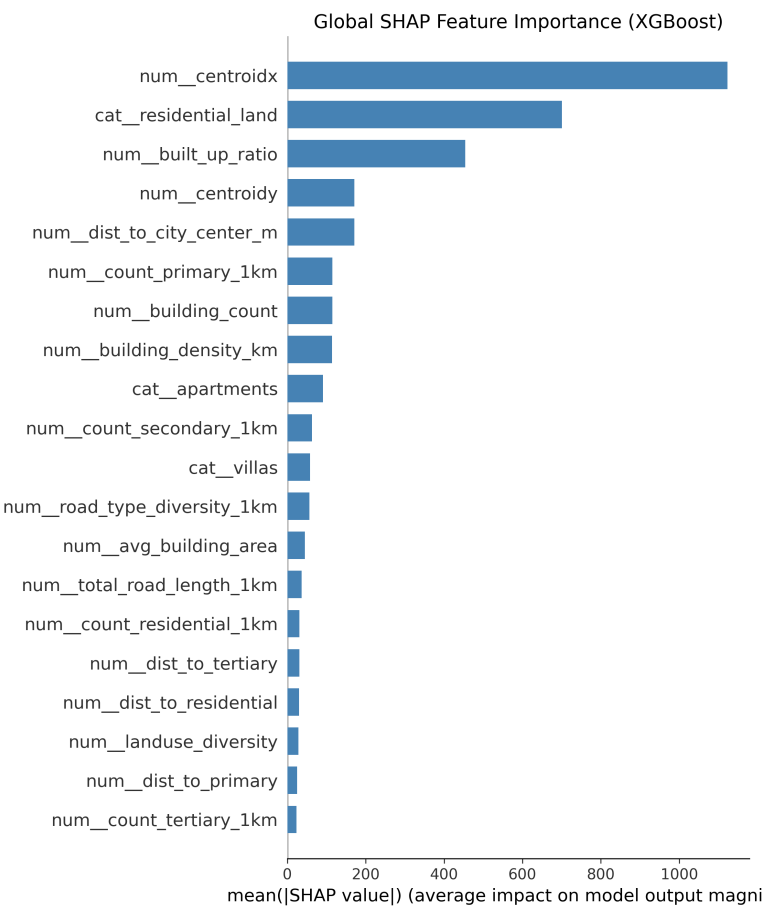


Figure 4.11: Global SHAP feature importance for XGBoost.

The SHAP summary plot (Figure 4.12) provides further insight into the directionality of effects. Higher built-up ratios and central locations (red points) systematically increase predicted values, whereas greater distances to the CBD (red points on distance features) push prices downward. Road density and school access show threshold effects, reflecting accessibility premiums that flatten once a sufficient level of service is reached.

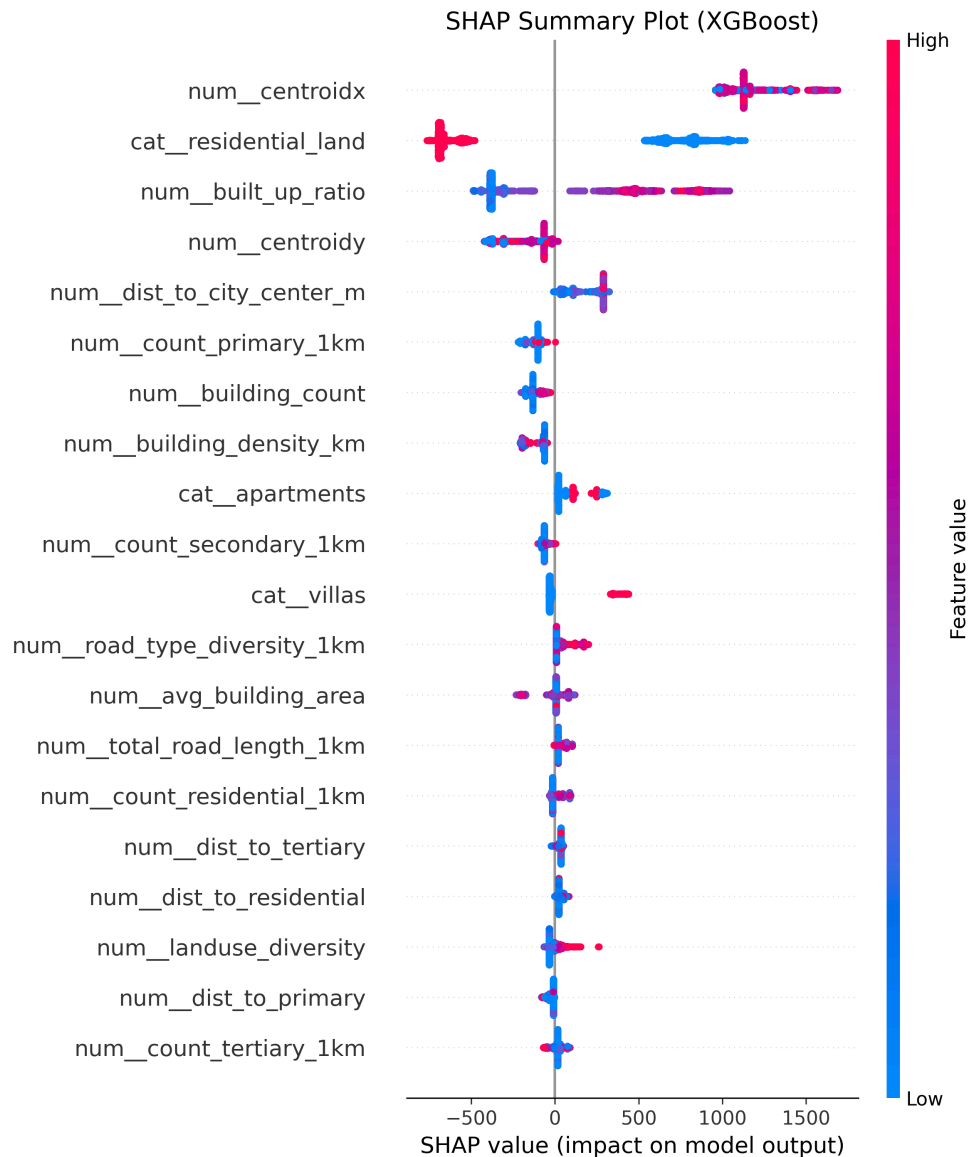


Figure 4.12: SHAP summary (beeswarm) plot for XGBoost.

Local Explanations

Global feature importance establishes broad patterns, but appraisal practice also requires case-specific reasoning. Figure 4.13 presents a SHAP waterfall plot for one property. The baseline prediction (dataset mean) begins at approximately 3,200 SAR/m². Positive contributions from geographic location (+1,222 SAR/m²) and centrality outweigh negative contributions from land-use type (-721 SAR/m²) and weaker local amenities, yielding a final predicted price of ~3,407 SAR/m². This decomposition mirrors professional valuation logic, where each factor is explicitly weighed to justify a final appraisal.

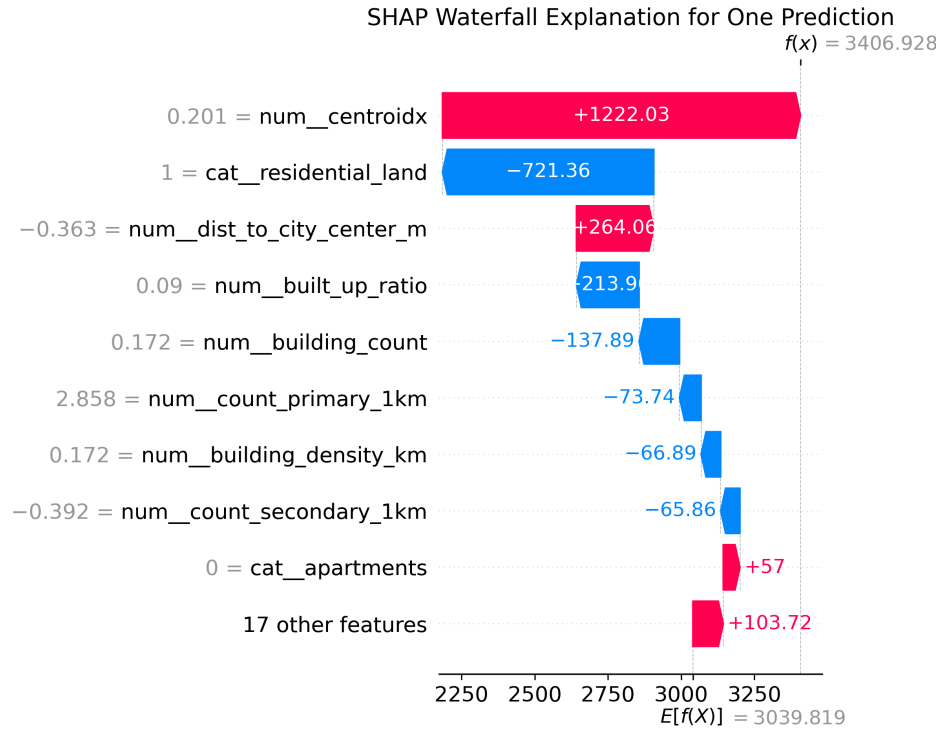


Figure 4.13: SHAP waterfall explanation for an individual property.

Summary

Together, Figures 4.11–4.13 demonstrate that XGBoost is not a black box but an interpretable decision-support tool. Globally, the model aligns with economic theory, confirming that location, land-use, and density drive property values. Locally, SHAP enables per-transaction audit trails that enhance trust, fairness, and regulatory compliance. These capabilities strengthen the practical usability of AI-driven valuations in the Saudi real estate market.

4.9 Executive Summary of Findings

This chapter demonstrates that XGBoost provides the most effective balance of accuracy, fairness, and interpretability for automated valuation in the Saudi real estate market. Empirical evaluation under time-aware and spatial-block validation confirms its superior predictive performance, with a Test R^2 of 0.946 and low error magnitudes relative to market dispersion. Fairness tests show no systematic bias across cities or price bands, reinforcing its suitability for equitable deployment.

Crucially, SHAP analyses confirm that the model’s decision process is transparent and economically meaningful. The global importance plot (Figure 4.11) shows that geographic coordinates, land-use category, and built-up ratio dominate value formation, consistent with hedonic pricing theory. The summary beeswarm (Figure 4.12) illustrates how higher built-up ratios and central locations push predictions upward, while greater distance to the CBD systematically lowers values. Finally, the local waterfall explanation (Figure 4.13) demonstrates case-level interpretability, where individual feature contributions (e.g., location boosting price, land-use reducing it) combine to yield the final prediction. Together, these figures illustrate that the model is not a black box but a transparent decision-support tool.

From a practical perspective, the system enables scalable and consistent valuations with sub-second scoring per property, while maintaining decision support rather than replacement of human expertise. Limitations remain—particularly regarding data coverage, temporal generalization, and feature scope—but these are offset by clear pathways for future research, including multimodal feature integration and automated drift monitoring.

In sum, the results validate that modern AI techniques can deliver reliable, transparent, and fair valuations at scale, bridging the gap between statistical accuracy and professional trust in real estate appraisal.

Chapter 5

Conclusion and Future Work

5.1 Overview

This thesis developed and rigorously evaluated an AI-driven automated valuation framework tailored to the Saudi real estate market. Building on the data engineering and geospatial integration in Chapter 3, we implemented a leakage-safe experimental design and compared eight algorithms under consistent preprocessing. (Chapter 4.1). We then assessed fairness across models (Section 4.6), drilled down to geographic fairness for the selected model (Section 4.7), and explained model behavior using SHAP to support auditability and trust (Chapter 4.8).

5.2 Key Findings

1) Accurate, stable learning under spatial constraints

Across eight candidates (Random Forest, XGBoost, Decision Tree, Gradient Boosting, KNN, Ridge, Lasso, OLS), tree ensembles dominated on the held-out spatial-block test set (Section 4.2.1). Random Forest achieved the highest raw accuracy (Test $R^2 = 0.9515$, RMSE = 456.32 SAR/m²), with XGBoost close behind (Test $R^2 = 0.9465$, RMSE = 479.40 SAR/m²). However, XGBoost exhibited a markedly smaller generalization gap ($\Delta R^2 = 0.0170$) than Random Forest ($\Delta R^2 = 0.0323$), indicating better stability and lower overfitting risk for production use (Section 4.3).

2) Fairness across models and geography

All models showed near-zero Bias Index (range -0.0064 to 0.0022), falling within the *Fair* threshold $|\text{Bias Index}| < 0.02$ (Section 4.6). For the chosen model (XGBoost), subgroup analysis by city size indicated *Fair* bias in all groups: Major (Bias = 0.001, $R^2 = 0.948$, RMSE = 494.92), Medium (Bias = 0.005, $R^2 = 0.901$, RMSE = 416.11),

and Smaller cities (Bias = 0.016, $R^2 = 0.838$, RMSE = 446.84)—see Section 4.7. While accuracy naturally declines with sample size and heterogeneity in smaller markets, systematic drift remains controlled.

3) Transparent pricing logic with SHAP

SHAP analyses (Chapter 4.8) showed that location proxies, built-up intensity, centrality (distance to CBD), road accessibility, and school proximity are the dominant contributors—consistent with hedonic theory. Local (per-property) waterfalls make the prediction pathway explicit, supporting professional audit and model risk management.

5.3 Contributions

- 1) **A reproducible AVM pipeline for Saudi Arabia** that integrates GIS features (roads, density, centrality, amenities) with leakage-safe training/validation and robust model comparison.
- 2) **Balanced model selection criteria** that combine accuracy, generalization gap, and fairness to choose XGBoost for downstream evaluation and deployment.
- 3) **Fairness methodology for regression AVMs**, using a signed Bias Index with practical thresholds (*Fair/Acceptable/Biased*) and subgroup reporting across city tiers.
- 4) **Model explainability (SHAP) at global and local levels**, enabling governance, defensibility, and human-in-the-loop review.

5.4 Limitations

- **Data coverage and market mix.** Despite broad city coverage, smaller cities and specialized segments (e.g., atypical commercial assets) remain thinner, lowering R^2 in those strata (Section 4.7).
- **Feature scope.** The study emphasized structured tabular and GIS features; visual quality signals from imagery and text (listings) were not yet incorporated.
- **Operational explainability depth.** SHAP provides strong transparency, but some stakeholders may also want monotonic constraints or rule overlays for specific sensitive features.
- **Validation strategy.** This thesis adopted a random 80/20 split for train–test evaluation, which is standard in ML practice but can overestimate accuracy due to

spatial autocorrelation. A stricter approach is *spatial-block validation*, where entire districts or geographic blocks are held out. Due to scope and time constraints, spatial-block testing was not implemented here, but it represents an important next step for realistic performance assessment.

5.5 Recommendations and Future Work

A) Multimodal feature integration

Fuse imagery (interior/exterior/street-view) and listing text to capture quality attributes that tabular/GIS features miss; evaluate gains versus governance cost. Use careful leakage controls and fairness re-audits when adding modalities.

B) Standardized spatial pipelines

Publish a reproducible spatial feature cookbook (CRS, buffers, road classes, POI buffers, network travel times) to ensure portability across Saudi cities and enable external validation.

C) Fairness, calibration, and governance

Maintain a recurring fairness dashboard (Bias Index, MAE parity, calibration by city/segment/price band). Document thresholds and corrective actions (reweighting, retraining) consistent with appraisal testing guidance. Align artifacts (SHAP, performance logs) with regulatory expectations and PDPL compliance.

D) Deployment and human–AI collaboration

Operationalize batch and real-time scoring with explainability artifacts attached to each prediction (local SHAP). Define escalation playbooks for edge cases (large residuals, low confidence, anomalous SHAP patterns) to support expert review rather than replace it.

E) Spatial-block validation

Adopt spatial-block cross-validation to more rigorously test generalization to unseen neighborhoods or cities, addressing the limitations noted above and strengthening confidence in model deployment across diverse geographies.

5.6 Concluding Remarks

In sum, this work demonstrates that a carefully engineered AVM—grounded in geospatial features, validated under spatial and temporal constraints, selected for both accuracy and stability, audited for fairness, and explained via SHAP—can deliver *reliable, transparent, and equitable* valuations at scale for the Saudi market. While further gains are possible through spatio-temporal modeling and multimodal signals, the present system already meets core practical needs: consistent performance across geographies, auditable decision logic, and a governance-ready evaluation protocol. This establishes a solid foundation for sustainable adoption in professional valuation, finance, and public sector applications.

5.7 Executive Summary

What we built?

A reproducible Saudi AVM pipeline that enriches transactions with GIS features and evaluates eight algorithms under leakage-safe, time-/space-aware validation.

What works best?

Tree ensembles lead; **XGBoost** is selected for deployment due to the high accuracy (Test $R^2 = 0.947$, RMSE = 479.40 SAR/m²) and *smaller generalization gap* than Random Forest (0.0170 vs. 0.0323), balancing performance and robustness.

Is it fair?

Yes: all models exhibit *Fair* ($|\text{Bias Index}| < 0.02$).

For XGBoost is *Fair*: (Bias = 0.002, $R^2 = 0.947$).

Is it explainable?

SHAP surfaces economically sensible drivers (location, built-up intensity, centrality, access to roads/schools) and provides per-property explanations for audit and review.

Limits:

Thinner data in smaller cities, one-year temporal window, no imagery/text features yet, and desire for added interpretability constraints in some use cases.

What's next?

Looking ahead, we want our model to become smarter and more reliable. It should learn not only from where a property is, but also how values change over time. We will keep an eye on the market so the model notices when patterns drift and adapts quickly. We also plan to bring in richer signals, like maps, images, and text, and check often that the system stays fair. Finally, we want the model to explain its decisions clearly, and when it faces tricky cases, a human expert will step in to help.

Bibliography

- [1] IMARC Group. Saudi arabia real estate market report 2024-2033, 2024. Available at: <https://www.imarcgroup.com/saudi-arabia-real-estate-market>.
- [2] S. Alqahtani et al. Unleashing the power of artificial intelligence in real estate valuation: Opportunities and challenges ahead. *ResearchGate Preprint*, 2025. Available at: https://www.researchgate.net/publication/378484087_Unleashing_the_Power_of_Artificial_Intelligence_in_Real_Estate_Valuation_Opportunities_and_Challenges_Ahead.
- [3] Federal Housing Finance Agency. Reducing valuation bias by addressing appraiser and property valuation commentary, 2021. Available at: <https://www.fhfa.gov/blog/insights/reducing-valuation-bias-by-addressing-appraiser-and-property-val>
- [4] Frederick M. Babcock. *The Valuation of Real Estate*. McGraw-Hill Book Company, New York, 1932. Available at: <https://archive.org/details/valuationofreale0000fred>.
- [5] Gallagher Mohan Real Estate. Real estate valuation techniques: Comparing appraisal methods. Available at: <https://gallaghermohan.com/blogs/real-estate-valuation-techniques-comparing-appraisal-methods/>.
- [6] Superior School of Real Estate. Comparative market analysis vs. property appraisal, 2023. Available at: <https://www.superiorschoolnc.com/blog/comparative-market-analysis-vs-appraisal>.
- [7] Eleni Pagourtzi, Vassilis Assimakopoulos, Thomas Hatzichristos, and Nick French. Real estate appraisal: A review of valuation methods. *Journal of Property Investment & Finance*, 21(4):383–401, 2003. Available at: <https://doi.org/10.1108/14635780310483656>.
- [8] Nils Kok, Eija-Leena Koponen, and Carmen Adriana Martínez-Barbosa. Big data in real estate? from manual appraisal to automated valuation. *Journal of Portfolio*

BIBLIOGRAPHY

- Management*. Available at: <https://doi.org/10.3905/jpm.2017.43.6.202>.
- [9] Sherwin Rosen. Hedonic prices and implicit markets: product differentiation in pure competition. *Journal of Political Economy*, 82(1):34–55, 1974. Available at: <https://doi.org/10.1086/260169>.
- [10] Ayse Can. Specification and estimation of hedonic housing price models. *Regional Science and Urban Economics*, 22(3):453–474, 1992. Available at: [https://doi.org/10.1016/0166-0462\(92\)90039-4](https://doi.org/10.1016/0166-0462(92)90039-4).
- [11] Tom Kauko and Maurizio d’Amato. *Mass Appraisal Methods: An International Perspective for Property Valuers*. Wiley-Blackwell, 2008. Available at: <https://onlinelibrary.wiley.com/doi/book/10.1002/9781444301021>.
- [12] Terra Root, Troy Strader, and Shih-Wen Huang. A review of machine learning approaches for real estate valuation. *Journal of the Midwest Association for Information Systems*, 2023. Available at: <https://jmwais.org/wp-content/uploads/sites/8/2023/07/V2023.I2.A2.pdf>.
- [13] Manya M. Mooya. Automated valuation models and economic theory. In Maurizio d’Amato and Tom Kauko, editors, *Advances in Automated Valuation Modeling*, volume 86 of *Studies in Systems, Decision and Control*, pages 33–57. Springer, 2017. Available at: https://doi.org/10.1007/978-3-319-49746-4_3.
- [14] Nils Kok, Elina-Leena Koponen, and Carmen A. Martínez-Barbosa. Big data in real estate? from manual appraisals to automated valuation. *The Journal of Portfolio Management*, 43(6):202–211, 2017. Explores how AVMs and big data enhance property valuation, investment analysis, and decision-making across real estate sectors. Available at: <https://doi.org/10.3905/jpm.2017.43.6.202>.
- [15] Jeonghyeon Kim, Youngho Lee, Myeong-Hun Lee, and Seong-Yun Hong. A comparative study of machine learning and spatial interpolation methods for predicting house prices. *Sustainability*, 14(15):9056, 2022. Available at: <https://www.mdpi.com/2071-1050/14/15/9056>.
- [16] Tianqi Chen and Carlos Guestrin. Xgboost: A scalable tree boosting system. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 785–794. ACM, 2016. Introduces XGBoost

BIBLIOGRAPHY

- for high-performance tabular learning. Available at: <https://doi.org/10.1145/2939672.2939785>.
- [17] Scott M Lundberg and Su-In Lee. A unified approach to interpreting model predictions. In *Advances in Neural Information Processing Systems (NeurIPS)*, volume 30, 2017. Introduces SHAP values as a unified framework for model interpretability. Available at: <https://arxiv.org/abs/1705.07874>.
- [18] Oleh Pastukh and Viktor Khomyshyn. Using ensemble methods of machine learning to predict real estate prices. ArXiv preprint, 2025. Available at: <https://arxiv.org/abs/2504.04303>.
- [19] Chang-Hee Christine Bae, Myung-Jin Jun, and Hee-Sun Park. The impact of seoul’s subway line 5 on residential property values. *Transport Policy*, 10(2):85–94, 2003. Available at: [https://doi.org/10.1016/S0967-070X\(02\)00048-3](https://doi.org/10.1016/S0967-070X(02)00048-3).
- [20] Elham Alzain, Ali Saleh Alshebami, Theyazn H. H. Aldhyani, and Saleh N. Alsubari. Application of artificial intelligence for predicting real estate prices: The case of saudi arabia. *Electronics*, 11(21):3448, 2022. Available at: <https://doi.org/10.3390/electronics11213448>.
- [21] Wei-Wei Du, Wei-Yao Wang, and Wen-Chih Peng. Dora: Domain-based self-supervised learning framework for low-resource real estate appraisal. In *Proceedings of the 32nd ACM International Conference on Information and Knowledge Management (CIKM)*, 2023. Available at: <https://doi.org/10.1145/3583780.3615470>.
- [22] Dubai Land Department. Instant Sale Feature Using AI-Based Smart Evaluation. Available at: <https://dubailand.gov.ae/en/news-media/dld-launches-instant-sale-feature-using-smart-evaluation-to-simp>
- [23] Saudi Data & AI Authority (SDAIA). National strategy for data & ai. <https://sdaia.gov.sa/en/SDAIA/SdaiaStrategies/Pages/NationalStrategyForDataAndAI.aspx>, 2025. Strategy linked to Vision 2030.
- [24] Real Estate General Authority (REGA). Real estate indicators — rega platforms. <https://rega.gov.sa/en/reg-services/platforms/real-estate-indicators/>, 2025. Digital indicators at Kingdom/city/neighborhood levels with comparative views;.

BIBLIOGRAPHY

- [25] Ministry of Justice (KSA). Real estate market — live transaction feed (service details). <https://www.moj.gov.sa/English/eServices/Pages/ServiceDetailsNew.aspx?itemId=433>, 2025. Lists transaction fields including price per square meter, location, area, and time; last modified 28 Jun 2025;.
- [26] Saudi Authority for Accredited Valuers (TAQEEM). About us. <https://taqeem.gov.sa/en/about-us>, 2025. Regulator of the valuation profession; role, mission, and standards in line with Vision 2030;.
- [27] Quara Holding. Wasalt achieves major growth milestone in saudi arabia’s proptech market. <https://www.quaraholding.com/wasalt-achieves-major-growth-milestone-in-saudi-arabias-proptech>, 2025. Reports 750,000+ registered users and 100,000 monthly leads;.
- [28] TradeArabia. Digital real estate platform wasalt posts significant growth. <https://www.tradearabia.com/News/429985/Digital-real-estate-platform-Wasalt-posts-significant-growth>, 2025. Summarizes AI features incl. instant valuation;.
- [29] Linna Zhu, Michael Neal, and Caitlin Young. Racial disparities in automated valuation models: New evidence using property condition and machine learning. *Cityscape: A Journal of Policy Development and Research*, 26(1):297–320, 2024. Available at: <https://www.huduser.gov/portal/periodicals/cityscape/vol26num1/ch15.pdf>.
- [30] CFPB, OCC, FRB, FDIC, NCUA, FHFA. Quality control standards for automated valuation models (final rule). <https://www.consumerfinance.gov/rules-policy/final-rules/quality-control-standards-for-automated-valuation-models/>, 2024. Final rule published Aug 7, 2024;.
- [31] Anastasios N. Angelopoulos and Stephen Bates. A gentle introduction to conformal prediction and distribution-free uncertainty quantification. *arXiv preprint arXiv:2107.07511*, 2023. Introductory tutorial on conformal prediction for valid prediction intervals without distributional assumptions. Available at: <https://arxiv.org/abs/2107.07511>.
- [32] John W. Tukey. *Exploratory Data Analysis*. Addison-Wesley, 1977. Available at: <https://onlinelibrary.wiley.com/doi/10.1002/bimj.4710230408>.

BIBLIOGRAPHY

- [33] Victoria Hodge and Jim Austin. A survey of outlier detection methodologies. *Artificial Intelligence Review*, 22(2):85–126, 2004. Supports robust outlier handling and winsorization. Available at: <https://doi.org/10.1023/B:AIRE.0000045502.10941.a9>.
- [34] OpenStreetMap contributors. Planet dump [data set]. <https://planet.openstreetmap.org/>, 2025.
- [35] Overture Maps Foundation. Overture maps documentation and data guides. <https://docs.overturemaps.org/>.
- [36] Microsoft Bing Maps Team. Global ml building footprints. <https://github.com/microsoft/GlobalMLBuildingFootprints>. Open dataset of worldwide building polygons.
- [37] Wikipedia contributors. Haversine formula. https://en.wikipedia.org/wiki/Haversine_formula, 2025.
- [38] RW Sinnott. Virtues of the haversine. *Sky and Telescope*, 68:159, 1984. Reference for great-circle (haversine) distance.
- [39] Fabian Pedregosa, Gaël Varoquaux, Alexandre Gramfort, et al. Scikit-learn: Machine learning in python. *Journal of Machine Learning Research*, 12:2825–2830, 2011. Python ML library used for pipelines, preprocessing, and baseline models. Available at: <https://jmlr.org/papers/v12/pedregosa11a.html>.
- [40] Zhe Lim and Anthony Bellotti. Normalized nonconformity measures for automated valuation models. *Expert Systems with Applications*, 180:115165, 2021. Applies conformal prediction to AVMs to produce reliable user-specified prediction intervals (e.g., 80%, 90%, 95%). Available at: <https://doi.org/10.1016/j.eswa.2021.115165>.
- [41] Andrew Krause, Adam Martin, and Michael Fix. Uncertainty in automated valuation models: Error-based versus model-based approaches. *Journal of Property Research*, 37(4):308–339, 2020. Compares AVM prediction-interval methods and how to assess calibration/efficiency. Available at: <https://doi.org/10.1080/09599916.2020.1807587>.
- [42] Arne Johan Pollestad, Arild Brandrud Næss, and Are Oust. Towards a better uncertainty quantification in automated valuation models. *The Journal of Real Estate Finance and Economics*, 2024. Proposes and calibrates uncertainty estimators for non-linear AVMs; uses conformal calibration. Available at: <https://link.springer.com/article/10.1007/s11146-024-10002-7>.

BIBLIOGRAPHY

- [43] Rob J. Hyndman and Anne B. Koehler. Another look at measures of forecast accuracy. *International Journal of Forecasting*, 22(4):679–688, 2006. Discusses issues with MAPE and alternatives (e.g., sMAPE). Available at: <https://doi.org/10.1016/j.ijforecast.2006.03.001>.
- [44] Thomas Davidson. *Mass Appraisal Methods: An Introduction to Real Property Valuation*. Wiley, Hoboken, NJ, 2015. Provides methodological foundations for mass appraisal and quality testing of AVMs.
- [45] International Valuation Standards Council. International valuation standards 2022. Technical report, IVSC, London, 2022. Standards for professional valuation practice.
- [46] Federal Housing Finance Agency (FHFA). Automated valuation model (avm) testing guidelines. Technical report, FHFA, 2018. Establishes standardized practices for testing fairness and accuracy of AVMs.