Data-Driven Office Rental Price Prediction: An Empirical Study of Deep Learning vs. Gradient Boosting

Yixuan Yang

School of Business, Central South University, Changsha, China

1602230114@csu.edu.cn

Abstract. Predicting office rental prices is critical for urban planning and investment strategy. However, valuing commercial real estate remains challenging due to extreme data heterogeneity. Predictive models must seamlessly reconcile rigid structured attributes (e.g., floor area) with unstructured textual information (e.g., location tags). To address this, we introduce a novel multi-modal dataset of commercial office listings in Hangzhou, China, fusing web-scraped transaction records with rich geospatial Point of Interest (POI) metrics via mapping APIs. While recent literature heavily favors multi-modal deep learning for heterogeneous data, our empirical findings challenge this consensus. We demonstrate that Gradient Boosting Decision Trees (GBDT), specifically CatBoost, decisively outperforms both traditional statistical approaches and complex deep learning baselines, including BERT-based multi-modal fusion models. Beyond raw performance, feature importance analysis reveals that specific spatial metrics drive prices far more than generic textual embeddings. Ultimately, we propose a practical real estate analytics workflow, proving that in tabular-dominant tasks, rigorous feature engineering combined with robust tree ensembles effectively beats overparameterized neural networks.

Keywords: Commercial Real Estate · Rental Price Prediction · CatBoost · Heterogeneous Data · Empirical Study.

1 Introduction

Rapid urban economic shifts in China necessitate reliable commercial office rent valuation models for developers, investors, and policymakers [6]. Automated valuation facilitates rational resource allocation and market regulation in expanding metropolitan regions [1,4,13,26,42].

However, valuation remains notoriously difficult due to multifaceted influencing factors. Traditional models often overemphasize internal physical metrics [3], neglecting unstructured contextual data like surrounding transit patterns [19,45,55]. Even when incorporating textual cues, basic methods like TFIDF or simple one-hot encoding fail to capture deep semantic nuances and struggle with high-cardinality features [36,39,43].

Consequently, researchers have increasingly adopted sophisticated Natural Language Processing (NLP) models, evolving from basic embeddings [2,9] to complex Transformers (e.g., BERT, ERNIE) [12,15,41,46,47]. Yet, success on pure NLP benchmarks does not guarantee adaptability to tabular data—the primary format of real estate records. Tabular data remains notoriously challenging for deep neural networks [23]. While generative architectures attempt to bridge this gap [25,48], fine-tuned decision tree ensembles featuring regularized target encoding consistently deliver state-of-the-art predictive accuracy [35].

Deploying over-parameterized neural networks for rent prediction risks severe overfitting. Conversely, traditional gradient boosting models continuously demonstrate robustness across global housing markets [5,24,33]. Seeking an optimal balance—handling high-cardinality heterogeneous data without sacrificing tabular stability—we conduct a data-driven empirical study of the Hangzhou office rental market.

This paper makes three key contributions:

Dataset Construction: We construct a novel dataset ($N = 994$) fusing web-scraped office listings with API-derived spatial POI characteristics.

Methodological Benchmarking: We benchmark CatBoost against multimodal deep learning models, proving the superiority of its categorical encoding over neural alternatives.

Interpretability Analysis: We identify the dominant spatial and textual variables driving local office rents, opening the black box of valuation.

2 Related Work

2.1 Machine Learning in Real Estate and Airbnb Pricing

Algorithmic valuation methodologies are largely rooted in the short-term rental market (e.g., Airbnb). Tree-based ensembles, particularly XGBoost, have emerged as robust industry standards for estimating rents across global hubs [17,20,28, 29]. Researchers have continually refined these baselines by integrating amenity metrics [10,53], spatial analytics [32], and benchmarking against classical models like Support Vector Machines [16,44]. While Large Language Models (LLMs) are currently being explored for residential pricing [8], commercial office valuation demands significantly more customized, multi-modal feature engineering due to strict geographic and financial constraints.

2.2 Advancements and Limitations of Gradient Boosting

Beyond real estate, Gradient Boosting Decision Trees (GBDT) dominate complex tabular tasks, from imbalanced classification [56] to petrophysical forecasting [34] and flood risk mapping [30]. However, standard XGBoost struggles with high-cardinality categorical variables, as one-hot encoding inevitably yields prohibitively sparse memory matrices [37]. To circumvent this, CatBoost processes categorical features

natively via permutation-driven target encoding, effectively bypassing the sparsity trap and suppressing target leakage.

2.3 Deep Generative Models and Tabular Representation

Concurrently, deep learning architectures are increasingly applied to tabular urban modeling. Frameworks rooted in Auto-Encoding Variational Bayes (VAE) [27] are now synthesized for urban profiling [21], housing utility fairness [40], and real estate transaction decoding [22]. In parallel, NLP advancements—from scaling LLMs [52] to robust Chinese embeddings like C-Pack [50] and BGE M3-Embedding [7]—have drastically improved unstructured text processing. To ensure a rigorous empirical evaluation, we deliberately construct a heavy-duty, multi-modal deep learning baseline utilizing these advanced semantic embeddings [11,18], establishing a formidable benchmark against which to test CatBoost's tabular efficiency.

2.4 Data Acquisition and Feature Engineering

To capture the true drivers of Hangzhou's office rental market, we curated over 20 core features affecting commercial valuation, moving beyond superficial metrics.

2.5 Data Collection via Web Scraping and APIs

We scraped transaction records from Beike, a leading Chinese real estate platform, extracting internal physical and financial attributes for each listing (e.g., unit rent, usable area, workstation counts, and floor levels). To contextualize property value externally, we utilized the Gaode Map Web API to map Points of Interest (POIs) within a 1,000-meter radius (e.g., subways, bus stops) and monitored surrounding business density within a 100-meter radius.

2.6 Data Description and Descriptive Statistics

Post-cleaning, our finalized dataset comprises $N = 994$ valid commercial listings, partitioned into training ($N = 815$) and testing ($N = 179$) sets. Combining scraped attributes with API-based metrics yielded a 31-dimensional feature space, retaining 22 as continuous or label-encoded features (Table 1).

Training distributions (Table 2) reflect a highly diverse market. Unit rent averages 2.19 RMB/m^2/day (SD=0.73). Floor space ranges drastically from 15 m^2 micro-offices to 3,661 m^2 corporate floors, ensuring a representative market cross-section. Spatially, transit accessibility is high, with a mean subway distance of 689.86 meters.

Pearson correlations (Fig. 1) reveal a significant structural challenge: severe multicollinearity.

Three feature pairs exhibit aggressive collinearity (> 0.80): *Floor Area* and *Workstations* (0.87; larger spaces hold more desks), *Payment Cycle* and *Min Lease Term*

(0.83; strict landlord lease policies), and *Dist. to Bus* and *Dist. to Parking* (0.91; urban transit agglomeration).

Table 1. Description and Encoding Methods of Key Features

Feature Name	Data Type	Encoding Method / Description
Target Variable		
Unit Rent	Continuous	Daily rent per square meter (RMB/m^2/day).
Internal Physical Attributes		
Floor Area	Continuous	Total floor area in square meters (m^2).
Workstations	Continuous	Number of available workstations.
Floor Level	Ordinal	Encoded as: High (3), Medium (2), Low (1).
Orientation	Nominal	Encoded 1 to 8 (e.g., East, South, West).
Renovation	Ordinal	Fine (2), Simple (1), Roughcast (0).
External Spatial Features (POI)		
Subway POI	Continuous	Density/Distance metrics to nearest subways.
Bus Station POI	Continuous	Density/Distance metrics to nearest bus stops.
Catering POI	Continuous	Surrounding catering facilities metric.

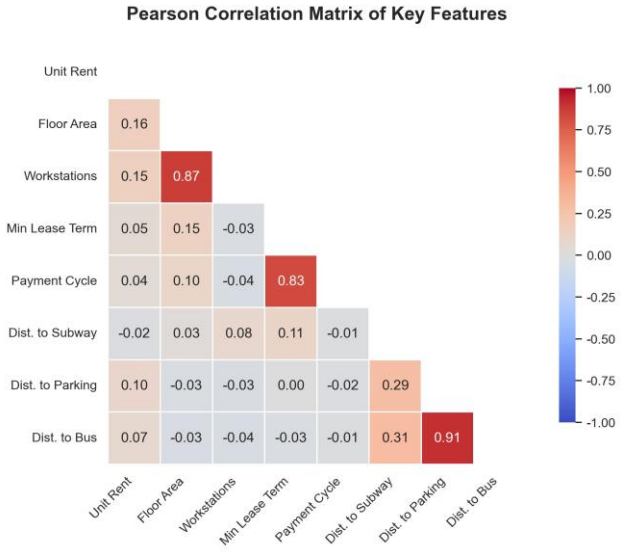

Fig.1. Pearson correlation matrix highlighting severe multicollinearity among spatial features.

Table 2. Descriptive Statistics of Core Numeric Features (Training Set, N=815)

Feature Name	Mean	Std. Dev	Min	Max
Unit Rent (RMB/m^2/day)	**2.19**	**0.73**	**0.90**	**16.00**
Floor Area (m^2)	252.31	291.44	15.00	3661.00
Workstations (Count)	29.49	33.96	2.00	428.00
Payment Cycle (Months)	4.59	1.65	1.00	6.00
Min Lease Term (Months)	19.08	7.40	1.00	24.00
Near Subway Tag (0/1)	0.93	0.25	0.00	1.00
Dist. to Subway POI (m)	689.86	752.60	7.00	2000.00
Dist. to Bus Station POI (m)	220.22	215.12	16.00	2000.00
Dist. to Parking POI (m)	135.42	242.99	6.00	2000.00

Collinearity violates Ordinary Least Squares (OLS) assumptions and destabilizes deep neural networks via exploding gradients. While Principal Component Analysis (PCA) could alleviate this for our DL baselines, it irreversibly destroys the physical interpretability of spatial features critical for actionable real estate insights. This constraint practically necessitates Gradient Boosting Decision Trees (GBDT). Tree-based ensembles inherently resist multicollinearity, dynamically selecting optimal split points without requiring destructive dimensionality reduction.

3 Methodology

Our predictive framework integrates raw multi-source data through rigorous feature engineering into a robust regression engine. Given the inherently tabular nature of real estate records, our primary solution centers on Gradient Boosting Decision Trees (GBDT), specifically **CatBoost**. To empirically validate this, we also constructed a sophisticated multi-modal deep learning architecture (DLCrossAttn) as a stringent comparative baseline.

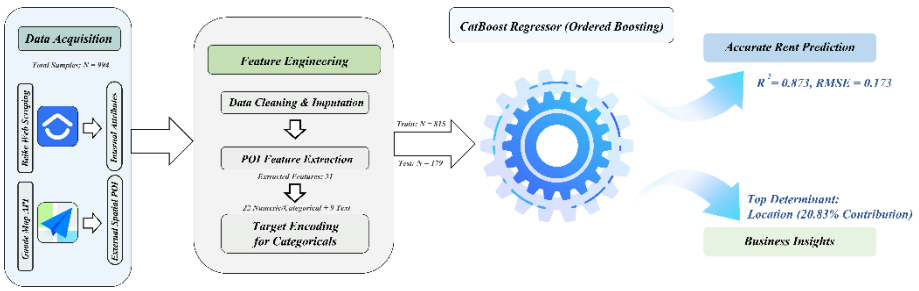

Fig.2. Overall framework integrating scraped attributes and POIs into CatBoost.

3.1 Gradient Boosting on Decision Trees (CatBoost)

CatBoost is exceptionally suited for this task due to its native categorical data architecture. Conventional GBDTs are highly susceptible to target leakage. CatBoost circumvents this via *Ordered Boosting*, actively inhibiting overfitting. For high-cardinality features, it maps raw text to target statistics via regularized Target Encoding:

$$\hat{x}_k^i = \frac{\sum_{j=1}^{n} \mathbb{I}_{\{x_j^i = x_k^i\}} \cdot y_j + a \cdot P}{\sum_{j=1}^{n} \mathbb{I}_{\{x_j^i = x_k^i\}} + a} \tag{1}$$

Here, $\hat{x}_k^i$ represents the numeric replacement of the k-th category of the i-th feature. The indicator $\mathrm{I}_{\{\cdot\}}$ identifies matching categories, y_j is the actual target, P is a global prior preventing rare-category memorization, and a is a smoothing weight.

Crucially, standard encoding computes statistics using the entire dataset, inevitably causing leakage. CatBoost mitigates this fatal flaw via *Ordered Target Encoding*. During training, it generates multiple random permutations and calculates the target statistic relying **only on observed history**—rows placed strictly before it. This chronological simulation ensures predictions remain strictly independent of their own target values, effectively eliminating prediction shift.

3.2 The Multi-Modal Deep Learning Baseline

To rigorously evaluate whether neural networks are requisite for commercial valuation, we engineered a complex multi-modal baseline, DL-CrossAttn (Fig.3).

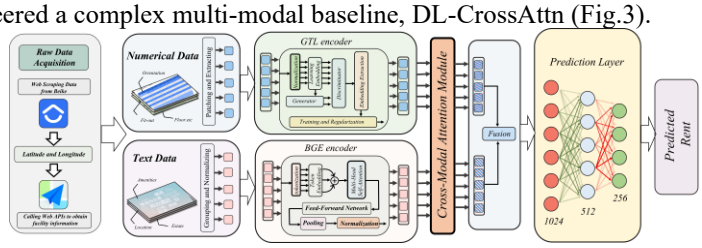

Fig.3. DL-CrossAttn baseline fusing text embeddings with tabular data.

It employs the BGE model to compress unstructured descriptions into dense semantic vectors, while a Generative Tabular Learning (GTL) encoder maps numeric/categorical data into a unified latent space. To reconcile modalities, we implemented a **Cross-Attention mechanism**: numerical embeddings act as Queries, while text embeddings serve as Keys and Values. This scaled dotproduct attention aligns physical statistics with textual narratives, passing the fused representation

through a Multi-Layer Perceptron (MLP) for final prediction. DL-CrossAttn serves strictly as a formidable upper-bound proxy of multimodal trends, utilized to empirically demonstrate that our leaner GBDT approach decisively outperforms complex fusion networks.

3.3 Implementation Details and Training Strategy

Experiments utilized an RTX 4090 GPU, Core i9 CPU, and 64GB RAM. To ensure robust, leak-free evaluation uninfluenced by random seeds, we employed strict 5-fold cross-validation [14,49,51]. To mitigate severe overfitting risks inherent to deep learning [54], aggressive early stopping was implemented [31,38], terminating training if the out-of-fold RMSE failed to improve over 50 successive rounds.

4 Experiments and Results

4.1 Evaluation Metrics and Baselines

We evaluate model performance using Root Mean Squared Error (RMSE) and the Coefficient of Determination (R^2), prioritizing robust prediction of raw RMB/m^2/day values alongside overall variance explanation.

To ensure baseline fairness, gradient boosting models utilized TF-IDF and Truncated SVD for text dimensionality reduction, whereas deep learning baselines natively ingested high-dimensional textual embeddings (e.g., BGE M3). Furthermore, we integrated **TabNet**—a dedicated tabular neural architecture— to represent state-of-the-art deep learning. Table 3 summarizes the 5-fold crossvalidation performance.

Table 3. Average Performance Metrics Across 5 Cross-Validation Folds

Model	**RMSE ↓**	R^2 ↑	**Avg. Time (s)**
TabNet (Tabular DL)	0.254	0.702	18.20
DL-CrossAttn (Baseline)	0.282	0.659	3.86
XGBoost + TFIDF	0.182	0.858	9.45
CatBoost (Proposed)	**0.173**	**0.873**	54.66

4.2 Main Results: Model Comparison

Both neural architectures (DL-CrossAttn R^2 = 0.659; TabNet R^2 = 0.702) lag significantly behind tree-based methods. This highlights how forcing highcardinality tabular data into deep networks obscures optimal decision boundaries. Conversely, **CatBoost** dominates with an R^2 of **0.873** and minimum RMSE (**0.173**).

Beyond averages, fold-by-fold stability (Fig. 4) reveals the DL baselines' high sensitivity to training subsets, whereas CatBoost maintains unwavering robustness, credited to its permutation-driven *Ordered Boosting* engine.

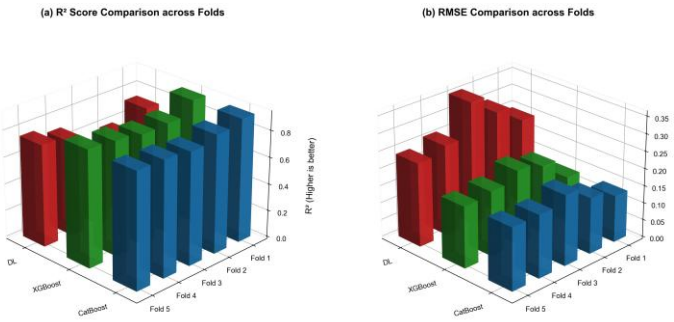

Fig.4. 3D performance comparison across 5 cross-validation folds. CatBoost demonstrates clear superiority and stability in R^2 and RMSE.

Efficiency Trade-offs: While GPU-accelerated DL-CrossAttn trains rapidly (3.86s), CatBoost requires 54.66s due to dynamic target coding. However, for commercial valuation—updated weekly or monthly—exchanging less than a minute of training for a 32% R^2 improvement is a highly practical business trade-off. Furthermore, the CatBoost optimal split (Fold 5) demonstrates exceptional predictive alignment along the diagonal (Fig. 5).

4.3 Feature Importance and Business Insights

Feature importance extraction from the optimal CatBoost model (Fig. 6) reveals key valuation drivers: **1. Location Dominance:** *Location* (20.83%) and precise geographic coordinates form the valuation baseline. **2. Brand over Space:** High-cardinality textual features like *Building Name* (8.44%) heavily influence price, proving corporate prestige often overrides raw floor area metrics. **3. Commuter Premium:** High reliance on *Distance to Parking* (9.24%) and *Bus Station* (6.07%) quantifies the commercial premium placed on accessibility.

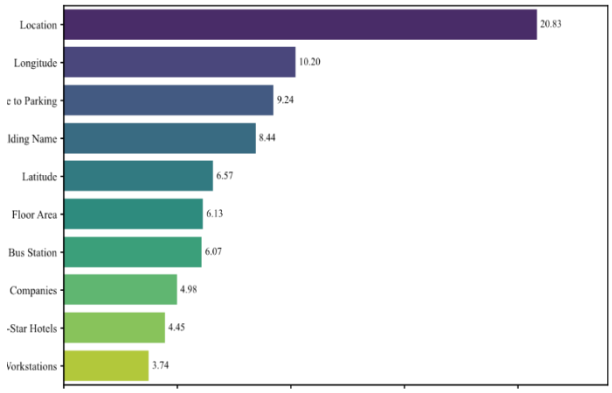

Fig.5. Scatter plot of Actual vs. Predicted rental prices (Fold 5), demonstrating exceptional predictive alignment.

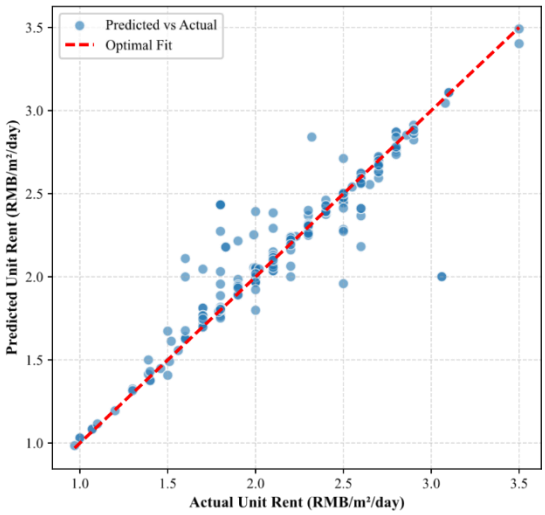

Fig.6. Top 10 feature importances (Fold 5). Spatial coordinates, POI proximity, and unstructured textual attributes dominate predictive contributions.

4.4 Ablation Analysis on Feature Modalities

An ablation study validates our multi-source strategy. Removing unstructured text (*Building Name, Location, Tags*) severely degraded R^2 by ~0.12 [3, 55]. Similarly, excluding external POI metrics increased RMSE by > 18% [10,17,20]. This establishes that relying solely on internal physical attributes is inadequate; robust commercial pricing strictly requires the fusion of tabular, spatial, and textual modalities [8,19].

5 Discussion, Limitations, and Future Work

Despite the prevailing trend of applying deep generative models and LLMs to tabular data [18,21,22], our empirical findings strongly challenge this paradigm. For structured commercial real estate valuation, robust tree ensembles utilizing rigorous categorical encoding decisively outperform over-parameterized neural networks [23].

However, we must acknowledge critical experimental limitations. First, deep neural networks are notoriously "data-hungry." The underperformance of our DL baselines is partially attributable to our medium-scale dataset (N = 994), which restricted generalized representation learning and induced overfitting. Conversely, CatBoost demonstrated exceptional data efficiency optimal for localized urban constraints. Second, our cross-sectional data intrinsically omits macroeconomic temporal cycles.

Furthermore, our API-derived POIs strictly measure spatial proximity, disregarding the operational capacity or quality of these facilities.

Future research will proceed along two axes. To circumvent neural datahunger, subsequent work should explore LLM-driven tabular augmentation and Generative Adversarial Networks (GANs) [27,40] to synthesize massive arrays of high-fidelity commercial records [52]. Simultaneously, incorporating time-series transaction data into recurrent architectures will awaken the model to temporal market fluctuations, enabling dynamic rent forecasting.

Declaration of AI Usage. The author utilized Large Language Models solely for language polishing. The author assumes full responsibility for all content, experimental design, and analytical conclusions.

References

1. Abdi, H.: Z-scores. In: Encyclopedia of measurement and statistics, vol. 3, pp. 1055–1058. SAGE Publications (2007)
2. Athiwaratkun, B., et al.: Probabilistic fasttext for multi-sense word embeddings. arXiv:1806.02901 (2018)
3. Baur, K., et al.: Automated real estate valuation with machine learning models. Expert Syst. Appl. **213**, 119147 (2023)
4. Buranasiri, J., Laokulrach, M.: Rental price prediction using ann approach (2025)
5. Caroni, C.: Rental price prediction using machine learning (2022)
6. Chen, G., et al.: Introducing chinese private enterprise survey. Nankai Bus. Rev. Int. **10**(4), 501–525 (2019)
7. Chen, J., et al.: Bge m3-embedding: Multi-lingual text embeddings. arXiv:2402.03216 (2024)
8. Chen, T., Si, S.: Predicting rental price with machine learning and llms. arXiv:2405.17505 (2024)
9. Church, K.W.: Word2vec. Nat. Lang. Eng. **23**(1), 155–162 (2017)
10. Costa, J.R.G.A.: Predicting Airbnb prices using machine learning algorithms. Master's thesis, ISEG (2025)
11. Cui, L., et al.: Tabular data augmentation for machine learning. arXiv:2407.21523 (2024)
12. Devlin, J., et al.: Bert: Pre-training of deep bidirectional transformers. In: NAACL. pp. 4171–4186 (2019)
13. Filippova, O., et al.: Pricing office rents in sydney cbd. Prop. Manag. **40**, 230–246 (2022)
14. Fushiki, T.: Estimation of prediction error by using k-fold cv. Stat. Comput. **21**, 137–146 (2011)
15. Gao, T., et al.: Simcse: Simple contrastive learning of sentence embeddings. arXiv:2104.08821 (2021)
16. Gao Jr, Y.: Unravelling airbnb price prediction based on machine learning (2025)
17. Ghosh, I., et al.: An ensemble machine learning framework for airbnb pricing. Int. J. Contemp. Hosp. Manag. **35**, 3592–3611 (2023)
18. Gu, R., Lin, L.: Application of lda and autoencoder to real estate datasets. J. Supercomput. **81**, 118 (2025)
19. Heidari, M., Rafatirad, S.: Semantic cnn model for safe business investment by using bert. In: SNAMS. pp. 1–6 (2020)
20. Islam, M.D., et al.: Airbnb rental price modeling based on lda and mesf-xgboost. Mach. Learn. Appl. **7**, 100208 (2022)

21. Johnsen, M., et al.: Population synthesis for urban resident modeling. Neural Comput. Appl. **34**, 4677–4692 (2022)
22. de Jong, B.: Modeling tabular real estate data with deep generative models (2023)
23. Kadra, A., et al.: Well-tuned simple nets excel on tabular datasets. In: NeurIPS. vol. 34, pp. 23928–23941 (2021)
24. Karanko, L.: Rental Price Prediction on Greater Helsinki Apartments. Master's thesis, Univ. of Helsinki (2022)
25. Kim, D.K., et al.: Generative models for tabular data: A review. J. Mech. Sci. Technol. **38**, 4989–5005 (2024)
26. Kim, T.Y., et al.: Determinants of housing rental prices in seoul: explainable ai. Spatial Econ. Anal. **20**, 312–332 (2025)
27. Kingma, D.P., Welling, M.: Auto-encoding variational bayes. arXiv:1312.6114 (2013)
28. Lektorov, A., et al.: Airbnb rental price prediction using machine learning. In: IEEE CCWC. pp. 0339–0344 (2023)
29. Liu, Y.: Airbnb pricing based on statistical machine learning models. In: CONFSPML. pp. 175–185 (2021)
30. Ma, M., et al.: Xgboost-based method for flash flood risk assessment. J. Hydrol. **598**, 126382 (2021)
31. Mahsereci, M., et al.: Early stopping without a validation set. arXiv:1703.09580 (2017)
32. Medpalliwar, A., et al.: Airbnb rental price prediction using ml techniques. In: AIP Conf. Proc. vol. 3233 (2025)
33. Ming, Y., et al.: Prediction and analysis of housing rent based on xgboost. In: Int. Conf. Big Data Tech. pp. 1–5 (2020)
34. Pan, S., et al.: An optimized xgboost method for predicting reservoir porosity. J. Pet. Sci. Eng. **208**, 109520 (2022)
35. Pargent, F., et al.: Regularized target encoding outperforms traditional methods. Comput. Stat. **37**(5), 2671–2692 (2022)
36. Poslavskaya, E., Korolev, A.: Encoding categorical data: Is there yet anything 'hotter' than one-hot encoding? arXiv:2312.16930 (2023)
37. Prates, P., et al.: Limitations of xgboost in predicting material parameters. In: MATEC Web of Conf. vol. 408 (2025)
38. Prechelt, L.: Automatic early stopping using cross validation. Neural Networks **11**, 761–767 (1998)
39. Qaiser, S., Ali, R.: Text mining: use of tf-idf. Int. J. Comput. Appl. **181**(1), 25–29 (2018)
40. Qu, T.: Optimizing urban housing utility fairness via vae-gan models (2023)
41. Reimers, N., Gurevych, I.: Sentence-bert: Sentence embeddings using siamese bertnetworks. arXiv:1908.10084 (2019)
42. Senthilkumar, V.: Enhancing house rental price prediction models (2023)
43. Shantal, M., et al.: A novel approach for data feature weighting. Symmetry **15**(12), 2185 (2023)
44. Srianomai, S., et al.: Predicting prices of airbnb accommodations by svm and xgboost. Prog. Appl. Sci. Tech. **14**, 16–23 (2024)
45. Sun, Y., et al.: Ernie: Enhanced representation through knowledge integration. arXiv:1904.09223 (2019)
46. Sun, Y., et al.: Ernie 2.0: A continual pre-training framework. In: AAAI. vol. 34, pp. 8968–8975 (2020)
47. Sun, Y., et al.: Ernie 3.0: Large-scale knowledge enhanced pre-training. arXiv:2107.02137 (2021)

48. Wen, X., et al.: From supervised to generative: tabular deep learning with llms. In: KDD. pp. 3323–3333 (2024)
49. Wong, T.T., Yeh, P.Y.: Reliable accuracy estimates from k-fold cross validation. IEEE Trans. Knowl. Data Eng. **32**, 1586–1594 (2019)
50. Xiao, S., et al.: C-pack: Packed resources for general chinese embeddings. In: SIGIR. pp. 641–649 (2024)
51. Yadav, S., Shukla, S.: Analysis of k-fold cross-validation on colossal datasets. In: IEEE IACC. pp. 78–83 (2016)
52. Yang, Y., et al.: Lesa: Learnable llm layer scaling-up. arXiv:2502.13794 (2025)
53. Yao, S.: Use xgboost to predict the rental based on airbnb data (2022)
54. Yao, Y., et al.: On early stopping in gradient descent learning. Constr. Approx. **26**, 289–315 (2007)
55. Zhang, H., et al.: House price prediction with textual description data. Nat. Lang. Eng. **30**, 661–695 (2024)
56. Zhang, P., et al.: Research and application of xgboost in imbalanced data. Int. J. Distrib. Sens. Netw. **18** (2022)