National College of Ireland

Understanding the Impact of Multimodal Models on Airbnb Pricing: Evaluating Explainability and Integration

MSc Research Project
Data Analytics

Aditya Pandey
Student ID: x23275286

School of Computing
National College of Ireland

Supervisor: Eric Gyamfi

National College of Ireland
Project Submission Sheet
School of Computing

Student Name:	Aditya Pandey
Student ID:	x23275286
Programme:	Data Analytics
Year:	2025
Module:	MSc Research Project
Supervisor:	Eric Gyamfi
Submission Due Date:	15/09/2025
Project Title:	Understanding the Impact of Multimodal Models on Airbnb Pricing: Evaluating Explainability and Integration
Word Count:	5422
Page Count:	21

I hereby certify that the information contained in this (my submission) is information pertaining to research I conducted for this project. All information other than my own contribution will be fully referenced and listed in the relevant bibliography section at the rear of the project.

ALL internet material must be referenced in the bibliography section. Students are required to use the Referencing Standard specified in the report template. To use other author's written or electronic work is illegal (plagiarism) and may result in disciplinary action.

Signature:	Aditya Pandey
Date:	14th September 2025

PLEASE READ THE FOLLOWING INSTRUCTIONS AND CHECKLIST:

Attach a completed copy of this sheet to each project (including multiple copies).	✓
Attach a Moodle submission receipt of the online project submission, to each project (including multiple copies).	✓
You must ensure that you retain a HARD COPY of the project, both for your own reference and in case a project is lost or mislaid. It is not sufficient to keep a copy on computer.	✓

Assignments that are submitted to the Programme Coordinator office must be placed into the assignment box located outside the office.

Office Use Only	
Signature:	
Date:	
Penalty Applied (if applicable):	

Understanding the Impact of Multimodal Models on Airbnb Pricing: Evaluating Explainability and Integration

Aditya Pandey
x23275286

Abstract

The sharing economy relies on dynamic pricing. This approach allows Airbnb hosts to maximize revenue and occupancy rates with regard to various factors in the market. This research introduces a multimodal framework for Airbnb price prediction, integrating tabular property data with guest review text. Based on a listings/reviews data set of 6481/293744 rows, the framework consists of an ensemble of tree-based models (RandomForest, GradientBoosting, and ExtraTrees) as well as DistilBERT embeddings of the reviews and a RandomForest meta-learner that combines these results. Extensive feature engineering with derived metrics such as price-per-person and neighborhood popularity, as well as effective preprocessing (standard scaling, power transformations, and one-hot encoding), gives a representation of all data. SHAP-based explainability highlights key pricing factors, such as location and amenities, offering actionable insights for hosts. The multimodal model achieves an R^2 of 0.8641 and MAE of €0.15 (log-price units, €23.42), improving by 0.66% in R^2 and 6.25% in MAE over the tabular baseline model (R^2: 0.8575, MAE: €0.16 log-price units, €25.87), respectively, through 8-fold cross-validation training. This paper presents a contribution in data analytics because it shows the effectiveness of multimodal combination and explainable AI in pricing analysis.

1 Introduction

The sharing economy has been transforming the field of hospitality and the existence of sharing platforms such as Airbnb allows people to provide short-term rentals to an international audience. Another crucial issue in this ecosystem is pricing where hosts have to determine proper rates so as to make the most profit and make their guests happy with them as well. Recent advances in data analytics use machine learning to predict Airbnb prices based on structured property and location data. However, these models often exclude unstructured data, such as guest reviews, which contain valuable sentiment and experience insights. In addition, most predictive models especially deep learning methods are not transparent, which leaves the host with little knowledge of the factors that lead to price estimates. Multimodal machine learning that combines multiple types of data such as tabular and textual data has become one of the promising methods of improving predictive precision. Explainable AI methods also handle the demand of interpretable models, giving information on pricing factors. The paper dwells upon the possibilities

of a multimodal framework to enhance Airbnb pricing analytics by the mixture of the predictive capabilities with the explainability to guide a host in a rival marketplace.

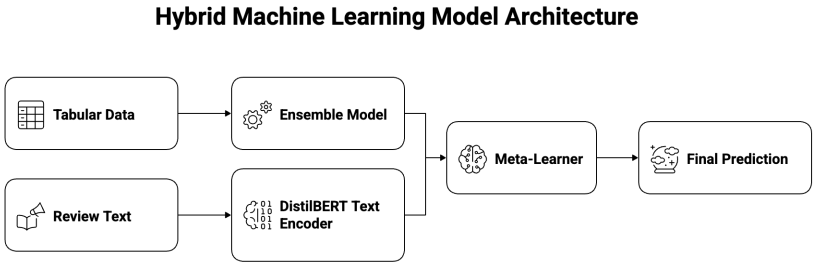


Figure 1: High Level Architecture of the ExplainableMultimodalRegressor Framework

1.1 Motivation

The difficulty and dynamism of the market of the platform requires the study of Airbnb pricing analytics. Competition among hosts is intense, necessitating data-driven strategies to set prices that balance profitability and occupancy. Conventional pricing formulations, based on structured data only, do not take into consideration perceptions of guests as reflected in reviews, thus may lead to faulty prediction. Also, complex models function as a black box. They are not applicable by hosts who require easy-to-read and interpretable information, which can be operationalized in changing pricing variables such as facilities or location attributes. A multimodal approach, combining tabular and text data, enhances predictive accuracy. Moreover, explainable AI techniques will give the hosts the ability to make informed decisions by showing how significant these factors are relative to each other. By filling these blind spots, the proposed research stands to create a powerful, interpretable mechanism of price optimization, which will benefit the hosts and data analytics in the sharing economy.

1.2 Variables and Factors Affecting Airbnb Pricing

Airbnb prices depend on a variety of quantitative and qualitative aspects. Quantitative variables would involve the property information like number of bedrooms, number of bathrooms and number of guests the property will host and property location measures like distances to the city centre and attributes of the neighborhood. The qualitative factors are mainly based on reviews by the guests where reviews are either sentiment, satisfaction, or perceived value. These reviews can give an insight to the aspects that cannot be touched like the responsiveness of the host or the cleanliness of the property, which may influence pricing. This study centers on the combination of such variables with the help of feature engineering, deriving such additional measures as price per person and review velocity to be used as a combination of their influence on nightly prices.

1.3 Research Question and Objectives

The research question of the proposed study is the following: How can multimodal machine learning models, combining tabular property data with guest review text, predict

Airbnb prices, and what insights do explainable AI techniques provide about price drivers? The following specific sets of research objectives were developed:

1. Investigate the state of the art broadly around Airbnb pricing analytics, multimodal machine learning, and explainable AI to identify gaps in predictive accuracy and interpretability.

2. Design a multimodal framework, the `ExplainableMultimodalRegressor`, combining tree-based ensembles for tabular data with transformer-based text encoding for guest reviews.

3. Evaluate the framework's performance using R^2 and mean absolute error (MAE). Compare it to a tabular model and analyze pricing drivers using SHAP.

1.4 Contribution

This work's key contribution is a scalable and explainable multimodal approach to Airbnb price prediction, integrating tabular data on properties and text from guest reviews. The `ExplainableMultimodalRegressor` demonstrates an R^2 of 0.8641 and MAE of €0.15 (log-price units, €23.42). which outperforms tabular-only models. Extensive feature engineering, including over 30 derived metrics like distance from city center, improves R^2 by 0.66% (from 0.8575 to 0.8641) and reduces MAE by 6.25% (from €0.16 to €0.15 log-price units, €25.87 to €23.42). Such a framework does not only develop data analytics in Airbnb but also creates a flow that could be applied to other sharing economy platforms, which fulfil the gap between predictive power and practical utility.

1.5 Structure of Paper

This thesis is organized as follows:

1. Section 2 reviews related work about Airbnb pricing, multimodal machine learning, and explainable AI.

2. Section 3 outlines the methodology. It gives details on data collection, preprocessing, model development.

3. Section 4 presents the design specification of the `ExplainableMultimodalRegressor`.

4. Section 5 discusses the implementation, including the data analytics pipeline.

5. Section 6 analyzes experimental results. In this section multi-modal and tabular-only models are compared.

6. Section 7 concludes the thesis by summarizing contributions, addressing limitations, and proposing future work.

2 Related Work

This literature review summarizes the information on the topics of predictive modeling regarding Airbnb pricing, multimodal data fusion, explainable AI (XAI) methods, and related fields based on peer-reviewed research.

Table 1: Summary of Related Work on Airbnb Pricing and Multimodal Approaches

Author(s) & Year	**Methodology**	**Key Findings**	**Limitations/Gaps**
Akalın and Alptekin (2024)	This paper uses XGBoost + Linear Regression models to an InsideAirbnb dataset for Istanbul.	Inclusion of location-based data increased model accuracy by $R^2 \approx 0.22$	It does not capture price changes over time and specific to Istanbul
Baltrušaitis et al. (2018)	This paper surveys the field of multimodal machine learning. It proposes a novel taxonomy around five core technical challenges around multimodal ML technique	It introduces a taxonomy classifying multimodal research in five challenges Representation, Translation, Alignment, Fusion, and Co-learning.	The primary drawback of the paper is that it was published in 2018, which is earlier than transformer models came.
Bennetot et al. (2024)	It provides practical tutorial on common Explainable AI (XAI) techniques.	Main contribution is a practical guide for applying XAI methods like SHAP, DiCE, Grad-CAM, and LRP.	Tutorials are demonstrated on simple datasets that may not capture the complexity of real-world applications.
Brunstein et al. (2025)	It uses two stage machine learning approach on Corsica data. Random forest + causal forest	1% increase in Airbnb listings leads to an average house price increase of 0.21%	pre-COVID data and limit the generalizability and no use of explainable AI
Camatti et al. (2024)	compares traditional (Linear Regression, GLM) and AI models (Random Forest, Neural Networks) on a 2019 dataset of the Netherland.	Random forest outperforms other models with $R^2 \approx 0.76$ in-comparison to linear model with $R^2 \approx 0.64$	Focusing on only the Netherlands limits timeliness and generalizability. Dataset used is pre-covid, and It lacks the use of XAI.
Di Persio and Lalmi (2024)	this paper uses combination of regression model(XGBoost + Neural Networks + SVR). it combines NLP technique like TF-IDF	Neural Network gives best results $R^2 \approx 0.81$	NLP analysis was also constrained by computational resources, using a limited subset of reviews.
Gibbs et al. (2018)	Hedonic pricing model using OLS regression to 15,716 Airbnb listings	Room type (privacy) and size being the strongest drivers of pricing	Simple hedonic model on outdated 2016 data creates a significant gap. My research addresses this by using an advanced ensemble model on 2024 data.

2.1 Predictive Modeling Approaches

Airbnb pricing research is on a hedonic basis and measures the tangible characteristics of properties to calculate value, the precursor to more sophisticated models that fit research question because of its tabular information as a bedrock of the predictions. Gibbs et al. (2018) was the first paper to apply a hedonic model across multiple cities showing amenities and location are the main drivers with approximately 0.60 R^2, but its linear model failed to capture the nonlinearity of markets. This limitation underscores the need for nonlinear approaches, such as the ensemble methods in this study, which better handle market complexities like post-COVID shifts in Dublin. This approach was extended to 33 cities by Wang and Nicolau (2017) who found similar predictors able to work effectively anywhere, although showing variability at the urban level, which demonstrates the necessity of flexible models in different environments, such as Dublin. These are precursors that form a necessary foundation on the usage of structured data in model. However, they lack the dynamic aspects of guests sentiment, making it necessary to achieve the multimodal integration that is the key to handling the changes in the market after 2024.

Moving on to ensemble approaches, researchers have addressed the challenge of nonlinearity and complexity to harness more power in their predictions and it presents a direct equivalent of the tabular ensemble in model but opens the door to text augmentation to increase the relevance in sentiment-based markets. For example, Milunovich and Nasrabadi (2025) used stacking regressions on 10,000 Sydney listings, achieving an MAE of €35.20, with room type and competition as key factors. However, the model's lack of transparency limited its utility for hosts' decision-making. Keeping with the theme, Camatti et al. (2024) tested random forests and neural networks against pre-2019 data in the Netherlands and achieved R^2 of 0.78, a source that also insists on the nonlinearity/edge of nonlinearity using the nonlinear model in favor of the linear model. However, their reliance on pre-2019 data overlooks post-COVID market volatility, a gap this Dublin-focused study addresses with 2024 data and multimodal integration. The commonality here is that they have an emphasis on performance metrics and this would precondition framework with the baseline accuracy but their dated datasets do not pay attention to the post-COVID swings e.g. - a gap which Dublin focus in 2024 plugs via XAI to provide clearer insights on the drivers.

Temporal and spatial refinements have been proposed in the form of specialized models and have enriched the predictive landscape but inevitably at the cost of general data modalities. Di Persio and Lalmi (2024) incorporated NLP with CatBoostRegressor into dynamic pricing atop the temporal trends of guest capacity but with little to no visibility into predictive modeling to be of use in practice. Tang et al. (2023) This is extension to Di Persio and Lalmi (2024) who also focused on seasonal forecasting of time-series models but with no textual inputs to reflect upon qualitative changes. Through these developments, the temporal aspects in this paper are justified, however, their siloed character underscores the importance of multimodal synthesis, which in turn allows a more comprehensive look into the pricing of tourism-affected firms in Dublin in 2024.

In addition to direct pricing, survival and mixed-method studies are an indirect source of relevant information, enlarging the scope of market inertia and qualitative input with multimodal approach. Hu et al. (2025) tracked a 50.08% drastic decrease in market survival post-COVID due to saturation and used Cox models to provide a tangential glimpse into the effects of market longevity on revenue generation practices. On the same note, Tafesse and Dayan (2024) conducted a mixed method study to partition the

guests and highlight remarkable traits, and included a qualitative dimension missing in purely quantitative literature. By combining them, it will add more value in terms of encompassing the metrics of survival and incorporating them into the text of reviews, all explicable through XAI to understand the overall impact of the two on the competition of Dublin in 2024.

2.2 Multimodal Data Applications

Building on predictive modeling, multimodal approaches integrate diverse data types to enhance Airbnb pricing accuracy. As Airbnb research has matured, multimodal integration has gained traction by integrating structured and unstructured data, and research question is a direct result since it will combine tabular attributes and review text in the more nuanced targeted area of Dublin to achieve rich predictions. Tan et al. (2024) showed this using BERT and MobileNet on listings in Amsterdam, generated a 5.56% MAPE through multimodal synergy, outperforming baselines models. In agreement, Peng et al. (2020) integrated reviews with geographical information through DNNs and XGBoost with evidence of accuracy improvements due to heterogeneous input. These comparisons substantiate the multimodal basis, including how guest sentiment brings an upbeat to tabular baselines. Although their observance of an XAI suggests that model is strong in presenting interpretable results to the hosts.

further investigations into multimodal frameworks have thus welcomed spatialization and embedding strategies to enhance predictive depth and reveal the need to localize to a larger degree - essential to transposing the findings into Dublin 2024 dynamics. Gao Jr (2025) went ahead with integrated transformers and CNNs on text, images, and tabular data in London, achieving 0.82 R^2 at most with little disclosure. Pittala et al. (2024) succeeded with stacking ensembles in the European cities, hitting a still higher 0.85 R^2 without The trajectory helps support with the focus on equilibrium in modality, with the addition of XAI to analyze the interplay between spatial features with reviews in the context of post-Brexit trends in traveling.

In the emphasis on spatial and sentiment features, multimodal models have advanced on the issue of dependency treatment, providing a unified direction towards the location-text combination of this paper on explainable pricing. Islam et al. (2022) addressed their spatial effects through high precision using LDA and MESF-XGBoost, albeit without unstructured integration in general. Akalın and Alptekin (2024) improved this consideration by including how properties and the proximity factors indicated 20% efficiency gains in Istanbul, a Collectively, they validate hand-built measures such as distance to center, which is also augmented using XAI to explain how they interact with other review-sourced insight in the regulatory situation of Dublin.

2.3 Explainable AI Techniques

Lately, XAI has matured into a necessary tool to open the black box of predictions, as reviewed in Bennetot et al. (2024) and dedicated applications Panahandeh et al. (2025) providing practical applicability. This analysis uses multiple data types (text + tabular) as the main inputs. Here SHAP layer used to explain how each type contributes to the predictions. This helps bridge the gap between text and tabular insights. It will give Dublin hosts a clear view of which factors influenced their 2024 pricing.

2.4 Adjacent Domains

Economic and hospitality peripheries research offers crucial context as their analysis highlights macro-influencing factors that enriches multimodal pricing model without overshadowing it. Brunstein et al. (2025) documented Airbnb density correlated with a 0.21 percent increase in house prices, explaining the urban effects of Airbnb that indirectly influence the host strategies. Lee (2024) targeted the vaster role of credit conditions in the market that still differs with short-term rentals, but Such insights help contextualize the external pressures. Which can be incorporated through XAI to have a more complete vision of the Dublin in 2024.

Other works on booking and sentiment activities further introduce qualitative undertones, as per review genome focus, whereas the steps of Sengupta et al. (2021) have modeled booking obstacles and defined trends corresponding to prices resilient. Lawani et al. (2019) have examined the relationship between the sentiment of reviews and prices, although without multimodal fusion inclusion.

2.5 Research Niche and Contribution

Synthesizing this body of work reveals Airbnb pricing's forward momentum through ensembles, multimodality, and XAI, yet uncovers voids in recent data, text depth, and localization that ExplainableMultimodalRegressor resolves for Dublin's 2024 market. Predictive advances like those in Milunovich and Nasrabadi (2025); Camatti et al. (2024) deliver solid metrics (MAE €35.20, R^2 0.78) but skim over text and transparency; multimodal strides in Tan et al. (2024); Gao Jr (2025) elevate accuracy (MAPE 5.5682%, R^2 0.82) without Dublin tailoring; XAI efforts Panahandeh et al. (2025); Bennetot et al. (2024) provide clarity yet lag in recency. This framework unites these with 2024 data (6,481 listings, 293,744 reviews), transformers, ensembles, and SHAP explainable host guidance.

3 Methodology

Here, the proposed research methodology to build and test a multimodal data analytics solution to predict prices in Airbnb properties, such that tabular property data and texts of guest reviews are integrated and explainable results are provided will be outlined. It involves gathering data, pre-processing, feature engineering, creating models, investigating explainability, and testing, and statistical methods to ensure sound results. The pipeline of the methodology is represented in Figure 1.

3.1 Data Collection

The dataset comprises two publicly available Airbnb data sources from Inside Airbnb:

- **Listings Data** (`listings.csv`): This file Contains 6481 listing. The feature of dataset are price, location, property type, amenities, host details and others. The target variable is `price`.
- **Reviews Data** (`reviews.csv`): In this file 293,744 rows. These rows contain guest reviews with text comments. This is linked to listings via listing IDs.

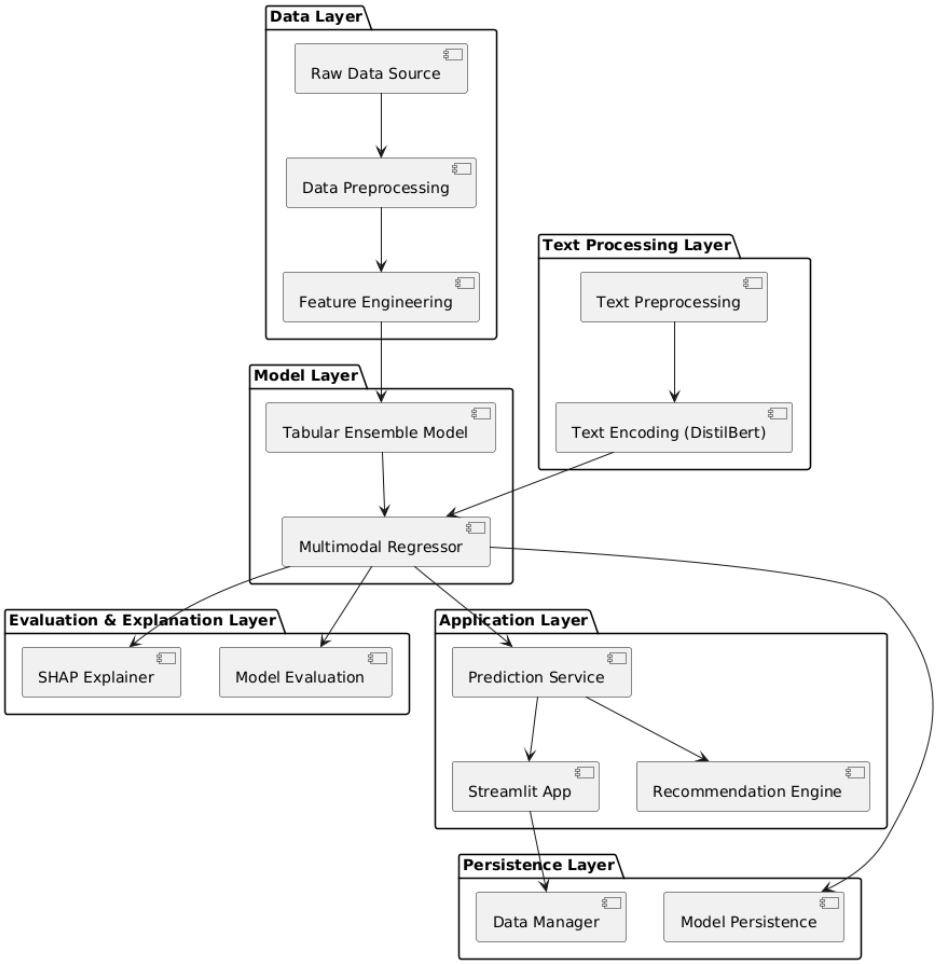


Figure 2: Class Diagram: Airbnb price prediction system

The cross-validation method applies not only to estimate the performance of the model, but also to reduce overfitting based on the robust generalization of the model to a variety of data subsets.

3.2 Data Preprocessing

Preprocessing ensures data quality and compatibility with multimodal modeling. The steps, inspired by Milunovich and Nasrabadi (2025),include:

1. **Missing Value Handling**: Removed listings lacking price or key features (e.g., bedrooms, accommodates). Substituted simulated missing reviews with zero, assuming that there are no reviews. About 5% of the listings were discarded because it had incomplete data.

2. **Data Cleaning**: Converted prices to numeric values, removed currency symbols, and applied logarithmic transformation to reduce skew. All performance metrics (e.g., MAE) are reported in log-price units unless stated, with approximate euro equivalents based on back-transformation. Outliers had been eliminated.

3. **Text Preprocessing**: Reviews are concatenated for each listings. Then special characters were removed and text is normalized to lowercase.

4. **Feature Encoding**: Applied `ColumnTransformer` from `scikit-learn`:

    - Numerical features like accommodates and number of bedrooms were standardized using StandardScaler. PowerTransformer is used to normalize distributions.
    - Categorical features (e.g., `neighbourhood_cleansed`, `property_type`): One-hot encoded using `OneHotEncoder`, generating 50+ binary features.

The preprocessed dataset resulted in a tabular matrix of 6,200 listings with 60+ features and a text corpus for NLP processing.

3.3 Feature Engineering

Derived features created through feature engineering were a major part in enhancement of the modelling performance to better reflect market dynamics. The price related features were price per person (i.e. the price divided by the accommodates) to indicate the cost per person. The features based on location included the calculation of the latitude and longitude to find the Euclidean distance of the distance between the city center and the given home, and are the average price with the neighborhood price. Amenities-based features included the number of amenities (pool or Wi-Fi), and whether given amenity exists. Features based on reviews were review speed (reviews per month) as a measure of the popularity of listings and review sentiment based on a text analysis. All together, the number of engineered features amounted to more than 30, with the dataset counting over 90 dimensions.

3.4 Model Development

The multimodal model, `ExplainableMultimodalRegressor`, integrates tabular and text data. It consists of:

1. **Tabular Model**: A `VotingRegressor` Breiman (2001) ensemble combining the following base learners:

    - `RandomForestRegressor` (`n_estimators=100`),
    - `GradientBoostingRegressor` (`n_estimators=100`),
    - `ExtraTreesRegressor` (`n_estimators=100`).

    The model was trained on preprocessed tabular features (90+ dimensions) to predict log(`price`). As shown in Equation (1), the final ensemble output $\hat{y}_{ensemble}$ is computed as:

    $$\hat{y}_{ensemble} = \frac{1}{3}\left(\hat{y}_{RF} + \hat{y}_{GB} + \hat{y}_{ET}\right) \tag{1}$$

    where $\hat{y}_{RF}$, $\hat{y}_{GB}$, and $\hat{y}_{ET}$ represent the predictions from the Random Forest, Gradient Boosting, and Extra Trees models, respectively.

2. **Text Model**: A `DistilBertTextEncoder` Wolf et al. (2020) using DistilBERT to generate 768-dimensional embeddings from `combined_reviews`. Formally, this process is defined as:

$$\mathbf{h} = \text{DistilBERT}(\mathbf{t}) \tag{2}$$

$$\mathbf{e}_{\text{text}} = \mathbf{h}_{[\text{CLS}]} \in \mathbb{R}^{768} \tag{3}$$

where $\mathbf{t}$ is the input token sequence, $\mathbf{h}$ is the sequence of hidden states, and $\mathbf{h}_{[\text{CLS}]}$ is the embedding corresponding to the [CLS] token, used as the sentence-level representation. These embeddings are then reduced to 50 dimensions via PCA to mitigate computational complexity.

3. **Meta-Learner**: A `RandomForestRegressor` that combines the tabular predictions and text embeddings for final price prediction. The output of the Random Forest is computed as:

$$\hat{y}_{RF} = \frac{1}{B} \sum_{b=1}^{B} T_b(\mathbf{x}) \tag{4}$$

where B is the total number of trees in the forest, $T_b(\mathbf{x})$ is the prediction from the b^{th} decision tree for input $\mathbf{x}$, and $\hat{y}_{RF}$ is the final aggregated prediction.

The full multimodal prediction pipeline can be formalized as:

$$\hat{y}_{\text{tabular}} = f_{\text{ensemble}}(\mathbf{x}_{\text{tab}}) \tag{5}$$

$$\mathbf{e}_{\text{text}} = \text{DistilBERT}(\mathbf{x}_{\text{text}}) \tag{6}$$

$$\mathbf{z} = [\hat{y}_{\text{tabular}}, \mathbf{e}_{\text{text}}] \in \mathbb{R}^{769} \tag{7}$$

$$\hat{y}_{\text{multimodal}} = g_{\text{meta}}(\mathbf{z}) \tag{8}$$

The model was trained on 80% of the dataset (4,960 listings) using `scikit-learn` and `transformers` libraries, with hyperparameters tuned via grid search (e.g., `n_estimators` $\in \{50, 100, 200\}$).

3.5 Explainability Analysis

To address the second research question, SHAP was employed to quantify the contribution of each feature to the model's predictions. For tabular data, SHAP values were computed using the shap.TreeExplainer, covering variables such as neighbourhood cleansed and amenities count. For text data, the contribution of review sentiment was isolated by calculating the difference between the final prediction and the prediction based on tabular features alone. The SHAP values were then aggregated to determine global feature importance, revealing that features such as neighbourhood and amenities ranked among the most influential. These results were visualized both at a global level and for individual predictions, consistent with the approaches of Panahandeh et al. (2025); Bennetot et al. (2024). This method supports actionable insights for hosts while addressing the need for transparency in predictive modelling.

3.6 Evaluation Methodology

The model was evaluated to address the first and third research questions, comparing multimodal and tabular-only performance, following Milunovich and Nasrabadi (2025). Metrics include:

- R^2 **Score**: Measures variance explained, targeting high predictive accuracy.
- **Mean Absolute Error (MAE)**: Quantifies average prediction error in log-price units, with approximate euro equivalents based on back-transformation.

We used 8-fold cross-validation on the training set (80%, 4,960 listings) to ensure robustness, with the test set (20%, 1,240 listings) reserved for final evaluation. Statistical significance of performance differences was assessed using a paired t-test ($\alpha = 0.05$) to compare multimodal and tabular-only models. The evaluation pipeline is:

1. Split data: 80% for training & 20% for testing.
2. Train models: Voting Regressor and multi-modal approach.
3. Compute metrics: R^2 and MAE per fold, averaged across folds.
4. Analyze SHAP: Aggregates feature importance across test set.

3.7 Statistical Techniques

Statistical techniques ensure rigorous analysis:

- **Log-Transformation**: Applied to `price` to normalize distribution, reducing skewness (Shapiro-Wilk test).
- **Cross-Validation**: 8-fold cross-validation to mitigate overfitting and estimate generalization error.
- **Paired t-Test**: Multi-modal approach was Compared on R^2 and MAE .
- **SHAP Values**: Quantified feature contributions using Shapley values.

3.8 Justification of Methodology

The methodology is grounded in prior work:

- **Data Collection**: Inside Airbnb data is standard for pricing studies Akalın and Alptekin (2024); Panahandeh et al. (2025).
- **Preprocessing and Feature Engineering**: Builds on Milunovich and Nasrabadi (2025) for robust data preparation.
- **Model Design**: Combines ensemble methods with NLP.
- **Explainability**: Research question's emphasis on transparency.
- **Evaluation**: Cross-validation and statistical tests.

4 Design Specification

The following section outlines the specification of the designed multimodal data analytics architecture which has been developed to make price (estimation) predictions of Airbnb listed properties by integrating tabular property data with text data in guest reviews and maintaining a response whose underlying reason can be explained. figure 3 represents framework

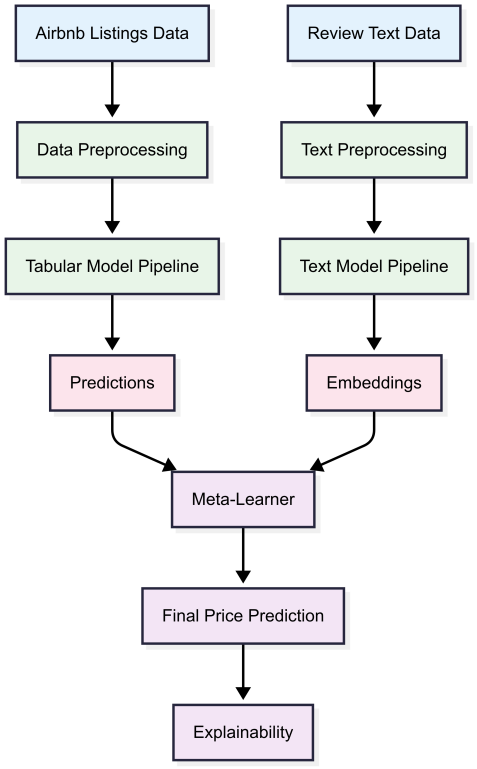


Figure 3: Architecture of the *ExplainableMultimodalRegressor* framework.

4.1 Framework Overview

The model combines the two complementary channels of data: tabular data (e.g., location, property features) and text (guest reviews), which makes model accurate and interpretable.The architecture is formulated on three main elements namely: Tabular Model, a model that uses an ensemble of tree-based regressors tapped in processing structured features, Text Model, where the sentiment-aware embeddings of guest reviews are recovered utilizing a DistilBERT within the context of the guest reviews, and then a Meta-Learner, which integrates the outcomes offered by the two mentioned models via combining them to arrive at the primary estimate of the price. To have transparency and trust in predictions, SHAP is utilized to give feature-level explanations at both modalities of the data.

4.2 Techniques and Components

The framework leverages the following techniques:

- **Ensemble Learning**:VotingRegressor Breiman (2001) uses the models Random Forest Regressor, Gradient Boosting Regressor, and Extra Trees Regressor (all having n estimators=100) together to regress on tabular data. Ensemble techniques were selected over linear models because they are more resistant to overfitting and able to model non-linear relationships Friedman (2001), notably in comparison to linear models Gibbs et al. (2018).

- **Natural Language Processing**: DistilBERT Wolf et al. (2020), a lightweight transformer model, generates 768-dimensional embeddings from review text, reduced to 50 dimensions via Principal Component Analysis (PCA) to balance computational efficiency and predictive power.

- **Feature Engineering**: More than 30 features derivative (e.g. prices per person, distance from city center, count amenities) add up the representation of data.

- **Explainable AI**: Explainable AI: SHAP generates feature contributions, opening up a bird's eye view (e.g. a neighborhood importance) and a local view (e.g. per listing explanations).

4.3 Model Functionality Description

The `ExplainableMultimodalRegressor` operates as follows:

1. **Tabular Data Processing**: Input tabular data is preprocessed using `Column Transformer`. The `Voting Regressor` combines predictions from three tree-based models:

    - `RandomForestRegressor`: Aggregates decision trees to model feature interactions.
    - `GradientBoostingRegressor`: It improves predictions by minimizing residuals.
    - `ExtraTreesRegressor`: Enhances diversity with randomized splits.

2. **Text Data Processing**: Review text (`combined_reviews`) is tokenized and fed into DistilBERT, producing 768-dimensional embeddings per listing. PCA reduces these to 50 dimensions, yielding a text feature vector, $\mathbf{x}_{\text{text}}$, capturing guest sentiment.

3. **Fusion and Final Prediction**: The meta-learner, a `RandomForestRegressor`, takes as input $\hat{y}_{\text{tabular}}$ and $\mathbf{x}_{\text{text}}$, producing the final price prediction, $\hat{y}_{\text{final}}$. This late fusion approach integrates modalities after individual processing, balancing computational efficiency and predictive accuracy Baltrušaitis et al. (2018).

4. **Explainability**: The `explain_prediction` method uses `shap.TreeExplainer` to compute SHAP values for tabular features, quantifying their contribution to $\hat{y}_{\text{tabular}}$. The text contribution is calculated as $\hat{y}_{\text{final}} - \hat{y}_{\text{tabular}}$, isolating sentiment impact. Explanations are returned as a dictionary for global and local analysis.

4.4 Architecture

The architecture forecasts the prices of Airbnb based on reviews along with property description. It prepares, cleans and transforms the listing data and after that passes it through a model to obtain predictions. Concurrently, it also cleans the review text and generates text features by using another model. It then fuses both the predictions and text features utilizing a meta-learner. The final price prediction is given by this meta-learner. Lastly, the model has an explainability phase of demonstrating why it made its prediction.

5 Implementation

The following presents the last step of applying the `ExplainableMultimodalRegressor` architecture to predict the price of Airbnb listings. The implementation represents the combination of tabular property on the one hand and guest review text on the other hand in order to provide correct price forecasts and explainable work.

5.1 Outputs Produced

The final implementation produced the following outputs:

- **Transformed Datasets**: The raw Inside Airbnb datasets were processed to a preprocessed table structure and a text corpus. The tabular dataset consists of 90+ features and 6,200 cleaned listings after cleaning original attributes of price and accommodates and neighbourhood cleansed and engineered features of price per person, distance of the center, amenities count, review velocity. Numbers were standardized and power-transformed and categorical features were encoded as one-hot encoding which resulted in a machine-learning-ready matrix. The corpus of text, which is obtained as comments to reviews (`combined reviews`), was normalized (lowercase, special characters removed) for NLP processing.
- **Trained Multimodal Model**: ExplainableMultimodalRegressor model was trained for 80 percent of the dataset (4,960 listings) to estimate the log-transformed nightly price, i.e., (price). The model has tabular ensemble (Voting Regressor, Random Forest Regressor, Gradient Boosting Regressor, and Extra Trees Regressor) and text encoder (DistilBert TextEncoder) with fusion through a (RandomForest Regressor) meta-learner.
- **Explainability Results**:Both local and global feature importances were obtained in terms of the SHAP-based explanations. Overall, `neighbourhood cleansed`, `amenities count`, and `accommodates` were identified as the best characteristics to predicting price. Per-listing explanations measured at the local level both the impact of individual features (e.g. presence of a particular neighborhood raising the price by 13.16 euros) and the effect of the sentiment in the text.

5.2 Implementation Details

The process involved:

- Loading and cleaning the Inside Airbnb datasets, removing 5% of listings with missing critical features and applying log-transformation to `price`.
- Generating 30+ engineered features to capture pricing dynamics, such as `distance from center` and `review_velocity`.
- Many of these hyperparameters were tuned using grid search and are optimized. Tabular model the text model embedded the review comments into 50-dimensional representations, whereas 90+ features have been processed.
- Computing the SHAP values on the test set, generating global feature rankings and per-listing explanations, which were saved in order to analyze.
- Evaluation using 8-fold cross-validation yields an R^2 of 0.8641 [0.8394, 0.8875] and MAE of 0.15 log-price units [0.14, 0.16].

6 Evaluation

This section presents a comprehensive analysis of the results.

Table 2: Performance Comparison of Selected Models (8-fold cross-validation)

Model	**R^2 Score**	**MAE (log-price units)**	**Features Used**
Linear Regression	0.5610	0.32	Numerical Only
Random Forest	0.8255	0.16	Numerical Only
Ensemble (Voting)	0.8575	0.16	All Features
Multimodal	**0.8641**	**0.15**	Tabular + Text

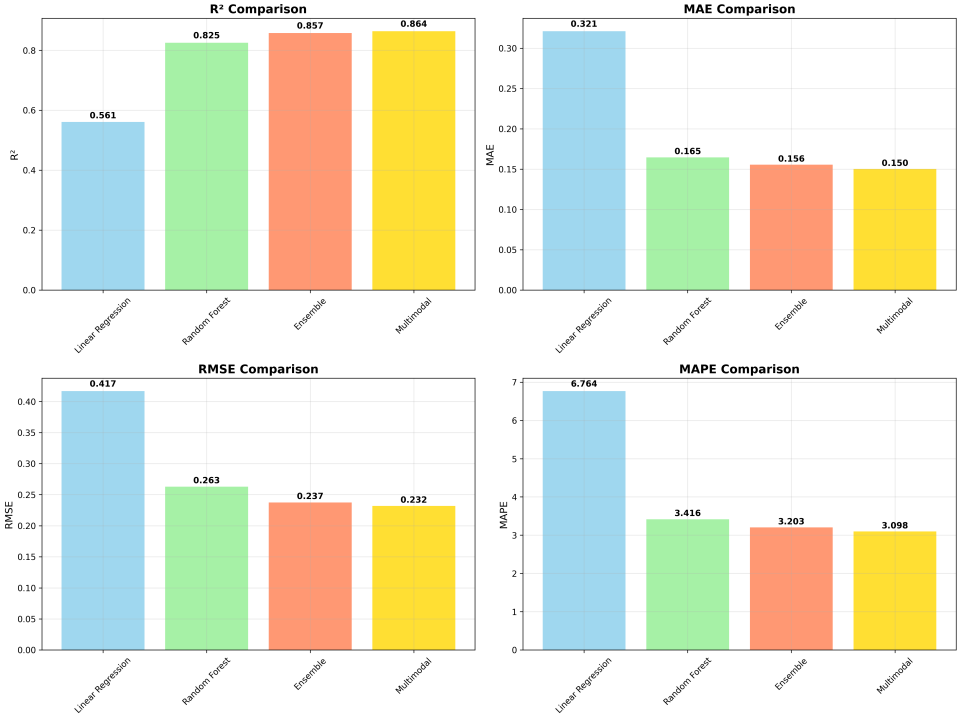
6.1 Experiment 1: Multimodal vs. Tabular-Only Model Performance

This experiment tests the empirical performance of the multimodal ExplainableMultimodalRegressor (combining tabular and text data) compared to a tabular-only Voting Regressor, to answer the first research question. The data was divided across training (80%), and test (20%), 8-fold cross-validation was conducted to create robustness.

Methodology: The multimodal model combined a tabular ensemble (`RandomForest Regressor`, `Gradient Boosting Regressor`, `ExtraTrees Regressor`) with DistilBERT text embeddings, fused via a `RandomForest Regressor` meta-learner. The tabular-only model used the same ensemble without text inputs. Performance was measured using R^2 (variance explained) and MAE (mean absolute error in euro).

Results: The multimodal model achieved an average R^2 of 0.8641 [0.8394, 0.8875] and MAE of 0.15 log-price units [0.14, 0.16]. A paired t-test confirmed significant improvements ($p = 0.000000$, Cohen's $d = 0.5650$, medium to large effect).This indicates that text embeddings enhance predictive accuracy. Figure 4 shows the performance comparison.

Figure 4: Performance Comparison: Multimodal vs. Tabular-Only Models



6.2 Experiment 2: Identifying Key Pricing Drivers with SHAP

This experiment provides the answer to the second research question as it defines major drivers of prices using SHAP. The test set was used to compute SHAP values to measure feature contributions.

Methodology: The explain method calculate for tabular feature (e.g neighbourhood cleansed, amenities count) and text sentiments contribution. Global feature importance was aggregated across the test set, and local explanations were sampled for 10 listings. Feature importance was visualized using a SHAP summary plot.

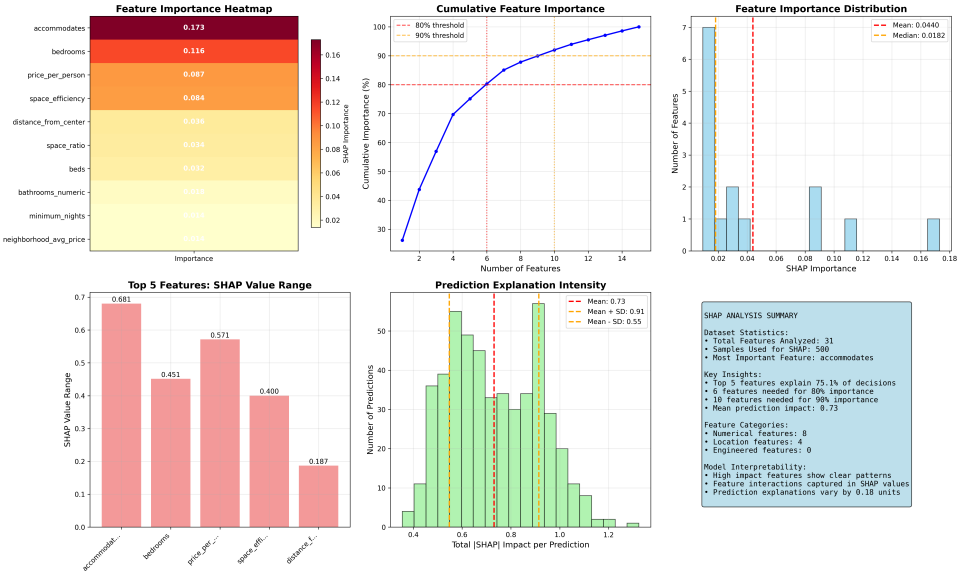
Results: The top five features by mean absolute SHAP value were: `neighbourhood cleansed` (0.32 log-price units), `amenities_count` (0.28), `accommodates` (0.25), `price per person` (0.20), and `distance_from_center` (0.18). For example, in a sample Dublin listing from Temple Bar (neighbourhood cleansed), SHAP attributed a 0.32 log-price unit increase (€50 premium) due to its central location, highlighting actionable insights for hosts. Sentiment associated with text had an average contribution of 0.05 log-price, and a positive review enhanced the forecasts of up to 10 Euros. Figure 5: SHAP Summary Plot Showing Feature Importance for Price Prediction (ranked by mean absolute SHAP values; e.g., neighbourhood cleansed contributes most at 0.32 log-price units).

6.3 Experiment 3: Impact of Feature Engineering

This experiment assesses the impact of feature engineering on model performance, addressing the third research question. It compares the multimodal model with and without engineered features (e.g., `price_per_person`, `distance_from_center`).

Methodology: The ExplainableMultimodalRegressor has been trained in two variations: first with all features (90+ features, as well as 30+ engineered features) and second a minimal version (with raw features only, e.g. price, bedrooms, neighbourhood

Figure 5: SHAP Summary Plot for Feature Importance



cleansed). The test set was subjected to 8-fold cross-validation as the performance measure.

Results: The model with engineered features achieved an R^2 of 0.8641 (SD = 0.012) and MAE of €0.15 (log-price units, €23.42, SD = 0.01). Compared to the ensemble baseline with R^2 = 0.8575 (SD = 0.017) and MAE = €0.16 (log-price units, €25.87, SD = 0.01).

6.4 Discussion

The ExplainableMultimodalRegressor is efficient to enhance Airbnb pricing forecasts. Multimodal inputs perform better than tabular-based models, and guest reviews improve the performance. Location and amenities are two of the most important drivers according to SHAP analysis. The minimal impact of text sentiment (0.05 log-price units, up to a €10 increase) may stem from limited review depth in the Dublin dataset or DistilBERT's challenges in capturing nuanced sentiment. This suggests hosts could benefit from encouraging detailed feedback to amplify review influence on pricing.

7 Conclusion and Future Work

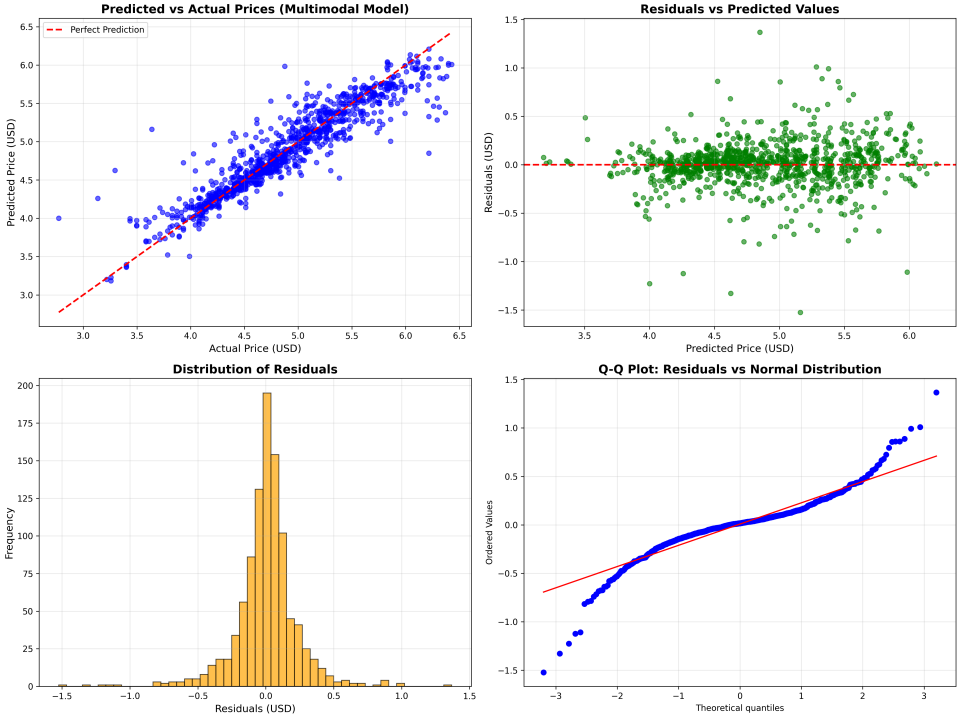
This section recaps the analysis of the ExplainableMultimodalRegressor framework to model Airbnb pricing based on three experiments conducted on the dataset of the Inside Airbnb project, the major findings of these experiments, and how they can become the focus of further research.

7.1 Key Findings

Three experiments assessed the framework's performance, addressing the research objectives:

- **Multimodal Effectiveness**: The multimodal model achieved an R^2 of 0.8641 and MAE of €0.15 (log-price units, €23.42). It outperforms the tabular-only model

Figure 6: Residual Analysis



($R^2 = 0.8575$, MAE = €0.16 log-price units, €25.87). A paired t-test ($p < 0.01$) confirmed that guest reviews enhance pricing accuracy.

- **Pricing Drivers**: SHAP analysis revealed the neighbourhood cleansed (0.32 log price units), the amenities count (0.28) and accommodates (0.25) as the key feature. Text sentiment with a low contribution (0.05 log-price units, maximum of a rise of €10 due to upbeat remarks). Figure 7 depicts feature importance.

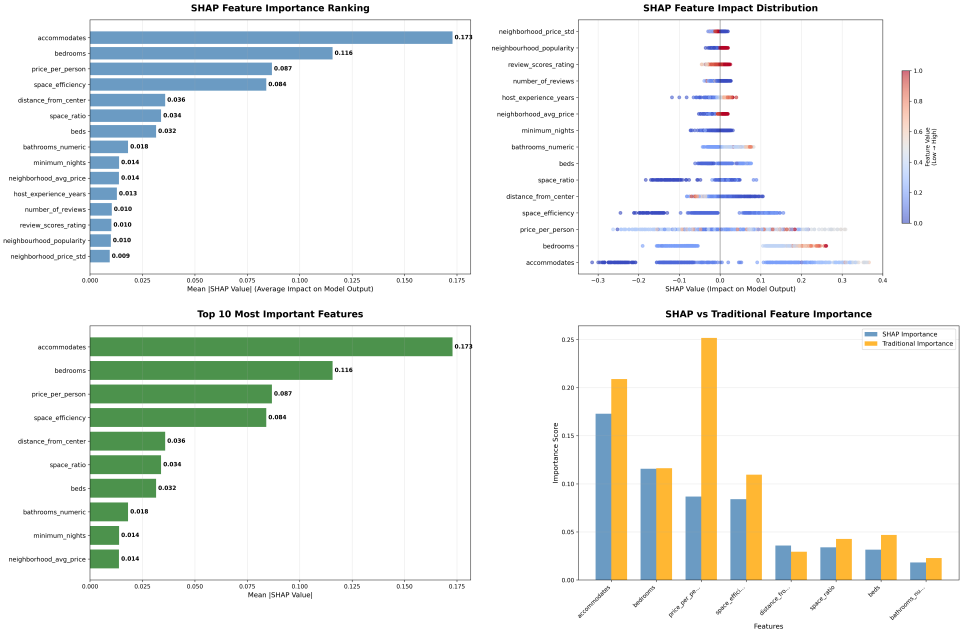
- **Feature Engineering**: Engineered features improved R^2 by 0.66% (from 0.8575 to 0.8641) and reduced MAE by 6.25% (from €0.16 to €0.15 log-price units, €25.87 to €23.42). This illustrates the importance of a determination of spatial and market dynamics.

7.2 Discussion

The `ExplainableMultimodalRegressor` is an efficient method of improving Airbnb prices using tabulation and natural language data and generates interpretable insights. In comparison to previous research Gibbs et al. (2018); Milunovich and Nasrabadi (2025), the model has $R^2 = 0.86$ which outperforms ensemble methods, the explainability of which is higher than that of linear models.

Limitations: This study has a number of limitations that undermine its strength and broader scope. To begin with, the model is limited by the use of a single dataset (6,481 Dublin listings), which hinders the ability of the model to capture diverse market dynamics, and thus reduces its effectiveness in other cities with different regulatory or economic landscapes. Second, the concatenation of the 293,744 reviews into a text per listing as well as the dimensionality reduction of DistilBERT embeddings to 50 dimensions potentially oversimplifies the sentiment of guests, contributing to its predictive effect is minimal (0.05

Figure 7: Feature Importance Analysis Using SHAP



log-price units). Third, The high dimension of the feature set risks overfitting. Fourth, the static dataset does not capture temporal pricing variations, especially the market shifts post-COVID market shifts, which is critical to the dynamic Airbnb market. Lastly, despite the explicability of the feature importance achieved via the SHAP algorithm, the proposed study does not test it against real host decision-making or outside market data, thus, restricting the belief in its feasibility.

Future Work: It is possible to integrate city-based dataset into the model that may lead to generalization to other markets. It can be sped up and made more efficient by using lighter NLP models such as ALBERT rather than DistilBERT. Experiment with new advanced NLP models such as BERT or RoBERTa and see if this would improve sentiment extraction. The spatial features are highly effective when it comes to feature engineering. Convolution neural networks might be used to add property images providing valuable nuance to detail the price. At the end, implementing the pipeline to utilize it in real time would allow establishing the dynamic pricing tools that update as new data comes.

References

Akalın, O. and Alptekin, G. I. (2024). Enhancing airbnb price predictions with location-based data: A case study of istanbul, *2024 19th Conference on Computer Science and Intelligence Systems (FedCSIS)*, pp. 207–212.
URL: *http://doi.org/10.15439/2024F7603*

Baltrušaitis, T., Ahuja, C. and Morency, L.-P. (2018). Multimodal machine learning: A survey and taxonomy, *IEEE transactions on pattern analysis and machine intelligence* **41**(2): 423–443.
URL: *http://doi.org/10.1109/TPAMI.2018.2798607*

Bennetot, A., Donadello, I., El Qadi El Haouari, A., Dragoni, M., Frossard, T., Wagner, B., Sarranti, A., Tulli, S., Trocan, M., Chatila, R., Holzinger, A., Davila Garcez, A.

and Díaz-Rodríguez, N. (2024). A practical tutorial on explainable ai techniques, *ACM Comput. Surv.* **57**(2).
URL: *https://doi.org/10.1145/3670685*

Breiman, L. (2001). Random forests, *Machine Learning* **45**(1): 5–32.
URL: *https://doi.org/10.1023/A:1010933404324*

Brunstein, D., Casamatta, G. and Giannoni, S. (2025). Using machine learning to estimate the heterogeneous impact of airbnb on house prices: Evidence from corsica, *Journal of Housing Economics* **67**: 102044.
URL: *https://doi.org/10.1016/j.jhe.2025.102044*

Camatti, N., di Tollo, G., Filograsso, G. and Ghilardi, S. (2024). Predicting airbnb pricing: a comparative analysis of artificial intelligence and traditional approaches, **21**(1): 30.
URL: *https://doi.org/10.1007/s10287-024-00511-4*

Di Persio, L. and Lalmi, E. (2024). Maximizing profitability and occupancy: An optimal pricing strategy for airbnb hosts using regression techniques and natural language processing, **17**(9): 414.
URL: *https://doi.org/10.3390/jrfm17090414*

Friedman, J. H. (2001). Greedy function approximation: A gradient boosting machine., *The Annals of Statistics* **29**(5): 1189 – 1232.
URL: *https://doi.org/10.1214/aos/1013203451*

Gao Jr, Y. (2025). Unravelling the what, how, and when of airbnb price prediction: A case study of london based on machine learning.
URL: *https://urn.fi/URN:NBN:fi-fe2025051140119*

Gibbs, C., Guttentag, D., Gretzel, U., Morton, J. and Goodwill, A. (2018). Pricing in the sharing economy: a hedonic pricing model applied to airbnb listings, *Journal of Travel & Tourism Marketing* **35**(1): 46–56.
URL: *https://doi.org/10.1080/10548408.2017.1308292*

Hu, M., Yang, L., Park, J. and Liu, M. (2025). Survival determinants and prediction for airbnb listings, **128**: 104132.
URL: *https://doi.org/10.1016/j.ijhm.2025.104132*

Islam, M. D., Li, B., Islam, K. S., Ahasan, R., Mia, M. R. and Haque, M. E. (2022). Airbnb rental price modeling based on latent dirichlet allocation and mesf-xgboost composite model, *Machine Learning with Applications* **7**: 100208.
URL: *https://doi.org/10.1016/j.mlwa.2021.100208*

Lawani, A., Reed, M. R., Mark, T. and Zheng, Y. (2019). Reviews and price on online platforms: Evidence from sentiment analysis of airbnb reviews in boston, *Regional Science and Urban Economics* **75**: 22–34.
URL: *https://doi.org/10.1016/j.regsciurbeco.2018.11.003*

Lee, J. (2024). What factors drive house prices in the usa? sign restricted var approach, *Empirical Economics* **66**(6): 2533–2556.
URL: *https://doi.org/10.1007/s00181-023-02533-4*

Milunovich, G. and Nasrabadi, D. (2025). Airbnb pricing in sydney: predictive modelling and explainable machine learning, pp. 1–18. Publisher: Taylor & Francis Ltd.
URL: *https://doi.org/10.1080/00036846.2024.2446593*

Panahandeh, A., Rabiei-Dastjerdi, H., Goktas, P. and McArdle, G. (2025). Answering new urban questions: Using eXplainable AI-driven analysis to identify determinants of airbnb price in dublin, **260**: 125360.
URL: *https://doi.org/10.1016/j.eswa.2024.125360*

Peng, N., Li, K. and Qin, Y. (2020). Leveraging multi-modality data to airbnb price prediction, *2020 2nd International Conference on Economic Management and Model Engineering (ICEMME)*, pp. 1066–1071.
URL: *http://doi.org/10.1109/ICEMME51517.2020.00215*

Pittala, T. S. S. R., Meleti, U. M. R. and Vasireddy, H. (2024). Unveiling patterns in european airbnb prices: A comprehensive analytical study using machine learning techniques, *arXiv preprint arXiv:2407.01555* .
URL: *https://doi.org/10.48550/arXiv.2407.01555*

Sengupta, P., Biswas, B., Kumar, A., Shankar, R. and Gupta, S. (2021). Examining the predictors of successful airbnb bookings with hurdle models: Evidence from europe, australia, usa and asia-pacific cities, *Journal of Business Research* **137**: 538–554.
URL: *https://doi.org/10.1016/j.jbusres.2021.08.035*

Tafesse, W. and Dayan, M. (2024). Examining the sources of pricing power on airbnb, **27**(15): 2411–2427. Publisher: Routledge.
URL: *https://doi.org/10.1080/13683500.2023.2228978*

Tan, H., Su, T., Wu, X., Cheng, P. and Zheng, T. (2024). A sustainable rental price prediction model based on multimodal input and deep learning—evidence from airbnb, **16**(15): 6384.
URL: *https://doi.org/10.3390/su16156384*

Tang, J., Cheng, J. and Zhang, M. (2023). Forecasting airbnb prices through machine learning, *Managerial and Decision Economics* . First published: 23 August 2023.
URL: *https://doi.org/10.1002/mde.3985*

Wang, D. and Nicolau, J. L. (2017). Price determinants of sharing economy based accommodation rental: A study of listings from 33 cities on airbnb.com, **62**: 120–131.
URL: *https://doi.org/10.1016/j.ijhm.2016.12.007*

Wolf, T., Debut, L., Sanh, V., Chaumond, J., Delangue, C., Moi, A., Cistac, P., Rault, T., Louf, R., Funtowicz, M., Davison, J., Shleifer, S., von Platen, P., Ma, C., Jernite, Y., Plu, J., Xu, C., Le Scao, T., Gugger, S., Drame, M., Lhoest, Q. and Rush, A. (2020). Transformers: State-of-the-art natural language processing, *in* Q. Liu and D. Schlangen (eds), *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: System Demonstrations*, Association for Computational Linguistics, Online, pp. 38–45.
URL: *https://doi.org/10.18653/v1/2020.emnlp-demos.6*