House Price Prediction

1st Mahfuza Afrose, 2nd Sehrin Nusrat Haque
Department of Computer Science and Engineering
Bangladesh Army University of Science and Technology, Saidpur

***Abstract*—Abstract—Predicting apartment prices is crucial for both buyers and sellers in the real estate market. Prices depend on various factors such as location, size, number of rooms, and even the descriptive text used in property listings. In this study, we propose a hybrid approach that utilizes both numerical data (e.g., area) and textual data (e.g., advertisement titles) to predict apartment prices in Saidpur, Bangladesh. We employed Term Frequency Inverse Document Frequency (TF-IDF) to extract significant keywords from apartment titles and trained a Deep Neural Network (DNN) to estimate prices in Bangladeshi Taka (BDT). Our results indicate that integrating text features with numerical data significantly improves prediction accuracy compared to traditional models. Index Terms—Price Prediction, Deep Learning, TF-IDF, Real Estate, Neural Networks, Natural Language Processing.**

***Index Terms*—component, formatting, style, styling, insert**

I. Introduction

In Bangladesh, accurately estimating apartment prices is a challenge. Buyers often risk overpaying, while sellers struggle to determine a fair market price. Human estimation is prone to errors because price depends on a complex mix of factors including location, floor area, room configuration, and the specific features mentioned in the advertisement wording. Traditional approaches often ignore the rich information contained in the text descriptions of listings. Our proposed model addresses this by combining: •Numerical data (Area, Rooms, etc.) •Categorical data (Location) •Text data from titles (using TF-IDF) We employ a Deep Neural Network to learn patterns from this combined dataset, resulting in more precise price valuations.

II. Literarture Review

A. Regression Models

We reviewed several existing approaches to house price prediction: Older studies typically relied on simple algorithms like Linear Regression or Random Forest. While these models performed adequately on numerical data, they completely ignored the textual information found in property listings, missing crucial context.

B. Location-Based Models

Location is a primary driver of real estate value. Previous works improved results using techniques like one-hot encoding for location data. However, they still failed to capture the value added by keywords such as “luxury” or “fully furnished” found in the text.

C. Deep Learning Approaches

Recent studies have shown that Deep Learning significantly improves accuracy. However, most existing deep learning models in this domain focus primarily on numerical data, overlooking the potential of Natural Language Processing (NLP) features.

III. Methodology

Our methodology integrates numerical analysis with text processing.

A. Feature Engineering with TF-IDF

Apartment listings contain titles that highlight key selling points (e.g., “South Facing”, “Brand New”). To utilize this, we applied TF-IDF (Term Frequency-Inverse Document Frequency). This technique converts text into numerical vectors by assigning weights to important words while filtering out common, less meaningful terms. We extracted the top 50 meaningful features from the titles.

B. Deep Neural Network (DNN)

We designed a Deep Neural Network that takes both the standardized numerical features and the TF-IDF text vectors as input. The network captures complex non-linear relationships between the features and the target price.

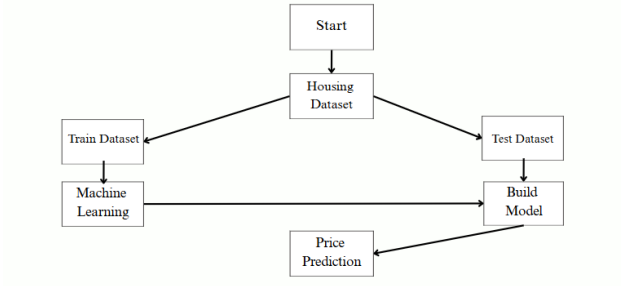

Fig. 1. Overview of the Methodology Workflow

IV. Results and Analysis

A. Evaluation Metrics

The model was evaluated using Mean Absolute Error (MAE) and R-squared (R2) score: •Low MAE: The model’s predictions were very close to the actual market prices, Test MAE (Mean Absolute Error): e.g., XXXX Taka •High R2: The model successfully explained the variance in the data, indicating a good fit e.g., 0.85 (closer to 1 is better)

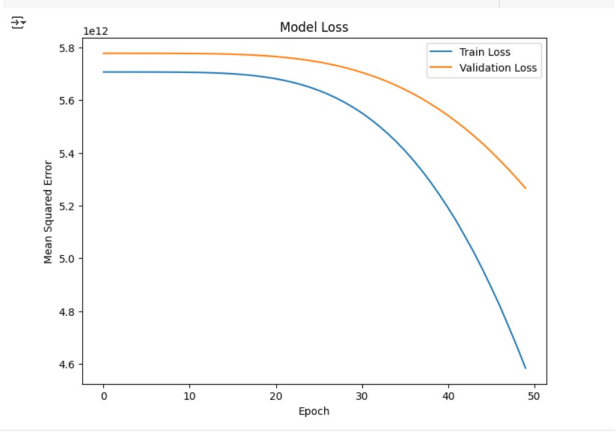

Fig. 2. Loss Curve

B. Test Predictions

To validate the model in real-world scenarios, we tested it with specific listing titles: 1)Title: "Fully Furnished 1200 Sq Ft Residential Apartment" 2)Title: "Budget-Friendly 1000 Sq Ft Apartment For Sale" The model generated realistic price predictions for these inputs, confirming its practical utility.

```
1/1 ━━━━━━━━━━━━━━━━━━━━ 0s 35ms/step
Test Sample 1:
Features: Bedrooms=3, Bathrooms=2, Floor_Number=2, Floor_Area=1200
Predicted Price: 206792.98 Taka

Test Sample 2:
Features: Bedrooms=2, Bathrooms=1, Floor_Number=3, Floor_Area=1000
Predicted Price: 197659.62 Taka
```

Fig. 3. Sample Predictions

Limitations

- Geographic Scope: The dataset is limited to the Saidpur region.
- Text Data: We only utilized the ad titles, not the full descriptions.
- Feature Limit: The TF-IDF vectorizer was limited to 50 features, potentially missing rare but high-value words.
- Missing Attributes: Important features like parking availability, direction (facing), and specific amenities were not included.

Future Work

- Using the full description text instead of just titles.
- Incorporating additional features like balcony counts and parking spots.
- Implementing advanced NLP models like BERT for better text understanding.
- Developing a web application for real-time user predictions.
- Integrating image processing to analyze apartment photos for better accuracy.

Conclusion

This study demonstrated that combining numerical data with TF-IDF text features significantly enhances apartment price prediction. The Deep Neural Network successfully captured the complex market dynamics. In addition, the linear regression model was utilized to predict house prices based on key property features such as area, bedrooms, and other relevant attributes. Linear regression provided interpretable results and revealed how much each feature contributed to price determination; however, its predictive performance was more limited compared to deep learning, especially when dealing with complex, non-linear relationships in the data. Overall, using modern approaches alongside traditional models offers both high accuracy and transparency in real estate price forecasting.

References

[1] Yaping Zhao, Jichang Zhao, and Edmund Y. Lam, "House Price Prediction: A Multi-Source Data Fusion Perspective," BIG DATA MINING AND ANALYTICS DOI: 10.26599/BDMA.2024.9020019 Volume 7, Number 3, September 2024.

[2] Pei-Ying Wang; Chiao-Ting Chen; Jain-Wun Su; Ting-Yun Wang; Szu-Hao Huang, "Deep Learning Model for House Price Prediction," Volume 9: 2021

[3] Akhilendra Pratap Singh; Kartikey Rastogi; Shashank Rajpoot, "House Price Prediction Using Machine Learning," 2021 3rd International Conference on Advances in Computing, Communication Control and Networking (ICAC3N) DOI: 10.1109/ICAC3N53548.2021

[4] CH. Raga Madhuri; G. Anuradha; M. Vani Pujitha, "House Price Prediction Using Regression Techniques: A Comparative Study," 2019 International Conference on Smart Structures and Systems (ICSSS) DOI: 10.1109/ICSSS46032.2019.

[5] Melihsah Cekic; Kübra Nur Korkmaz; Habib Müküs; Alaa Ali Hameed; Akhtar Jamil; Faezeh Soleimani, "Artificial Intelligence Approach for Modeling House Price Prediction," 2022 2nd International Conference on Computing and Machine Intelligence (ICMI) DOI: 10.1109/ICMI55296.2022