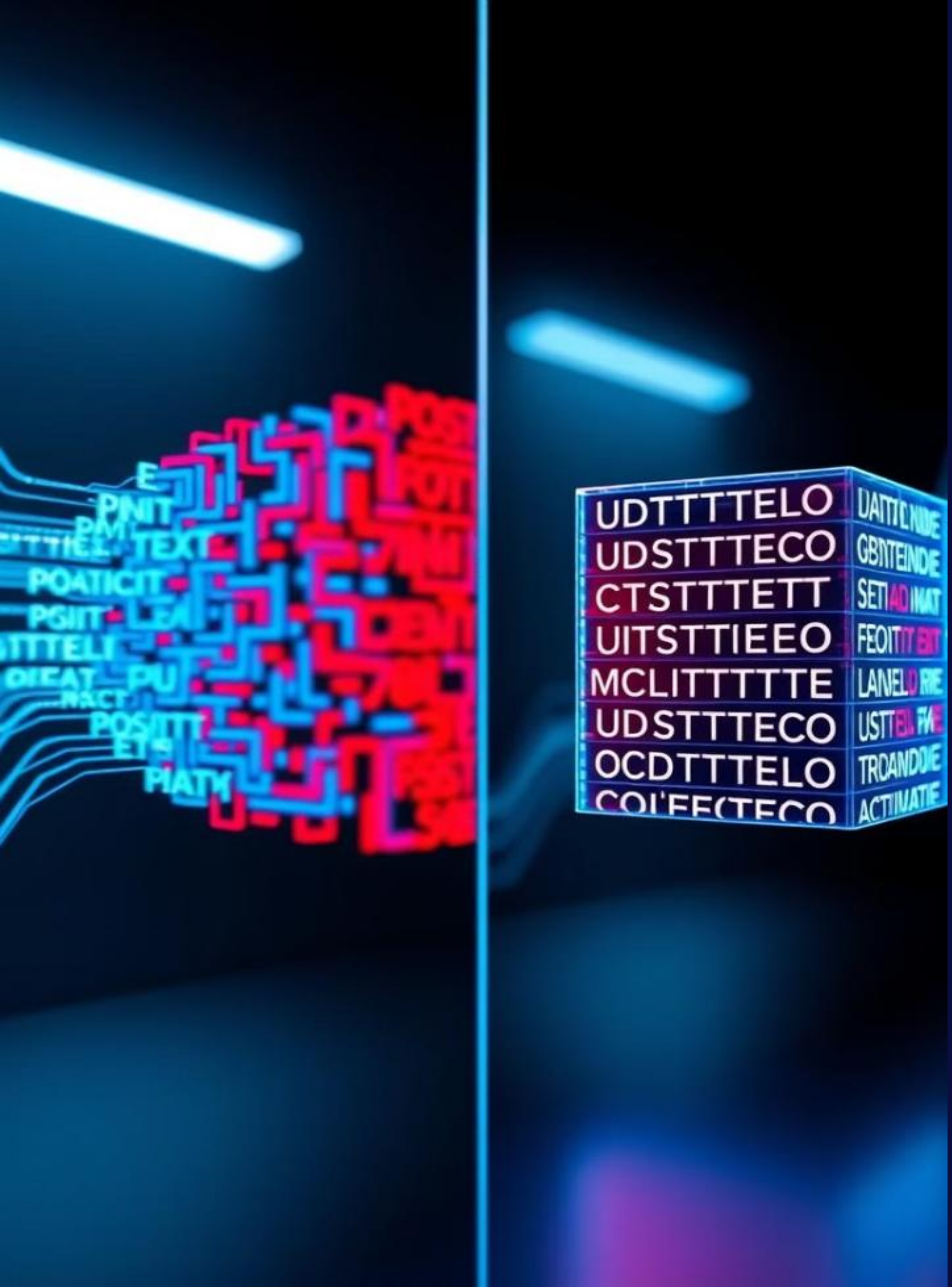
House Sales Price Prediction Prediction

Hybrid Deep Learning and Classical Machine Learning
Kaggle Australian Property Sales Dataset
Graduation Project

By: Haneen Nabil, Sama Samer, Nadine Ali, El Menshawy, Mostafa Tamer



Project Objectives

Robust Forecasting

Use LSTM for time-series prediction of property prices.

Insightful Analysis

Identify seasonal and structural sales patterns.

Smart Strategies

Enable inventory and pricing decisions with data.

Scalable Tool

Deploy AI model with user-friendly interface.

overview

Technologies & Tools

Used Python – core language for data processing and modeling
Pandas / NumPy – data manipulation and numerical operations
Matplotlib / Seaborn – data visualizationScikit-learn – **classical ML models and preprocessing utilitiesTensorFlow / Keras** – for building and training LSTM deep learning models
Streamlit – web framework for deploying an interactive prediction
appMinMaxScaler / TimeseriesGenerator – data normalization and sequence generation for **LSTMGoogle Colab / Jupyter** Notebook – development environment

Scope of Work

Data Acquisition & Preprocessing

Clean data, encode features, normalize values.

Exploratory Data Analysis

Visualize trends, detect outliers, analyze features.

Model Design & Training

Build and train LSTM and classical ML models.

Evaluation & Deployment

Assess models, deploy with Streamlit interface.

Installation

Installation and setups

Access the Application
The application will be available at:
http://localhost:8501

Install Dependencies

Install the required Python packages:

```bash
pip install pandas numpy matplotlib seaborn scikit-learn tensorflow keras streamlit
```

Prepare Model

Ensure your trained LSTM model is saved as:

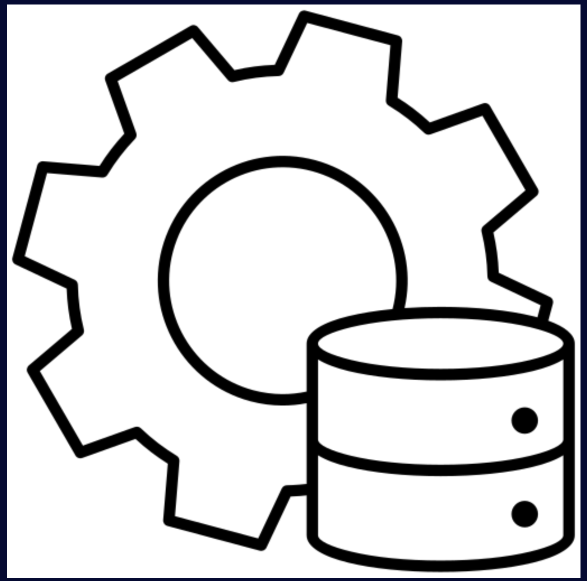
```bash
model.h5
```

Place it in your project directory for prediction or deployment.

Run the Streamlit App

Start the web interface using:

```bash
streamlit run app.py
```



Analysis

Exploratory Data Analysis

Seasonal Trends

Strong cyclical price patterns and market recovery after drops.

Correlation Insights

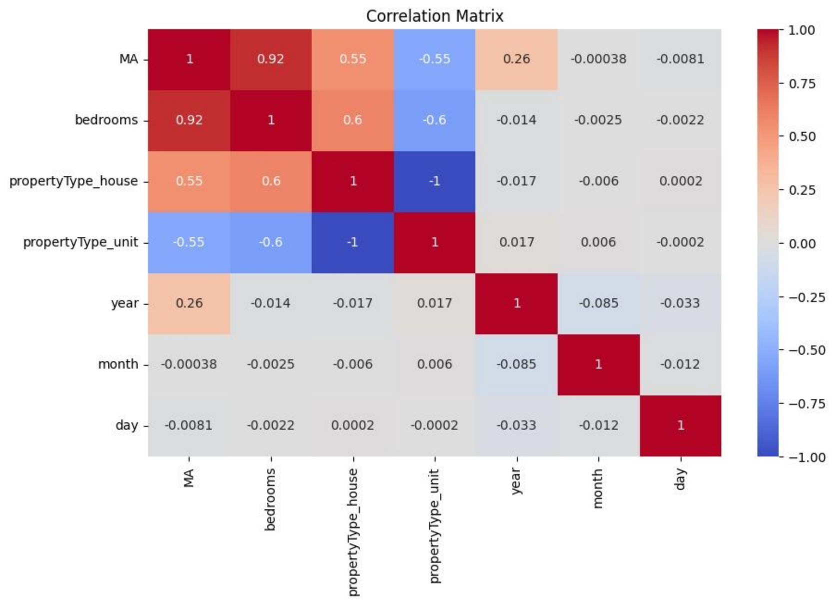
Bedrooms strongly correlate (0.92) with moving average prices.

Property Type Impact

Houses have higher prices than units, confirmed by correlations.

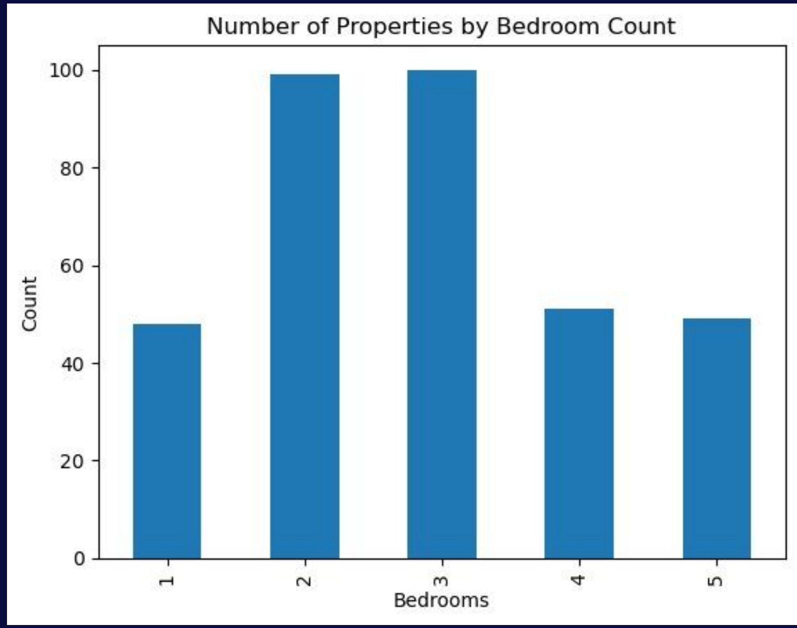
Visualization the data

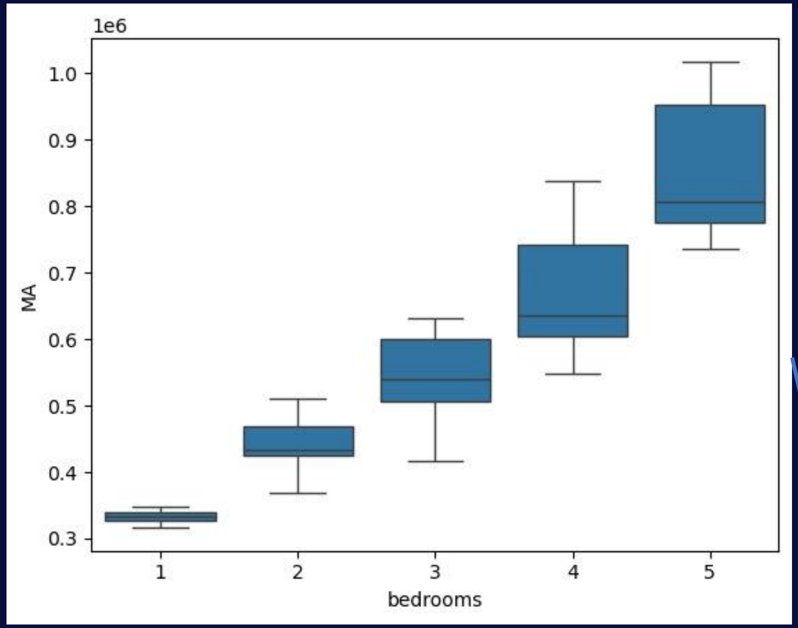
the correlation between the feature



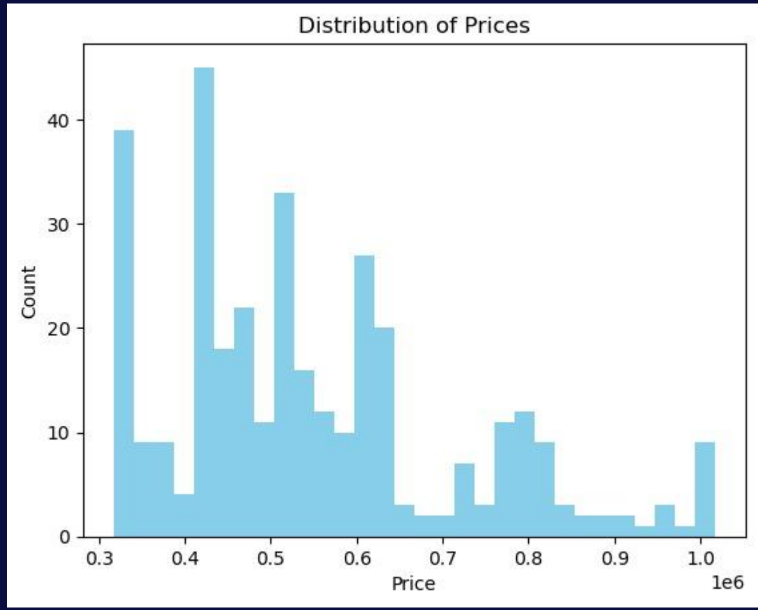
"MA" (possibly Moving Average) is highly positively correlated with "bedrooms" (0.92) and "propertyType_house" (0.55), but strongly negatively correlated with "propertyType_unit" (-0.55).

balance between the number of rooms

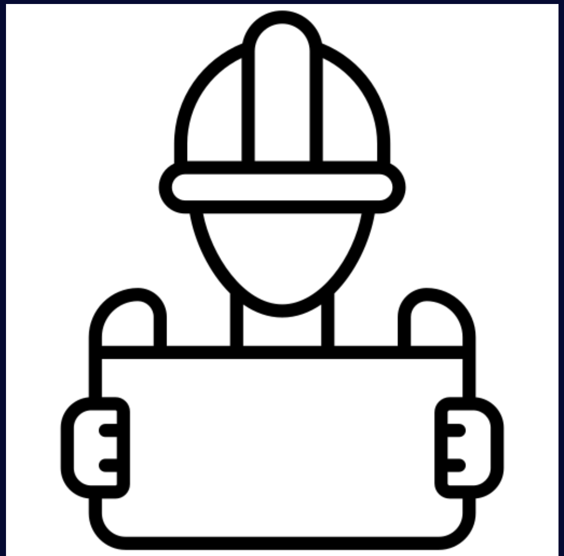




Distribution of the MA



the MA (likely a housing price metric) rises consistently, with both the median and spread of values growing. Notably, homes with 5 bedrooms show the widest price range, suggesting greater variability in high-end properties.



modeling

Modeling Approach

Layer (type)	Output Shape	Param #
lstm (LSTM)	(None, 12, 200)	163,200
dropout (Dropout)	(None, 12, 200)	0
lstm_1 (LSTM)	(None, 12, 150)	210,600
dropout_1 (Dropout)	(None, 12, 150)	0
lstm_2 (LSTM)	(None, 100)	100,400
dense (Dense)	(None, 50)	5,050
dropout_2 (Dropout)	(None, 50)	0
dense_1 (Dense)	(None, 20)	1,020
dense_2 (Dense)	(None, 1)	21

Classical Models

- Linear Regression baseline
- Random Forest & XGBoost for nonlinear patterns

Deep Learning

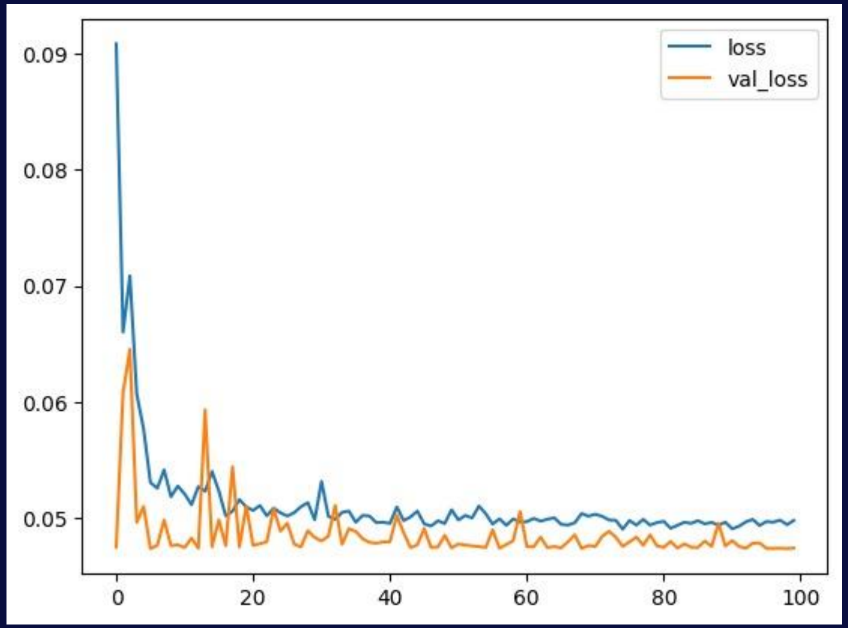
LSTM with 3-5 layers, Adam optimizer, time-series validation.

Hybrid Model

Ensemble averaging to improve accuracy.

The result of some model

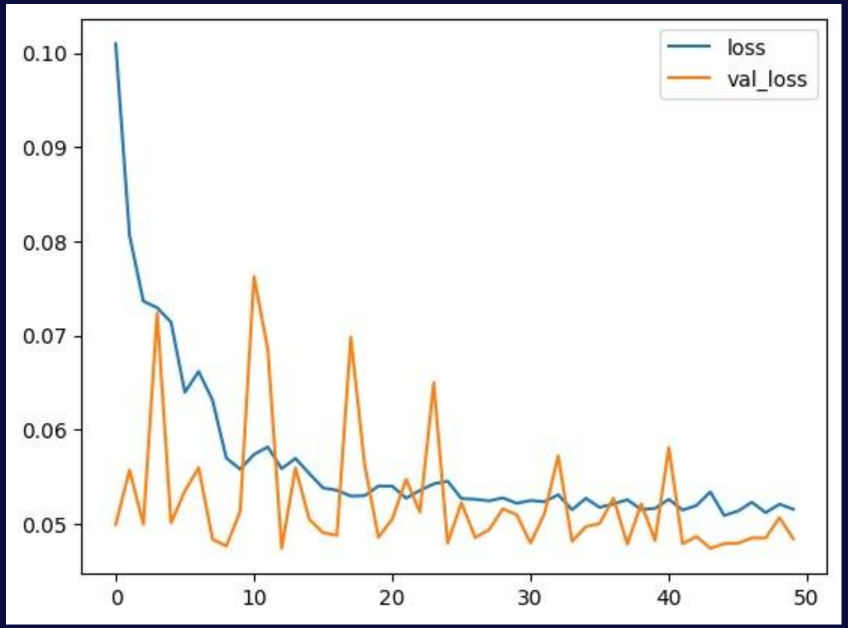
3 layer of LSTM and fully connected



Loss :

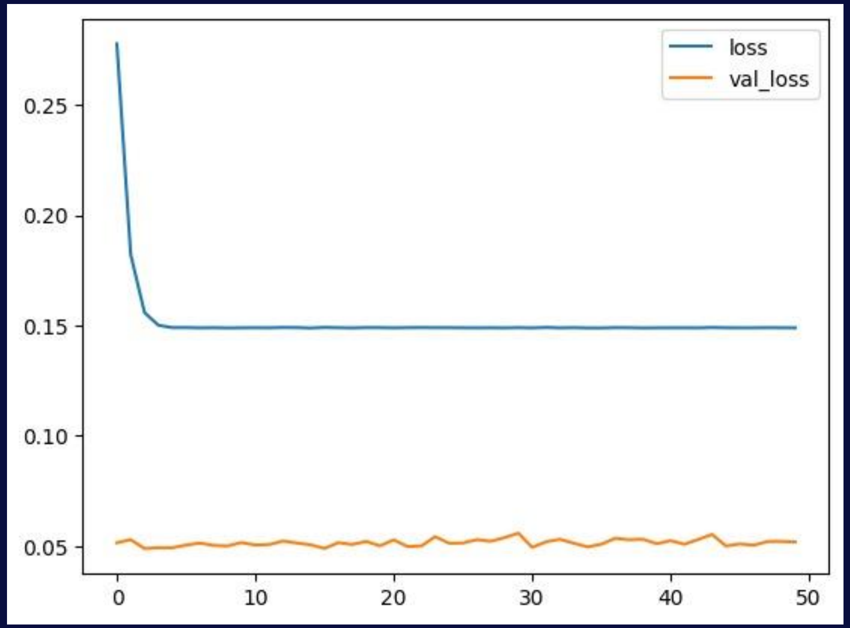
- Train loss : `0.0475`
- Validation loss : `0.0474`

5 layer of LSTM and fully connected



Loss :

- Train loss : `0.0472`
- Validation loss : `0.0484`



Loss :

- Train loss : `0.1515`
- Validation loss : `0.0519`

Deploy and GUI

Evaluation & Deployment

Evaluation Metrics

Used MSE, MAE, RMSE to assess models.

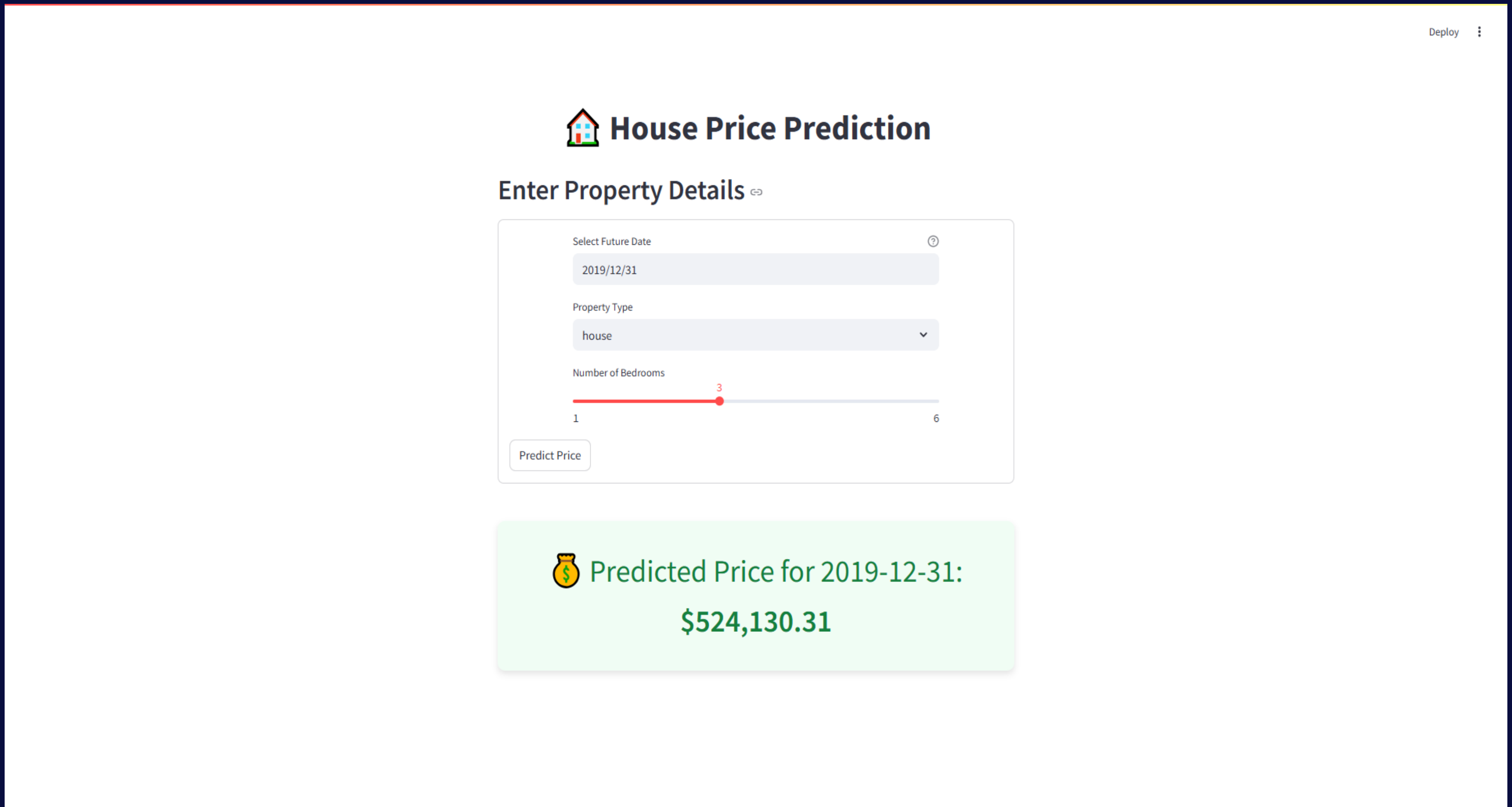
Deployment

Streamlit app for real-time price prediction.

Model Export

Saved final model in .h5 format for inference.

GUI


Deploy
House Price Prediction
Enter Property Details
Select Future Date
2019/12/31
Property Type
house
Number of Bedrooms
3
1
6
Predict Price
Predicted Price for 2019-12-31:
$524,130.31

Conclusion

Conclusion & Future Work

1

Conclusion

Hybrid models combine temporal and structural insights effectively.

2

Future Work

Add geospatial data and attention mechanisms for better forecasts. Data might not be applicable in Egypt

3

Integration

Enable API connections with real estate platforms.

Conclusion

At the end, this project demonstrates the effectiveness of combining classical and deep learning models for time-series forecasting in the real estate domain. The hybrid model benefits from both LSTM's temporal awareness and tree-based models' structural feature handling. The deployment via Streamlit adds practical utility for end-users and stakeholders.