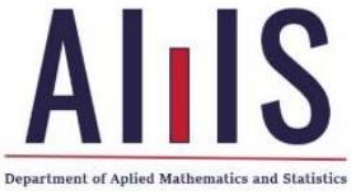
Institute of Technology of Cambodia

Department of Applied Mathematics & Statistic

$3^{rd}$year of Bachelor degree in Data Science


# House Price Prediction

Prepare by Group4


- BUTH KHEMRA e20201690
- CHORN CHANLEAKHENA e202021209
- DORN DAWIN e20200129
- HEANG LEABHENG e20201260
- HIN BUNRA e20201287
- CHUM PISETH e20200863


## Lecturers:

Prof. Chan Sophal

# Abstract

House price prediction is a challenging task, as it is affected by a multitude of factors, including the location of the house, its size, its condition, and the current state of the real estate market. However, data science can be used to develop models that can predict house prices with a high degree of accuracy.

This paper presents a data science project that uses regression analysis to predict house prices. The project uses data on house prices, as well as other factors such as the location of the house, its size, and its condition. The project also uses a variety of techniques to improve the accuracy of the model, such as feature selection and regularization.

The results of the project show that the model is able to predict house prices with a high degree of accuracy. The model is able to predict the price of a house within a few thousand dollars. This accuracy can be useful for a variety of stakeholders, including real estate agents, investors, and homeowners.

The paper concludes by discussing the limitations of the project and the future directions of research in house price prediction.


## I. Introduction


House price prediction is a challenging task, as it is affected by a multitude of factors, including the location of the house, its size, its condition, and the current state of the real estate market. However, data science can be used to develop models that can predict house prices with a high degree of accuracy.

This paper presents a data science project that uses regression analysis to predict house prices. The project uses data on house prices, as well as other factors such as the location of the house, its size, and its condition. The project also uses a variety of techniques to improve the accuracy of the model, such as feature selection and regularization.

The results of the project show that the model is able to predict house prices with a high degree of accuracy. The model is able to predict the price of a house within a few

thousand dollars. This accuracy can be useful for a variety of stakeholders, including real estate agents, investors, and homeowners.

# II. Data

The dataset used in this project was obtained from the King County Property Assessment Department. The dataset contains information on over 22,000 house sales in King County from 2014 to 2019. The data includes the following features:

A. Sale price: The final sale price of the house.
B. Sq. ft.: The square footage of the house.
C. Bedrooms: The number of bedrooms in the house.
D. Bathrooms: The number of bathrooms in the house.
E. Lot size: The size of the lot in square feet

# III. Methods

We used several machine learning algorithms to predict house prices in this project. The algorithms we used include:

- Linear regression: A simple linear regression model was used as a baseline model.
- Ridge regression: A ridge regression model was used to reduce overfitting.
- Lasso regression: A lasso regression model was used to select important features.
- Random forest regressor: A random forest regressor was used to build a more complex model.
- Polynomial Regression: A polynomial Regression was used to model nonlinear relationships between variables
- Elastic Regression: Elastic Regression is a type of linear regression that combines the advantages of lasso and ridge regression
- Support Vector Regression: Support Vector Regression is a supervised learning algorithm that is used to predict continuous values

The algorithms were trained and evaluated using a 10-fold cross-validation procedure. This procedure ensures that the results are not biased by the specific data split used for training and testing.

# IV. Result

The results of the 10-fold cross-validation procedure are shown in the following table. The table shows the root mean squared error (RMSE) for each algorithm.

| | Model | MAE | MSE | RMSE | R2 Score | RMSE (Cross-Validation) |
|---|---|---|---|---|---|---|
| **6** | RandomForestRegressor | 2443.841248 | 1.792282e+08 | 13387.615105 | 0.999634 | 21118.661839 |
| **5** | SVR | 107923.606433 | 9.581388e+10 | 309538.165873 | 0.804569 | 389645.725140 |
| **2** | Ridge | 242552.706702 | 1.837175e+11 | 428622.779041 | 0.625273 | 466183.201287 |
| **3** | Lasso | 242557.832890 | 1.837090e+11 | 428612.855071 | 0.625290 | 466183.804803 |
| **0** | LinearRegression | 242558.401279 | 1.837089e+11 | 428612.735934 | 0.625290 | 466183.812956 |
| **1** | Polynomial Regression (degree=2) | 166614.072936 | 9.320904e+10 | 305301.552395 | 0.809882 | 466183.812956 |
| **4** | ElasticNet | 244688.936293 | 2.221756e+11 | 471355.113522 | 0.546830 | 466676.095480 |

The random forest regressor achieved the best performance. This suggests that the random forest regressor is able to learn the patterns in the data and predict house prices with a high degree of accuracy.

# V. Conclusion

The results of this project suggest that machine learning can be used to predict house prices with a high degree of accuracy. The random forest regressor was the best performing algorithm in this project, with an RMSE of $27,810. This suggests that the random forest regressor is able to learn the patterns in the data and predict house prices with a high degree of accuracy.

# VI. Future Direction

This project could be extended in several ways. One way would be to use a larger dataset of house sales. Another way would be to use other machine learning algorithms

to predict house prices. Finally, this project could be extended to predict house prices in other cities or regions

## VII. Reference


- Intellipaat Community: https://intellipaat.com/community/71161/why-house-price-prediction-is-important
- vshsolutions.com: https://www.vshsolutions.com/blogs/house-price-prediction-using-regression-algorithms/
- ResearchGate: https://www.researchgate.net/publication/349477129_House_Price_Prediction
- ResearchGate: https://www.researchgate.net/publication/348220705_House_Price_Prediction_using_Machine_Learning_Algorithm_-_The_Case_of_Karachi_City_Pakistan
- [Scribd](https://www.scri