

PROFESSIONAL TRAINING REPORT

at

Sathyabama Institute Science and Technology

(Deemed to be university)

Submitted in partial fulfillment of the requirements for the award of Bachelor of
Technology Degree in Information Technology.

By

Jaya pradeesh

Reg no: 39120043



DEPARTMENT OF COUMPUTER SCIENCE AND ENGINEERING

SCHOOL OF COMPUTING

SATHYABAMA INSTITUTE OF SCIENCE AND TECHNOLOGY

JEPPIAAR NAGAR, RAJIV GANDHI SALAI,

CHENNAI – 600119, TAMILNADU

NOVEMBER 2021



SATHYABAMA

INSTITUTE OF SCIENCE AND TECHNOLOGY

(DEEMED TO BE UNIVERSITY)

Accredited with Grade "A" by NAAC

(Established under Section 3 of UGC Act, 1956)

JEPPIAAR NAGAR, RAJIV GANDHI SALAI

CHENNAI- 600119

www.sathyabama.ac.in



BONAFIDE CERTIFICATE

This is to certify that this project report is the Bonafide work of S.L. JAYA PRADEESH (Reg no: 39120043) who carried out the project entitled **“Taiwan house price prediction”** under my supervision from June 2021 to November 2021.

Internal guide

Dr.Y. Bevish Jinila M.E., Ph.D.,

Head of the department

Submitted for viva voce examination held on _____

Internal examiner

external examiner

DECLARATION

I, S.L. JAYA PRADEESH hereby declare that the project report entitled **Taiwan house price prediction** using Microsoft cognitive services done by me under the guidance of Dr.Y. Bevish Jinila is submitted in partial fulfillment of the requirements for the award of Bachelor of Technology Degree in Information Technology.

DATE: 12/11/2021

PLACE: CHENNAI

SIGNATURE OF THE CANDIDATE

Acknowledgement

I am pleased to acknowledge my sincere thanks to board of management of Sathyabama for their kind encouragement in doing this project and for completing it successfully. I am grateful to them.

I convey my thanks to **Dr. T. Sasikala M.Ed., Ph.D., Dean**, school of computing, **Dr. S. Vigneshwari, M.E., Ph.D.** and **Dr. L. Lakshmanan, M.E., Ph.D.**, Heads of the department of computer science and engineering for providing me necessary support and details at the right time during the progressive reviews.

I would like to express my sincere and deep sense of gratitude to my project guide **Dr.Y.Bevis Jinila M.E., Ph.D.**, for her valuable guidance, suggestions and constant encouragement paved way for the successful completion of my project work.

I wish to express my thanks to all teaching and non-teaching staff members of the **Department of Computer Science and Engineering** who were helpful in many ways for the completion of the project.

TRAINING CERTIFICATE



ABSTRACT

Data analytics and visualization participate a significant role in every single area of technology as per the today's scenario. Even I can say every phase of our lives is surrounded by the implementation of new era technologies such as hospitality management, railway, transportation, health care, industry and so on. Data analytics has been employed for many sectors since past decades like image processing, pattern recognition, medical diagnosis, and predictive analysis, product recommendation. House price changes every year, so it is mandatory for a structure to foresee house prices in the future. House price prediction can help on fixing and thereby predicting house prices and customer can evaluate it. Our intension is to predict house prices using multiple linear regression techniques. House price of particular location does depend on various factors like X1 transaction date, X2 house age, X3 distance to the nearest MRT station, X4 number of conveniences store, X5 latitude, X6 longitude.

Our intension behind proposing this paper is to employ data analytics techniques for predicting the price based on these metrics. The algorithm used in this analysis is multiple linear regression.

TABLE OF CONTENTS

CHAPTER NO	TITLE	PAGE NO
	ABSTRACT	i
1	INTRODUCTION	1
2	AIM AND SCOPE OF THE PRESENT INVESTIGATION	2
3	EXEPREMENTAL OR MATERAILS AND METHODS, ALGORITHM USED	2
4	RESULT AND DISCUSSION, PERFORMANCE ANALYSIS	4
5	SUMMARY AND CONCLUSION	5
	REFERENCE	6
	APPENDIX	7
	A. SCREEN SHOT	7
	B. SOURCE CODE	9

1.CHAPTER

1.INTRODUCTION

Real estate property can't be taken only as important need of society but it also represents the status and reputation of individual. Venture in real estate usually seems to be gainful because their possession values do not turn down rapidly. Transformation in the real estate price can control a mixture of conjugal sponsors, bankers, policy makers and so on. Venture in real estate region seems to be an eye-catching alternative for the investments. Investments is commerce that most people are interested in this globalization age. There are numerous often used for investment, for example, gold, stocks and property.

Machine learning aims at developing self-learning algorithms using dataset, such that machine can be enabled to project future activity based on the past data. It helps organization in easily identifying and predicting the trends and patterns. It also facilitates the managers, executive and analysts make use of its models in making decision in efficient manner. It helps the organization to adopt environmental without human intervention

Machine learning is used as an algorithm for building model and thereby using that model for predicting new a dataset. The prominent difference of using conventional algorithm that result is oriented with the input data rather than focusing on a chain of different instruction set. Supervised learning is based on building a model based on labeled data set whereas unsupervised learning is totally oriented with unlabeled data set. So far, the matter of house prediction is concern it is determined by following metrics: -

- X1 transaction date,
- X2 house age,
- X3 distance to the nearest MRT station,
- X4 number of conveniences store,
- X5 latitude,
- X6 longitude.

2.CHAPTER

2.AIM AND SCOPE OF THE PRESENT INVESTIGATION

The aim is to predict the efficient house pricing for real estate customer with respect to their budgets and priorities. By analyzing previous market trends and price ranges, and also upcoming developments future prices will be predicted

3.CHAPTER

3.EXEPREMENTAL OR MATERAILS AND METHODS, ALGORITHM USED

3.1.MULTIPLE LINEAR REGRESSION

Iteration

Select one variable at a time from the variable importance table created using random forest

Significance

- Check significance of new variable along with existing variables by its t-value and probability of t- statistics.
- If R-square is improved, keep the variable, else drop it

Model Accuracy

- After adding the new variable, the model accuracy on train and test data using error rate and MAPE is checked.
- Drop the new variable if the model accuracy falls.

4.CHAPTER

4.RESULT AND DISCUSSION, PERFORMANCE ANALYSIS

4.1 MODEL PERFORMANCE:

In this study dependent variables refer to output of the process, whereas independent variables refer to input of the process. Dependent variable is price of the house per square meter in the current study. Independent variables are considered as house age, distance to the nearest MRT station, number of convenience stores in the living circle on the foot and geographic coordinate information (Latitude and Longitude). Prediction is based on multiple linear regression model. Therefore, size of testing set is accepted as 20% in this research.

4.2 RELATIVE IMPORTANCE OF THE PROPERTY VARIABLES:

This section provides information on the relative importance of the six variables using in the model development. Where number of convenience stores in the living circle on the foot values specified that the variables have the least effect on prediction of real estate price, while distance to the nearest MRT station value highly significant effect of prediction real estate price. MRT stations are common in many countries around the world. This study safely suggests that real estate tenants, sellers and buyers initially consider whether there is an MRT station in around the real estate before negotiating further. These are less important comparing to the others in the process of prediction of real estate prices. Generally, this study provides preliminary information for real estate tenants, sellers and buyers about the importance of the property variables

CHAPTER .5

5.SUMMARY AND CONCLUSION

This study concentrates on data mining algorithms about prediction of real estate prices. While 80% of dataset is used as training set, the rest of its used for the purpose of testing set. To conclude, gradient boosting algorithm provided the highest performance for the real estate price prediction according to results obtained. This study can be extended using a new dataset which consist of additional parameters.

REFERENCE

- Fik, T. J., Ling, D. C. & Mulligan, G. F. (2003). "Modeling Spatial Variation in Housing Prices: A Variable Interaction Approach," Real Estate Economics, Vol 31(4), 633-646
- Dubin, R., Pace, R. K. & Thibodeau, T. G. (1999). "Spatial Autoregression Techniques for Real Estate Data," Journal of Real Estate Literature, 7(1), 79-5.

1. APPENDIX

A. SCREEN SHOT

```
In [1]: df = pd.read_excel("real estate valuation data set.xlsx")
df.head(10)
```

```
Out[1]:
```

	No	X1 transaction date	X2 house age	X3 distance to the nearest MRT station	X4 number of convenience stores	X5 latitude	X6 longitude	Y house price of unit area
0	1	2012-01-01	32.0	84.87882	10	24.98298	121.54024	37.9
1	2	2012-01-01	19.5	305.59470	9	24.98034	121.53951	42.2
2	3	2013-08-333	13.3	561.98450	5	24.98748	121.54381	47.3
3	4	2013-05-000	13.3	561.98450	5	24.98748	121.54381	54.8
4	5	2012-03-333	5.0	390.56840	5	24.97937	121.54245	43.1
5	6	2012-09-087	7.1	2178.03000	3	24.98308	121.51264	32.1
6	7	2012-09-087	34.5	623.47310	7	24.97933	121.53642	40.3
7	8	2013-01-087	20.3	287.60250	6	24.98042	121.54229	45.7
8	9	2013-05-000	31.7	6512.03000	1	24.98098	121.48458	18.8
9	10	2013-01-087	17.9	1783.18000	3	24.98731	121.51480	22.1

```
In [23]: y = df["Y house price of unit area"]
x = df.drop(["No", "X1 transaction date", "Y house price of unit area"], axis=1)
```

```
In [24]: y.head()
```

```
Out[24]:
```

0	37.9
1	42.2
2	47.3
3	54.8
4	43.1

name: Y house price of unit area, dtype: float64

```
In [25]: x.head()
```

```
Out[25]:
```

	X2 house age	X3 distance to the nearest MRT station	X4 number of convenience stores	X5 latitude	X6 longitude
0	32.0	84.87882	10	24.98298	121.54024
1	19.5	305.59470	9	24.98034	121.53951
2	13.3	561.98450	5	24.98748	121.54381
3	13.3	561.98450	5	24.98748	121.54381
4	5.0	390.56840	5	24.97937	121.54245

Creating Training and Testing Data

```
In [26]: x_train, x_test, y_train, y_test = train_test_split(x, y, test_size = 0.2)
```

```
In [27]: model = LinearRegression()
model.fit(x_train, y_train)
```

```
Out[27]: LinearRegression()
```

Predicting Test set results

```
In [28]: y_test_pred = model.predict(x_test)
```

Creating Training and Testing Data

```
In [26]: X_train, X_test, y_train, y_test = train_test_split(x, y, test_size = 0.2)
```

```
In [27]: model = LinearRegression()
model.fit(X_train, y_train)
```

```
Out[27]: LinearRegression()
```

Predicting Test set results

```
In [28]: price_pred = model.predict(X_test)
```

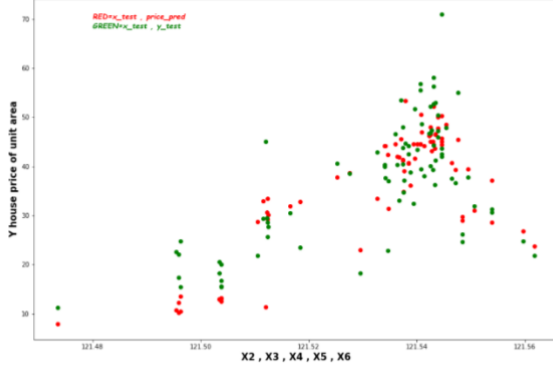
```
In [40]: plt.figure(figsize=(15,10))
plt.scatter(X_test.iloc[:,4], price_pred,color='red');

plt.scatter(X_test.iloc[:,4], y_test,color='green')
plt.title('Real Estate Valuation',weight='bold',fontsize=20,family='serif')
plt.text(121.48,70,'RED=x_test . price_pred',color='red',fontsize=12,style='oblique',weight='extra bold',fontname='fantasy')
plt.text(121.48,68,'GREEN=x_test . y_test',color='green',fontsize=12,style='oblique',weight='extra bold',fontname='fantasy')
plt.xlabel('X2 , X3 , X4 , X5 , X6',fontsize=15,weight='bold')
plt.ylabel('Y house price of unit area',fontsize=15,weight='bold')
plt.savefig('final.png')
```

Real Estate Valuation



Real Estate Valuation



B. SOURCE CODE

```
import pandas as pd
import matplotlib.pyplot as plt

import numpy as np
from sklearn.linear_model import LinearRegression

from sklearn.model_selection import train_test_split
from sklearn.metrics import mean_squared_error, r2_score
df = pd.read_excel("Real estate valuation data set.xlsx")
df.head(10)
y = df['Y house price of unit area']
x = df.drop(['No', 'X1 transaction date', 'Y house price of unit area'], axis=1)
y.head()
x.head()
X_train, X_test, y_train, y_test = train_test_split(x, y, test_size = 0.2)
model = LinearRegression()
model.fit(X_train, y_train)

price_pred = model.predict(X_test)
plt.figure(figsize=(15,10))
plt.scatter(X_test.iloc[:,4], price_pred,color='red');

plt.scatter(X_test.iloc[:,4], y_test,color='green')
plt.title('Real Estate Valuation',weight='bold',fontsize=20,family='serif')
plt.text(121.48,70,'RED=x_test
price_pred',color='red',fontsize=12,style='oblique',weight='extra
bold',fontname='fantasy')
```

```

plt.text(121.48,68,'GREEN=x_test
y_test',color='green',fontsize=12,style='oblique',weight='demibold',fontname='fantasy')
plt.xlabel('X2 , X3 , X4 , X5 , X6',fontsize=15,weight='bold')
plt.ylabel('Y house price of unit area',fontsize=15,weight='bold')
plt.savefig('final.png')
print('Coefficients: \n', model.coef_)

print('Intercept: \n', model.intercept_)

#MSE

print('Mean squared error: %.2f'
      % mean_squared_error(y_test, price_pred))

#cod

print('Coefficient of determination: %.2f'
      % r2_score(y_test, price_pred))
y_pred=model.predict(X_test)
print(y_pred)
pred_y_df=pd.DataFrame({'Actual Value':y_test,'Predicted Value':y_pred, 'Difference':
y_test-y_pred})
pred_y_df[0:30]

```