Melbourne Estates Company

Sundeep Baath
Id- 0680218
ssbaath@lakeheadu.ca

Aswin Suresh Nair
Id-10911474
asuresh1@lakeheadu.ca

Adityasingh Thakur
Id- 0904521
athakur1@lakeheadu.ca

ABSTRACT

In this paper, we describe how to deploy a machine-learning model on an IAAS (Infrastructure as a service).

INTRODUCTION

The proceedings are description of the Melbourne Estate Company website whose purpose is to give an estimation of properties in the suburbs of Melbourne, Australia. There is a total of 297 suburbs accounted in this project. The Dataset used in this project is referred from a website named Kaggle.com. To get an estimation of a particular area we can choose number of bedrooms, bathrooms, cars and the desired area (in square meter). The key concept used to develop this website is a machine-learning model which gives the estimated prices of the area in the suburbs of Melbourne based on the dataset used.

System Description

Overview- In this project the following tools have been used to develop the ML model and the website

- Pycharm -python
- Html; CSS; FLASK framework
- JS-Java Script
- Google Compute Engine
- Google Kubernetes Engine
- Dockerhub

Description- Pycharm and flask have been used to create the main codes for the website.

Flask is a small and powerful web framework for Python.

Html is used to design the webpages.

JS (Java Script) is used for the back handling of the website (i.e. to run and develop).

Google Compute Engine/Google Kubernetes Engine are used as Virtual Machines through which we can run the website.

Dockerhub is used to create the docker image by which the installation of the app is done on GCE, GKE.

Machine Learning Model- To train the model, the app **uses RandomForestRegressor Model** (Used when dependent variable/variables are continuous, Random Forest Regressor is mainly used when there is a lot of data and an average is to be taken of the unique dataset. RandomForest maintains average of data even if there are fluctuations of data); **DecisionTreeRegressor Model** (Used to classify, variable screening and feature selection). **Linear regression model** gives the statistical calculations for the **dependent** and the **independent** variables. (Independent variables are – **bedroom, bathroom, car, area**; Dependent variable is **estimated price**.)

Other tools used for this project **are price predict function, columns json, price prickle and the CSV files (referred data)** all help to achieve price prediction.

Scope and Limitations – The real companies in the world have their data is reserved. Hence in this project the dataset referred from Kaggle.com had to be cleaned (i.e. reduction of noise from the data). A lot of information had to be removed from the dataset as we did not require the entire dataset. Refer to fig.1 and fig.2 for noise reduction of the dataset.

The calculations are also corrected with the help of Google collab as there was reduction of data.

Please refer to fig.3 and fig.4

The machine-learning model used in this project is 70-75 % reliable because of the dataset provided is a made up one for understanding and learning purpose only.

1. FIGURES

Suburb	Address	Rooms	Type	Price	Method	SellerG	Date	Distance	Postcode	Bedroom2	Bathroom	Car	Landsize	BuildingAr	YearBuilt	CouncilAre	Latitude	Longitude	Regionname	Propertycount
Abbotsford	85 Turner St		2 h	1480000	S	Biggin	3/12/2016	2.5	3067	2	1	1	202			Yarra	-37.7996	144.9984	Northern Metropolitan	4019
Abbotsford	25 Bloomburg St		2 h	1035000	S	Biggin	4/02/2016	2.5	3067	2	1	0	156	79	1900	Yarra	-37.8079	144.9934	Northern Metropolitan	4019

fig.1

		Suburb	Price	Bedroom2	Bathroom	Car	BuildingAr	price_pe
0		Abbotsford	1035000	2	1	0	79	13101.2
1		Abbotsford	1465000	3	2	0	150	9766.66

fig.2

```
df.Bedroom2.unique()
```

```
array([2., 3., 4., 1., 6., 5., 9., 0., 7., 8.])
```

```
[ ] df.pivot_table(index=['Bedroom2'], aggfunc='size')
```

```
Bedroom2
0.0       5
1.0     356
2.0    1910
3.0    2964
4.0    1494
5.0     328
6.0      35
7.0       4
8.0       2
9.0       3
dtype: int64
```

```
[ ] df1=df[(df.Bedroom2 != 0) & (df.Bedroom2 != 8) & (df.Bedroom2 != 9) ]
```

```
[ ] df1.pivot_table(index=['Bedroom2'], aggfunc='size')
```

```
Bedroom2
1.0     356
2.0    1910
3.0    2964
4.0    1494
5.0     328
6.0      35
7.0       4
dtype: int64
```

```
[ ]  def predict_price(Suburb,Bedroom2,Bathroom,Car,BuildingArea):
        loc_index = np.where(x.columns==Suburb)[0][0]

        z = np.zeros(len(x.columns))
        z[0] = Bedroom2
        z[1] = Bathroom
        z[2] = Car
        z[3] = BuildingArea
        if loc_index >= 0:
            z[loc_index] = 1

        return lr.predict([z])[0]
```

```
[ ]  predict_price('Abbotsford',2, 1,0,79)
```

```
852190.8567475517
```

Price predict function

2. Deployment Results

The heading of a section should be in Times New Roman 12-point bold in all-capitals flush left with an additional 6-points of white space above the section head. Sections and subsequent sub-sections should be numbered and flush left. For a section head and a subsection head together (such as Section 3 and subsection 3.1), use no additional space above the subsection head.

3. CONCLUSION

The Objective of the project to implement a machine-learning model and to implement it on IAAS model has been tested and successful. Docker image of the app created on Pycharm is successful. The model is implemented on Google compute engine and Google Kubernetes engine.

4. REFERENCES

[1] Pino, T. (2018, October 14). Melbourne Housing Market.Retrieved November 22, 2020, from https://www.kaggle.com/anthonypino/melbourne-housing-market

[2] Google Colaboratory. (n.d.). Retrieved November 22, 2020, from https://colab.research.google.com/drive/1KcS5w-cUnxhL9BOl9mfTilLFbdzO98yU

[3] Ssbaath. (n.d.). Ssbaath/Melbourne. Retrieved November 22, 2020, from https://github.com/ssbaath/Melbourne

[4] How TO - Social Media Buttons. (n.d.). Retrieved November 22, 2020, from https://www.w3schools.com/howto/howto_css_social_media_buttons.asp