A short project Proposal on

Machine Learning Application- Predicting Rent of House in Dublin

Trinity College Dublin
Coláiste na Tríonóide, Baile Átha Cliath
The University of Dublin

WEEK 5 ASSIGNMENT

Submitted to:

Prof. Douglas Leith

Submitted by: Group 14

Ayush Kalra	Student Id: 20327429
Jing Wang	Student Id: 21330849
Kaaviya Karunanidhi	Student Id: 21331515

Motivation :- In today's world, before buying or making a purchase a person has an idea of the price for the same but that is not the case in renting or buying a property and that too when you are an individual who has landed directly from a plane. One can find various websites to find a home or apartment for rent but none of them will tell the actual price in the area and that too given the set of requirements by the individual. With some of the listings being fake as well it is very tough to rent a home. So we propose a Machine learning model that will take the requirements of the individual and based on that we will be generating an estimated price of different areas. It will not only help the individual to get an idea of the price but will also help to negotiate a bit as well.

DataSet :- We plan to crawl data from daft.ie. A Web crawler, sometimes called a spider or spiderbot and often shortened to crawler, is an Internet bot that systematically browses the website. Every property has 34 features, and the target to be predicted is the rent price per month. The details of the 34 are shown below.

1. Dublin's postal districts

(Dublin 1, Dublin 2, Dublin 3, Dublin 4, Dublin 5, Dublin 6, Dublin 6W, Dublin 7, Dublin 8, Dublin 9, Dublin 10, Dublin 11, Dublin 12, Dublin 13, Dublin 14, Dublin 15, Dublin 16, Dublin17, Dublin 18, Dublin 20, Dublin 22, Dublin 24. 23 districts in total, this will be saved as one hot data, so 23 features)

2. Number of bedrooms (1 room: 1, 2 rooms: 2, 3 rooms: 3 ……)

3. Number of bathrooms (1 room: 1, 2 rooms: 2, 3 rooms: 3 ……)

4. Furnished or not (Not furnished: 0, Furnished: 1)

5. Studio or apartment (Studio: 0, Apartment: 1)

6. Whether is only for students: (No: 0, Yes: 1)

7. Lease (None: 0, With lease: 1)

8. Whether is near to public transport (No: 0, Yes: 1)

9. Whether is near to shops (No: 0, Yes: 1)

10. Whether has parking lot (No: 0, Yes: 1)

11. Whether has the Internet (No: 0, Yes: 1)

12. Whether has Garden / Patio / Balcony (No: 0, Yes: 1)

Method :- To train the dataset from Daft and to predict the price of houses in Dublin we are planning to use the below Regression models:

Multivariate Linear Regression is used as it is the basic model which fits training data sets in order to predict the house prices from test data.

SVR (Support Vector Regression) is used where the hyperplane line is used for separation between classes with higher dimension and actual dimension and is used to predict the target value (continuous output) of the house prices from the data.

Lasso Regression is used to train our dataset in order to reduce the overfitting, model complexity and in feature selection and Ridge Regression is used where regularization parameters can be used for house price prediction.

XGBoost model is also used as it is considered to have good speed and the algorithm provides more stable prediction with lesser variance.

Intended Experiments :- Using the above regression methods we are going to train and test the dataset using google colab or jupyter notebook.

Optimal hyperparameters will be chosen for training the data in order to control the overfitting and underfitting of the models and primary metrics - Accuracy, Precision, Mean Squared Error, Root mean square, Mean absolute percentage error will be calculated.

Above regression models are compared with each other to find out the best model and also it will be compared with a Baseline Regressor in order to get the best model that accurately predicts the house prices in Dublin.