

HOUSE PRICE PREDICTION

SUBMITTED BY: NANSHIN LAZAK

ACKNOWLEDGEMENT

Fliprobo Live Session expositions on Deep Learning

Stackoverflow

Vidyaanalytics

kaggle

INTRODUCTION

The Dataset considered many attributes that affect the cost of a house. The project analysis these attributes to find how they contribute to the selling price of a house. The analysis will be utilizing the past data to be able to predict house prices in the future.

The domain related to this project is the business industry, specifically the property market.

REVIEW OF LITERATURE

Both Linear Regression Models and Deep Learning was applied. With sales price as the target variable, Linear Regression was most appropriate for the continuous variable. The resources researched on this topic include;

Deep Learning in python by edureka YouTube video which explained Deep learning models.

The Kaggle competition submission for House price prediction using keras.

MOTIVATION

This Project may serve as a guide for decision towards the future selling prices of houses. Such decisions are critical to both buyers and sellers so that the risk involved is greatly reduced. Investment decision need substantial information else devastating consequences may evolve.

MODELLING

The Linear Regression, DecisionTreeRegressor and Keras models where used for the predictions.

DATA SOURCES

The source of data is Fliprobo tech. it was obtained as the House price prediction project as a csv file.

DATA PROCESSING

Data had numerous missing values, which were replaced. Columns with more than 50% of missing values were dropped. Data also contained both categorical and numerical values. Categorical values were replaced using the one-hot encoding. Data was also normalized using the standard scaler.

HARDWARE AND SOFTWARE USED

A windows 10 laptop is utilized.

The software utilized for this project include;

The Python programming language

Jupyter Notebook which gives a platform for input and output of the programs.

Pandas library which is a tool for analyzing and manipulating data in the python programming language

Numpy library which is a tool for the analysis of numerical data in the python programming language

Matplot library which is a tool for analysis graphical data in the form of charts and plots

Statsmodels Library which is a tool for statistical analysis of the data

Seaborn library for data visualization

Scikit Learn library for data processing

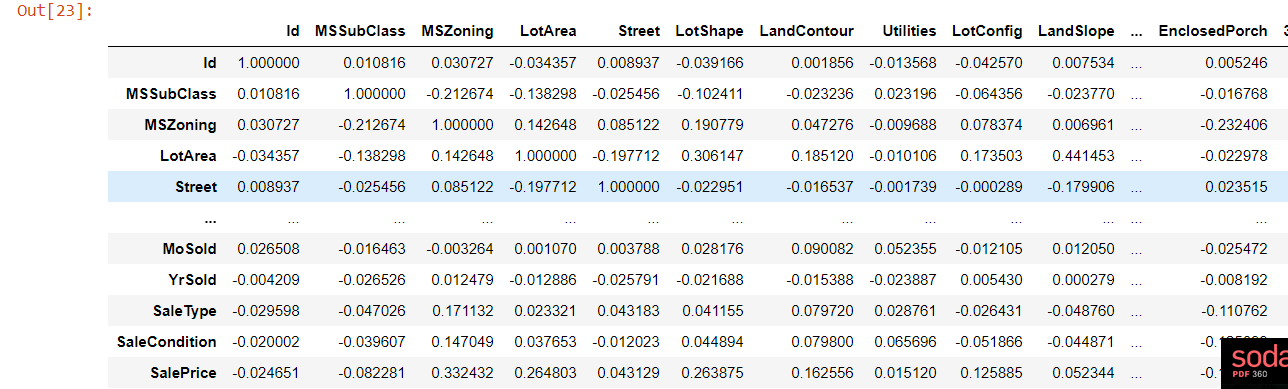
Tensorflow Library used for deep learning models such as keras and sequential.

Google: used for stuffing the internet for obtain information over grey areas that arouse in the cause of the project execution.

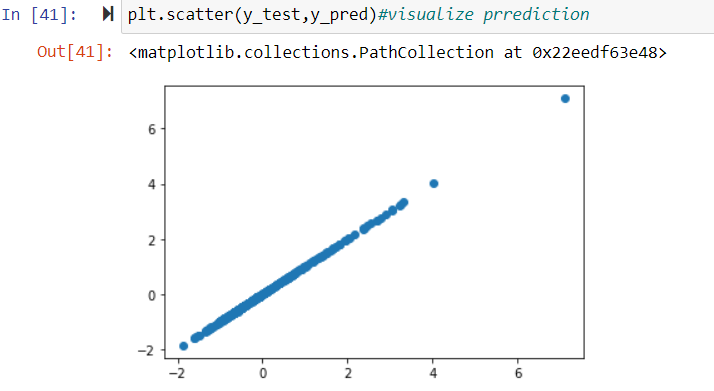
Microsoft Office Package specifically the Microsoft word used for the compilation of the project report

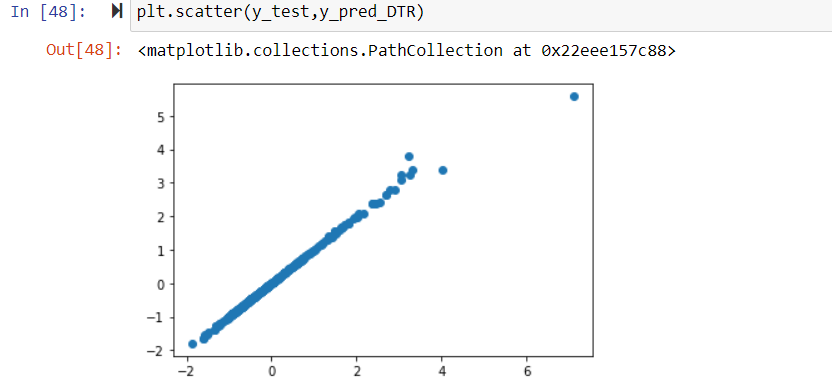
MODELS/DEVELOPMENT AND EVALUATION

These problem was solved by uploading the dataset set given unto the jupyter notebook, than visualizing the data and other details such as shape, description, datatypes, information, missing values etc. the data was plotted to check correlation between variables.



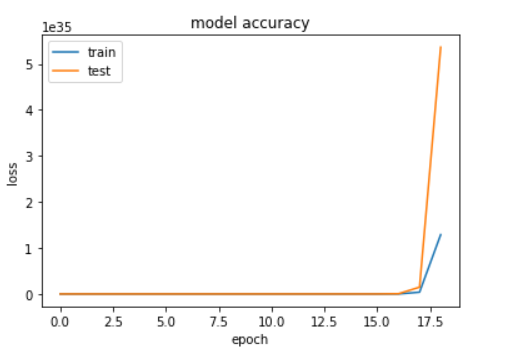
Data was normalized into machine language and split into train an tests sets to be fitted into the models. The linear regression was able to predict with an accuracy of 99%, while the DecisionTreeRegressor predicted with an accuracy of 99%.





Both models predict with 99% accuracy and the regularization models also prove the same.

The Deep learning models i.e tensorflow is then applied using keras ad sequential, with the activation functions of relu and softmax. The training sets are fitted to the model in batches and then a prediction is made.



CONCLUSION

In this article, data on attributes that contribute to the selling price of houses in the past was used to predict selling prices of houses in the future. The models applied for the prediction performed well.

* I learnt more about data preprocessing.
* I learnt the implementation of deep learning models i.e tensorflow, keras, sequential.
* I did not clearly comprehend how to visualize the accuracy of my models.
* I did not also comprehend how to interprete the output from my predictions with the deep learning models.