**Zillow Prize: Zillow’s Home Value Prediction (Zestimate)**

Applied Artificial Intelligence - CS514

Project – 5

Codename: EZIO

**Rules and Guidelines:**

The Zillow Prize contest competition, sponsored by Zillow, Inc. (“Sponsor”) is open to all individuals over the age of 18 at the time of entry. The competition will contain two rounds, one public and one private. Each round will have separate datasets, submission deadlines and instructions on how to participate. The instructions on how to participate in each round are listed below. Capitalized terms used but not defined herein have the meanings assigned to them in the Zillow Prize competition Official Rules.

**Overview:**

Submissions are evaluated on Mean Absolute Error between the predicted log error and the actual log error. The log error is defined as

Logerror = log(Zestimate)−log(SalePrice)logerror=log(Zestimate)−log(SalePrice)

and it is recorded in the transactions training data. If a transaction didn't happen for a property during that period, that row is ignored and not counted in the calculation of MAE.

For each property (unique parcelid), you must predict a log error for each time point. You should be predicting 6 timepoints: October 2016 (201610), November 2016 (201611), December 2016 (201612), October 2017 (201710), November 2017 (201711), and December 2017 (201712). The file should contain a header and have the following format:

ParcelId,201610,201611,201612,201710,201711,201712

10754147,0.1234,1.2234, -1.3012,1.4012,0.8642-3.1412

10759547,0,0,0,0,0,0

etc.

**DATA descriptions:**

*properties\_2016.csv* - all the properties with their home features for 2016. Note: Some 2017 new properties don't have any data yet except for their parcel id’s. Those data points should be populated when properties\_2017.csv is available.

*properties\_2017.csv* - all the properties with their home features for 2017 (released on 10/2/2017)

*train\_2016.csv* - the training set with transactions from 1/1/2016 to 12/31/2016

*train\_2017.csv* - the training set with transactions from 1/1/2017 to 9/15/2017 (released on 10/2/2017)

*sample\_submission.csv* - a sample submission file in the correct format

**Model:**

* The model is built using different types of neural networks.
* Makes use of various hyper parameters which are decided on an experimental basis.
* Data is processed to retain a few necessary columns which are then fed to the model.
* The model predicts the logerror based on the training and test data.

**Requirements:**

* Python 3.6
* Tensorflow latest version
* Keras latest verion (tensorflow backend)

• Sklearn, pandas, numpy

• Jupyter Notebook

**Usage:**

* Set up the required environment
* Open the notebook and ensure all dependencies are satisfied
* Ensure that the required data is present in the correct path
* Run the entire notebook and observe the outputs of each cell
* The final predication values will be written to the output csv file that will be generated.

**Results:**

As per kaggle’s score estimator, this model obtained a private score of 0.0768606 and public score of 0.0657930.