Author response to reviews of

Manuscript Paper 147

Urban and socio-economic correlates of property price evolution: Application to Dublin's costal area

submitted to *Special Session-Environmental and Geo-spatial Data Analytics (EnGeoData) of the 7th IEEE International Conference on Data Science and Advanced Analytics (DSAA'2020).*

RC: *Reviewer Comment*   AR: Author Response   ☐ Manuscript text

Dear Antonio, David, Mathieu and Maguelonne,

Thank you to consider the manuscript entitled "Urban and socio-economic correlates of property price evolution: Application to Dublin's costal area" in EnGeoData 2020. We would also like to take this opportunity to express our thanks to the reviewers for the positive feedback and helpful comments for correction or modification.

We believe have resulted in an improved revised manuscript, which you will find uploaded alongside this document. The manuscript has been revised to address the reviewer comments, which are appended alongside our responses to this letter.

1. Reviewer #1

RC: **Page 1 : « 10389 houses » in the Abstract and « 10395 properties » in the Method section and « 10387 Properties » in Discussion section ?**

RC: **Page 2 : « an eXtreme Gradient Boosting (XGBoost) regression model was calculated to examine the extent to which house prices rely on urban and on socio-economic features ». Could you explain wether you worked on the XGBoost algorithm or not ? What was your contribution to develop/adapt/or parameterize this algorithm ?**

RC: **I suppose that Urban features importance is computed by the XGBoost model (Table I). Could you explain this stage and comment results ? Embassy importance is detected thanks to statistical computations. Is it possible to comment and validate it by qualitative interviews with experts ?**

RC: **Page 4 : Similarly, I couldn't understand how Socio-Economic features importance is computed (Table II). Could you explain this stage and comment results ? Religion importance is detected thanks to statistical computations. Is it possible to comment and validate it by qualitative interviews with experts ?**

RC: **« Embassies are in general located in the most expensive areas of cities ». Are the ambassies installed in expensive areas or are the areas expensive because of their embassies ?**

AR: \RC{Globally, it is difficult to evaluate the added-value of this research work :

- Is it a proposal of a new regression model ? If yes, you should better explain XGBoost and its originality
- Is it a experiment of an existing regression model applied to housing prices ? If yes, you should comment the results on the 20% of the original dataset : are the price predictions relevant ? Is this approach better than existing ones for price prediction ? Number of parameters ? Computation speed ? etc.}

RC: **What was not possible before and what is possible now thanks to this work ?**

RC: **Typos : Page 1 : « 10395 property »**

AR:

2. Reviewer #2

RC: **This paper presents machine learning experiments to estimate price changes in the real estate market.**

The main method is the XGBoost algorithm. Perhaps more methods could have been tried, including some simple baseline to put the presented results into perspective.

The focus of the paper in on the contribution of the proposed features.

There in not much novelty, but the experiments are well executed and the results are convincing. Real data is used for training and testing.

AR:

3. Reviewer #3

RC: **The topic of the paper fits well with the special session main topics. Below some suggestions of improvement that can be made to this contribution.**

RC: **Concerning the following statement: "techniques used usually include artificial neural networks due to the volatility of the market", even though it is backed up by bibliographic references, it would be worth further explaining it. The same applies to the following statement "While these methods are efficient in a non-restricted space, they have limitations when they are used in coastal areas.", explain why.**

RC: **There are some issues related to the data acquisition process and the possible presence of errors. You noted that errors may occur when the data is being filed, could you provide some calculated information about the quality of the data used based on missing values, incoherent values, etc. Also, property addresses were geocoded relying on OSM geocoder, can you estimate accuracy of the process? The quality of OSM data is usually weak, what are the sources reported on the OSM urban features you extracted (source key: https://wiki.openstreetmap.org/wiki/Key:source)?**

RC: **Input data is not available in the github repository given by the authors, can we suppose it is due to restrictions to share the data, if so, this should be said clearly.**

RC: **Add URL to point to the Central Statistics Office and All-Island Research Observatory where data is available.**

PAR: explicit acronym "SD".

AR: