Comparing the determinants of house prices in selected Irish and English cities using Machine Learning Techniques

Kian Sweeney
kian.sweeney27@mail.dcu.ie

School of Computing, Dublin City University, Ireland

Abstract. Being able to accurately and efficiently value properties and understand the determinants of these prices is key for anyone looking to purchase a home. In Ireland the housing market has seen a resurgence in the last decade with prices rising rapidly year on year. Focusing on areas where prices are increasing noticeably can give a good understanding of what the market strongly desires when purchasing a home over time. Using machine learning techniques we build a price prediction model for areas of Dublin with increasing market prices from 2014 to 2021 and explain what is driving these price increases through explainable artificial intelligence (AI). We attempt to develop both an accurate prediction firstly and understand what heavily influences our respective model to see the determinants of house prices in this area. We take this model and apply it to another major Irish city and also try a novel cross country comparison to a similarly sized city to further validate our model and try to see do buyers in different cities / countries look for similar features when buying a property in their respective city.

Keywords: Machine Learning · Housing · Prediction · AutoML · Explainable AI

1 Introduction

The Irish house market has seen an marked rise in prices since prices reached a nadir in 2013. [1] notes that since a 54% decrease in house prices from 2007 to 2013, prices since 2013 have increased by 50% to the year ending 2021, akin to similar prices rises before the economic crash in 2008. This is in large part due to increased demand for housing from 2013 onwards but supply has been an issue for the Irish housing market. [2] highlights that with an ever increasing population housing supply needs to keep up but in Ireland this appears not to be the case. Since the beginning of the coronavirus pandemic the housing market supply has been dwindling and prices in major Irish cities reflect that [3]. The Irish housing market sees a preference for home ownership and what can be seen as a lack of repeat sales, with home ownership rates at 70.6% at the end of 2020 [4] and the average house changing hands once every 60 years [5]. However, with the current state of the Irish housing market, this high rate of home ownership is unlikely to continue. [6] highlights that many young people in Ireland will never be able to purchase their own homes and that even with increased house supply, new supply is being sold or rented at extortionate prices for a majority of those looking to purchase a home. Housing prices have a direct impact on our standards of living, with it being noted that rising house prices can create a greater societal divide and make it difficult for non home-owners (typically young people) to make a living [7]. Similarly, a downturn in house prices can lead to economic recession due to loans on properties built by large companies and affect consumer confidence in the economy [7], [8]. As a result, this paper postulates to examine the following:

- Develop an automated valuation method (AVM) for publicly available Irish housing data.
- Accurately predict property prices in a selected city in Ireland with rising real-time prices.
- Explain the determinants between these rising costs and compare to another city in Ireland to see if these determinants have a similar impact. Examine if our model has similarly accurate results here.
- Compare to a similarly sized city in the UK to further validate the predictive performance of our model and further examine house price determinants in another country.

2 Related Works

There is limited literature related that applies data science to the Irish housing market or the UK housing market. There is some very relevant work in [9], which focuses on a dataset comprised of apartment sales in county Dublin for 2018. A generalized linear model (GAM) is deployed here to accurately predict the prices of these apartments as they seek to explain the impact an address can have on apartment prices in Dublin. Traditional hedonic models along with more modern ML approaches are compared here with a GAM and extensive text mining is done to extract more features from the dataset. We see the use of machine learning approaches applied to a classification task in Fairfax, Virginia, USA in [10]. This combined three datasets into one, focusing on structural attributes, mortgage ratings dataset and public school ratings dataset. Overall, there is a high number of features used here (28). There was originally 76 features but through t-tests and stepwise logistic regression, the number of features was cut down based on feature importance. It also contained a lot of manually inputted data thus it needed a thorough cleaning process similar to Irish property data. The first and most important attribute extracted, which subsequently determined the class of each record, was the field indicating if the closing price was higher or lower than the listing price. A 10 fold cross validation approach was used with a very strong error rate of below 25% on average using the Ripper algorithm. AdaBoost and Bayesian were also used here. In [11] a sample of data from Alicante, Spain is used and seeks to investigate the non-linearaties in the determinants of house prices in this area. It has over 50 features and these range from locational attributes such as longitude, latitude and provincial codes among others. There is also the influence of other locational attributes such as nearby amenities and there is a strong influence of structural attributes used here but the paper also looks at socio-economic attributes such as population and economic activity. ML algorithms such as XgBoost, Linear Regression, Random Forest (RF), Neural Networks (NN) and AdaBoost are utilised here. This work also focuses on the explainability of machine learning models through Shapley (SHAP) values. We see in [12] a study of Turkish house market price determinants from 2004. It's research question looks at comparing the typical hedonic model vs NN's to find the determinants of house prices in the Turkish house market. This paper focuses on the structural attributes of a property in addition to the column rural/urban of where the property is located. It highlights feature importance using t-values here in addition to the coefficient of correlation. It uses these feature importance metrics along with splitting the data into urban and rural in helping understand and determine feature importance with regards housing prices in Turkey. It then compares the performance of the algorithms using MSE, RMSE and MAE. An ANN is proven to be far more accurate with regards to house price prediction rather than a traditional linear model in this instance.

3 Data Processing

3.1 Initial Dataset

Our initial dataset was the Irish property price register dataset which is freely available to download by the public[1]. This was initially established in 2010 and includes basic information such as the date of sale, sale price and address of all residential properties purchased in Ireland since the 1st January 2010, as declared to the revenue commissioners in Ireland for stamp duty purposes. It is manually inputted and often contains missing or faulty values. There is generally no related works covering this specific dataset. Our data was downloaded at the start of 2022 and contains all records from 2010 right through to 2021. There were approximately 500,000 rows in the dataset at the time of downloading. As per Table 1, this was the information our data contained to begin with and would need extensive processing. Similarly, we downloaded our UK housing data from the UK government website[2]. This data was well maintained and would not require the cleaning our process we outline for our Irish data below.

3.2 Data Cleaning

Our initial phase of cleaning began with removing duplicate addresses and extracting information from our date and address columns. We extracted the year

[1] https://www.propertypriceregister.ie/
[2] https://www.gov.uk/government/statistical-data-sets/price-paid-data-downloads

Table 1. Irish Property Price Register Dataset Columns

Variable	Description
Date Of Sale	Column with date of sale in format dd/mm/yyyy.
Address	The full address provided to the revenue commissioners here. No set format.
Postal Code	An approximate value for postcode area (e.g. D6W). Only covers a section of Dublin based properties sold, containing a lot of null values.
County	The county in which the property was sold.
Price (€)	The sale price of a property in string format provided to the revenue commissioners.
Not Full Market Value	If property is sold at what is seen as market price.
VAT Exclusive	Is VAT rate of 13.5% included in sale price.
Description of Property	Specifies whether a property is second hand or newly built.
Property Size Description	Gives rough details of size of property. Has three values: less than $38m^2$, between $38m^2$ and $125m^2$ and greater than $125m^2$.

and month for our date of sale column and created our month and year columns, removing date of sale in the process. We removed the postal code column due to the presence of extensive null values and began trying to construct our town column. Typically, Irish house addresses are in the following format:

House Number / House Name, Townland, Town/City, County

However, as mentioned earlier the address column is manually inputted with no common format. We firstly took the last part of the address string and then treated each county on a county by county basis. We took a list of all towns and villages in each county found on Wikipedia[3] and if our original town value matched this it remained unchanged otherwise we search the second to last string and if that matches a value in the counties list of towns we designate the town as that. Addresses we fail to identify are labelled like:

County Name + County

[3] `https://en.wikipedia.org/wiki/Category:Towns_and_villages_in_Dublin_(city)`

Rows containing such values are then later removed from the data. Next we looked to sort faulty values in each county. We found the presence of addresses with faulty counties such as the following:

No102 Aughnaharna, Summerhill, Portlaoise, Co.Leitrim

We removed these faulty town values by checking if the town it assigned (in this case Portlaoise), actually matched the list of towns in the county (in this case Leitrim) provided beforehand. If there was no match, all rows containing this specific town and county combination would be removed. We also handled cases when a town name appeared in multiple counties such as Blackrock, and assigned the name of this town plus the county it was located in as the town value.

3.3 Text Mining and Feature Creation

Once our data was cleaned we began the process of trying to extract more information from the dataset. The first step of this was to extract structural information from our dataset. We found what properties sold were apartments from text mining firstly. We searched for the presence of strings such as "flat", "apartment" or "apt" among others and were able to label these properties as apartments as a result. We used text mining to create a "neighbourhood" feature also. This involved mining the address column once again. To do this our data required further cleaning. We fixed abbreviations such "rd" (road) among others in our address column and misplaced capital values in our strings (e.g. "12 St Finian'S Way") which would be treated separately if we did not address the capital "s" used here. Once this was done, we would take the start of the address and create a column based on this value. The idea being that properties in the same estate / area would carry similar values in price. Looking at the example below:

5 Braemor Drive, Churchtown, Co.Dublin

We would extract the part of the string that is next to what the town is denoted as. So in this case here we would have the town as Churchtown and our function would extract "5 Braemor Drive" from the string. If we had an address that contained "4 Braemor Drive" also we could then extract this from the string also and we would have a common neighbourhood once the number of the property was removed. As per [9], it is noted that property agents base their price estimates on opinion and prices of similar nearby sold properties. Our next step was to remove any neighbourhoods with numbers at the start of string to create groupings of unique values. We then looked at some major towns and cities with incorrect values for neighbourhoods within them. Due to the formatting of some addresses with some strings ending in "Cork" and "Cork City" for example some addresses with Cork may have been from around county Cork and not the city so some of these needed to be removed from our dataset.

We then created an eircode column. We gathered the list of towns and cities used in our data and matched them to their eircode routing key. This was the first three values of an eircode in Ireland and would give properties in the same locality but not necessarily the same town a means of regionalizing our data. Lastly, we attempted to adjust these prices for current market price. A property sold in 2010 is highly likely to be a significantly lower value as a similar property sold in a similar area in 2021 due to inflation and changing economic situations. To adjust our prices we would use the residential property price index[4]. The Irish residential property price index measures different regions and types of property but the majority of this only dates back to 2017 on a quarterly basis. Instead we would take the quarterly values for all properties sold in Ireland going back to the start of 2010 and adjust accordingly. The formula for this adjustment can be seen below:

Sale Price x (Current Quarterly House Price Index ÷ Quarterly House Price Index from time of sale) = Current Market Value

We created a quarter column for all our values and took the latest value for our residential property price index (2021 Q4) and created our "PriceInflation" column. Once this was concluded, we commenced our exploratory analysis.

3.4 Exploratory Analysis

The first step of our exploratory analysis was to deal with outliers at both the upper and lower end. We plotted a box plot and seen a small selection of values over €1.75 million and below €30,000 with our lowest value being €5,080 and our highest being over €8 million. We removed values less than €30000 and above €1.75 million from our data. This left us with approximately 250,000 rows of data to explore further. We then began some basic exploratory analysis. We investigated the mean and median prices for both standard and adjusted prices nationwide and on a county by county basis to gain an understanding of the Irish house market from 2010 to 2021. Using these two plots we saw the nadir of house prices in Ireland between 2013 and 2014 for real prices and adjusted prices in 2014 alone. We saw that our adjusted house prices peaked in 2010 and dropped to their nadir in 2014 and these prices never returned to near the level of prices seen at the start of the decade. From this we investigated what counties showed an increase in mean prices from the period 2014 to 2021 for both actual sale prices and adjusted values. Looking at actual sale values there was nothing interesting to note with all counties a minimum mean price rise of approximately €50000. However, our prices adjusted using the residential property price index saw some interesting results which can be seen in Figure 1. Counties we typically associate with higher house prices in Ireland such as Dublin showed noticeable declines in mean current market prices. An overwhelming majority of counties based in rural Ireland had actually seen an increase in the mean of house prices over the period 2014 to 2021.

[4] https://www.cso.ie/en/statistics/prices/residentialpropertypriceindex/

Fig. 1. Counties with increasing / decreasing current market value means 2014-2021

We delved further into this to see how many towns in each county seen an a mean increase in adjusted prices from 2014 to 2021. We would express this as a percentage of the number of towns in a given county and this would allow to see if prices were actually increasing across a county or if there were outliers in some of these counties data influencing the results. We plotted our findings on a choropleth map using the GeoPandas[5] library. From this we found some interesting values here. As per 6, 6, Dublin county had one of the lowest values for this across Ireland despite it being regarded as "seriously unaffordable" in [13], which one would assume would mean increasing real-time prices and demand. With a lower number of areas in Dublin seeing real-time price increases, we felt it would be worth investigating these areas to see what is determining these rises in house prices. Figure 2 shows the areas in question, along with the number of properties sold in each of these areas from 2014-2021 and the mean current market price increase of properties sold in these areas from 2014 to 2021.

Our next step was to give this selected subset of data longitude and latitude information. There are extensive solutions for geocoding data such as "geocode.xyz"[6], which achieve strong accuracy but entail considerable costs. Due to limited resources we utilized the Nominatim[7] python library which is based off OpenStreetMap data. This libraries accuracy can vary but this is the most accessible geocoding library and does not entail any costs. We started with approximately 7,500 rows of data and after removing rows we could not obtain longitudinal and latitudinal co-ordinates for, we had approximately 5,600 rows of data to work with.

In [9], locational distances to schools among others were used as extra features to improve predictions and here we look to use some of these features and investigate other points of interest. Using publicly available data taken from the

[5] https://geopandas.org/en/stable/
[6] https://geocode.xyz/
[7] https://nominatim.org/

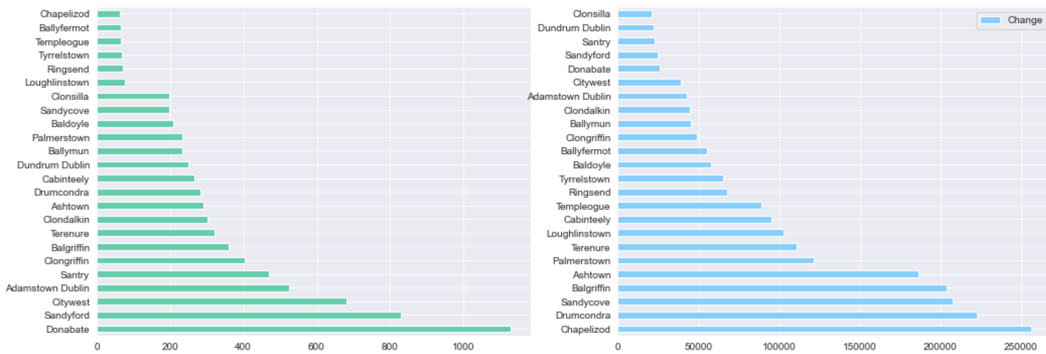

Fig. 2. Left figure shows number of properties sold per area over period 2014-2021, right figure shows mean price change over same time period.

Irish government website[8], we created distance columns to the internation financial services center (IFSC), distance to nearest primary and secondary schools, distance to nearest LUAS stop and distance to nearest park, similar to the work of [9]. In addition to this we investigated the distance to nearest beach, shopping center and multi-storey car park. We tested for outliers and any values over 25km away from the IFSC were removed from the data. We had no issues with incorrect co-ordinates after this. After examining correlation coefficients with sale prices, we selected the features that had the most predictive power for our experiments. These can be seen in Table 2 along with a brief description of the features below.

4 Results and Evaluation

Instead of deploying a typical training to test split on our dataset, we utilize a fixed estimation window scheme, similar to that seen in [14]. Due to our dataset spanning several years we felt it would be a more appropriate approach. This was because if we used a typical training to test split, we could have data from 2021 used in our training set to be tested on 2014 housing sales. We instead postulate using the previous three years of data to predict a year ahead. This would be using 2014, 2015, 2016 sales data to predict 2017 sales and so forth. We target encoded all categorical columns in our dataset as this would have the best performance for our model. We then began to trial some of the algorithms trialled in [10], [11] and [12]. These included LightGBM, CatBoost, AdaBoost, MLR and Bayesian Regression. We also wanted to trial an AutoML approach to house price prediction. During our literature review we noticed there was very little works related to AutoML in house price prediction so we trialled three different

[8] https://data.gov.ie/

Table 2. Processed Dataset Columns

Variable	Description
Year	Year property sold.
Month	Month property sold.
Eircode	An approximate value for postcode area based on area code, the first three characters of postal code. Categorical.
Town	The area of the city where property is located. Categorical.
Neighbourhood	The area within a town in which the property is located. Categorical.
UsedProperty	If property was second hand. 1 if a second hand property.
Apartment	Denotes if property is an apartment, 1 if it is, 0 otherwise.
Longitude	Longitudinal point for property.
Latitude	Latitude point for property.
DistanceIFSC	Distance to IFSC from property in kilometers.
DistancePark	Distance to nearest park from property in kilometers.
DistanceSecSchool	Distance to nearest secondary school from property in kilometers.
DistanceSchool	Distance to nearest primary school in kilometers.

AutoML libraries: Ml-Jar AutoML[9], AutoViml[10] and TPOT AutoML[11]. We evaluated our results using root mean square error (RMSE), median absolute error (MEDAE), mean absolute percentage error (MAPE) and accuracy. The accuracy metric is typically used in classification tasks but here it is the number of properties that fall within $\pm 10\%$ of actual value, similar to that seen in [9] and [15]. We took the average results for each of these metrics across the five years tested on. Our results can be seen in Table 3 below:

We can see a a noticeable disparity in results achieved by two of our AutoML libraries here to the rest of our algorithms. These algorithms were all run on default setting and even with hyper-parameter tuning it is difficult to see how it could better some of these results achieved by AutoML. Both our top performing AutoML libraries here use hyper-parameter tuning to achieve the best results, in contrast to AutoViml which is tuned by the user. We enabled cat boosting here among others for our best results. Ml-Jar produces stable results and found the ensemble method to produce the best results on our data. TPOT AutoML

[9] https://supervised.mljar.com/
[10] `https://github.com/AutoViML/Auto_ViML`
[11] http://epistasislab.github.io/tpot/

Algorithm	RMSE	MEDAE	MAPE	Accuracy within 10%
TPOT AutoML	€106813	€25455.41	15.3%	58.3%
ML-Jar AutoML	€128734.05	€37786.15	19.4%	46.7%
CatBoost	€132601.20	€37183.15	19.7%	43.6%
LightGBM	€129980.10	€38752.37	19.9%	43.1%
Bayesian Ridge	€150545.72	€45116.86	22.8%	37.6%
MLR	€150171.31	€45189.11	22.9%	37.3%
AutoViml	€145852.68	€51010.50	24.1%	34.6%
AdaBoost	€147314.62	€52364.82	22.4%	33.1%

Table 3. Results for Dublin Prediction

results can vary with each run, with it being noted that for reproducibility of results this library is not ideal. This led to us running this experiment five times and taking the best results which can be seen above. TPOT AutoML found that the RF algorithm produced the best results across the five years tested. This was the same for each of the five trials we did. A detailed breakdown of our best results can be seen in Figure 3.

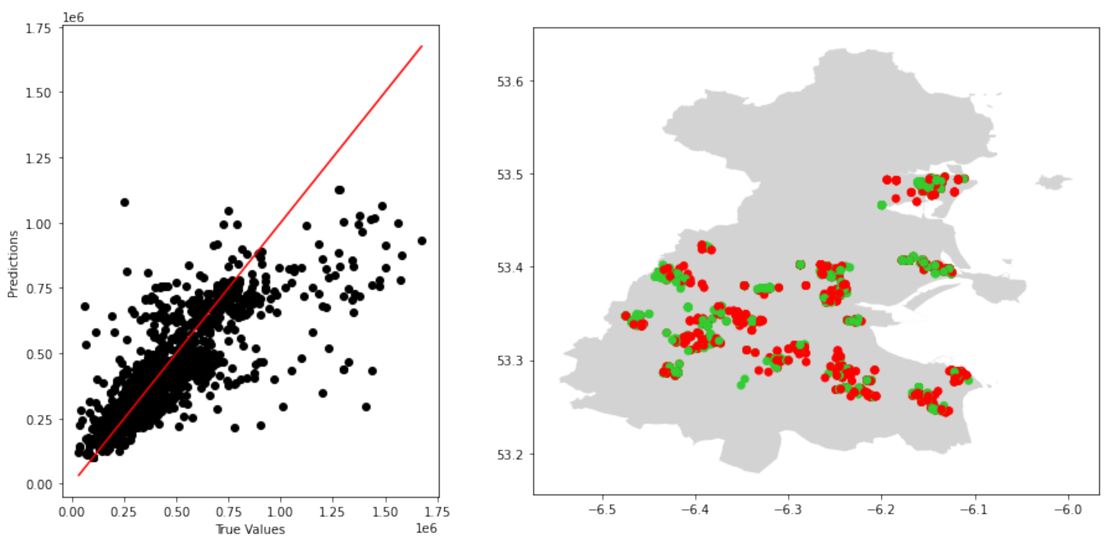

Fig. 3. Actual and Predicted results in Dublin with location of accurate predictions in green above.

It is interesting to note that our model struggles as real prices increase as can be seen in Figure 3, a trend also seen in [9], and that it appears to struggle with properties located in the south of Dublin. This would be plausible as south Dublin house prices tend to be more expensive than the rest of Dublin, as noted in [16]. We then utilized SHAP values to try to explain our best performing model's outputs. SHAP values find the variables with the highest average impact on model output. We can see our SHAP values for our Dublin predictions in Section 6, Figure 7. Across the five years we tested on the "neighbourhood" column we created was by far the most important variable in our results. As per [9], the Irish house buyer tends to place greater emphasis on the address of the property they are buying and with these SHAP values it shows our model is putting more emphasis on this column for accurate predictions. Eircode is another feature with high importance across the five years and further highlights that certain addresses in Dublin appear to have ever-increasing values. Our model achieved slightly higher accuracy than the linear GAM used in [9], which is the most comparable related works here due to it's work on the city of Dublin also.

This leads us on to investigating another major Irish city. We want to further test the validity of AVM for the Irish market and also investigate if features such as "Neighbourhood" hold the same level of importance in other locations. The city selected was Cork city, which has a much smaller urban population of around 208,000[12]. Due to the lack of similarly sized cities in Ireland compared to Dublin and also for cross country comparison we looked at a city in the UK. The UK has seen similar home-ownership rate patterns to Ireland, such as being below the average home ownership rate in the EU bloc as seen in [17] and [18]. These two works highlight that UK home ownership rates remain a few percent below their Irish counterparts, with [17] looking at before the economic crash in 2009 and after in 2013 in [18]. We selected the city of Leeds as a comparison due to it's similar population as Dublin, with an estimated urban population of 1.9 million[13]. We are only comparing performance here of our best model (TPOT AutoML) instead of all the algorithms tested to see does this model have similarly effective performance. Our results can be seen in Table 4 below.

City	RMSE	MEDAE	MAPE	Accuracy within 10%
Cork	€131184	€47487.69	32.5%	27.8%
Leeds	£77951.13	£17589	16.7%	58.7%

Table 4. Results for Cork City and Leeds

This was again run with five trials of AutoML ran and the best results taken. The RF regressor was the best model AutoML found for these cities also. It is very telling the huge difference in results obtained here. Our Cork dataset comprised of approximately 6,600 rows, similar in size to the Dublin market we focused on. We used the same features with the exception of distance to nearest secondary school. This was due to the fact there was a lack of data available to create this feature with our Cork city data. We used a feature called

[12] `https://en.wikipedia.org/wiki/Cork_(city)`
[13] `https://en.wikipedia.org/wiki/Leeds`

"DistanceCar" which was distance to nearest car park instead after inspecting correlation co-efficients on the dataset. We can see in Figure 4, that our AutoML model struggles on a wide range of values, particularly rising values once again.

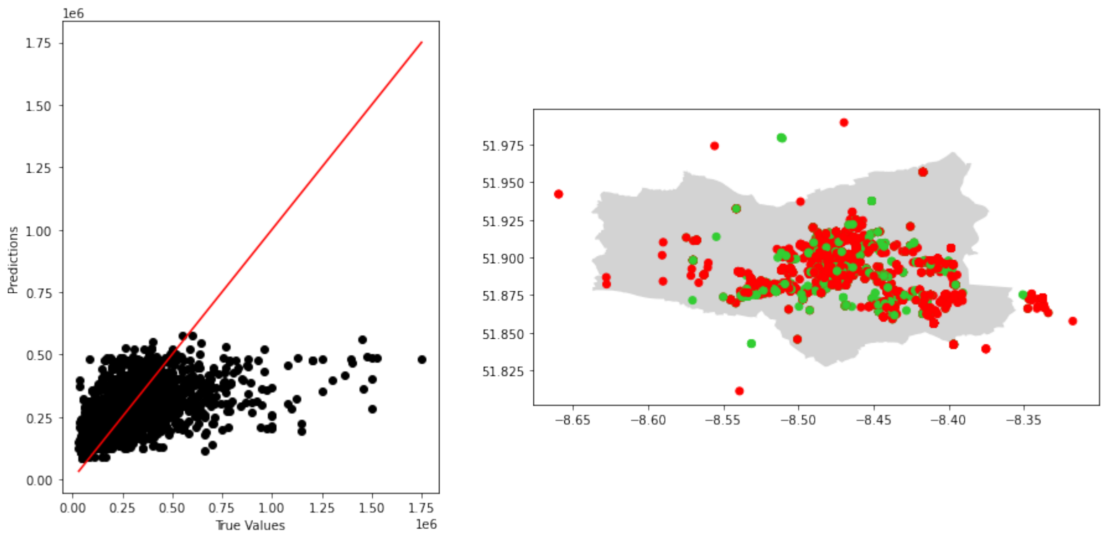

Fig. 4. Actual and Predicted results in Cork City with location of accurate predictions in green above.

We can see in Section 6, Figure 8, that our neighbourhood feature holds great importance here, other locational aspects do not hold the same importance as before in our Cork dataset. This would definitely indicate that in Dublin buyers are more willing to pay for seemingly more affluent addresses as per [9], and also that areas of Dublin have more defined valuations in specific areas than in Cork. This means that a specific neighbourhood or postcode may only contain properties of higher values in Dublin whereas in Cork property sale prices in neighbourhoods can be more varied. This may explain in part the limitations of our model in accurately predicting prices in Cork city. We see in [9] where it is noted that property valuations are commonly based on similar properties sold in the same area. Without similar properties this could lead to varying sale prices in a given area and make predictions difficult, similar to what we see in our Cork results. In contrast we see even stronger results after five trials on our data for Leeds. It achieved higher accuracy than Dublin and had similar results when currencies are converted for RMSE and MEDAE. This was based on a much larger sample size than both our other datasets also. There is more data here as we took all of the city of Leeds not a subset, we were able to geocode all the data and that houses in the UK appear to change hands more than in Ireland at once roughly every 22 and a half years for outright ownership [19].

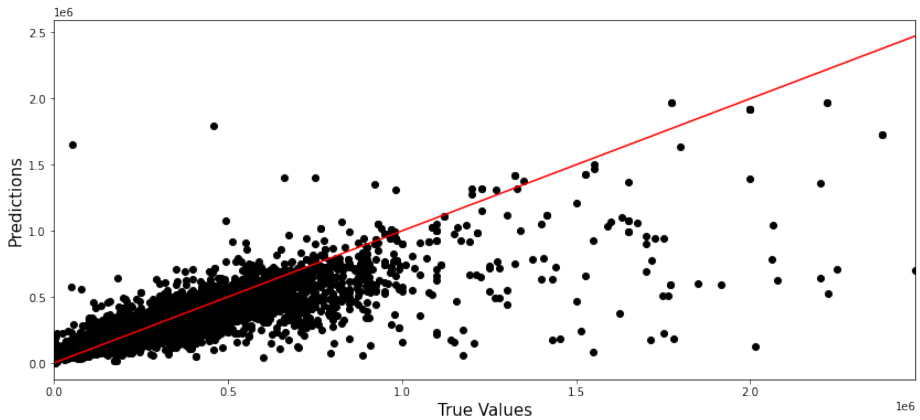

Fig. 5. Actual and Predicted results in Leeds.

As we can see in Figure 5, our TPOT RF model struggles with values at the higher end. It works well with lower values but more outliers appear as values increase. The lack of suitable shapefiles on the UK government website meant we could not plot the location of accurate / inaccurate points. When we examine our SHAP values in Section 6, Figure 9 shows the importance of location with our "Add2" feature (equivalent of neighbourhood column in our Irish data) being the most important feature here. Also, it is interesting to note that a house being semi-detached is of great importance here. We one-hot encoded the property type column in the data and this was hugely significant here. This data appears to align better with our Dublin data and suggests that larger cities appear to have more defined prices in each neighbourhood in contrast to a smaller city like Cork.

5 Conclusion and Future Work

In this paper we proposed an AVM for publicly available Irish and UK property data. We investigated the ability of machine learning algorithms in accurately predicting house prices in three major cities without detailed structural attributes. We were also able to understand the determinants of these prices through explainable AI using SHAP values that showed us what variables heavily influenced our model. We acknowledge it may be difficult to extend this approach to rural Ireland as a result of poorly inputted addresses, something also noted by [9]. As per [9], explainability of some algorithms such as RF are not entirely possible but with TPOT AutoML these are compatible with the SHAP[14] library and make explaining the model feasible. Lastly, we saw the the power of AutoML in accurate predictions over standard algorithms typically

[14] `https://shap.readthedocs.io/en/latest/index.html`

used in these house price prediction tasks. TPOT AutoML achieved accuracy 12% higher than Ml-Jar Supervised AutoML, and even this library achieved approximately 3% higher prediction accuracy than commonly used algorithms for house price prediction tasks. We achieved stronger accuracy results for all properties sold in a wide range of years than in [9], which focuses exclusively on apartments in one calendar year of sales. This dataset also utilized far more descriptive variable of properties than our dataset which is certainly a scope for future work with our AutoML approach. The limitations of our approach could be seen in our results for Cork city where it is generally regarded addresses are less well defined outside of Dublin and due to the importance of locational details in our model our results suffer as a result. We saw even with limited structural information that in our UK data, the one-hot encoded detached house column ranked highly in SHAP values highlighting it's importance in model output. By utilizing more detailed datasets such as those used in [9], [10], [11] and [12] better predictions and analysis of determinants could be achieved. We could also investigate the importance of a specific feature in properties in greater detail. For example, in [20] the importance of strong energy ratings is assessed in properties sold in Dublin and something along these lines could be worth investigating in our future work. With more detailed datasets, we could also extend this to a wider range of towns, cities and rural areas in Ireland for example.

References

1. Kieran McQuinn, Teresa Monteiro, and Conor O'Toole. House price expectations, labour market developments and the house price to rent ratio: A user cost of capital approach. *The Journal of Real Estate Finance and Economics*, 62(1):25–47, 2021.
2. Thomas Conefrey, David Staunton, et al. Population change and housing demand in Ireland. *Central Bank of Ireland Economic Letter*, 14:1–16, 2019.
3. Lalaine C. Delmendo. Ireland's housing market holding up, amidst limited supply. [Online]. Available from: `https://www.globalpropertyguide.com/Europe/Ireland/Price-History`. Accessed: 06-04-22.
4. Eurostat. Ireland home ownership rate. [Online]. Available from: `https://tradingeconomics.com/ireland/home-ownership-rate`. Accessed: 23-04-22.
5. Phil Maguire, Robert Miller, Philippe Moser, and Rebecca Maguire. A robust house price index using sparse and frugal data. *Journal of Property Research*, 33(4):293–308, 2016.
6. Dr. Rory Hearne. Why fixing Ireland's housing crisis requires a change of policy. [Online]. Available from: `https://www.maynoothuniversity.ie/research/spotlight-research/why-fixing-irelands-housing-crisis-requires-change-policy`. Accessed: 25-04-22.
7. Tejvan Pettinger. How the housing market affects the economy. [Online]. Available from: `https://www.economicshelp.org/blog/21636/housing/how-the-housing-market-affects-the-economy/`. Accessed: 25-04-22.
8. Christopher Kent and Philip Lowe. Property-price cycles and monetary policy. In *BIS Conference Papers*, volume 5, pages 239–63, 1998. [Online]. Available from: `https://www.bis.org/publ/confp05.pdf#page=251`.

9. Aoife K Hurley and James Sweeney. Irish property price estimation using a flexible geo-spatial smoothing approach: What is the impact of an address? *arXiv preprint arXiv:2108.09175*, 2021.
10. Byeonghwa Park and Jae Kwon Bae. Using machine learning algorithms for housing price prediction: The case of Fairfax County, Virginia housing data. *Expert systems with applications*, 42(6):2928–2934, 2015.
11. Juan Ramón Rico-Juan and Paloma Taltavull de La Paz. Machine learning with explainability or spatial hedonics tools? An analysis of the asking prices in the housing market in Alicante, Spain. *Expert Systems with Applications*, 171:114590, 2021.
12. Hasan Selim. Determinants of house prices in Turkey: Hedonic regression versus artificial neural network. *Expert systems with Applications*, 36(2):2843–2852, 2009.
13. Demographia. 13th Annual Demographia International Housing Affordability survey. 2017. [Online]. Available from: `http://www.demographia.com/dhi2017.pdf`.
14. George Milunovich. Forecasting Australia's real house price index: A comparison of time series and machine learning methods. *Journal of Forecasting*, 39(7):1098–1118, 2020.
15. Rotimi Boluwatife Abidoye and Albert PC Chan. Improving property valuation accuracy: A comparison of hedonic pricing model and artificial neural network. *Pacific Rim Property Research Journal*, 24(1):71–83, 2018.
16. Mirko Moro, Karen Mayor, Seán Lyons, and Richard SJ Tol. Does the housing market reflect cultural heritage? a case study of greater dublin. *Environment and Planning A*, 45(12):2884–2903, 2013.
17. Steven C Bourassa, Donald R Haurin, Patric H Hendershott, and Martin Hoesli. Determinants of the homeownership rate: An international perspective. *Journal of Housing Research*, 24(2):193–210, 2015.
18. Sylvain Bouyon et al. *Recent trends in EU home ownership.* ECRI, 2015.
19. Statista Research Department. Average number of years in current home in england between 2010 and 2021, by tenure. [Online]. Available from: `https://www.statista.com/statistics/755763/number-of-years-in-current-home-england/`. Accessed: 26-04-22.
20. Marie Hyland, Ronan C Lyons, and Seán Lyons. The value of domestic building energy efficiency—evidence from ireland. *Energy economics*, 40:943–952, 2013.

6 Appendix

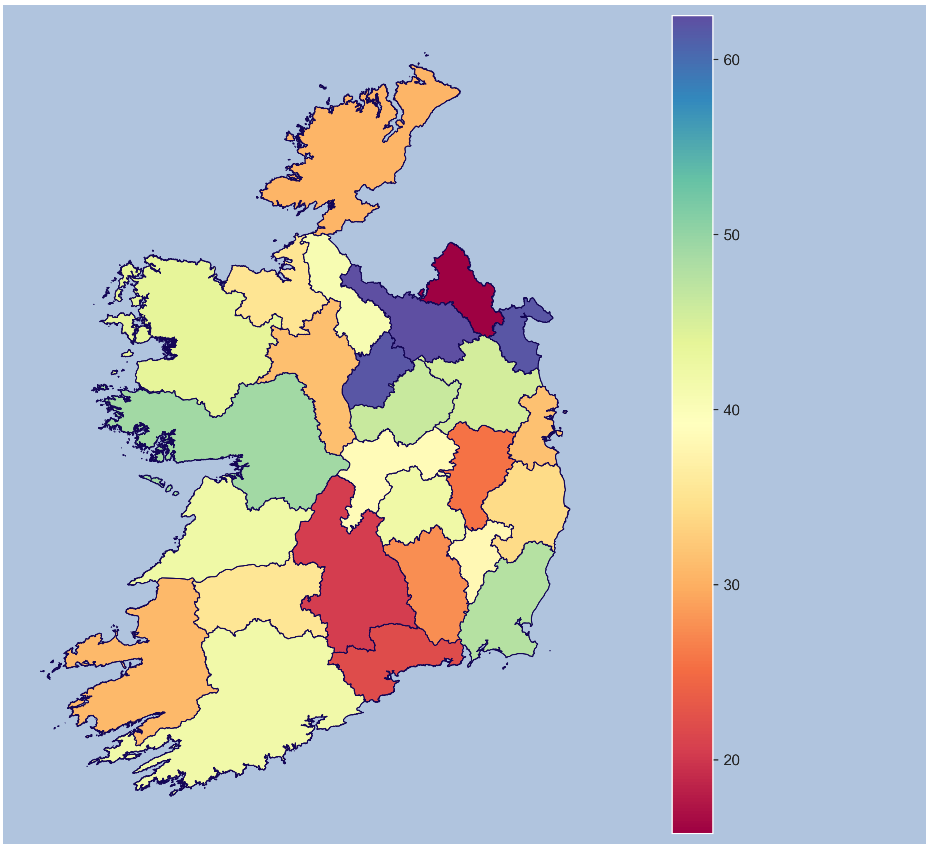

Fig. 6. Percentage of Towns per County with current market value increases from 2014-2021.

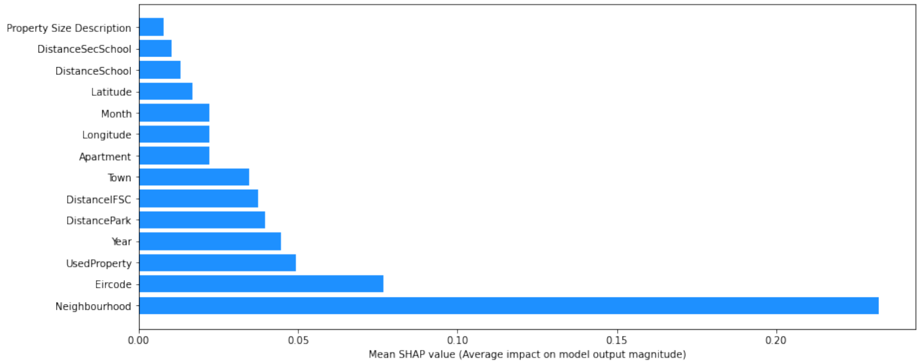

Fig. 7. Mean SHAP values over 5 year period on Dublin TPOT AutoML results.

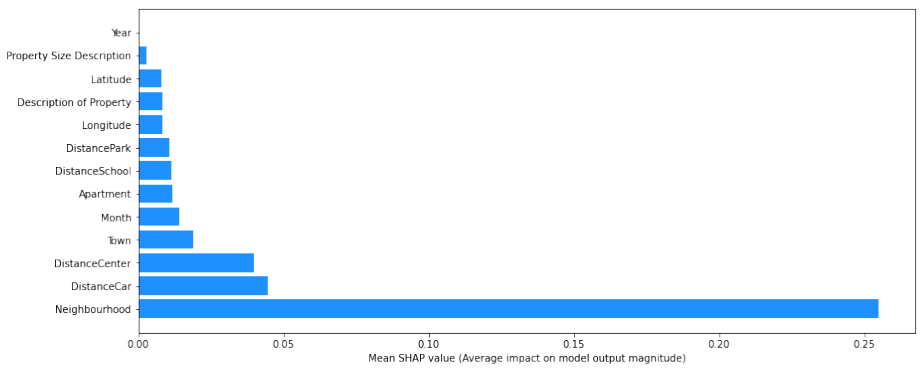

Fig. 8. Mean SHAP values over 5 year period on Cork TPOT AutoML results.

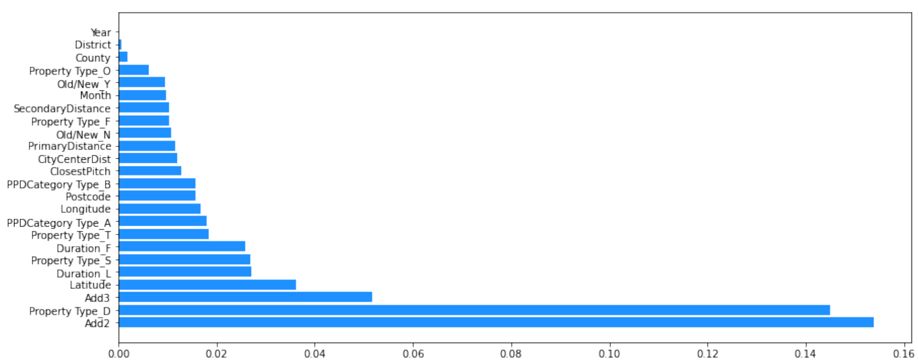

Fig. 9. Mean SHAP values over 5 year period on UK TPOT AutoML results.