

Developing Real Estate Automated Valuation Models by Learning from Heterogeneous Data Sources

Francesco Bergadano^{1*}, Roberto Bertilone¹, Daniela Paolotti², Giancarlo Ruffo¹

¹Department of Computer Science, University of Turin, Corso Svizzera 185, Turin, Italy
²ISI Foundation, Via Chisola 5, Turin, Italy

*Corresponding author's email: francesco.bergadano@unito.it

Article history: Received: 16 February 2021 Received in revised form: 23 April 2021
Accepted: 17 May 2021 Published online: 23 June 2021

Abstract

In this paper we propose a data acquisition methodology, and a Machine Learning solution for the partially automated evaluation of real estate properties. The novelty and importance of the approach lies in two aspects: (1) when compared to Automated Valuation Models (AVMs) as available to real estate operators, it is highly adaptive and non-parametric, and integrates diverse data sources; (2) when compared to Machine Learning literature that has addressed real estate applications, it is more directly linked to the actual business processes of appraisal companies: in this context prices that are advertised online are normally not the most relevant source of information, while an appraisal document must be proposed by an expert and approved by a validator, possibly with the help of technological tools. We describe a case study using a set of 7988 appraisal documents for residential properties in Turin, Italy. Open data were also used, including location, nearby points of interest, comparable property prices, and the Italian revenue service area code. The observed mean error as measured on an independent test set was around 21 K€, for an average property value of about 190 K€. The AVM described here can help the stakeholders in this process (experts, appraisal company) to provide a reference price to be used by the expert, to allow the appraisal company to validate their evaluations in a faster and cheaper way, to help the expert in listing a set of comparable properties, that need to be included in the appraisal document.

Keywords: Automated Valuation Models, real estate appraisal, open data, Machine Learning, Web Crawling

© 2021 Penerbit UTM Press. All rights reserved

1.0 INTRODUCTION

Real estate appraisal is a complex and important task, that can be made more precise and faster with the help of automated valuation tools. Usually, the value of some property is determined by taking into account both structural and geographical characteristics. However, while geographical information is easily found, obtaining significant structural information requires the intervention of a real estate expert, a professional appraiser.

The correct evaluation of house prices plays a fundamental role in our economy and affects all the participants to the real estate market, including:

- professional appraisers, who are expert in the evaluation of properties and normally perform *in loco* visits and off-site paperwork;
- real estate appraisal companies, who request, harvest, standardise and verify the work of professional appraisers;
- financial institutions, needing to (1) set a justified property price prior to offering a mortgage loan or (2) evaluate a property portfolio, e.g., in the context of NPL (non-performing loan) management;
- notaries/solicitors, needing to verify property values prior to guaranteeing the validity of some public transaction (e.g., a deed of purchase or the handling of inheritance issues);
- homeowners and buyers, and real estate agents, wanting to assess the reasonable market price of a property.

In relatively recent years, the concept of an Automated Valuation Model (AVM) has emerged in this industry (Kok et al., 2017): an AVM is a software system, often based on online data and resources, that can produce a property evaluation in a semi-automatic way (Downie & Robson, 2007; Jaen, 2002; Kok et al., 2017). More recently, artificial intelligence and Machine Learning approaches to AVM construction have been adopted (Bahia, 2013; Moosavi, 2017; Ng, 2015; Núñez Tabales et al., 2013; Pow et al., 2014; Yang et al., 2016). This strategy is becoming increasingly useful in two wide areas of business application:

- Appraisal professionals and companies produce a precise and authoritative evaluation, but with relatively high cost and time requirements - they can reduce such costs and times by using an AVM as a verification system (for the appraisal company) or as a helping tool (for the professional appraiser);

- Some applications and stakeholders need a quick, or even real time property evaluation, that cannot be produced with the traditional process involving an expert. For example, a street-level bank office may want to immediately propose a draft mortgage offer to an incoming customer, before a formal expert appraisal is available.

When designing an AVM, one must consider the fact that the appraisal of a real estate property is a very difficult task, due to the high heterogeneity in both structural and geographical data. Moreover, the price can be influenced by macroeconomic factors, that obviously change over time. In fact, the creation of a model able to predict real prices needs to deal with several problems caused by the complexity and the dynamics of the real estate market, and to the difficulty of obtaining reliable and objective data.

1.1 Research Objectives

This research addresses the issue of predicting property prices in a specific business context, where the AVM has to be useful for real estate appraisal companies and for the experts appointed by these companies. We do not address the needs of the final consumer, e.g., by providing an online tool for evaluating the house where they live, or the properties they are willing to buy. In particular, we consider the business processes relevant in this context, that are related to the property evaluation that is mandatory before a mortgage application can be approved. There is normally a contract between the real estate evaluation company and the bank, where a large number of such evaluations are performed each month. Especially when the bank is a large financial institution, the rules to be followed are tightly defined in such contracts, and evaluations are assigned via Web services and automated procedures. The real estate appraisal companies will, in turn, appoint subcontractors, or individual experts, who will actually perform the evaluation, providing a complete report that must respond to precise requirements. In particular, the report will include the estimated price, considerations about the state of the property and its surroundings, and a specific list of nearby and comparable properties. The expert's report will then undergo a so-called 'validation' phase, where the appraisal company checks for possible errors and bias. The validated price will then be submitted to the banking institution via an automated procedure and as specified in the contract.

Based on the above context, we consider the following research objectives:

- the AVM's prediction will be used by experts for a reference and comparison price; therefore, it has to be close to similar valuations: we therefore address the prediction of expert valuations, not the prediction of prices from deeds of purchase or online advertisements. This research wants to determine if such prediction problem can be solved with an accuracy that is adequate for its business purpose. The answer will be shown to be positive, based on the experimentation detailed in Section 6.
- with this specific prediction problem in mind, and in the context of the Italian market, we wanted to select a number of suitable feature sets, and compare the corresponding performances obtained via different Machine Learning algorithms. In particular, we will compare three different feature sets (hedonic, reduced, and reduced with the addition of comparables).
- a method has been sought for obtaining so-called 'comparable' properties, that are useful for two purposes: (1) for computing the corresponding features and (2) for helping the expert by providing a ready-to-use list of properties to be included in their report.

1.2 Our Contribution

In this study, in order to address the above-mentioned objectives, we first selected features computed from diverse sources of data, including open data available on the Web, unstructured data that are obtained via Web Crawling, and private data from previous expert appraisals. We then generated an AVM using Machine Learning methods, where the selected features are automatically used to define a model that can output a predicted price. The generated model is thus adaptive and able to deal with the complexity and fast-changing characteristics of the real estate market. This AVM can run using readily available input data and is capable of producing real-time valuations.

The novelty and importance of the proposed approach lies mainly in two aspects:

- when compared to AVM tools as available to real estate operators, the present methodology is highly adaptive, non-parametric, and based on Machine Learning techniques, with the possibility of integrating diverse data sources.
- when compared to Machine Learning literature that has addressed real estate applications, our approach is more directly linked to the actual business processes of appraisal companies.

A case study is provided, where this proposed Adaptive AVM is applied to the residential real estate market in Turin, Italy. Data and application context were obtained and defined in collaboration with a leading Italian real estate appraisal company.

Three different approaches to feature selection and Machine Learning have been experimented with, yielding surprising results, where many features normally considered important have turned out to be irrelevant. We have used a set of previously appraised residential properties in Turin, Italy, obtaining additional relevant data from heterogeneous Web sources. The results, measured on independent test sets, have shown that the model is predictive and practically effective.

The next section on related work discusses previous approaches to AVM optimisation and construction and discusses in more detail the differences with respect to the proposed methodology.

■ 2.0 RELATED WORK AND INNOVATIONS

In this section, we give an overview of existing literature relevant to our research, highlighting the ideas we adopted in our approach. Given the highly application-oriented impact of this work, we also make references to papers that have been published in a multi-disciplinary range of journals and conference proceedings, that spans from fields such as economic and econometric, to finance, property management, and so

on. In particular, here we focus mainly on how different models have been adopted and assessed using methods from computer science and information science, including references to the role of big/open data for price predictions and value creation. At the end of the section, we describe briefly the novelties we adopted in addition to those already proposed in the literature.

2.1 Hedonic Models

Previous research on Automated Valuation Models for real estate has been initially led by so-called “hedonic” models (Goodman & Thibodeau, 2003; Hu et al., 2013; Lisi & Iacobini, 2013; Loberto et al., 2018). “Hedonic” literally suggests that the buying of the target property, and hence living in it, is a source of “pleasure”. The better (and hence more expensive) the property, the higher this specific notion of real estate pleasure, stemming from property characteristics leading to such sensations: a nice view, proximity to services and pleasant life, the presence of an elevator, parking lots, a concierge. We will call all such pleasure-giving characteristics our “hedonic” features.

In older approaches hedonic features were mainly derived from intrinsic characteristics of the property, e.g., number of rooms and square meters, the number of bathrooms, the floor number. Again, in those traditional approaches, there was generally a linearity assumption - the value $V(P)$ of some property P would depend linearly on the corresponding hedonic features $[f_1(P), \dots, f_n(P)]$:

$$V(P) = \eta + \sum_{1 \leq i \leq n} \beta_i \times f_i(P) + \epsilon(P) \quad (1)$$

where β_i is the “ i -th” coefficient, $\epsilon(P)$ is some correction to be applied to this particular property, and η is a property-independent correction that applies to some geographical perimeter or application context.

2.2 Non Linear and Even Non-Parametric Models

In more recent research, a number of novelties come into place: we no longer assume the valuation depends linearly on the selected features, as in Equation 1. Some studies suggest that this is not in fact the case for real estate valuations (Nghiep & Al, 2001). Others (e.g., Abidoye & Chan, 2017, 2018) have proved that more complex Automated Valuation Models perform better than traditional regression methods. In many cases, we just do not know what kind of dependency exists between the features and the sought valuation - we thus follow a non-parametric approach, where the type of regressor is unknown (Chopra et al., 2007). In the present study we also follow this approach: we do not assume linearity nor the correspondence to a particular form of classifier/regressor, and we recognize the complex nature of real estate appraisal. Moreover, selected features are not individually “good” or “bad”, but rather a part of a more complex analysis.

2.3 Implementation with AI and Machine Learning

The availability and increased performance of Machine Learning approaches has led to a widespread use of such technologies in AVMs for real estate (Baldominos et al., 2018; Caplin et al., 2008; De Nadai & Lepri, 2018; Gao et al., 2019; Jaen, 2002; Ng, 2015). This includes the use of artificial neural networks (Bahia, 2013; Limsombunchai, 2004; Nghiep & Al, 2001; Núñez Tabales et al., 2013; Tibell, 2014), decision trees (Jaen, 2002), random forests (Čeh et al., 2018; Kok et al., 2017; Moosavi, 2017), gradient boosting (Kok et al., 2017), Bayesian compressed vector auto-regressive model (Gupta et al., 2019), and support vector machines (Bui et al., 2017; Mu et al., 2014).

Finally, multitask learning has been used both for AVMs (Gao et al., 2019) and for DOM (Days on the Market) prediction (Zhu et al., 2016). Other AI techniques may be relevant, such as language classification and even semantic NLP for any text that can be referred to the property neighbourhoods. Customers' opinions and user generated content's sentiment can be exploited to extract appraisal expression patterns (Zheng et al., 2014). Image classification and recognition has also been used in the real estate AVM context (Ahmed & Moustafa, 2016).

2.4 Extrinsic Features from Web and Open Data

Not only the features directly related to the target property are important, but also the ones derived from neighbouring amenities and services, as well as linked to totally external information. Space and location-dependent external information has often been found to be important (Caplin et al., 2008; Kok et al., 2017). Geographic information that can contribute significantly to improve location's description can be retrieved and mined efficiently from the Web (Zhang et al., 2019). Neighbouring area information has also been used, including criminality rate, population density and average income, pollution, services and transportation, and the distance from local riverbanks (Bahia, 2013; De Nadai & Lepri, 2018; Pow et al., 2014). Recent results on prices and sales trends prediction show that the analysis of data available on the Internet, such as search queries, can be useful to understand the house market status and its indicators (Lee et al., 2019).

In the present study we address this issue in a structured and general way, by defining a notion of “point of interest” (PoI) concerning some subject or service (e.g., transportation, entertainment, sports). Such PoIs are sought for and geographically mapped, based on available open data (see “online resources” at the end of the paper: open data for the Turin municipality - “aperTO” (<http://aperto.comune.torino.it/>), and nationwide (Agenzia per l’Italia Digitale, 2019), Foursquare, Google Maps, OMI – “Osservatorio del Mercato Immobiliare” – Italian Real-Estate Observatory (Festa et al., 2018).

2.5 Locally-Oriented Context

A number of approaches were tailored to a particular geographic area, or evaluated in such local perimeters, e.g., Montreal (Pow et al., 2014), Beijing (Yang et al., 2016; Zhu et al., 2016), London (Ng, 2015), Los Angeles (Caplin et al., 2008), Hanoi (Bui et al., 2017), Zurich (Moosavi, 2017), Stockholm (Tibell, 2014), Singapore (Ibrahim et al., 2005), Nigeria (Abidoye & Chan, 2017, 2018), Italy (Fregonara et al., 2012),

and Slovenia (Čeh et al., 2018). This could be seen as a limitation, as if a general methodology was out of hand. Realistically, though, it makes very good sense because (1) valuation practices and regulations differ from one nation to another, (2) different kinds of open data are available and (3) good feature selection is essential and sets of best available features differ geographically. One important contribution of the present paper was to develop a feature set that performs well in the selected geographical area.

2.6 Target Price

For a real estate appraisal company, the goal is to obtain an authoritative valuation for some property. Our target property price is thus corresponding to valuations performed by a registered expert, not to the prices as obtained from Web advertisements, as, e.g., in (Loberto et al., 2018; Moosavi, 2017). We however did use such advertised prices, but only as so-called “comparable” properties (see Section 4 below), that are routinely used by appraisal experts. The comparable property prices are thus part of the feature set but are not the target of the AVM output.

3.0 DATA SET, OPEN DATA AND COMPUTATION OF RELEVANT FEATURES

For the purpose of this research, we have used three different data sources: (1) a corpus of professional and validated property appraisal documents and corresponding data base, (2) geographical and open data obtained from heterogeneous public Web sources, and (3) advertised prices for comparable properties obtained via Web Crawling. We discuss such data sources in the next subsections.

3.1 Data Set of Available Appraisals

The data used to perform the analysis contains the information summarize in Table 1 and was provided by a multinational real estate appraisal company, with a subsidiary having significant activities in Italy. The initial data set consisted of 7988 property valuations, performed by professional appraisers, and validated by the company. These properties are all located in the city of Turin, and the valuations were performed between 2011 and 2016.

Table 1 Information in the available appraisal data set
(Source: Multinational real estate appraisal company, from 2011 to 2016)

Variable	Type of data	Range
Year of construction	Integer (year)	1966 to 2020
Number of bathrooms	Integer (number)	0 to 7
Floor	Integer (number)	-1 to 11
Surface	Integer (square meters)	40, 4-249
Elevator	Boolean	yes/no
Maintenance status	Ordinal	low/medium/high
Quality of Installations	Ordinal	low/medium/high
Finishing Quality	Ordinal	low/medium/high
View	Ordinal	low/medium/high
Energy Efficiency	Discrete	A4, A3, A2, A1, B, C, D, E, F, G
Registered Use	Discrete	residential/office/warehouse
Orientation	Discrete	east/west/south/north
Latitude/Longitude	Geographical	Turin Metropolitan Area
Address	Geographical	Turin Metropolitan Area
City area	Geographical	Central/Near-central /Larger-City/Suburbs
Valuation	Target variable	20k€-700k€

As a first step, this data set was anonymized, removing personal buyer or mortgage application information, as well as bank details. For the most important variables in Table 1, we add the following remarks:

- Number of bathrooms: 0 to 7, mean value = 1.2
- Floor: -1 to 11, mean value = 3
- Surface: 40.4 to 249 square meters, mean value = 94.7
- Energy efficiency classification¹: this value is mandatory for selling or renting a residential property.
- Registered use: based on the Italian land register (“catasto”²), the possible values include “residential”, “office”, “warehouse”.
- City Area: this is a general classification including the values "Central, Near-central, Larger City Boundary, Suburbs" - a more detailed area classification is obtained via Latitude/Longitude and OMI areas.

- Valuation: valuation in Euros, as assessed by a professional appraiser and stored in the appraisal document. This is the target variable, i.e., the value we will want to predict and the output of our AVM when we will use it on new, yet to be evaluated properties.

This data set contained a wealth of information that is normally difficult to acquire in such quantity and detail. However, it was too diversified, and a data cleaning and selection process was performed. In particular:

- in order to have a homogeneous data set, we only considered properties with a "residential" registered use, because this was the most common case; by contrast, properties with different registered uses are difficult to compare;
- we excluded houses with a surface greater than 250 square meters, as they are rare and belong to a peculiar market sector;
- we only considered houses with a global valuation between 20,000 € and 700,000 €, as the other cases are considered exceptional for the same reasons;
- we also excluded the valuations of complex properties, as for example houses with a garage, because of their limited number and more complex description structure.

After this selection process, the 7988 valuations were reduced to 3983, still a significant number for our practical purposes and prediction targets. Finally, we performed a *feature standardisation* activity. For example, some features were normalised so as to have values between -1 and 1, e.g. construction year and floor, so as to make them easier to process in subsequent phases.

3.2 Geographical and Open Data

Using information available on the Web, indexed with the geographical location of the properties, we have extended the features contained in the appraisal data set, with new and useful information. This consists mainly of two data categories: OMI areas and values, and nearby area information.

3.2.1 OMI Areas

The Italian Revenue Agency is responsible for OMI (Festa et al., 2018), see links at the end of the paper. For each registered use of properties, the surface of each Italian municipality is divided into different areas, called OMI areas, having homogeneous real estate characteristics and valuation schemes.

Every six months, OMI provides an update of the price range (minimum and maximum price), for each OMI area. For example, Figure 1 provides a graphical representation of the 41 OMI areas in the municipality of Turin, with the corresponding average prices. For the purpose of easier comparison, we report next, in Figure 2 the valuations as predicted by our AVM - as it can be seen the predicted prices are quite close to the actual values.

The price ranges in an OMI zone may vary according to the dynamics of the real estate market or to specific changes regarding some geographic location. The OMI area associated to some property may be obtained from OMI, using its geographical coordinates. OMI also provides the formal description of "polygons" that define the OMI areas. We have retrieved OMI polygons and price ranges from the links referred to at the end of the paper in January 2020, by means of an automated data acquisition procedure, that can be considered as a separate component of our AVM. Although changes in the OMI repository are usually not significant, this data acquisition phase should be repeated for every new release, i.e., every six months.

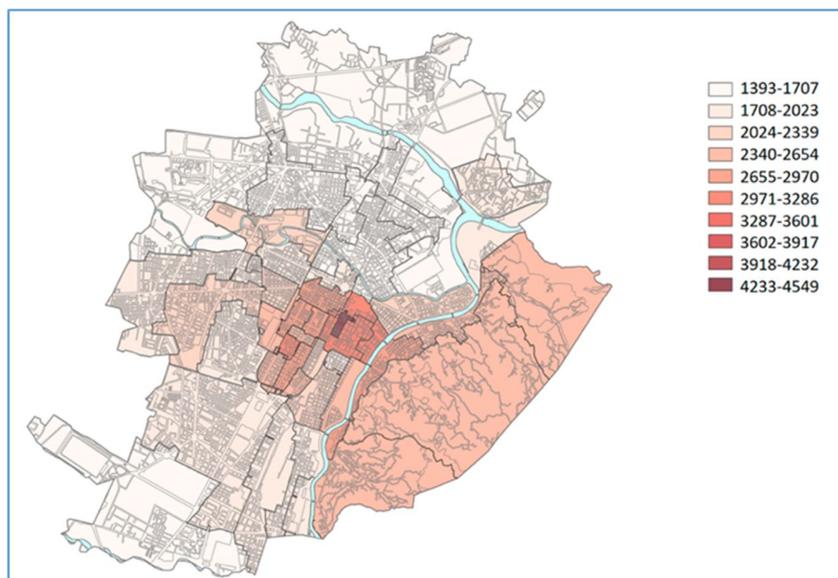


Figure 1 Graphical representation of the Turin real estate data set divided by OMI area - average prices (Euros per square meter)
(Data source: OMI prices and geographical polygons - www.agenziaentrate.gov.it)

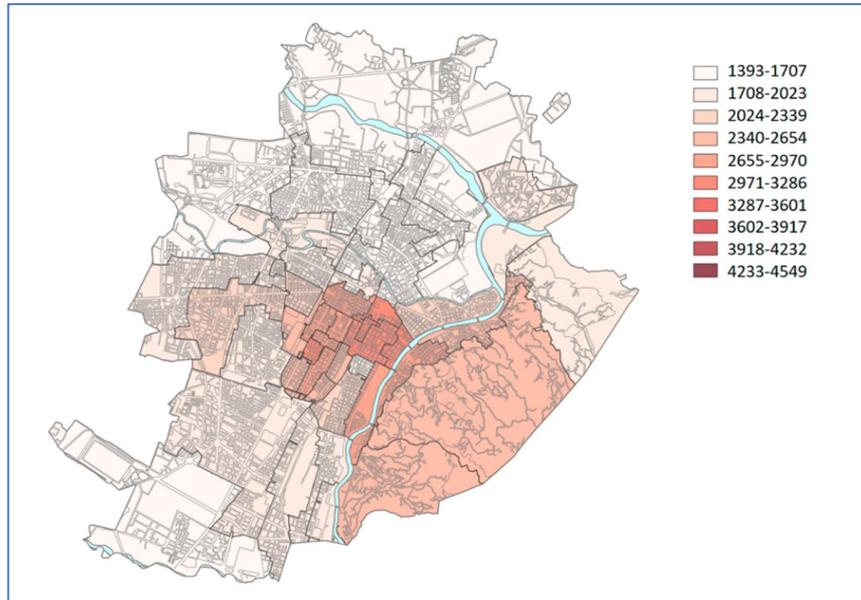


Figure 2 Graphical representation of the Turin real estate data set divided by OMI area - predicted prices (Euros per square meter)

We have followed two different approaches in this research:

1. In the first approach (**OMI names**), we used the name of the OMI area as an additional, constructed feature. This has some drawbacks. First, as previously explained, the price range in a specific OMI area may change over time. This implies that if the model has been trained using valuations performed in a certain period of time, the valuation of a new property in the future could be affected by the changes of price in the corresponding OMI area. Second, the feature will be effective only in the valuation of properties that were located in OMI areas that are represented in the training set. Another issue could be represented by the creation of new OMI areas or the disappearance of old OMI areas over time. Finally, by using the OMI area name, a discrete value, we lose important geographical information, such as the distance between different OMI areas.
2. In the second approach (**OMI min/max values**), we introduce as new features the upper and the lower limit of the price range of the OMI area at the time of the appraisal. The idea behind this choice is that of separating the model from specific OMI areas, trying to transform the OMI area name into ordinal features. As the criteria used by OMI for the creation of the price range are the same for each municipality and remain constant over time, this choice allows us to use the model to evaluate properties located in OMI areas not even contained in the training set, and to perform evaluations long after its creation. Please, also notice that the minimum and maximum property price in any given area is evaluated by OMI itself as a paid service.

We have performed experiments with both feature construction approaches, and the latter (OMI min/max values) has produced superior results, as discussed later.

3.2.2 Nearby Area Information (Points of Interest - PoI)

Starting from the geographical position of the property (latitude, longitude), as contained in the appraisal data, we construct new features using open data available from the Web and related to corresponding surrounding areas (nearby areas). In particular, we built a set of so called "Points of Interest" (PoI). Points of interest correspond to activities and resources that are present in the territory, are geo-localised, and have a potential positive influence on the price of surrounding properties. This corresponds to best practices in real estate appraisal, where experts normally produce valuation reports that include features such as nearby metro and bus stops, schools, and museums. We have grouped our PoIs into 13 categories: Arts, Business&Services, Entertainment, Food&Beverage, Healthcare&Wellness, Instruction, Landmarks, Religious services, Retail, Security, Sport&Recreation, Transportation, Travel. A data set of PoIs has been created by aggregating information from Foursquare, Google Maps and the Turin Open Data "AperTo" Web Site. An example of a PoI is provided in Figure 3 below.



Figure 3 Example of a POI: Mole Antonelliana, category: Arts; coordinates: 45.067999728, 7.68916391

The idea is to construct a new property feature for each PoI category, by counting the number of PoIs in that category that are within a threshold distance. However, in order to avoid the on/off effect of a strict threshold, we define 4 circles around the property, and associate descending weights to the PoIs falling within these circles, thus giving more importance to the PoIs that are closer to the property. We obtain the following formula, defining the property feature f_j for PoI category j ($1 \leq j \leq 13$):

$$f_j = \sum_{d_k \leq \frac{r}{8}} 1 + \sum_{\frac{r}{8} < d_k \leq \frac{r}{4}} \frac{1}{2} + \sum_{\frac{r}{4} < d_k \leq \frac{r}{2}} \frac{1}{4} + \sum_{\frac{r}{2} < d_k \leq r} \frac{1}{8} \quad (2)$$

where d_k is the distance of the property from PoI number k in the j^{th} category, and r is a threshold distance, initially set at 1 km. Finally, an additional feature f_0 was used, defined simply as the distance from the city centre, a special kind of PoI. Feature f_0 is fed to the Machine Learning algorithms, just like the other features f_1, \dots, f_{13} .

4.0 ADDITIONAL FEATURES: COMPARABLE PROPERTIES

Besides the feature set described in the previous section, we have also used an additional input feature: the price per square meter of known properties that are similar to the valuation target. This stems from business practices adopted by appraisal companies and professionals (discussed in the next subsection). The same has been done in our AVM (discussed in Section 4.2 below).

4.1 Use of Comparable Properties in the Appraisal Process

According to best appraisal practices in Italy, the expert's valuation document normally comprises a description and identification of so-called "comparable properties", i.e., properties that are geographically near the target property and have similar characteristics. The price per square meter will then be similar and the professional appraiser will use it as an important starting point in order to reach a final valuation. Professional appraisers can obtain some such comparable property data from the appraisal company's database of previous valuations. They cannot use purchase deeds from notaries as this is not publicly available in Italy (as it is in France, and, partially, in the UK). Using previous appraisal company valuations will however introduce some bias, as the same experts and the same company standards were used.

In recent years, appraisers also use real estate offers as advertised on the Web. There are many such publicly available services in Italy, offered to real estate agents as well as private individuals. These open services normally allow for sophisticated search interrogations, that may be filtered by, e.g., square meter range, distance from a specified location, property type, price range, floor, maintenance status. The expert can then download the advertised property description and include a selection of relevant data in her valuation document. The advertised price of such comparable properties will also be used as a reference in setting the valuation of the target property. It must be noted, however, that the advertised price is not a sale price, but rather an upper bound, awaiting further negotiations. The expert must consider this when using the advertised price as an input. As a consequence, the expert's goal will be to use all available data (including advertised price of similar properties), in order to set an authoritative valuation price, not to "guess" a price that would be similarly advertised for the target property. This is an important distinction, that has inspired our work, as we want our AVM to follow real experts' behaviour.

4.2 Use of Comparable Properties in the AVM

We have simulated the expert best practices for comparable data acquisition, as described in the previous subsection, while letting Machine Learning do the rest and decide how to use such data. In particular, our system is able to acquire three distinct types of comparable properties:

1. properties that are the target of previous appraisals in the data set: we have such data available in a structured data base, and the corresponding valuation is labelled as a valuation price, as it was produced by some appraisal expert sometime in the past - the corresponding valuation date is stored and should be taken into account. This was already mentioned in section 3, but in that case the valuation was used as a target variable. Here, instead, it will be used to compute an additional input feature: the average price per square meter of comparable properties.
2. properties that are advertised for sale on the Web.

For the latter category of comparable properties, we simulated the behaviour of the expert, who will normally perform a manual Web search and navigation using a limited number of well-known Italian real estate portals. The expert will filter advertised properties by geographical area (close to the target property), and characteristics (that must be similar to the ones of the target property).

In order to do this, we implemented a robot with a “controlled Web Crawling strategy”:

- the crawler runs at delayed and randomized time intervals;
- the available real estate portals are used with a randomized round robin strategy;
- search is limited w.r.t. to the number of visited pages and required filters, using a limited distance from the target property (a tuneable parameter);

This is done in order to avoid bot detection and blocking, and to ensure that the crawler's behaviour is factually similar to human search. If too many properties are found, the distance threshold is reduced and additional filters are set, with the purpose of finding properties that have similar characteristics (e.g., floor and maintenance status).

Once an appropriate number³ of comparable properties is found, we must retrieve the corresponding prices per square meter, and compute their average value. There are two cases, corresponding to the enumeration at the beginning of this section:

1. properties that are the target of previous appraisals in the data set: the valuation price and the square meters are readily available, as they are included in the records of this structured database;
2. properties that are advertised for sale on the Web: what was obtained by Web Crawling is unstructured information with variable formats (normally a portion of an html page) - as a consequence the relevant information (price and square meters) is extracted automatically on a case-by-case basis.

The two types of comparable properties, and the corresponding prices per square meter, are integrated by computing a simple average of the prices per square meter, regardless of the information source.

If the demonstrator is scaled up and used in a production and commercial service, a number of issues will have to be addressed, including legal concerns about the use of a robot to retrieve possibly proprietary information (though actually publicly available on the Web). Some technical issues should also be analysed, such as dealing with anti-automation (e.g., Captchas), and robot detection/classification (Perna & Tagarelli, 2018). Finally, frequent Web format changes in real estate portals will require manual Crawler adaptation.

The demonstrator is now able to retrieve a set number of comparable properties as described above, with corresponding attributes and advertised or valued price. The result of the process is a new constructed feature for the target property: the average price per square meter of comparable properties.

5.0 MACHINE LEARNING METHODOLOGY

We will now describe how the described data set was used, in order to train our AVM. First, we will describe three different approaches that we have used for feature selection, and subsequently we will describe how the data set was partitioned and which Machine Learning algorithms were used. The experimentation was carried out using the well-known Machine Learning tool Scikit Learn (Pedregosa et al., 2011) providing a wealth of different learning algorithms and feature evaluation functions. The solution we have developed is a demonstrator software, and it was built for academic and research purposes, and was not used in a commercial application at the present time.

5.1 Feature Selection

The data set described in the previous section is complex and involves significant amounts of correlated information. We have followed three distinct approaches to the use of such data, through the semi-automated selection of three different subsets of the available features:

- **Hedonic features.** In this approach, we select hedonic features that seem to naturally influence the value of the property, in either positive or negative ways (Limsombunchai, 2004; Lisi & Iacobini, 2013). Specifically, we have used:
 - a) all the features contained in the appraisal data set (Table 1)
 - b) the POI features for the categories described in Section 3.2.2 (i.e., f_1, \dots, f_{13})
 - c) the distance from the city centre (f_0)
 - d) the OMI area name, as defined in Section 3.2.1.

- **Reduced feature set (OMI-centered features).** By using hedonic features in the learning step, with different algorithms and hyper parameters, it turned out that many features were irrelevant and seldom used by the output regressors. We then adopted a new approach, with a substantially reduced feature set, including:
 - a) OMI minimum and maximum price per square meter valuations for the OMI area where the property is located (OMI-min, OMI-max)
 - b) surface in square meters. The identification of the correct OMI area is obtained from “polygons” that are made available by OMI and based on the property location (latitude and longitude).
- **Reduced feature set plus comparable prices.** In this case, we used:
 - a) the same features as in the previous approach (reduced feature set)
 - b) the average price per square meter of comparable properties.

As it will be evident from the results described in the next section, the latter approach provides the best prediction.

The definition of the above three different set of features was partially automated and partially manual. In particular, we followed the steps listed below, that we will subsequently discuss in more detail:

- we automatically computed the relative importance of all the available features
- we carefully analysed the results, realizing that square meters and OMI areas were by far the most important features
- we obtained further features related to OMI areas, namely the minimum and maximum price per square meter in each area
- we manually assembled the feature sets 1, 2, and 3 listed above

Some of the available learning algorithms were used for the analysis of the feature set. For example, with Random Forests (Breiman, 2001), Extra Trees (Geurts et al., 2006) and Gradient Boosting (Friedman, 2001), it is possible to measure the relative importance of features. In particular, in Random Forests, once a target feature is set, one can compute, for every node in each tree of the forest, the obtained decrease of impurity, and then compute the average of all such values (this is called “mean decrease impurity” (Scornet, 2020)). For classification tasks impurity can be computed as Gini impurity or information gain, and for regression it can be based on variance. This is what was done in Figures 4 and 5. For other measures of feature importance, see Saarela and Jauhiainen, 2021.

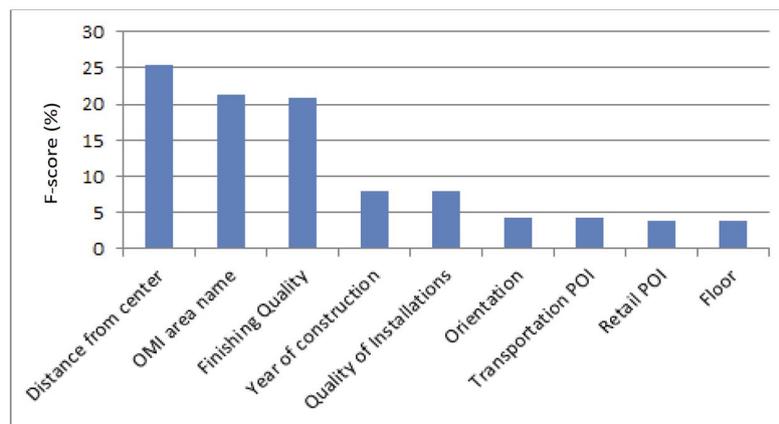


Figure 4 Impurity-based feature importance of the Hedonic feature set, using the Random Forest regressor (best 9)

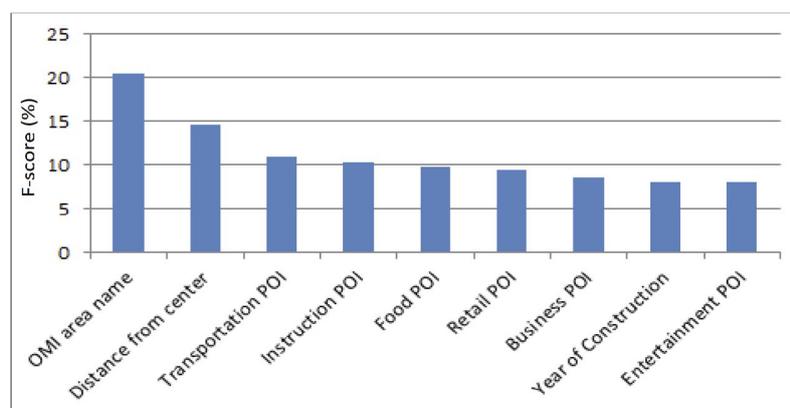


Figure 5 Impurity-based feature importance of the Hedonic feature set, using the Gradient Boosting regressor (best 9)

For the Hedonic feature set (as defined in Section 5.1), we obtained the rankings described in Figures 4 and 5, for the Random Forests and the Gradient Boosting regressors, respectively. An even more important feature consists in the property surface, in square meters. This feature was however not used for computing the feature importance in Figures 4 and 5, because it would have been significantly more important than the others, making the figure difficult to read. Instead, we ran the algorithms without the surface feature, in order to compare more easily the relative importance of the other features. In the figures, the ranking values reported on the y axis sum up to 100 for the 9 most important features, that are reported on the x axis.

When using other algorithms, we obtained partly similar rankings - in particular, the OMI area name always ranks first or second. This prompted us to a radical change in addressing feature importance. In fact, features that we initially thought to be relevant, such as the POI indexes for Arts and Security, were rarely used. This is maybe due to the fact that the target price is an expert valuation, and not the real deed of purchase value. However, we had to stick to this prediction problem, as it was the only possible, and face the fact that the Hedonic feature set was far from perfect. The presence of many irrelevant features could have led to over-fitting phenomena. At the same time, all learning algorithms gave great importance to OMI areas and distance from the city centre, that are obviously correlated.

Further analysis highlighted the fact that OMI areas, in the Hedonic feature set, were coded by names. As a consequence, they do not carry any geographical properties and correlations, e.g., proximity between different OMI areas. To correct this issue, we found that it was possible to obtain from OMI, as a payed service, the minimum and maximum property price in any given area, as evaluated by OMI itself (OMI-min and OMI-max). By training regressors after adding these two features to the hedonic feature set, it turned out that they are by far the most important features, and significantly superior even to the OMI area name.

As a consequence, the use of the regressors as feature ranking tools, based on the training set, led us to the addition of the OMI-min and OMI-max features, and to the use of the reduced feature sets as listed at the beginning of this section (OMI-centered, and OMI-centered with comparables).

5.2 Learning Step: Obtaining the Valuation Model

The available data set was initially shuffled, so as to avoid order dependence and position-based regularities. Then it was partitioned into three separate subsets:

- learning set (70% \Rightarrow 2789 valuations)
- validation set (15% \Rightarrow 596 valuations)
- test set (15% \Rightarrow 596 valuations)

where the learning set is used to train a regressor, the validation set is used to choose the best hyper-parameter values needed by the learning algorithm, and an independent test set is used to evaluate and report the results.

The learning task is a supervised regression problem, where the target feature to be predicted is the property valuation. We have experimented, depending on the selected feature set, with up to six Machine Learning regression algorithms: KNN - K Nearest Neighbour (Shakhnarovich et al., 2006), Random Forests, Extremely Randomised Trees (Extra Trees), Gradient Boosting, AdaBoost (Schapire, 2013), and Bagging (Breiman, 1996).

The learning process described above was repeated for each of the three feature sets listed at the beginning of Section 5.1, and the corresponding results are reported in the next section.

6.0 RESULTS

We now describe the results for the three different feature sets. The results reported in this section are measured on the test set. The data in the training set are instead used during the learning phase, and the data in the validation set are used to measure the performance of the developed AVMs when learning with different hyper-parameters (e.g., number of trees in a random forest, or the maximum tree depth). The best hyper-parameters are then used to develop the final AVM. The evaluation and performances that are reported in this section, are measured on the separate and independent test set. The results for the Hedonic feature set are reported in Table 2, whereas the results for the reduced, OMI-centred feature sets are reported in Tables 3 and 4 and in Figures 6 and 7.

Table 2 Results for the hedonic feature set

Algorithms	ME	MSE	R^2
Bagging Regressor	46.4	2155	0.774
AdaBoost	52.8	2787	0.707
Gradient Boosting	43.9	1935	0.797
K-Nearest Neighbour	54.4	2964	0.689
Extra Trees	41.6	1738	0.810
Random Forest	43.1	1864	0.807

Table 3 Results for the OMI-centred feature set

Algorithms	ME	MSE	R ²
Bagging Regressor	37.74	1425	0.838
AdaBoost	39.29	1543	0.824
Gradient Boosting	47.41	2248	0.744
K-Nearest Neighbour	46.24	2139	0.757
Extra Trees	34.84	1214	0.862
Random Forest	35.79	1281	0.854

Table 4 Results for the OMI-centred feature set with the addition of comparables

Algorithms	ME	MSE	R ²
Bagging Regressor	21.61	467	0.957
Extra Trees	21.12	446	0.959
Random Forest	21.48	461	0.958

**Figure 6** Results for OMI-centered features (random forest)**Figure 7** Results for OMI-centered features plus comparables (random forest)

6.1 Results for the Hedonic Feature Set

For the hedonic feature set, with the addition of the OMI area name, we mainly consider the mean error (ME):

$$ME = \sqrt{\frac{1}{n} \sum_{[x_i, y_i] \in TS} (y_i - \hat{y}_i)^2} \quad (3)$$

where \hat{y}_i is the predicted value, and $[x_i, y_i]$ is an example in the test set TS , with the target feature being the property valuation y_i and $|TS| = \mathfrak{N}$. This is expressed in thousands of Euros, and the results are reported in Table 2, where the Mean Square Error ($MSE = ME^2$) and the coefficient of determination R^2 are also shown.

Since the mean property value in the data set was 189.400 Euros, the mean errors in Table 2 cannot be considered a very good result, though it can be of aid to a professional appraiser as a rough indication and a starting point. Extra Trees and Random Forests seem to perform best.

6.2 Results for the Reduced Feature Set

For the reduced feature set (OMI-centred), and for the different learning algorithms that were previously listed, we have obtained the results shown in Table 3.

This is significantly better than the hedonic feature approach. The best results (see Figure 4) are obtained after introducing comparable values. The mean error is now around 21 K€, that has been considered acceptable by the domain experts, leading to a tool that would be very useful in the appraisal process.

A comparison of results for OMI-centred features, with and without comparables, can be seen in different detail in Figures 6 and 7. Each dot in these scatter plots represents a real estate property in the test set, with the projection on the x axis representing its expert-valuated price, and the projection on the y axis representing the error in our AVM prediction. It is immediately evident that the accuracy of the AVM with comparables (Figure 7) is significantly better, making the corresponding Web Crawling activity worthwhile. This can also be observed on the concentration graph on the right of both figures.

7.0 CONCLUSIONS

The presented research has addressed the development of an AVM with Machine Learning capabilities, following previous research in this area (see detailed discussion in Section 2.3). But it has done this from a novel perspective, with the goal of providing a tool that can be used effectively by professional appraisers and appraisal companies.

A crucial instrument to achieve this goal was the use of a structured database of previous appraisals, and the adoption of the price indicated therein as a target variable for the Machine Learning algorithms. Previous Machine Learning approaches to AVM development in this context had used either the price obtained from the deed of purchase, in countries where this is available, or, more commonly, the publicly advertised price (e.g., on consumer real estate web sites).

The AVM described here can help the stakeholders in this process (experts, appraisal company) in the following ways:

- provide a reference price, to be used by the expert as a starting point of the evaluation - since the AVM is trained on previous appraisals, the obtained price is relevant in this specific context;
- allow the appraisal company to do what they call a “validation step” in a faster and hence cheaper way, by using the AVM's output as a first and automated check;
- help the expert in listing a set of comparable properties, that need to be included in the appraisal document.

The above advantages can help save money and time, because the expert's work will be made easier, and the company's validation will be partly automated. It will also avoid human errors, that are sometimes a source of important incidents (if the expert's evaluation is significantly mistaken, the AVM will point this out immediately). The experimentation has shown promising results, demonstrating the appraisal price can be predicted with sufficient accuracy. In particular, the use of the reduced feature set with the addition of comparable property prices, has shown the best results: the observed mean error was around 21 K€, for an average property value of about 190 K€. This was measured on an independent test set.

On the negative side, the appraisal system will sometimes work against these innovations, and change management approaches will be necessary. In particular, experts and verifiers may see the AVM as a possible threat to their profession. Some may even fear “replacement” by the AVM. The appraisal company might then face difficulties in the practical adoption of the methodology.

Notes

- (1) With possible values being A4, A3, A2, A1, B, C, D, E, F, and G, where A4 has the highest efficiency and G the lowest, according to European Union directive 2002/91/CE and Italian laws D.lgs. 102/05 and D.lgs. 192/15.
- (2) www.agenziaentrate.gov.it/portale/Schede/FabbricatiTerreni/Visura+catastale/
- (3) This is also a tunable parameter, that was set to 5 in our AVM, so as to make it consistent with best practices followed by the appraisal company.

Acknowledgements

The presented research results would not have been achieved without the technical support by NetAtlas s.r.l. We would also like to thank the Lagrange Project of Fondazione Cassa di Risparmio di Torino (Fondazione CRT), that financially supported Roberto Bertilone to perform part of the experimental analysis.

Availability of Data and Material

We used the following on-line resources:

OMI (Italian Real Estate Observatory): <https://www.agenziaentrate.gov.it/wps/content/nsilib/nsi/schede/fabbricatiterreni/omi>

OMI area perimeters: https://www.agenziaentrate.gov.it/geopoi_omi/index.php

Open Data Torino: <http://aperto.comune.torino.it/>

Foursquare: <https://www.foursquare.com/>

Google Maps APIs: <https://developers.google.com/places/web-service/intro>

Other data from real estate businesses are confidential and cannot be re-distributed.

References

- Abidoye, R. B., & Chan, A. P. C. (2017). Critical review of hedonic pricing model application in property price appraisal: A case of Nigeria. *International Journal of Sustainable Built Environment*, 6(1), 250-259.
- Abidoye, R. B., & Chan, A. P. C. (2018). Achieving property valuation accuracy in developing countries: the implication of data source. *International Journal of Housing Markets and Analysis*, 11(3), 573-585.
- Agenzia per l'Italia Digitale. (2019). *Open data from the Italian public administrations*. Retrieved from <https://www.dati.gov.it/content/contatti>
- Ahmed, E. H., & Moustafa, M. N. (2016, November). House price estimation from visual and textual features. In J. J. M. Guervós, F. Melício, J. M. Cadenas, A. Dourado, K. Madani, A. E. B. Ruano & J. Filipe (Eds.), *IJCCI 2016. Proceedings of the 8th International Joint Conference on Computational Intelligence* (Vol. 3, pp. 62-68). Setúbal: Science and Technology Publications.
- Bahia, I. S. H. (2013). A data mining model by using ANN for predicting real estate market: Comparative study. *International Journal of Intelligence Science*, 3, 162-169.
- Baldominos, A., Blanco, I., Moreno, A. J., Iturrarte, R., Bernárdez, O., & Afonso, C. (2018). Identifying real estate opportunities using machine learning. *Applied Sciences*, 8(11), 2321.
- Breiman, L. (1996). Bagging predictors. *Machine Learning*, 24(2), 123-140.
- Breiman, L. (2001). Random forests. *Machine Learning*, 45(1), 5-32.
- Bui, Q.-T., Do, N.-H., & Hoang, H.-P. (2017, May 29-June 2). *House price estimation in Hanoi using artificial neural network and support vector machine: In considering effects of status and house quality*. Paper presented at the FIG Working Week 2017, Helsinki, Finland.
- Caplin, A., Chopra, S., Leahy, J., LeCun, Y., & Thampy, T. (2008, December 14). *Machine learning and the spatial structure of house prices and housing returns* (Technical report). Retrieved from <http://yann.lecun.com/exdb/publis/pdf/caplin-ssm-08.pdf>
- Čeh, M., Kilibarda, M., Lisec, A., & Bajat, B. (2018). Estimating the performance of random forest versus multiple regression for predicting prices of the apartments. *ISPRS International Journal of Geo-Information*, 7(5), 168.
- Chopra, S., Thampy, T., Leahy, J., Caplin, A., & LeCun, Y. (2007). Discovering the hidden structure of house prices with a non-parametric latent manifold model. In P. Berkhin, R. Caruana, X. Wu & S. Gaffney (Eds.), *Proceedings of the Thirteenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (pp. 173-182). New York, NY: Association for Computing Machinery.
- De Nadai, M., & Lepri, B. (2018, October). The economic value of neighborhoods: Predicting real estate prices from the urban environment. In F. Bonchi (Ed.), *Proceedings of the 2018 IEEE 5th International Conference on Data Science and Advanced Analytics (DSAA)* (p. 323-330). Los Alamitos, CA: IEEE Computer Society.
- Downie, M. L., & Robson, G. (2007). *Automated valuation models: An international perspective*. London: Council of Mortgage Lenders.
- Festa, M., Longhi, S., Cantone, G., & Papa, F. (2017). *Manuale della banca dati quotazioni dell'osservatorio del mercato immobiliare*. Rome: Agenzia delle Entrate. Retrieved from <https://www.agenziaentrate.gov.it/wps/content/nsilib/nsi/schede/fabbricatiterreni/omi/manuali+e+guide>
- Fregonara, E., Rolando, D., & Semeraro, P. (2012). The value spatial component in the real estate market: The Turin case study. *Aestimim*, 60, 85-113.
- Friedman, J. H. (2001). Greedy function approximation: A gradient boosting machine. *The Annals of Statistics*, 29(5), 1189-1232.
- Gao, G., Bao, Z., Cao, J., Qin, A. K., Sellis, T., & Wu, Z. (2019). Location-centered house price prediction: A multi-task learning approach. Retrieved from <https://arxiv.org/pdf/1901.01774.pdf>
- Geurts, P., Ernst, D., & Wehenkel, L. (2006). Extremely randomized trees. *Machine Learning*, 63(1), 3-42.
- Goodman, A. C., & Thibodeau, T. G. (2003). Housing market segmentation and hedonic prediction accuracy. *Journal of Housing Economics*, 12(3), 181-201.
- Gupta, R., Lau, C. K. M., Plakandaras, V., & Wong, W.-K. (2019). The role of housing sentiment in forecasting U.S. home sales growth: Evidence from a Bayesian compressed vector autoregressive model. *Economic Research-Ekonomiska Istraživanja*, 32(1), 2554-2567.
- Hu, G., Wang, J., & Feng, W. (2013). Multivariate regression modeling for home value estimates with evaluation using maximum information coefficient. In R. Lee (Ed.), *Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing 2012. Studies in Computational Intelligence* (Vol. 443, pp. 69-81). Heidelberg: Springer.
- Ibrahim, M. F., Cheng, F. J., & Eng, K. H. (2005). Automated valuation model: An application to the public housing resale market in Singapore. *Property Management*, 23(5), 357-373.
- Jaen, R. D. (2002). Data mining: An empirical application in real estate valuation. In S. Haller & G. Simmons (Eds.), *FLAIRS-02. Proceedings of the Fifteenth International Florida Artificial Intelligence Research Society Conference* (pp. 314-317). Menlo Park, CA: AAAI Press.
- Kok, N., Koponen, E.-L., & Martínez-Barbosa, C. A. (2017). Big data in real estate? From manual appraisal to automated valuation. *The Journal of Portfolio Management*, 43(6), 202-211.
- Lee, K., Kim, H., & Shin, D. H. (2019). Forecasting short-term housing transaction volumes using time-series and internet search queries. *KSCE Journal of Civil Engineering*, 23(6), 2409-2416.
- Limsombunchai, V. (2004, June). *House price prediction: Hedonic price model vs. Artificial neural network*. Paper presented at the 2004 New Zealand Agricultural and Resource Economics Society (NZARES) Conference, Blenheim, New Zealand.
- Lisi, G., & Iacobini, M. (2013). Real estate appraisals, hedonic models and the measurement of house price dispersion. *Journal of Economics and Econometrics*, 56(1), 61-73.
- Roberto, M., Luciani, A., & Pangallo, M. (2018, April). *The potential of big housing data: An application to the Italian real-estate market* (Working Paper No. 1171). Rome: Bank of Italy. Retrieved from https://www.bancaditalia.it/pubblicazioni/temi-discussione/2018/2018-1171/en_tema_1171.pdf

- Moosavi, V. (2017). Urban data streams and machine learning: A case of Swiss real estate market. Retrieved from <https://arxiv.org/ftp/arxiv/papers/1704/1704.04979.pdf>
- Mu, J., Wu, F., & Zhang, A. (2014). Housing value forecasting based on machine learning methods. *Abstract and Applied Analysis*, 2014, 648047.
- Ng, A. (2015). *Machine learning for a London housing price prediction mobile application* (Bachelor's thesis). Imperial College, London, United Kingdom. Retrieved from https://www.doc.ic.ac.uk/~mpd37/theses/2015_beng_aaron-ng.pdf
- Nghiep, N., & Al, C. (2001). Predicting housing value: A comparison of multiple regression analysis and artificial neural networks. *Journal of Real Estate Research*, 22(3), 313-336.
- Núñez Tabales, J. M., Caridad y Ocerin, J. M., & Rey Carmona, F. J. (2013). Artificial neural networks for predicting real estate price. *Revista de Metodos Cuantitativos para la Economía y la Empresa*, 15, 29-44.
- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., ... Duchesnay, E. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12, 2825-2830.
- Perna, D., & Tagarelli, A. (2018, July). Learning to rank social bots. In *HT '18. Proceedings of the 29th ACM Conference on Hypertext and Social Media* (pp. 183-191). New York, NY: Association for Computing Machinery.
- Pow, N., Janulewicz, E., & Liu, L. (D). (2014). *Applied machine learning project 4 prediction of real estate property prices in Montreal*. Retrieved from http://rl.cs.mcgill.ca/comp598/fall2014/comp598_submission_99.pdf
- Saarela, M., & Jauhiainen, S. (2021). Comparison of feature importance measures as explanations for classification models. *SN Applied Sciences*, 3(2), 272.
- Schapiro, R. E. (2013). Explaining AdaBoost. In B. Schölkopf, Z. Luo & V. Vovk (Eds.), *Empirical Inference* (pp. 37-52). Heidelberg: Springer.
- Scornet, E. (2020). Trees, forests, and impurity-based variable importance. Retrieved from <https://hal.archives-ouvertes.fr/hal-02436169v2/document>
- Shakhnarovich, G., Darrell, T., & Indyk, P. (Eds.). (2006). *Nearest-neighbor methods in learning and vision: Theory and practice*. Cambridge, MA: MIT Press.
- Tibell, R. (2014). *Training a multilayer perceptron to predict the final selling price of an apartment in co-operative housing society sold in Stockholm City with features stemming from open data* (Unpublished master's thesis). KTH Royal Institute of Technology, Stockholm, Sweden.
- Yang, Y., Liu, J., Xu, S., & Zhao, Y. (2016). An extended semi-supervised regression approach with co-training and geographical weighted regression: A case study of housing prices in Beijing. *ISPRS International Journal of Geo-Information*, 5(1), 4.
- Zhang, Y., Ma, Q., Chiang, Y.-Y., Knoblock, C., Zhang, X., Yang, P., ... Hu, X. (2019). Extracting geographic features from the Internet: A geographic information mining framework. *Knowledge-Based Systems*, 174, 57-72.
- Zheng, X., Lin, Z., Wang, X., Lin, K.-J., & Song, M. (2014). Incorporating appraisal expression patterns into topic modeling for aspect and sentiment word identification. *Knowledge-Based Systems*, 61, 29-47.
- Zhu, H., Xiong, H., Tang, F., Liu, Q., Ge, Y., Chen, E., & Fu, Y. (2016, August). Days on market: Measuring liquidity in real estate markets. In *KDD '16. Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (pp. 393-402). New York, NY: Association for Computing Machinery.