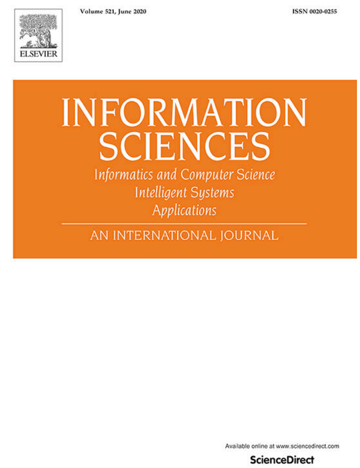
Journal Pre-proofs



Real estate price estimation through a fuzzy partition-driven genetic algorithm

Barbara Cardone, Ferdinando Di Martino, Sabrina Senatore

PII: S0020-0255(24)00355-4
DOI: https://doi.org/10.1016/j.ins.2024.120442
Reference: INS 120442

To appear in: *Information Sciences*

Received Date: 21 November 2023
Revised Date: 29 January 2024
Accepted Date: 3 March 2024

Please cite this article as: B. Cardone, F. Di Martino, S. Senatore, Real estate price estimation through a fuzzy partition-driven genetic algorithm, *Information Sciences* (2024), doi: https://doi.org/10.1016/j.ins.2024.120442

This is a PDF file of an article that has undergone enhancements after acceptance, such as the addition of a cover page and metadata, and formatting for readability, but it is not yet the definitive version of record. This version will undergo additional copyediting, typesetting and review before it is published in its final form, but we are providing this version to give early visibility of the article. Please note that, during the production process, errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

© 2024 Published by Elsevier Inc.

Real estate price estimation through a fuzzy partition-driven genetic algorithm

Barbara Cardone[a], Ferdinando Di Martino[a*], Sabrina Senatore[b]

[a] Dipartimento di Architettura, Università degli Studi di Napoli Federico II, Via Toledo 403, Napoli, Italy

[b] Dipartimento di Ingegneria dell'Informazione ed Elettrica e Matematica applicata, Università degli Studi di Salerno,

Via Giovanni Paolo II, 132, 84084 Fisciano, Italy

Abstract. Evaluating the actual price of a residential property is a critical issue in the real estate market. Real estate market practitioners gauge a property's price by considering features such as property type and residential area. Subsequently, they evaluate the property's intrinsic features, such as condition, sun exposure, scenic views, and ancillary amenities. Finally, extrinsic features such as the proximity of services and infrastructure are assessed. This paper proposes a new genetic approach for selecting residential properties that meet the purchase offer and the intrinsic and extrinsic characteristics desired by the client. Since the real estate market's changes can influence extrinsic features, the method introduces price fluctuations of properties. Extrinsic features are modelled as fuzzy partitions: each fuzzy set describes a qualitative aspect of the corresponding feature that, expressed in a linguistic term, has a human-like interpretation. Then, a deviation value (fluctuation) from the average price of the property is considered for each fuzzy set in the partition. All the property features, extrinsic and intrinsic, are encoded in the chromosome genes of the genetic algorithm. The fitness function calculates the distance between the unit price of the property and the purchase offer. Some case studies were conducted in various Italian municipalities, using the average price per square meter of residential properties the *Osservatorio del Mercato Immobiliare* (OMI) assigned. Depending on customer requirements and preferences, different OMI zones were selected using additional characteristics such as type, location, conservation, and proximity to various urban services. The results demonstrated the effectiveness of the proposed approach for all the case studies, showing how the optimal solution represents a good compromise between customer preferences and market offerings.

Keywords: fuzzy set, fuzzy partition, linguistic terms, fuzzy entropy, fuzziness, genetic algorithm, real estate, property price assessment, reliability measure.

Journal Pre-proofs

*Corresponding author.

Email addresses: b.cardone@unina.it (Barbara Cardone), fdimarti@unina.it (Ferdinando Di Martino), ssenatore@unisa.it (Sabrina Senatore)

Journal Pre-proofs

Journal Pre-proofs

1. Introduction

The real estate sector plays a dominant role in many countries, evaluated in terms of volume, the share of the economy, and the labor force; it contributes to the development of the national economy by catalyzing investments.

One of the most critical aspects affecting this market is related demographics, i.e., population growth and migration, which can influence the demand for various real estate types. Interest rates and government tax policy can also impact the market, affecting demand and supply behaviour.

The real estate market in the European Union faced some critical situations, during the financial crisis of 2008–2009, from which it successfully recovered and then the situation due to the COVID-19 pandemic [13]. Specifically, the pandemic has led to considerable volatility in terms of uncertainty, which has affected investment decisions, construction timelines, and the market's overall stability [2]. The assessment of the urban housing market during the COVID-19 pandemic required careful analysis of various factors, such as changes in buyer preferences, job losses, government policies, and economic conditions. The pandemic has resulted in shifts in housing preferences, with increased demand for larger homes with dedicated office spaces and outdoor areas, as remote work and stay-at-home orders became more prevalent [12, 42].

The housing demand in everyday situations is driven by some common factors, such as the house size (for example, area covered and number of rooms) and location, by considering as critical factors the proximity to public transportation or, in general, to services (school, hospital, shoppers).

Although the main factor determining the value of houses is undoubtedly demand [3, 19], there are unpredictable features, such as location considered “fashionable” whose final price can deviate from their expected value.

The price estimate of a property's price is determined by numerous features that can be classified as objective (extrinsic) features, such as the distance of the nearest bus stop, or subjective (intrinsic) features, such as artistic value or building style. The difficulty in standardizing valuation methods is mainly due to the different metrics for valuing each nation's objective and subjective/hedonic features. In assessing the health of the urban housing market, deviation from the average price of a property is an important metric that provides insights into the market's overall health [44]. A positive deviation from the average price indicates that the market is expanding; a negative deviation

Journal Pre-proofs

indicates a market contraction. Deviation from the average cost is a valuable tool for real estate professionals, investors, and analysts evaluating the current state and prospects of the urban housing market. Statistical analysis also plays a crucial role in interpreting these deviations and their impact on the real estate market [46]. A positive or negative deviation from this price helps the real estate agent find an actual price close to the user's requirement.

The paper proposes a novel approach to real estate estimation; it combines, on the one hand, the genetic algorithm [18] for finding the property whose price meets the customer expectation and on the other hand, it introduces human-like modelling of property features through the fuzzy partitions.

More specifically, properties are genes in the genetic algorithm; each gene encodes the quantitative property features modelled as fuzzy partition: each fuzzy set in the fuzzy partitions represents a qualitative feature evaluation.

Fuzzy logic provides a framework to handle uncertainty and imprecision in data, allowing for a more nuanced and flexible real estate appraisal process. The appraisal model can accommodate imprecise data, such as subjective expert opinions or property features, using fuzzy sets and membership functions. Fuzzy logic allows practitioners to work with linguistic variables, such as "high," "low," "very good," or "fair," which are very common in natural language and then more immediate in real estate assessments. For example, it is easier for the expert to describe the feature "distance from some bus service" using a linguistic expression such as very close, close, not very close, far, etc., rather than provide a quantitative evaluation of the feature, often tricky to describe to the customer. By constructing a fuzzy partition on the domain of a feature, the expert expresses the value of a feature using natural language expressions. As in the real estate market domain, the expert can estimate a possible deviation (i.e., the variability) of certain features from the average price.

The genetic algorithm's fitness function aims to minimize the gap between the asking price and the assessed price for the solution. The trustworthiness of the solution is evaluated by the measure of reliability, introduced as a parameter of the fitness function, using the fuzzy entropy measure developed by De Luca and Termini [14].

In a nutshell, the paper's contribution can be summarized as follows:

- A genetic algorithm for urban housing market assessment is introduced: it works on the intrinsic and extrinsic features desired by the client and finds the property closest to the client’s request.

Journal Pre-proofs

- Extrinsic features are modeled as fuzzy partitions to facilitate human understanding. In the previous example of the feature: “distance from some bus service”, the expert could define a fuzzy partition consisting of three fuzzy sets with the labels: *close, not very close*, and *far*. The linguistic approach helps appraisers intuitively grasp the proximity condition from bus service, fostering greater transparency and confidence in the appraisal process. As stated, in the fuzzy partition modelling, a deviation in property features is also considered.

- The synergy between the genetic algorithm and fuzzy partitions used in gene modeling makes the approach innovative and effective in finding the optimal solution. The approach, indeed, introduces the evaluation of the reliability of the computed solution, measuring its fuzziness. Reliability assessment provides an estimate of the property's price that is typically complex to gauge because of the challenge of predicting fluctuations in economic values assigned to property characteristics, also influenced by the seller's expected profits.

Our approach presents a hybrid solution that leverages a genetic algorithm to model property features as fuzzy partitions. Expert-defined fuzzy partitions assign a positive or negative deviation from the average unit price of the property. While traditional machine learning algorithms rely on considerable data size for a practical learning phase, our approach eliminates the need for parameter tuning or a time-consuming training phase.

The paper is organized as follows. The next section introduces related literature, focusing on real estate modelling and applications.

Then, in Section 3, the preliminary theoretical concepts related to fuzzy partition, fuzzy entropy, and fuzziness are introduced. The overall framework is presented and discussed in Section 4. Section 5 is instead devoted to presenting some case studies in the Italian territory; finally, conclusions and future perspectives close the paper.

2 Related works

Journal Pre-proofs

The real estate market is constantly evolving, influenced by several aspects such as demographics, interest rates, government regulation, and, in short, global economic health. In particular, the changing nature of the real estate market strongly affects the valuation process of real estate, which can require some effort on the part of qualified real estate appraisers.

In market research, conjoint analysis [10] has become of critical interest since it represents a form of statistical analysis that firms use to study how customers assess different components or features of their assets. The idea behind it is that any product can be made up of a set of attributes that influence the user's perception of the value of an item or service. This analysis in real estate marketing [5, 6], focuses also on the geospatial components [17]. However, several parameters in evaluating the property price complicate predicting market price behaviour and purchase decisions.

The high price variability of some specific assets due to the instability of the local market, with, for example, market segments evolving at different rates, such as high-end luxury condominiums [3].

Traditional models for real estate valuation rely on hedonic regression, which identifies the constituent features of an asset to extract relations between them and the asset's value. This way, the property price is learned from the specified features [23, 29, 35]. Some approaches are based on multiple regression analysis to evaluate the value of a property by subjective features [40, 49].

Machine Learning approaches are widely employed in real estate to predict property prices [28, 43, 45, 47].

In [7] a single imputation Chained Paasche method for building real estate price indices is proposed. The approach shows that the prediction accuracy is higher for ML-based models than for linear ones. However complex black-box machine learning algorithms do not provide interpretable and explainable results that can find a natural answer in systems with fuzzy logic [50] and explainable AI (XAI) capabilities [26, 48]. In [30], for instance, fuzzy systems and fuzzy rules offer interpretable insights into feature-price relations through linguistically readable rules.

Several learning models, such as neural networks, are well suited to assess a property's economic value considering objective and subjective features. An adaptive neuro-fuzzy inference system (ANFIS) [21] for real estate appraisal was proposed for exploiting features such as location, year of construction and sale, square footage on each floor, number of baths, etc. The approach combines

Journal Pre-proofs

neural networks with fuzzy inference, outperforming multiple regression analysis (MRA), using only some features instead of all.

Neural network-based pricing models in mass real estate are proposed by [16, 32, 36]. In contrast, in [11], a neural parametric model of the real estate market value in the EU countries depending on the impact of a set of macroeconomic indicators is proposed.

Similar performance is obtained by comparing a neural network model to a multiple regression model using the Principal Component Analysis technique to reduce the number of parameters [49]. In [22], a case-based reasoning approach overcomes Hedonic price approaches, multivariate regression analysis and neural networks.

Several studies focus on defining automated approaches that can support real estate appraisers in accurately valuing individual buildings and a mass valuation of properties in an area to increase the accuracy of the assessments.

Table 1: Synthetic description of the main works in the real estate assessment context

Research	Method/Model Used	Primary Objective	Key Findings
[10]	Conjoint Analysis	Evaluate how customers assess components of assets.	Customer preferences for specific attributes influence market behavior.
[23], [28], [29], [35], [38], [40]	Traditional Real Estate Valuation using ML Models (Hedonic Regression, Multiple Linear Regression)	Assess property value based on various features.	Better performance with ML models with respect to traditional regression methods in predicting property prices.
[7]	Chained Paasche Method	Create real estate price indices using predictions.	Higher prediction accuracy compared to linear models.
[32], [21]	Hybrid approaches (Neural network and fuzzy logic systems)	Assess the economic value of properties.	Fuzzy neural networks and ANFIS combine to improve property valuation accuracy.

Journal Pre-proofs

[16, 36]	Neural Networks	Predict property prices on a large scale.	Neural network-based pricing models prove effective for mass real estate valuation.
[11]	Neural Parametric Model	Model real estate market value in EU countries and provide price forecasts	Forecast real estate prices, demonstrating the impact of macroeconomic variables on property values
[49]	Regression-Based and AI-Based Methods	Improve real estate price predictions.	Comparative study among traditional and non-traditional regression methods with neural networks methods. Use of PCA.
[22]	Case-Based Reasoning	Quantitative Comparative Approach for estimating correction coefficients.	Case-based reasoning outperforms classical Hedonic price approaches, multivariate regression analysis and neural networks.
[31]	Clustering	Automate property valuation.	Crisp and fuzzy clustering algorithms automate property valuation in urban areas.
[34], [37], [38], [41]	Genetic Algorithm	Combined to other methods to enhance prediction accuracy	Improve the features selection; mitigate uncertainty in real estate investments
[50]	Fuzzy Expert systems (comparative study)	Enhance precision in property value assessment (by feature reduction)	Fuzzy logic and memory-based reasoning alongside neural networks and regression in property value assessment.
[2], [27]	(Fuzzy) Decision Trees,	Support experts in housing appraisal.	Fuzzy decision trees provide interpretable results for housing appraisal.

Journal Pre-proofs

[40]	Support Vector Machine, Gradient Boosting, Random Forest (comparative study)	Evaluate property prices, with ML techniques.	RF and GBM outperform SVM in terms of prediction accuracy for property prices.
[24]	Ant Colony Optimization (ACO), Least Square Support Vector Machine (LS-SVM), Computable General Equilibrium (CGE) Model	Optimized real estate valuation model and assess the impact of real estate price fluctuations on macroeconomic equilibrium.	Ant colony algorithm optimization improves SVM convergence speed for property price modeling.

In [31] a clustering-based approach automates property valuation according to sales comparison. Properties located in a city are partitioned by crisp and fuzzy clustering algorithms for property valuation. Fuzzy logic-based expert systems are an alternative to model housing appraisal due to the similarity between such methods and the human approach to decision-making [49].

Further learning models devoted to assisting experts in housing appraisal range from decision trees [2], fuzzy systems [27] to neural [36], and hybrid methods [33], also based on comparable sales method [25]. Moreover, in [40] three machine learning algorithms, including support vector machine (SVM), gradient boosting machine (GBM), and random forest (RF), were used for property price evaluation, demonstrating better performance, in terms of prediction, with RF and GBM, compared to SVM. To address the slow convergence speed of SVM and the proper selection of design parameters, in [24], a least squares support vector machine model based on ant colony algorithm optimization is proposed.

In line with our approach, genetic algorithms are often combined with other methods, such as Hellwig's method, to mitigate uncertainty in real estate investment decisions [38] or for improving feature selection to enhance prediction accuracy [34]; similarly in [41], a genetic algorithm-optimized neural network model for real estate assessment is proposed; the model in [37] combines deep belief restricted Boltzmann machine and genetic algorithm to aid construction companies to assess the market before initiating new projects.

The main works introduced in this section are summarized in Table 1.

The continuous evolution of the housing market makes it challenging to analyze steady-state relationships with the labor market and possible interdependencies between income and different property prices, such as housing, rents, and property prices. The difficulties in accurately estimating the value of a property are due to many factors, including idiosyncratic personal circumstances that influence the transaction price, that are tricky to capture systematically.

Most of the approaches proposed in the literature need to be revised to manage the high variability of many parameters and implement qualitative features that are subject to constant change. For example, such high variability and possible unpredictable factors (for example, one neighborhood considered more attractive than another) could lead to a particular asset price deviating from the expected value.

In addition, for these reasons, collecting massive amounts of real estate data to train models could be complicated because of the difficulty of evaluating features depending on indicators and prices in a dynamic environment. The proposed approach would overcome the need to train a model (with attendant costs in terms of computational resources and time) precisely to avoid using feature values that can vary due to sudden changes in the real estate market and relies on the experience of human appraisers in defining fuzzy partitions reflecting current real estate feature valuation.

3. Preliminaries

This section is devoted to presenting two preliminary notions of the theoretical background that underlie the proposed approach:

- the definition of the fuzzy partition to model the features of interest. In our approach, the fuzzy partition is designed by the real estate expert to get a partitioning of the feature ranges providing a human-like description (by using linguistic terms modeled by fuzzy sets, such as “high”, “good”, “far from”, etc.), suitably reflecting the meaning desired by the client. This way, the expert can adequately set the swing around the basic price per unit, to widen the purchase opportunities satisfying the buyer's demands.
- A measure assessing the degree of fuzziness: it can be regarded as an entropy measure in the sense that it evaluates the uncertainty about the presence or absence of a specific feature, described as fuzzy partition. In particular, the degree of fuzziness of the element (in our case, a feature value) depends on the fuzzy partition designed. Our approach will exploit the fuzziness degree of the feature values in the design of the genetic algorithm's fitness function.

3.1 Fuzzy partition

Let FP = {F_1, ..., F_N} be a family of N fuzzy sets, defined on a universe of the discourse U. According to [39], FP is *a fuzzy partition* of U if the following conditions hold:

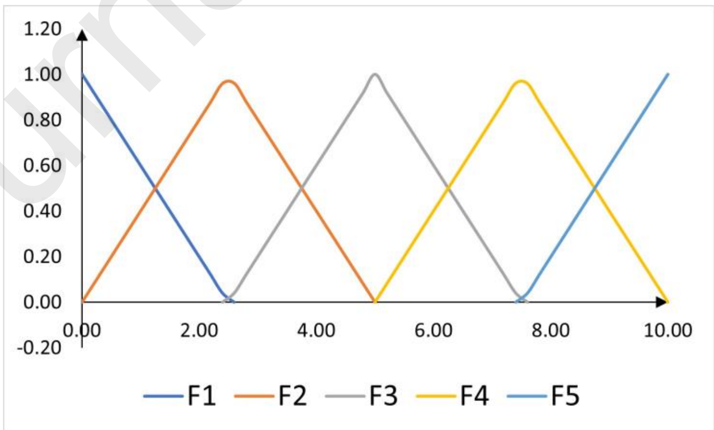
$$\forall F_j \in FP \;\; \exists u \in U : \mu_j(u) \neq 0 \quad (1)$$

$$\sum_{j=1}^{N} \mu_j(u) = 1 \;\; \forall u \in U \qquad (2)$$

where $\mu_j : U \rightarrow [0, 1]$ is the membership function of the j^{th} fuzzy set in FP.

The constraints (1) and (2) claim that all fuzzy sets in FP are not empty and the union of the membership degree of any element u in U to a fuzzy set in FP is always equal to 1.

The Ruspini definition [39] does not require that fuzzy sets necessarily be disjoint. In Figure 1, an example of a fuzzy partition formed by five triangular fuzzy sets is defined in the universe of discourse U = [0, 10].



Journal Pre-proofs

Figure 1. Example of fuzzy partition with triangular fuzzy sets.

A fuzzy partition FP having N triangular fuzzy sets defined on U = [a, b], with a, b $\in R$ can be described as follows:

$$\mu_j(u) = \begin{cases} 0 & u < u_{j-1} \\ \frac{u - u_{j-1}}{h} & u_{j-1} \le u \le u_j \\ \frac{u_{j+1} - u}{h} & u_j \le u \le u_{j+1} \\ 0 & u > u_{j+1} \end{cases} \qquad j = 1, \ldots, N \tag{3}$$

where h = (b-a) / (N-1), $u_0 = u_1 = a$, $u_j = u_{j-1} + h$ for each j = 2,…,N and $u_{N+1} = u_N$.

In the example of Figure 1, a = 0, b = 10, N = 5 and h = 2.5. So, we have $u_0 = u_1 = 0$, $u_2 = 2.5$, $u_3 = 5$, $u_4 = 7.5$, $u_5 = u_6 = 10$.

3.2 Fuzzy entropy and fuzziness

Let FP = {F_1, ..., F_N} be a fuzzy partition of a universe of discourse U. Let $\mu_j : U \to [0, 1]$ be the membership function of the j^{th} fuzzy set F_j. To measure the *fuzziness degree* of belonging of an element u ∈ U to F_j, De Luca and Termini [14] proposed the following fuzzy entropy function:

$$h(\mu_j(u)) = \begin{cases} 0 & \text{if } \mu_j(u) = 0 \\ -\mu_j(u)\log_2(\mu_j(u)) - (1\text{-}\mu_j(u))\log_2(1-\mu_j(u)) & \text{if } 0 < \mu_j(u) < 1 \\ 0 & \text{if } \mu_j(u) = 1 \end{cases} \quad (4)$$

where the continuous function *h*: [0,1] → [0,1] has the following properties:

- is monotonically increasing in [0, ½];
- is monotonically decreasing in [½, 1];
- h(0) = h(1) = 0;
- $h(\mu_j) = h(1 - \mu_j)$;
- the maximum value $h(\mu_j) = 1$ is obtained for $\mu_j = 0.5$.

A measure of the degree of fuzziness of the element *u* is defined in [14] as:

$$H(u) = \frac{1}{N}\sum_{j=1}^{N} h(\mu_j(u)) \quad (5)$$

and constraint (2) holds as FP is a Ruspini fuzzy partition.

The proposed measure of fuzziness was introduced in [8, 15] to assess the fuzziness of fuzzy clusters: a fuzzy cluster was described by a fuzzy set whose data points assumed the corresponding values of the elements in the partition matrix as their degree of membership. In [9] the authors used the fuzziness measure (5) to evaluate the accuracy of the classification of documents according to the prevailing emotional category.

The fuzziness *H(u)* ranges in [0, 1]; the minimum value is 0 and it is obtained when the element belongs only to a fuzzy set with membership degree 1.

The degree of fuzziness of the element depends on the fuzzy partition created.

Journal Pre-proofs

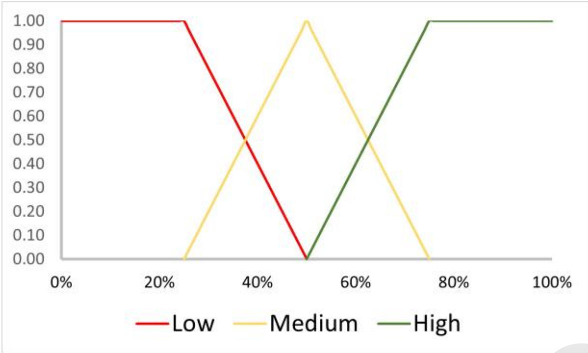


Figure 2. A fuzzy partition composed of a triangular and two semi-trapezoidal fuzzy sets.

In the example of Figure 2, a fuzzy partition of the domain U = [0%, 100%] described by three fuzzy sets is given. The first and last fuzzy sets are semi-trapezoidal fuzzy sets, and the second fuzzy set is a triangular fuzzy set.

Then, let us consider three elements u_1, u_2, and $u_3 \in U$. The fuzziness degree measured for the elements: u_1 = 20%, u_2 = 63%, u_3 = 71% is shown in Table 2.

Table 2. Fuzziness calculated for three elements using the fuzzy partition in Figure 2.

Element	u_j	μ_{Low}	μ_{Medium}	μ_{High}	Fuzziness
u_1	20.00%	1.00	0.00	0.00	0.00
u_2	63.00%	0.00	0.48	0.52	0.67
u_3	71.00%	0.00	0.16	0.84	0.44

Figure 3 shows a finer fuzzy partition of the domain U formed by seven fuzzy sets. The first and last fuzzy sets are semi-trapezoidal fuzzy sets; the other fuzzy sets are triangular fuzzy sets.

The fuzziness values measured for the three elements u_1, u_2 and u_3 are shown in Table 3.

Journal Pre-proofs

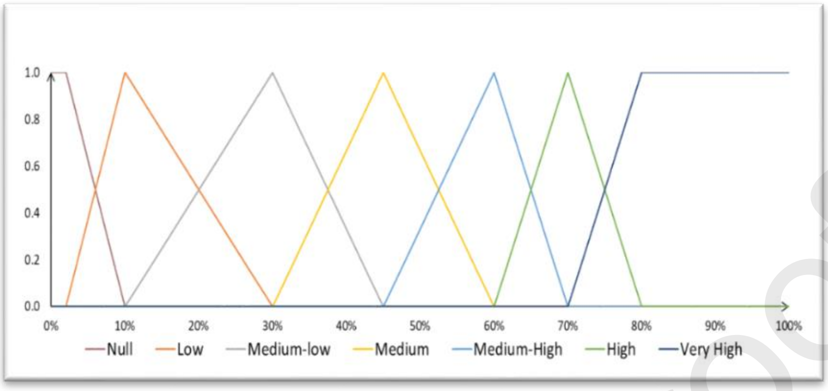


Figure 3. A fuzzy partition composed of five triangular and two semi-trapezoidal fuzzy sets.

Table 3. Fuzziness calculated for three elements using the fuzzy partition given in Figure 3.

Element	u_j	μ_{Null}	μ_{Low}	$\mu_{Medium\text{-}low}$	μ_{Medium}	$\mu_{Medium\text{-}High}$	μ_{High}	Fuzziness
u_1	20.00%	0.00	0.50	0.50	0.00	0.00	0.00	0.28
u_2	63.00%	0.00	0.00	0.00	0.00	0.70	0.30	0.25
u_3	71.00%	0.00	0.00	0.00	0.00	0.90	0.10	0.10

From Tables 2 and 3, let us notice the different fuzziness values calculated for the element u_1, which in the first case was zero (i.e., it fully belongs to the fuzzy set labeled "Low"), in the second case it is equal to 0.28; instead, the fuzziness values calculated for u_2 and u_3, are equal to 0.67 and 0.44, in Table 2 while they are smaller in Table 3, with values equal to 0.25 and 0.10, respectively. This aspect reveals that a more refined partition (i.e., the fuzzy partition shown in Figure 3) allows for a better association of the elements in each fuzzy set.

4. The proposed framework

The proposed framework aims to support a real estate agent in discovering the property whose features best meet the client's demands regarding available budget and preferences. Our model employs a genetic approach to address this challenge.

Formally, let X be the basic price per unit assigned to properties belonging to a specific residential category located in one particular urban area (e.g., the average price per square meter in the urban area of interest). Then, let Y be the price per unit the client intends to spend. Suppose the property to be searched for has n features, which may be intrinsic (e.g., the state or age of construction) or extrinsic (e.g., the distance to a public utility).

Let us consider c_i, with i = 1, ..., n, the i^{th} feature of the property and pr_i the corresponding preference value, which can assume the values 0 or 1. For example, if the feature is "proximity to hospital" has pr_i is 1 would mean that the client would like the proposed solution to be close to a hospital; or, if the feature is "presence of the central heating system," and pr_i is 1 would mean that the client would like the property to have a central heating system.

For each feature, the expert creates a fuzzy partition according to [39] composed of N_i fuzzy sets $\{F_{i1}, ..., F_{iNi}\}$. He assigns to each fuzzy set a deviation, i.e., a growth/decrease factor from the basic price X per unit for the i^{th} feature, Δx_i, depending on the client's preference.

For example, the fuzzy partition of the feature c_i = "proximity to a hospital" can be composed of 3 fuzzy sets with labels F_{i1} = "near to," F_{i2} = " medium distance," F_{i3} = "far from." Let us suppose that the expert assigns a value $\Delta x_{i1} = \Delta$ to the fuzzy set F_{i1}, $\Delta x_{i2} = 0$ to the fuzzy set F_{i2} and the value $\Delta x_{i3} = -\Delta$ to the fuzzy set F_{i3}. In this way, the base price per unit of the property in the selected urban area will be increased by the value Δ if the property is located near a hospital, and it will be decreased by the same value ($-\Delta$) if it is located far from the hospital; otherwise, it will remain unchanged if it is situated at an intermediate distance from the hospital facility.

The genetic algorithm finds the optimal solution to identify the appropriate feature values of the properties that, evaluated by price per unit, match the customer's purchase proposal Y as closely as possible.

According to the genetic algorithm, the optimal solution is selected by the population of best individuals over each generation and generates new by crossover and mutation.

Each individual consists of n genes representing the n features of the property, and each value of a gene is a feature value. As stated, the expert defines a fuzzy partition to a feature c_i and assigns a deviation Δx_i to the i^{th} gene in the solution. The value of Δx_i is set equal to the deviation assigned to the fuzzy set in the fuzzy partition defined for the feature c_i to which the individual belongs with highest membership degree. In other words, the expert sets the growth/decrease factor Δx_i, as the maximum of the membership values of the fuzzy sets enclosed in the fuzzy partition for the feature.

The price value of the k^{th} individual will be calculated as follows.

$$X_k = X + \sum_{i=1}^{n} \Delta x_i \quad (6)$$

where the summation term represents the overall deviation of the basic price X per unit of all; the term encloses the contribution Δx_i i = 1,…, n of all the n features.

The initial fitness function calculated for the k^{th} individual is:

$$f_k^0 = \frac{1}{|X_k - Y| + 1 \cdot 10^{-3}} \quad (7)$$

Let us notice that f_k^0 increases when the price value X_k of a property (calculated by the model) approaches the price Y offered by the customer. The value $1 \cdot 10^{-3}$ was empirically determined and added to prevent that f_k^0 from tending to infinity when X_k is equal to Y; in this case, $f_k^0 = 1 \cdot 10^3$, which is its maximum value.

The use of a distance metric as a fitness function to determine the difference between the unit price and the purchase offer could be a reasonable choice, particularly in terms of simplicity, direct comparison, and transparency. Firstly, a distance metric allows for a direct comparison between the unit price and the purchase offer. It provides a clear and objective way to quantify the difference between these two values, making it simple for appraisers to understand and interpret the results.

Different distance metrics, such as the Minkowski distance, Manhattan distance, and Euclidean distance, can be used, depending on the appraisal's specific requirements. This flexibility allows the fitness function to be tailored to the appraisal scenario and the nature of the variables involved. Furthermore, a straightforward fitness function can speed up the procedure because it is frequently easy to implement a distance metric and the computational cost is typically low. Finally, a distance metric adds transparency to the appraisal process by providing a precise way to gauge how close or far the purchase offer is from the unit price. This transparency can be crucial, especially when dealing with clients, buyers, or sellers, as it helps justify the appraisal value. However, the use of a distance metric also has some limitations such as the consideration of additional factors to ensure a comprehensive and accurate assessment of property values. Determining the appropriate weights/scaling factors for different variables in the distance metric can be challenging.

To this purpose and to guarantee the solution's reliability, the fitness function in (6) was weighted by considering additional factors for describing the fuzziness associated with each feature and the preferred features in a solution.

So, the fitness function included two multiplicative factors, namely H_k, the *fuzziness*, and λ_k, the *comprehensive preference* of the solution.

So, the fitness function included two multiplicative factors, namely H_k, the *fuzziness*, and λ_k, the *comprehensive preference* of the solution.

More formally, the revised fitness function calculated for the k^{th} individual is given by:

$$f_k = f_k^0 \cdot (1 - \mathrm{H_k}) \cdot \lambda_k \quad k = 1, \dots, \mathrm{N} \quad (8)$$

where H_k is the fuzziness of the k^{th} individual as defined in Section 3.1. This factor is inserted in the fitness function in the form (1- H_k) to consider the fuzziness of the solution (e.g., when the fuzziness H_k is 0, the contribution will be unitary).

The *comprehensive preference* λ_k of the solution represents the synthesis of the client's preferred features, i.e., those he selected as crucial in finding the desired property. It takes values in the range [0, 1], described by the following formula:

Journal Pre-proofs

$$\lambda_k = \frac{1}{n}\sum_{i=1}^{n} l_i \quad (9)$$

where:

$$l_i = \begin{cases} 1 \text{ if } pr_i = 0 \\ 1 \text{ if } pr_i = 1 \text{ and } \Delta x_i \geq 0 \\ 0 \text{ if } pr_i = 1 \text{ and } \Delta x_i < 0 \end{cases} \quad (10)$$

The value l_i equals 0 means that the customer selects the i^{th} feature (i.e., $pr_i = 1$) as preferred, but the solution is not satisfactory with that feature because the deviation Δx_i is negative; when li takes the value 1 it means that the solution leverages the feature, regardless of the customer's preference for the i*th* feature.

The comprehensive preference λ_k (9) equal to 0 means that although the customer may select one or all features as preferred, the solution has negative price deviations for all features. Instead, it will be equal to 1 if, for all the features selected by the customer, the price deviation is positive.

In the genetic algorithm, the selection process of the individual is achieved by using the roulette wheel selection method. The probability of selection (selection rate) p_k of the k^{th} individual is given by

$$p_k = \frac{f_k}{\sum_{h=1}^{N} f_h} \quad (11)$$

where N is the number of individuals and f_h (h= 1,…, N) are the computed fitness values given by (8). This method of selection allows for significant variability in the selection of individuals; in fact, it also allows weaker individuals to be selected with a lower probability of selection p_k: precisely, individuals with high fitness can also be generated from parents with low fitness.

Each time the roulette wheel is spun, a number between 0 and 1 is drawn randomly to select an individual; then, a copy of the selected individual is placed in the mating pool. The roulette wheel is spun twice to draw the parent pair. This process is repeated until N pairs of parents are extracted.

Then, the crossover process is executed by setting a crossover probability p_{cr}.

Each selected pair in the mating pool is randomly assigned a value in [0,1]. If this value is greater or equal to the crossover probability (crossover rate), the pair generates two offspring, given by the recombination of the chromosomes of the two parents; otherwise, the two individuals are not recombined, and then, the mutation process is applied to them. To reduce the execution time, a single-point crossover operator is adopted, setting the crossover point randomly.

After crossover, the mutation operator is applied to each individual in the mating pool. The mutation is applied to prevent the algorithm from evolving toward local optima, and it is applied by randomly changing a gene to produce a new offspring; if the randomly assigned probability of the gene is lower than the fixed mutation probability (mutation rate) p_{mu}, *random reset* is performed on this gene, setting a random value from the set of allowable values assigned to the corresponding feature.

The selection, crossover, and mutation processes are applied to the new population; the algorithm stops after a prefixed number of generations N_g.

Figure 4 sketches the workflow of the proposed framework.

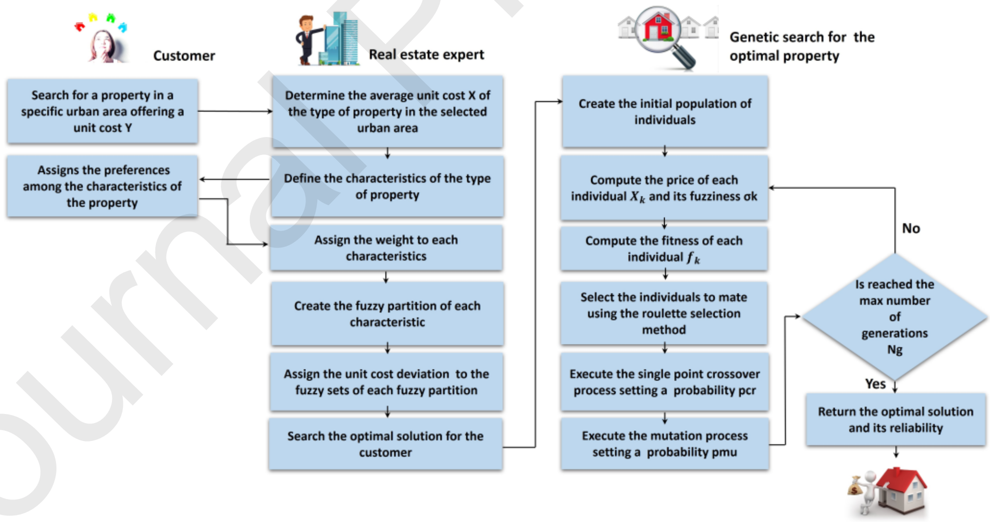


Figure 4. The overall workflow of the proposed framework.

The client intends to purchase a property in a particular urban area and proposes a unit purchase price, based on the available budget. After examining the client's requirements, the real-estate expert determines the average value of the type of property desired by the client and identifies the features of the property that affect its price. The customer selects the preferred features, assigning preference to each, and assesses if the solution (property price) meets his expectations. For example, suppose the i^{th} feature refers to the property's proximity to a transportation facility. In that case, the customer may assign a preference to this feature if the property he intends to purchase is close to a transportation facility; otherwise, if he does not consider the feature important, he will not assign a preference to the ith feature. The expert, considering the customer's preferences, sets the preference values $pr_1,\ldots, pr_n$ to the features, where pr_i will be set to 1 if the customer assigned a preference to the i^{th} feature, 0 otherwise.

Then, he defines the fuzzy partitions and assigns a unit price deviation to each fuzzy set. Let us recall that the deviation from the unit price is the highest membership value for the fuzzy set compared with the membership degrees of the other fuzzy sets in the fuzzy partition. At the end of this activity, the expert launches the genetic search algorithm, setting the number of individuals, the crossover and mutation probabilities p_{cr} and p_{mu} and the number of generations N_g.

The genetic algorithm returns the optimal solution, i.e., the individual with the highest fitness, and whose genes contain the values of the features that best approximate the property type sought. The fuzziness value associated with the selected individual allows the reliability of the current solution to be measured.

More formally, let $s = (s_1, s_2, \ldots, s_n)$ be the best solution selected, where s_i, i=1…n, is the i^{th} gene of the solution s. Let us assume that N_i is the number of fuzzy sets of the i^{th} fuzzy partition associated with the i^{th} gene and x_i is the value assumed by the i^{th} gene in the best solution (i.e., $s_i = x_i$), the fuzziness of this gene in the solution *s* is given as follows.

$$H(s_i) = \frac{1}{N_i}\sum_{j=1}^{N_i} h(\mu_j(s_i)) \quad (12)$$

where h(.) is the fuzzy entropy function defined in (4).

The fuzziness evaluation for the solution *s* will be given by

$$H(s) = \frac{1}{n}\sum_{i=1}^{n} H(s_i) \quad (13)$$

and the reliability of the solution *s* will be expressed as follows:

$$r(s) = 1 - H(s) \quad (14)$$

The maximum value of the reliability r(s) is 1, calculated when the fuzziness of the solution is zero.

Let us notice that our framework returns a ranked list of properties that meet the buyer's demand. However, further interactions with sellers, i.e., property owners, are required to negotiate the final price and obtain the final property proposals. That price must meet the buyer's requirements for features and, at the same time, satisfy the seller by approximating the desired sale price.

The proposed algorithm is schematized in the pseudocode below (Algorithm 1).

Algorithm 1: the proposed GA-based method

Input: The customer's purchase proposal Y

Output: The best solution s and its reliability r_s

---

1. Set the average unit cost X of the property
2. Set the n features $c_1, c_2, \ldots, c_n$
3. **For each** feature c_i $i = 1,\ldots n$:
4. Set the preference assigned by the customer pr_i
5. Create the fuzzy partition $\{F_{i1}, ..., F_{iN_i}\}$ formed by N_i fuzzy sets
6. **For each** fuzzy set F_{ij} $j = 1,\ldots N_i$:

Journal Pre-proofs

7. Set the unit cost deviation Δx_{ij}

8. **Next j:**

9. **Next i:**

10. Set the number of individuals N in the domain of the solutions

11. Set the crossover rate p_{cr}

12. Set the mutation rate p_{mu}

13. Set the maximum number of generations N_g

14. Initialize randomly the N individuals

15. g: = 1 //Generation number

16. **While** $g < N_g$

17. **For each** individual $s_k = (s_{k1}, s_{k2}, \ldots, s_{kn})$ $k = 1, \ldots, N$

18. Compute the comprehensive preference λ_k by (9)

19. Compute the fitness value f_k by (8)

20. **Next k:**

21. Execute the roulette wheel selection with a selection rate p_k given by (11)

22. Apply the single point crossover operator to the N individuals in the mating pool

23. Apply mutation operator to the N individuals in the mating pool

24. g:= g +1

25. **End while:**

26. **For each** individual $s_k = (s_{k1}, s_{k2}, \ldots, s_{kn})$ $k = 1, \ldots, N$

27. Compute the comprehensive preference λ_k by (9)

28. Compute the fitness value f_k by (8)

29. **Next k:**

30. Sort the N final individuals by their fitness in descending order

Journal Pre-proofs

31. s := s_1 // set the best solution to the first individual
32. Compute the fuzziness H_s of the best solution by (13)
33. Compute the reliability r_s of the best solution by (14)
34. **Return** s, r_s

The algorithm stops when the maximum number of generations N_g is reached. It returns the best solution s and its reliability r_s. The algorithm's computational complexity is $N \cdot N_g \cdot n$, where N is the number of individuals, N_g is the number of generations, and n is the number of features.

5. Case studies

The presented case studies concern properties located in Italian municipalities.

The study involves the periodic evaluation of property prices per unit by an organization called the "Osservatorio del Mercato Immobiliare" (OMI), the Italian institution of the Agenzia delle Entrate established as a permanent observatory of the real estate market.

OMI conducts real estate assessments every six months. These assessments determine the market values per unit area of residential buildings within specific homogeneous territorial areas. Based on these assessments, OMI defines what is known as an "OMI zone." An OMI zone represents a region with consistent urban planning and building characteristics.

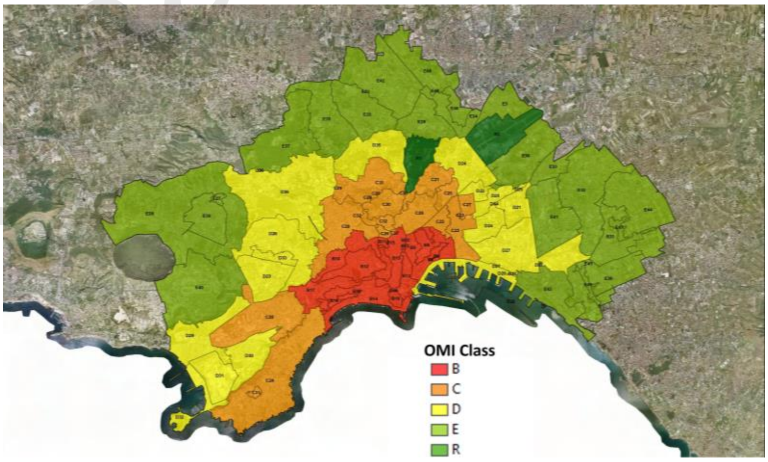
Each municipality is divided into OMI zones. These zones are critical for identifying and analyzing different areas within the municipality. Furthermore, an OMI zone belongs to an OMI class, which categorizes the zone into a broader macro-area within the municipality. There are five distinct OMI classes, as listed in Table 4.

Journal Pre-proofs

Table 4. Description of the five OMI zone classes

OMI Class	Description
B	Central
C	Semicentral
D	Peripherical
E	Suburban
R	Extra-urban

As an example of partitioning municipalities into OMI classes, Figure 5 is shown a thematic map of the OMI zones in the municipality of Naples; the areas of the city center are shown in red, the suburban and extra-urban areas of the city in light green, and dark green, respectively.



Journal Pre-proofs

Figure 5. Thematic map of the classes of the OMI zones in the city of Naples.

An OMI zone is characterized by both its class and a unique sequential number that distinguishes it from other zones within the same class. Every six months, estimates are conducted for each OMI zone to determine the range of minimum and maximum prices per square meter for various property types. These estimates encompass a range of different uses, such as residential use covering a variety of civil residential properties including affordable housing, garages, and parking spaces.

An example of estimation data for properties in the OMI zone is given in Table 5. Specifically, this example pertains to an OMI subzone falling under OMI class B, situated within the municipality of Naples, Italy.

Table 5. Example of estimation data referring to the OMI class B area, subzone 8 in the municipality of Naples (Italy).

Use	**Type of unit**	**State of preservation**	**Market value (€ /m^2)**	
			Min	**Max**
Residential	Civil housing	Normal	1750	2700
	Economic housing	Normal	1200	1850
	Garage	Normal	1500	2300
	Covered parking space	Normal	910	1400
Commercial	Warehouse	Normal	480	970
	Shop use	Normal	1650	3300

Journal Pre-proofs

Tertiary	Office	Normal	1250	2550
Productive	Laboratory	Normal	850	1700

The information presented in Table 5 refers to property valuations published by OMI during the first semester of 2022. These valuations encompass diverse property types located within subarea 8 of the OMI Class B zone, situated in the municipality of Naples. This particular subarea corresponds to the historical center of the city.

5.1 Data Description

Collected data are from the main Italian cities, selected to be representative of the whole national territory, during the first semester of 2022. Some of them were discarded, such as, for instance Venezia, since missing data. They are reported in Table 6, each one classified according to the four OMI classes, namely, B, C, D, E. Data related to Class R was also excluded due to the absence of data, which can be attributed to a lack of commercial interest from real estate agencies. For each class of a city, the data size (#samples) are provided, along with some key statistical parameters [41], including the minimum and maximum OMI class value (mOMI and MOMI, respectively), the minimum average value (mAV), maximum average value (MAV), minimum standard deviation (mSD), and maximum standard deviation (MSD).

From the data analysis, larger cities, and regional capitals, such as Milan and Rome, tend to have higher average minimum costs, while southern cities, like Bari and Palermo, sometimes, less economically developed have lower average minimum costs.

Table 6. Basic statistics of the OMI dataset for the main Italian cities.

Journal Pre-proofs

City	OMI Class	#samples	mOMI	MOMI	mAV	MAV	mSD	MSD
Torino	B	78	1100	1450	5305.77	7335.90	2615.18	3780.67
	C	125	650	900	3066.00	4338.40	1923.43	2658.23
	D	125	350	500	1702.40	2469.20	953.98	1348.63
	E	16	400	500	1818.75	2396.88	1112.94	1417.36
Milano	B	105	750	1000	4839.05	6707.14	2508.83	3579.35
	C	103	650	900	2893.20	4112.62	1995.98	2779.32
	D	228	350	500	1629.17	2321.93	897.05	1227.73
	E	50	300	450	1240.60	1696.00	652.52	771.95
Genova	B	80	465	920	1851.44	2963.38	834.31	1415.45
	C	229	370	730	1255.70	1946.81	601.26	872.24
	D	485	330	560	1088.62	1663.61	629.17	946.33
	E	6	360	650	828.33	1291.67	310.51	407.94
Bologna	B	16	1800	2700	2771.88	4256.25	710.86	1602.07
	C	72	1850	2200	2620.14	3419.44	572.98	954.93

Journal Pre-proofs

	D	181	300	450	1952.49	2520.72	825.83	967.49
	E	25	400	600	1472.00	1846.00	658.78	691.75
Firenze	B	36	900	1800	2252.78	3568.06	809.76	1104.87
	C	90	600	1200	1750.00	2632.78	723.72	853.32
	D	108	600	1200	1461.57	2263.43	649.46	802.55
	E	7	700	1350	1392.86	2157.14	640.59	689.46
Roma	B	72	840	1650	4142.36	5796.53	1571.23	2071.44
	C	303	440	900	2275.58	3242.08	944.37	1330.39
	D	402	690	600	1457.71	2096.02	690.23	990.16
	E	461	0	450	1219.13	1771.09	544.83	786.86
Napoli	B	105	460	930	2003.50	3431.17	1067.81	1947.52
	C	103	370	750	1406.96	2389.28	800.54	1290.78
	D	228	300	610	1028.63	1784.21	479.11	810.87
	E	50	240	480	750.92	1293.42	321.91	504.14
Bari	B	84	420	580	1211.19	1705.83	469.04	719.78

Journal Pre-proofs

	C	59	400	600	1315.76	1827.80	529.84	721.41
	D	58	300	400	909.31	1239.31	354.33	467.80
	E	87	300	400	871.26	1189.89	385.08	512.96
Palermo	B	106	250	360	846.32	1199.25	354.23	598.73
	C	56	225	320	740.18	1048.21	274.74	377.01
	D	67	230	485	724.40	1010.15	274.75	371.66
	E	102	240	330	760.74	1070.39	360.92	406.70

This trend is particularly evident in Class B, where central areas across all cities, including Turin, Genoa, and Florence, exhibit the highest average minimum costs. Conversely, Classes D and E, representing peripheral and suburban areas, generally have lower costs across the board. Similarly, maximal property costs vary across OMI classes and different cities, with some peaks in cities such as Turin, Milan, and Rome for the OMI class. Southern cities such as Bari and Palermo reveal almost homogeneous maximal costs across all the OMI classes, emphasizing that the value of the property is constant in all the cities.

On the other hand, the analysis of minimal and maximal standard deviation among all the cities reveals a similar situation: larger cities such as Milan, Turin, and Rome often exhibit in class B higher minimum and maximum standard deviations, i.e., greater variability in property prices, indicating a market that can be influenced by a wider range of factors. In central areas, instead, Genoa and Bologna tend to show more stable property cost ranges with lower standard deviations, suggesting a relatively consistent and predictable real estate market.

Also, in semi-central areas of Florence, Bologna, Genoa there are more stable cost ranges. Genoa shows instead that the standard deviations in Classes B and C are comparatively lower, indicating a more uniform real estate market.

In general, big cities such as Milan, Turin, Rome, and Naples exhibit a broad range of OMI classes, indicating varied real estate offerings. Class B consistently holds the highest maximum OMI values, suggesting premium properties in central areas. However, these cities also experience high standard deviations in Class B, reflecting the potential for significant variations in property costs in urban centers.

A similar analysis was conducted at the local level, with a focus on key municipalities within the province of Naples. Specifically, this analysis encompassed seven municipalities: Bacoli, Caivano, Capri, Casandrino, Quarto, Sorrento, and Torre del Greco.

Table 7. Key Statistics for the OMI Dataset in Select Municipalities within the province of Naples

City	OMI Class	mAV	MAV	mSD	MSD
Naples	B	2003.50	3431.17	1947.52	508.45
	C	1406.96	2389.28	1290.78	466.32
	D	1028.63	1784.21	810.87	462.11
	E	750.92	1293.42	504.14	657.45
Bacoli	B	1206.67	1980.00	748.87	426.22
	C	1140.00	1901.43	830.85	421.14
	D	936.84	1540.00	645.09	358.77
Caivano	B	1223.64	1971.82	743.96	419.75

	C	1275.83	2095.00	818.02	357.82
	D	896.25	1622.50	906.55	312.31
Capri	B	4875.00	8135.00	3301.69	822.32
	D	4283.33	6950.00	3147.22	540.21
Casandrino	B	538.33	926.67	369.36	433.61
	D	520.00	952.50	262.35	452.15
Quarto	B	940.00	1543.33	529.29	454.19
	C	2283.33	3771.11	1907.32	478.06
	D	689.29	1150.36	688.81	356.53
Sorrento	B	2853.33	4695.83	2028.26	708.66
	C	987.31	1599.23	688.81	442.81
	D	808.42	1328.95	540.54	407.15
Torre del Greco	B	908.21	1583.57	608.68	416.09
	D	806.88	1328.95	540.54	407.15

Table 7 presents a set of statistics derived from the OMI dataset for these municipalities. The data reveals distinct patterns, especially in tourist-centric

Journal Pre-proofs

locations like Capri and Sorrento. In Capri, it is evident that property values consistently remain high across all OMI classes. This phenomenon can be attributed to Capri's status as a renowned tourist destination, where all OMI zones exhibit elevated real estate costs. Even the peripheral areas in Capri, for each of the four parameters considered (minimum and maximum averages, as well as minimum and maximum standard deviations), display values surpassing those of Naples, the provincial capital.

Sorrento also stands out due to its notably high property values in OMI classes B, C, and D. This indicates that the central, semi-central, and peripherical areas of Sorrento boast consistently high and closely situated property values.

In the central OMI zones of Bacoli and Caivano, properties demonstrate relatively higher values. However, Bacoli appears to offer the greatest potential for high-value properties, particularly in peripherical areas. In contrast, Casandrino tends to exhibit lower property values at the upper end of the spectrum.

5. 2 Experimental setting

A comprehensive set of approximately 400 tests was conducted across various OMI subzones within different Italian municipalities. These tests involved manipulating factors such as the selection of features, client requisites, and preferences. The design of the fuzzy partitions and the assignment to each fuzzy set of the deviation from the unit price required the skills of a real estate expert.

The expert, indeed, defines the fuzzy partitions in the domain of each feature selected, assigning to each fuzzy set a positive or negative deviation from the average price of the property. This expert-driven approach helped tailor the fuzzy partitions and deviations to accurately reflect the nuances of the real estate market, enhancing the effectiveness of the valuation process.

The genetic algorithm is employed to identify the most optimal features for a property. This is done by considering both the price per square meter proposed by the client and the client's personal preferences.

If the estimated price of the property with the found features falls within the minimum and maximum OMI values for the selected subarea and property type, then the property will probably be in the desired area and the client will be satisfied; otherwise, it will be necessary to change subarea or property type.

Some test cases are proposed to assess our approach described in Section 3, according to the workflow in Figure 4.

To ensure a comprehensive exploration of the search space, high crossover and mutation probabilities are employed; however, they can lead to prolonged search times. Extensive experimentation was conducted to determine the appropriate values for crossover and mutation rates. A balanced compromise was achieved by selecting a crossover rate of 0.7 and a mutation rate of 0.01. Table 8 shows the genetic algorithm parameters setting for our tests.

Table 8. Genetic algorithm parameter settings in the experiments

Parameter	**Description**	**Value**
N	Number of individuals	500
N_g	Number of generations	100
p_k	Selection rate	Calculated by equation (10)
p_{cr}	Crossover rate	0.7
P_{mu}	Mutation rate	0.01

For the sake of conciseness, below are the summarized results of three test cases conducted on two distinct OMI zones: one situated in the central area and the other in the peripheral region.

Test case a

Journal Pre-proofs

This test case aims to evaluate the unit price in euros per square meter of residential properties of the "Civilian Housing" type.

The unit price of this property type in zone B, subzone 8 is X = 2225 euros per square meter, i.e., equal to the average between the minimum and maximum values found.

Six extrinsic features were selected based on the minimum distance of a service facility from the building. They are listed in Table 9.

Table 9. Description of the six features

ID	Features	Description	Fetaure domain (meters)
c_1	Public services	Presence of schools, banks, hospitals, post offices, etc.	[0, 1200]
c_2	Public transport	Presence of public transport stops (bus, tram, metro, etc.)	[0, 1200]
c_3	Essential commercial services	Presence of shops oriented to the trade of necessities, such as food, clothing, pharmacy and similar	[0, 1200]
c_4	Public green	Presence of public green areas	[0, 1200]
c_5	Main route infrastructure	Presence of important connection routes (motorway junctions, state roads, etc.)	[0, 1200]
c_6	Parking equipment	Provision of parking areas	[0, 1200]

Journal Pre-proofs

These features are referred to by the agency as positional factors and influence the property's price assessment. For the agency, these features represent positional factors that affect its price evaluation based on their distance from the property.

The domain of the six features (and also the gene domain in our genetic algorithm) is the integer interval [0, 1200]. For each feature, three classes are established: *Near*, *Far*, *and Absent*. Specifically, *Near* represents a distance of 300 to 400 meters from the property, *Fa*r represents 300 to 400 meters to one kilometer from the property, and *Absent* represents a distance greater than a kilometer.

The expert intervenes, based on the guidance provided by the agency, by defining the fuzzy partitions related to the six features considered and for each one models the corresponding fuzzy sets. For each fuzzy set, the expert establishes the corresponding unit price difference Δx based on the type of property, the OMI class, and the subzone in which it falls.

To estimate fluctuations from the average price, the expert considers the price of residential property in the selected subzone and, based on all characteristics, determines a variation between 1415 and 3035 euros per square meter, with a fluctuation of ± 810 euros per square meter from the average unit value of 2225 euros per square meter.

The fuzzy sets are modeled in this test case by triangular, Right shoulder, and Left shoulder membership functions (MFs). They are described as follows.

Triangular MF
$$\mu_A(x) = \begin{cases} 0 & x < a \\ \dfrac{(x-a)}{b-a} & a \leq x \leq b \\ \dfrac{b-x}{c-b} & b \leq x \leq c \\ 0 & x > c \end{cases} \tag{15}$$

Journal Pre-proofs

L-shoulder MF $\mu_A(x) = \begin{cases} 1 & x < b \\ \dfrac{(b-x)}{c-b} & b \le x \le c \\ 0 & x > c \end{cases}$ (16)

R-shoulder ML $\mu_A(x) = \begin{cases} 0 & x < a \\ \dfrac{(x-b)}{b-a} & a \le x \le b \\ 1 & x > b \end{cases}$ (17)

where a < b < c.

Table 10 lists the fuzzy sets of the fuzzy partitions of the six features.

Table 10. Description of the fuzzy sets of the six features

Features	**Fuzzy set**	**Type**	**a**	**b**	**c**	**Δx**
c_1	Near	R function		100	300	150
	Far	Triangular	100	300	1000	0
	Absent	L function	300	1000		-150
c_2	Near	R function		100	300	200

Journal Pre-proofs

	Far	Triangular	100	300	1000	-50
	Absent	L function	300	1000		-200
c_3	Near	R function		100	300	200
	Far	Triangular	100	300	1000	-50
	Absent	L function	300	1000		-200
c_4	Near	R function		100	400	50
	Far	Triangular	100	400	1000	0
	Absent	L function	400	1000		-50
c_5	Near	R function		100	400	50
	Far	Triangular	100	400	1000	0
	Absent	L function	400	1000		-50
c_6	Near	R function		100	400	150
	Far	Triangular	100	400	1000	0
	Absent	L function	400	1000		-150

The fuzzy partitions of the six features are shown below (Figures 6, 7 and 8).

Journal Pre-proofs

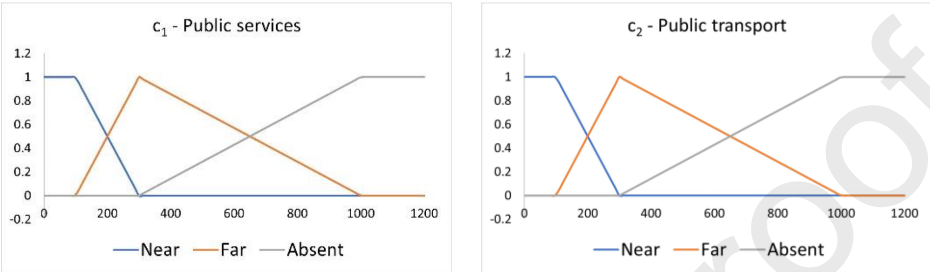


Figure 6. Fuzzy partitions of the features c_1 and c_2.

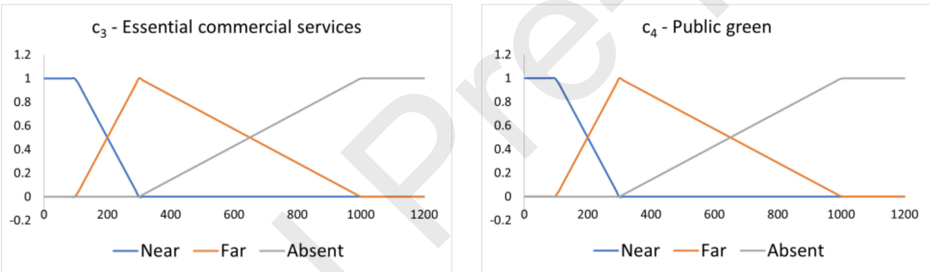
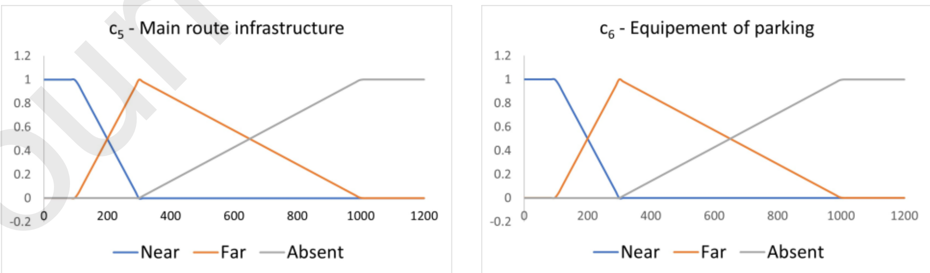


Figure 7. Fuzzy partitions of the features c_3 and c_4.



Journal Pre-proofs

Figure 8. Fuzzy partitions of the features c_5 and c_6.

The expert considered assigning a lower price change (+50 euros per square meter if Near, 0 if Far, and -50 euros per square meter if Absent) to features c4 (Public Green) and c5 (Main Roadway Infrastructure) because the OMI area in which the property is located in a downtown area and the presence of essential infrastructure or services is more significant than the presence of green areas or highway interchanges. Let us suppose the client wants to pay 2100 euros per square meter and prefers (i.e., expresses his preference for) features c1 (public services) and c2 (public transport). The maximum number of generations Ng is set to 100. Figure 9 shows the trend of the best fitness value in each generation. This trend has already reached a plateau after about 50 generations.

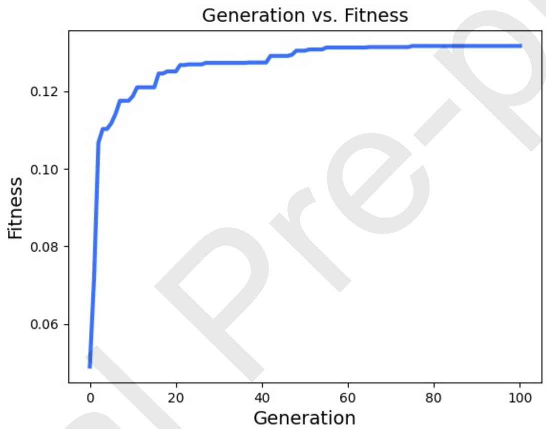


Figure 9. Test case a - Fitness function trend

The final best solution is shown in Table 11. For each selected feature, the distance value, the assigned fuzzy set and corresponding membership value, the deviation from the unit price and the calculated fuzziness value for each gene in the current solution are given.

Table 11. Test case a - The final best solution

Feature	Distance	Fuzzy set	Membership degree	Δx	Fuzziness
c_1	368 m	Far	0.90	0	0.31

c_2	84 m	Near	1.00	200	0.00
c_3	283 m	Far	0.92	-50	0.28
c_4	938 m	Absent	0.90	-50	0.32
c_5	1210 m	Absent	1.00	-50	0.00
c_6	896 m	Absent	0.83	-150	0.44

The fuzziness of the best solution (see Eq. 12) is 0.22, while the reliability is 0.78. The final unit price is 2125 euros per square meter, calculated by adding each Δx_i to the initial unit price X, which is approximately close to the client's 2100 euros per square meter proposal. It falls in the price range of 1750 - 2700 euros, assessed by the OMI as a "civilian housing" type of property in the subzone examined. Recall that the customer assigned preference to features c_1 and c_2; as can be seen in Table 9, the first preference is not fully satisfied by the solution, while the second is: in fact, distance to the nearest public services got the Far label, while transportation infrastructure got Near. This solution is also a good compromise for the expert who considers solutions that fully satisfy the two preferences too expensive compared to the user's demands; in fact, satisfying them would result in an average premium on the square meter value of the desired property of about 350 euros.

Test case b

The second test considers the same property type and in the same OMI area. In this case, the expert must also consider the intrinsic characteristics for assessing the property's price and the above-mentioned extrinsic characteristics.

The intrinsic characteristics that impact the property's price assessment most are the year of the building's last renovation, the building's wiring, and the property's exposure to sunlight. Considering all these characteristics, the expert estimated that the square meter value of residential property in the selected OMI area ranges between 1050 and 3325 euros per square meter, with a fluctuation of ± 1100 euros per square meter compared to the average unit value of 2225 euros per square meter.

A fuzzy partition was built for the extrinsic feature, c_7 – Year of renovation, as shown in Table 12. The domain of the feature c_7 (and the gene domain in our genetic algorithm) is the integer interval [1950, 2020]. The three relative fuzzy sets are shown in Figure 10 as well.

Table 12. Description of the fuzzy sets of the characteristic c_7 – Year of renovation

Feature	Fuzzy set	Type	a	b	c	Δx
c_7	Long ago	R function		1960	1990	100
	Some time	Triangular	1960	1990	2010	0
	Recent	L function	1990	2010		-100

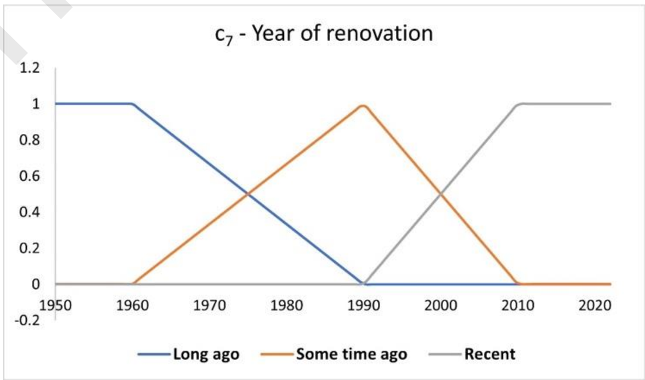


Figure 10. Fuzzy partitions of the feature c_7

For the other two new features, namely c8 - Cabling and c9 - Brightness, two crisp partitions are defined; the classes defined by the agency are described in Table 13 along with the gene value of the feature associated with the relative class (“Gene value” column).

Table 13. Description of the crisp sets of the features c_8 and c_9.

Feature	Class	Gene value	Description	Δx
c_8	Scanty	0	partial cabling and / or non-compliant with international standards.	-50
	Normal	1	structured cabling made in compliance with current international standards	50
c_9	Scanty	0	it is necessary to use artificial lighting for most of the daytime hours.	-150
	Medium	1	artificial lighting must be used for part of the day	0
	Good	2	there is no need to resort to artificial lighting, in the main rooms, during the day	150

The client's proposal is still 2100 euros per square meter. His preferences are for features c_1, c_2, c_7 and c_9. The genetic search for the optimal solution is performed by setting the maximum number of generations Ng to 100. Figure 11 shows the trend of the best fitness value for each generation. The trend has already reached a plateau after about 60 generations.

Journal Pre-proofs

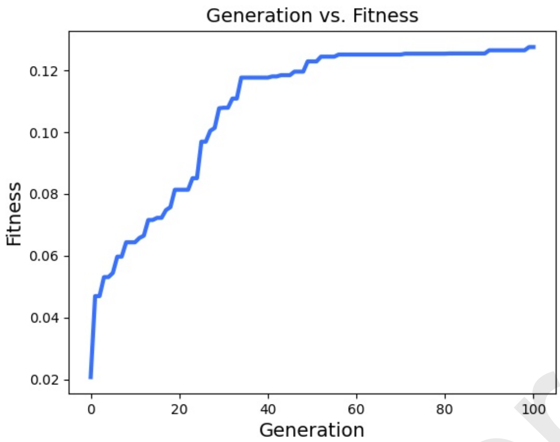


Figure 11. Test case b - Fitness function trend

The best solution is shown in Table 14. Each feature's value, the assigned fuzzy set and its membership value, the unit price difference, and the calculated fuzziness value for the corresponding gene in the solution are shown.

Table 14. Test case b - The final best solution

Feature	Value	Fuzzy set	Membership degree	Δx	Fuzziness
c_1	276 m	Far	0.88	0	0.35
c_2	284 m	Far	0.92	-50	0.27
c_3	319 m	Far	0.97	-50	0.12
c_4	950 m	Absent	0.92	-50	0.28
c_5	1188 m	Absent	1.00	-50	0.00

Journal Pre-proofs

c_6	902 m	Absent	0.84	-150	0.43
c_7	2008	Recent	0.90	100	0.31
c_8	Normal	-	-	50	0.00
c_9	Good	-	-	150	0.00

Considering the first seven characteristics, the fuzziness value of the best solution (Equation 12) is 0.25, while the reliability is 0.75.

As in the previous test, the final unit price, calculated by adding each Δx_i to the initial unit price X, is 2175 euros per square meter, quite close to the client's proposal of 2100 euros per square meter. The optimal solution is again in the price range of 1750 - 2700 euros, assessed by the OMI as a "civilian housing" type of property in the subzone examined.

The customer selected a preference for features c_1, c_2, c_7, and c_9; the preferences for features c_1, and c_2, are not completely satisfied by the solution; both features are related to the proximity of public services and public transport, respectively, and are classified as *Far*. On the contrary, the preferences for features c_7 and c_9, are fully satisfied, as c_7 (Year of renovation) is classified as *Recent* and c_9 (Brightness) assumes the crisp value *Good*. Then, the best solution represents a good compromise between the customer's demand and preferences. If the customer's preferences for features c_1 and c_2 had been fully met, the property price would have risen to a total value of 2575 euros, a price significantly higher than the customer's request, although within the price range of 1750 to 2700 euros.

Test case c

In this test case, the client wants to buy a residential property in a suburban area of the city, precisely in the OMI E 36 suburban area in the eastern part of the

Journal Pre-proofs

city. In this subarea, the price of "Civil dwelling" type real estate ranges between 720 and 1100 euros per square meter, with an average of X = 910 euros per square meter.

Based on the nine property features introduced, the expert estimated that the square meter value of residential property in OMI subarea E 36 is between 610 and 1210 euros per square meter, with a fluctuation of ± 300 euros per square meter from the average unit value of 910 euros per square meter.

Tables 15 and 16 show the parameters defined by the expert for both the fuzzy sets of fuzzy partitions related to features c_1, c_2, c_3, c_4, c_5, c_6, and c_7 for crisp sets created related to features c_8 and c_9, respectively.

Table 15. Description of the fuzzy sets of the first seven characteristic

Feature	Fuzzy set	Type	a	b	c	Δx
c_1	Near	R function		100	300	30
	Far	Triangular	100	300	1000	0
	Absent	L function	300	1000		-30
c_2	Near	R function		100	300	50
	Far	Triangular	100	300	1000	0
	Absent	L function	300	1000		-50
c_3	Near	R function		100	300	20
	Far	Triangular	100	300	1000	0
	Absent	L function	300	1000		-20

Journal Pre-proofs

c_4	Near	R function		100	400	30
	Far	Triangular	100	400	1000	0
	Absent	L function	400	1000		-30
c_5	Near	R function		100	400	50
	Far	Triangular	100	400	1000	0
	Absent	L function	400	1000		-50
c_6	Near	R function		100	400	30
	Far	Triangular	100	400	1000	0
	Absent	L function	400	1000		-30
c_7	Long ago	R function		1960	1990	30
	Some time	Triangular	1960	1990	2010	0
	Recent	L function	1990	2010		-30

Table 16. Description of the crisp sets of the characteristics c_8 and c_9.

Characteristic	**Class**	**Gene value**	**Δx**

Journal Pre-proofs

c_8	Scanty	0	-20
	Normal	1	20
c_9	Scanty	0	-50
	Medium	1	0
	Good	2	50

The client's proposal is 800 euros per square meter. The customer expressed preferences for features c_1, c_5 and c_9, as he would prefer a property not far from major transport infrastructure and road connections and mainly bright.

Figure 12 shows the fitness trend: the best fitness values measured in each generation are depicted, reaching a plateau already after about 65 generations.

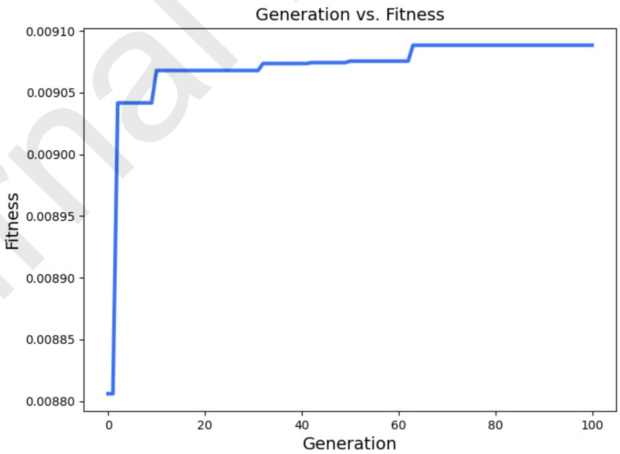


Figure 12. Test case c - Fitness function trend

Journal Pre-proofs

Table 17 shows the final best solution. For each feature, its value, the assigned fuzzy sets with their degrees of membership, the unit price difference, and finally, the fuzziness of the corresponding gene in the solution are given.

Table 17. Test case c - The final best solution

Characteristic	Value	Fuzzy set	Membership degree	Δx	Fuzziness
c_1	286 m	Far	0.93	0	0.24
c_2	112 m	Far	0.94	50	0.22
c_3	875 m	Absent	0.82	-20	0.45
c_4	1058 m	Absent	1.00	-20	0.00
c_5	391 m	Absent	0.97	0	0.13
c_6	887 m	Absent	0.81	-30	0.47
c_7	1962	Long ago	0.93	-30	0.24
c_8	Scanty	-	-	-20	-
c_9	Medium	-	-	0	-

The fuzziness of the best solution (Equation 12), considering the first seven features, is 0.25, while the reliability is 0.75.

Journal Pre-proofs

Like in the previous test, the final unit price, after adding each Δx_i to the initial unit price X, is 840 euros per square meter, a price close to the client's proposal of 800 euros per square meter. The best solution is within the price range of 720 to 1100 euros assessed according to the OMI as a "Civic Housing" type property in the selected suburban sub-zone.

The customer selected c_1, c_5 and c_9 as preferred features: c_1 and c_9 are not fully satisfied by the solution, while feature c_5 is satisfied. As in test case b, the solution represents a compromise between the customer's request and his preferences; if the customer's preferences for features c_1 and c_9 had been fully satisfied, the price of the property would have risen to a total value of 920 euros, much higher than the customer's proposed offer.

Test case results show that our approach provides valuable support in finding the property that best meets the client's requirements. The user-friendly framework lets users model property features in a fuzzy-based mode close to expert knowledge representation.

The expert can adjust the model's parameters by increasing the number and type of features assigned to the property or by constructing finer fuzzy partitions of the feature domains.

As noted, it is also possible to consider the modeled features as a crisp set in which a set of labels is used to classify the features. By estimating the reliability value of the best real estate price, calculated by measuring the fuzziness of the solution, it is possible to assess how reliable the evaluation of this price is. In all proposed tests, the reliability of the best solution is always higher than 0.7.

Comparison test

To assess the accuracy of the obtained results, a comprehensive evaluation was conducted. This evaluation involved comparing the unit price derived from the optimal solutions generated by the algorithm with the market values provided by a real estate company. These market values were associated with properties for sale within the same OMI area, possessing the same type and characteristics as the property sought.

Journal Pre-proofs

Journal Pre-proofs

For each of the approximately 400 test cases, properties with matching characteristics to the property under consideration were selected from the Tecnocasa real estate company's website (www.tecnocasa.it).

Subsequently, the average price per square meter of these identified properties was juxtaposed with the price generated by the solution. This average price per square meter was computed based on the collective prices of all properties currently available for sale through the real estate company. These properties shared the same type as the client's desired property and were situated within the identical OMI area.

All test cases refer to properties of different types, spanning various OMI areas across Italian municipalities. The tests were conducted utilizing the previously defined features c1 through c9.

To evaluate the performance of the proposed model, a comparison was made with three learning-based property price forecasting models as described in [40]. These models utilize Machine Learning techniques such as SVM, RF, and GBM to estimate the unit cost of property test cases. For the comparative tests, a training dataset was constructed, comprising more than 15,000 data points that encompass unit sales prices of properties in Italian cities spanning from 2013 to 2023. The training dataset includes 11 input features, encompassing the 9 property characteristics, the sale date, and the OMI area in which the property is situated. The creation of this training dataset involved importing data from the Tecnocasa databases.

Figure 13 displays the trend of the optimal solution obtained by running our framework, represented by the solid black line. This trend varies based on the average OMI price of the requested property type within the selected OMI area. The red dashed line illustrates the average sales price per unit of properties within the same OMI area while considering the same solution features. The solid blue, yellow, and green lines depict instead the projected trends of the solution obtained by running SVM, RF, and GBM, respectively.

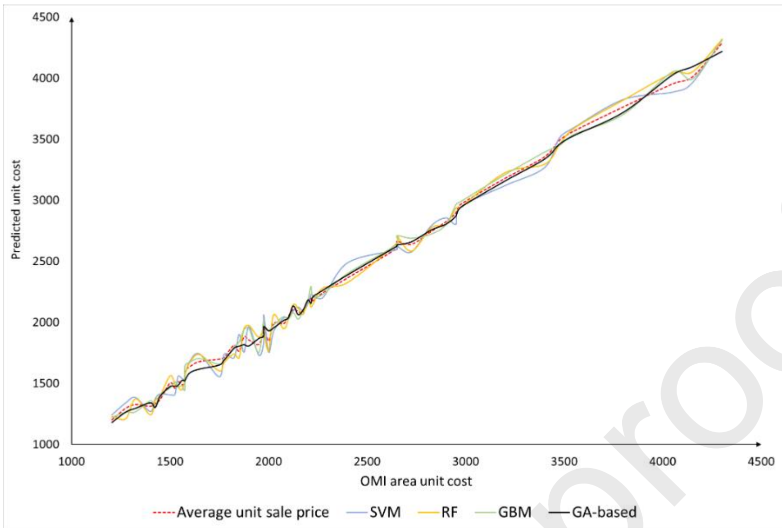
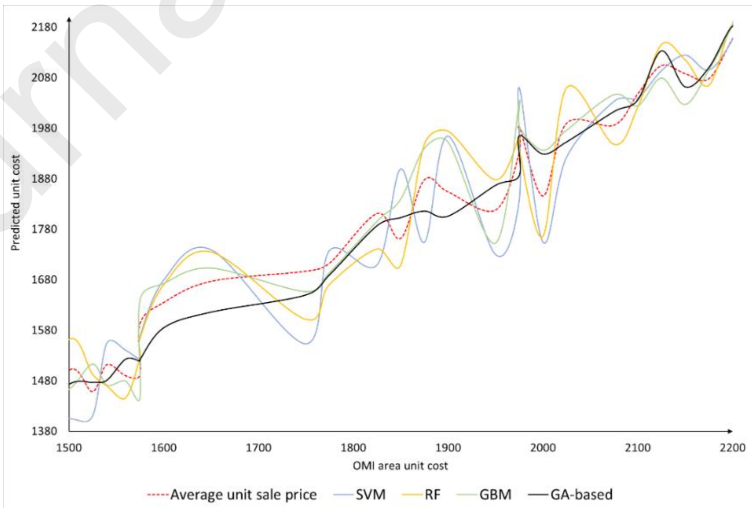


Figure 13. Trend of the predicted unit sale prices obtained by executing the four methods.

To emphasize the variations in unit sale price trends generated by the four methods, a closer look is provided in Figure 14. Specifically, the figure zooms in on the unit cost range from €1,500 to €2,200 per square meter. It becomes evident that the proposed method exhibits less pronounced fluctuations from real unit costs, in comparison to the SVM, RF, and GBM forecasting methods.



Journal Pre-proofs

Figure 14. Trend of the predicted unit costs in the range from 1,500 to 2,200 Euros.

To assess and compare the performance of our approach with SVM, RF, and GBM methods, we have considered several regression metrics. These metrics, which include Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE), and Root Mean Square Error (RMSE), are used to gauge the accuracy and quality of predictions generated by the models under analysis [46]. These indices are defined as follows:

$$MAE = \frac{1}{N}\sum_{\mathrm{j}=1}^{\mathrm{N}}\left|y_j - \hat{\mathrm{y}}_{\mathrm{j}}\right| \quad (18)$$

$$MAPE = \frac{1}{N}\sum_{\mathrm{j}=1}^{\mathrm{N}}\left|\frac{y_j - \hat{\mathrm{y}}_{\mathrm{j}}}{y_j}\right| \quad (19)$$

$$RMSE = \sqrt{\frac{\sum_{\mathrm{j}=1}^{\mathrm{N}}\left(y_j - \hat{\mathrm{y}}_{\mathrm{j}}\right)^2}{\mathrm{N}}} \quad (20)$$

where N represents the number of test cases, y denotes the best solution discovered by the proposed GA-based method for the j*th* test case and $\hat{\mathrm{y}}_{\mathrm{j}}$ is the corresponding average unit sale price.

Table 18 shows the values of the regression indices measured for the four methods.

Table 18. Performance indices of the four methods

Index	**SVM**	**RF**	**GBM**	**Our method (GA-based)**

Journal Pre-proofs

MAE	55.80 €/m^2	47.35 €/m^2	41.99 €/m^2	**35.10 €/m^2**
MAPE	2.85	2.42	2.10	**1.74**
RMSE	65.41 €/m^2	51.94 €/m^2	48.76 €/m^2	**39.27 €/m^2**

The best performance is achieved through the implementation of the proposed GA method, which yields the lowest values for MAE, MAPE, and RMSE. Specifically, the RMSE obtained from the GA-based method is 39.27 €/m^2. This highlights that, on average, the deviation between the solution price per unit and the actual sale price per unit for properties of the same type, located within the same OMI area, and possessing similar characteristics, is less than 40 €/m^2. In contrast, the values generated by the other three methods exceed this threshold. This outcome underscores that the solutions recommended by the proposed method are, on average, acceptable when compared to market prices and more accurate than those derived from SVM, RF, and GBM forecasting methods.

To comprehensively assess performance disparities among the selected methods, the Friedman and Nemenyi post-hoc statistical tests have been conducted. These analyses were based on RMSE measurements obtained from four distinct subsets, each corresponding to test cases associated with properties located in the OMI classes B, C, D, and E. The RMSE index values for each subset are presented in Table 19.

Table 19. RMSE of the four methods obtained for each subset (€/m^2)

OMI class	SVM	RF	GBM	Our method (GA-based)
B	63.11	52.63	48.52	**40.07**
C	64.47	50.25	50.36	**39.23**

D	65.43	52.24	47.85	**39.30**
E	68.10	52.37	48.31	**39.56**

The Friedman statistics is described as follows.

$$F_F = \frac{(N-1)\chi_F^2}{N(k-1)-\chi_F^2} \quad (21)$$

where N is the number of subsets, k is the number of methods and χ_F^2 is calculated by the following formula:

$$\chi_F^2 = \frac{12N}{k(k+1)}\left[\sum_{i=1}^{k} R_i^2 - \frac{k(k+1)^2}{4}\right] \quad (22)$$

where R_i denotate the average rank of the i*th* method.

The ranks assigned to each method for each subset are shown in Table 20, while the last row displays the average rank.

Table 20. Ranks assigned to the four methods

OMI class	**SVM**	**RF**	**GBM**	**Our method (GA-based)**

Journal Pre-proofs

B	4	3	2	1
C	4	2	3	1
D	4	3	2	1
E	4	3	2	1
R_i	4	2.75	2.25	1

After computation, the value F_F is 37. Using a significance level of $\alpha = 0.05$ and (k-1)(N-1) = 9 degrees of freedom, the critical value for F is established at 3.8625. Consequently, the null hypothesis suggesting that the four methods exhibit equivalent performance, can be rejected due to the computed F_F surpassing the critical value.

The Nemenyi post-hoc test was conducted to investigate significant differences among the four methods. The null hypothesis, suggesting equal performance between pairs methods, is rejected when the difference between their average ranks exceeds the threshold calculated as:

$$CD = q_\alpha \sqrt{\frac{\mathrm{k}(\mathrm{k}+1)}{6\mathrm{N}}} \quad (23)$$

At the significance level of $\alpha = 0.05$, for k = 4 the critical value q_α is 2.569.

Journal Pre-proofs

Consequently, our method exhibits significantly better performance than SVM (mean rank difference = 3) and slightly better than RF (mean rank difference = 1.75) and GVM (mean rank difference = 1.25).

6. Conclusion

The paper proposes a novel genetic algorithm for the real estate market estimation. It finds a trade-off between the client's request for a property purchase and the current availability of the real estate market. A fuzzy-based model describes the extrinsic features encoded as the genes of the algorithm. The model considers also the price fluctuations as the features change; more specifically, the real estate expert associates a fuzzy partition to each feature by defining some fuzzy sets that semantically describe each feature; he also assigns each fuzzy set a deviation price, i.e., a fluctuation value from the base price. The genetic algorithm's fitness function measures the gap between the requested price and the assessed price for the solution, returning the property that best meets the client's desires. Furthermore, the solution is assessed by a reliability measure that evaluates the fuzziness of the solution by leveraging the fuzzy entropy function. The reliability measure is robust to the fluctuations in property feature values.

Some test cases were carried out in the urban area of Naples (Italy), considering the OMI zones set by the Italian agency, namely "Agenzia delle Entrate" and the average price of the unit type in the selected subzone, updated semi-annually by the agency. The results show that the optimal solution represents a good compromise between customer demands and the market offerings in all the tests. Furthermore, the comparative experiments achieved with some ML methods aimed at evaluating the market prices of properties show evidence that our method shows less pronounced fluctuations from actual unit costs, and the price fluctuations concerning market prices are the minimum and less than € 40 per square meter. Our approach shows better results compared with the other ML methods used. Thanks to the fuzzy partition modelling, the framework can support the real estate expert in defining property features in a way that is very close to the natural language description and allows an intuitive understanding from the client by removing in this way the numerical values that are not always simple to interpret in their context (for instance, "far from the school" is more straightforward than "5 km from school" that could be not easy to quantize).

Journal Pre-proofs

Moreover, from a real estate expert's viewpoint, the framework allows him to include in the evaluation other attributes and mainly easily, he can refine the fuzzy partitions based on the client's demands and context knowledge (for instance, a rural area requires different features than an urban area).

This aspect outlines that the proposed approach can be employed for searching for different property types (e.g., residential apartments, villas, shops, garages) in various urban settlements (e.g., metropolitan areas, historic centers, villages, mountain towns, coastal tourist destinations). This flexibility is a strong point of our model as it does not require extensive learning sets to assess property unit costs.

Our future work will focus on improving valuation accuracy by integrating our model with advanced decision-support systems. The idea is to leverage the expertise of real estate professionals to refine the technical aspects of property evaluation through a decision-making process that is informed and reflective of the complex real estate market.

The integration of the expert-driven modeling of property features by fuzzy partition coupled with the monitoring of real estate market changes, including economic conditions, local regulations, and societal trends can provide a comprehensive overview of evolving market dynamics for decision support processes. By fusing real estate expertise with real-time data reflecting changes in the real estate market, our goal is to design a robust tool that can provide insights into property values and emerging market trends. In this way, our approach could be a reliable resource to help decision-makers drive the complexities of the ever-changing real estate landscape.

CrediT authorship contribution statement

Barbara Cardone: Conceptualization, Methodology, Validation, Writing – review & editing.

Ferdinando Di Martino: Conceptualization, Methodology, Validation, Writing – review & editing.

Sabrina Senatore: Conceptualization, Methodology, Validation, Writing – review & editing.

Journal Pre-proofs

Compliance with Ethical Standards

This research does not receive any source of funding and does not involve human participants and animals, moreover the authors have no potential conflicts of interest (financial or non-financial).

Acknowledgments

All authors approved the version of the manuscript to be published.

References

[1] Alm, J., Dronyk-Trosper, T., Larkin, S. (2023). Do opportunity zones create opportunities? The impact of opportunity zones on real estate prices. *Real* Estate Economics, doi: 10.1111/1540-6229.12434.

[2] Antipov, E.A., Pokryshevskaya, E.B. (2012). Mass appraisal of residential apartments: An application of random forest for valuation and a CART-based approach for model diagnostics, Expert Systems with Applications, 39,1772-1778, doi: 10.1016/j.eswa.2011.08.077.

[3] Baldominos A., Blanco I., Moreno A.J., Iturrarte R., Bernárdez Ó., Afonso C. (2018). Identifying Real Estate Opportunities Using Machine Learning, Applied Sciences, 8(11), 2321, doi: 10.3390/app8112321.

[4] Bas, M. (2022). The impact of the COVID-19 pandemic on the residential real estate market on the example of Szczecin, Poland, Procedia Computer Science, Volume 207, Pages 2048-2058, ISSN 1877-0509, doi:10.1016/j.procs.2022.09.264.

[5] Bravi M., Giaccaria S. (2005). Forecasting Consumer Choices: The Conjoint Analysis (CA) in Real Estate Market Research, ERES eres2005_125, European Real Estate Society conference in association with the International Real Estate Society, Dublin, Ireland. doi: 10.15396/eres2005_125.

Journal Pre-proofs

[6] Bravi M., Giaccaria S. (2009). The Conjoint Analysis (CA) in real estate, Aestimum, 48, 39-59, doi: 10.13128/Aestimum-6237.

[7] Calainho, F.D., van de Minne, A. & Francke, M.K. A Machine Learning Approach to Price Indices: Applications in Commercial Real Estate. J Real Estate Finan Econ (2022). https://doi.org/10.1007/s11146-022-09893-1.

[8] Cardone B., Di Martino F (2020). A Novel Fuzzy Entropy-Based Method to Improve the Performance of the Fuzzy C-Means Algorithm. Electronics, (9)554, doi: 10.3390/electronics9040554.

[9] Cardone B., Di Martino F., Senatore S. (2022). A fuzzy partition-based method to classify social messages assessing their emotional relevance, Information Sciences, 594, 60-75, doi: 10.1016/j.ins.2022.02.028.

[10] Cattin P., Wittink D. R. (1982). Commercial Use of Conjoint Analysis: An Update, Journal of Marketing, 46(3), 44-53, doi: 10.2307/1251701.

[11] Ćetković J., Lakić S., Lazarevska M., Žarković M., Vujošević S., Cvijović J., Gogić M. (2018). Assessment of the Real Estate Market Value in the European Market by Artificial Neural Networks Application. Complexity, vol. 2018, Article ID 1472957, 10 pages, doi:10.1155/2018/1472957.

[12] Cheshmehzangi, A. (2021). Housing and health evaluation related to general comfort and indoor thermal comfort satisfaction during the COVID-19 lockdown, Journal of Human Behavior in the Social Environment, 31:1-4, 184-209, doi: 10.1080/10911359.2020.1817225.

[13] Čirjevskis A. (2021). Value Maximizing Decisions in the Real Estate Market: Real Options Valuation Approach. J. Risk Financial Manag. 14, 278, doi:10.3390/jrfm14060278.

[14] De Luca A., Termini S. (1979). Entropy and Energy Measures of Fuzzy sets. In Advances in Fuzzy Set Theory and Applications, Gupta, M.M., Ragade, R.K., Yager, R.R., Eds.; North-Holland: Amsterdam, The Netherlands, 321-338.

[15] Di Martino F., Sessa S. (2020). A New Validity Index Based on Fuzzy Energy and Fuzzy Entropy Measures in Fuzzy Clustering Problems. Entropy 2020, 22, 1200. https://doi.org/10.3390/e22111200

[16] Fischer D., Lai P.P. (2006). Artifcial neural networks and computer assisted mass appraisal, in Proceedings of the 12th Annual Conference of the Pacifc Rim Real Estate Society (PRRES '06), Auckland, New Zealand.

Journal Pre-proofs

Journal Pre-proofs

[17] Geerts M., vanden Broucke S., De Weerdt J. (2023). A Survey of Methods and Input Data Types for House Price Prediction, ISPRS International Journal of Geo-Information, 12 (5), art. no. 200, doi: 10.3390/ijgi12050200.

[18] Goldberg D. (1989). Genetic Algorithms in Search, Optimization and Machine Learning. Reading. Addison-Wesley Professional,MA (USA), 432 pp. ISBN: 978-0201157673.

[19] Glaeser, E. L., Gyourko, J. (2005). Urban decline and durable housing. Journal of Political Economy, 113(2), 345-375.

[20] Goldberg D. (2013). The Design of Innovation: Lessons from and for Competent Genetic Algorithms, Springer Nature, NY inc. (USA), 276 pp. ISBN: 978-1475736458.

[21] Guan, J., Shi, D., Zurada, J.M., Levitan, A.S. (2014). Analyzing Massive Data Sets: An Adaptive Fuzzy Neural Approach for Prediction, with a Real Estate Illustration. Journal of Organizational Computing and Electronic Commerce, 24, 94-112, doi: 10.1080/10919392.2014.866505.

[22] I-Cheng Y., Tzu-Kuang H. (2018). Building real estate valuation models with comparative approach through case-based reasoning, Applied Soft Computing, Volume 65, Pages 260-271, ISSN 1568-4946, doi:10.1016/j.asoc.2018.01.029.

[23] Jiang L., Peter C. B. P., Jun Y. (2014). A New Hedonic Regression for Real Estate Prices Applied to the Singapore Residential Market, Social Science Research Network Electronic Paper Collection, Cowles Foundation Discussion Paper No. 1969, 23pp, doi:10.2139/ssrn.2533017.

[24] Jie Z., Qi C.H. (2023). An equilibrium analysis of the impact of real estate price volatility on macroeconomics based on ant colony algorithm. J Comb Optim 45, 70. doi: 10.1007/s10878-023-00995-x.

[25] Kim Y., Choi S., Yi M.Y., (2020). Applying Comparable Sales Method to the Automated Estimation of Real Estate Prices. Sustainability, 12(14), 5679, 19 pp, doi:10.3390/su12145679.

[26] Kucklick J.-P. (2023). HIEF: a holistic interpretability and explainability framework, Journal of Decision Systems, doi: 10.1080/12460125.2023.2207268.

[27] Kusan H., Aytekin O., Özdemir I., (2010). The use of fuzzy logic in predicting house selling price. Expert Systems with Applications 37(3), 1808-1813, doi: 10.1016/j.eswa.2009.07.031.

[28] Louati, A, Lahyani, R, Aldaej, A, Aldumaykhi, A, Otai, S. (2022). Price forecasting for real estate using machine learning: A case study on Riyadh city. Concurrency Computat Pract Exper. 34(6): e6748.

[29] Lord, A., Cheang, C.-W., Dunning, R. (2022). Understanding the geography of affordable housing provided through land value capture: Evidence from England. Urban Studies, 59(6), 1219–1237, doi:10.1177/0042098021998893.

[30] Edwin Lughofer, Bogdan Trawiński, Krzysztof Trawiński, Olgierd Kempa, Tadeusz Lasota, On employing fuzzy modeling algorithms for the valuation of residential premises, Information Sciences, Volume 181, Issue 23, 2011, Pages 5123-5142, ISSN 0020-0255, https://doi.org/10.1016/j.ins.2011.07.012.

[31] Malinowski, A., Piwowarczyk, M., Telec, Z., Trawiński, B., Kempa, O., Lasota, T., (2018). An Approach to Property Valuation Based on Market Segmentation with Crisp and Fuzzy Clustering. In: Nguyen, N., Pimenidis, E., Khan, Z., Trawiński, B. (eds) Computational Collective Intelligence. International Conference on Computational Collective Intelligence 2018. Lecture Notes in Computer Science, Springer, Cham, 11055, 534-548, doi:10.1007/978-3-319-98443-8_49.

[32] Mattos M., Simoes P., Zancam E., Ferreira N., Cechinel C., (2008). A neurofuzzy approach for property value prediction, in Proceedings of the 35th Conferencia Latinoamericana de Informatica (CLEI '08), 10 pp.

[33] Musa, A.G., Daramola, O., Owoloko, A., Olugbara, O. (2013). A neural-CBR system for real property valuation, Journal of Emerging Trends in Computing and Information Sciences, 4(8), 611-622.

[34] Muneer, S. M., Alvi, M. B., & Rasool, M. A. (2022). Genetic algorithm based intelligent system for estate value estimation. International Journal of Computational and Innovative Sciences, 1(1), 28-38.

[35] Obeng-Odoom F. (2022). Urban housing analysis and theories of value, Cities, Volume 126, 103714, ISSN 0264-2751, doi:10.1016/j.cities.2022.103714.

[36] Peterson, S., Flangan, A.B., (2009). Neural network hedonic pricing models in mass real estate appraisal, Journal of Real Estate Research, 31(2), 147-164, doi: 10.1080/10835547.2009.12091245.

[37] Rafiei, M. H., & Adeli, H. (2016). A novel machine learning model for estimation of sale prices of real estate units. Journal of Construction Engineering and Management, 142(2), 04015066.

[38] Renigier-Biłozor, M., Chmielewska, A., Walacik, M. et al., Genetic algorithm application for real estate market analysis in the uncertainty conditions. J Hous and the Built Environ 36, 1629–1670 (2021). doi: 10.1007/s10901-020-09815-8

[39] Ruspini E.H., (1969). A new approach to clustering, Information and Control 15(1), 22-32, doi: 10.1016/S0019-9958(69)90591-9.

[40] Sharma A. and T. Poongodi T., (2023). Prediction of Real-Time Estate Pricing using Train-Test Splitting Techniques, 2023 4th International Conference on Intelligent Engineering and Management (ICIEM), London, United Kingdom, pp. 1-6, doi: 10.1109/ICIEM59379.2023.10167318.

[41] Sun, Y. (2019). Real estate evaluation model based on genetic algorithm optimized neural network. Data Science Journal, 18, 36, pages 1-9.

[42] Tajani F., Liddo F.D., Guarini M.R., Ranieri, R., Anelli D. (2021). An Assessment Methodology for the Evaluation of the Impacts of the COVID-19 Pandemic on the Italian Housing Market Demand. Buildings 2021, 11, 592, doi:10.3390/buildings11120592.

[43] Truong Q., Nguyen M., Dang H., Mei B. (2020). Housing Price Prediction via Improved Machine Learning Techniques, Procedia Computer Science, Volume 174, Pages 433-442, ISSN 1877-0509.

[44] Tsai, I-C. (2019). European house price deviation: infectivity and the momentum effect, Economic Research-Ekonomska Istraživanja, 32:1, 1521-1541, doi:10.1080/1331677X.2019.1636698.

[45] Winky K.O. Ho, Bo-Sin Tang & Siu Wai Wong (2021). Predicting property prices with machine learning algorithms, Journal of Property Research, 38:1, 48-70, DOI: 10.1080/09599916.2020.1832558.

[46] Xingyu M. (2023). Forecasting Secondhand Tanker Price Through Wavelet Neural Networks Based on Adaptive Genetic Algorithm. Information Technology and Control, 52(2), 336-357. doi:10.5755/j01.itc.52.2.32804.

[47] Zhan C., Liu Y., Wu Z., Zhao M., Chow T.W.S., A hybrid machine learning framework for forecasting house price, (2023) Expert Systems with Applications, 233, art. no. 120981, DOI: 10.1016/j.eswa.2023.120981

[48] Zheng, Y., Xu, Z. & Xiao, A. (2023). Deep learning in economics: a systematic and critical review. Artif Intell Rev 56, 9497–9539, doi: 10.1007/s10462-022-10272-8.

[49] Zurada, J., Levitan, A.S., Guan, J. (2011). A comparison of regression and artificial intelligence methods in a mass appraisal context. Journal of Real Estate Research, 33(3), 349-388, doi: 10.1080/10835547.2011.12091311.

[50] Zurada J. M., Levitan A. S., Guan J., (2016). Non-Conventional Approaches To Property Value Assessment. Journal of Applied Business Research, 22(3), 14 pp., doi: 10.19030/jabr.v22i3.1421.

HIGHLIGHTS

- A hybrid solution that uses a genetic algorithm and models property features as fuzzy partitions is proposed for selecting the residential properties whose economic value is most consistent with the purchase offer
- Our framework does not need a learning phase such as in Machine Learning real estate price assessment models
- Expert knowledge is modeled using fuzzy partitioning of the feature domains and evaluating positive or negative deviation value from the average unit price
- The reliability of the solution is evaluated by measuring its fuzziness

Journal Pre-proofs