Revista Brasileira de Computação Aplicada, July, 2023

DOI: 10.5335/rbca.v15i2.13916
Vol. 15, Nº 2, pp. 120–128
Homepage: seer.upf.br/index.php/rbca/index

RBCA ISSN 2176-6649

ORIGINAL PAPER

Determination of house and apartment prices in the Juiz De Fora region, Minas Gerais, through intelligent algorithms

Leomar Santos Marques[1], Ricardo Rodrigues Magalhães[1], Danilo Alves de Lima[2], Juliana Santos Marques[3], João Moreira Neto[4], Jefferson Esquina Tsuchida[5]

[1]Programa de Pós-Graduação em Engenharia de Engenharia Agrícola (PPGENAGRI), UFLA, Campus Lavras, Lavras (MG), [2]Programa de Pós-Graduação em Engenharia de Sistemas e Automação (PPGESISA), UFLA, Campus Lavras, Lavras (MG), [3]Departamento de Engenharia de Alimentos (DEA), Universidade Federal de Lavras (UFLA), Campus Lavras, Lavras (MG) , [4]Departamento de Engenharia (DEG), UFLA, Campus Lavras Lavras (MG) , [5]Departamento de Física (DFI), UFLA, Campus Lavras, Lavras (MG)

*leomar.marques@engenharia.ufjf.br; ricardorm@ufla.br; danilo.delima@ufla.br; Juliana.marquesnutricionista@gmail.com; joao.neto@ufla.br; jefferson.tsuchida@ufla.br

Received: 2022-10-22. **Revised:** 2023-04-08. **Accepted:** 2023-07-16.

Abstract

Estate agents value properties, suggesting a price based on their experience and market data, whereas engineers evaluate properties, performing calculations, structural assessments, and property conservation. Nevertheless, specialised technology combining the technical expertise of both professionals can be created to define a valuation range for a specific property. This study presents the application of intelligent classification algorithms, namely, k-nearest neighbour (kNN), artificial neural network multilayer perceptron (MLP), and support vector machine (SVM) algorithms, and teaching-learning-based optimisation (TLBO) as the input selector. The algorithm developed in this study can be used in three neighbourhoods of the city of Juiz de Fora, Minas Gerais state, Brazil, to classify the price of a property in ranges varying from A to F with 75-millisecond classification speed and 82% accuracy. The system can serve to guide and assist estate agents and engineers in their assessment, thus facilitating the work of these professionals by incorporating the techniques of both in their sales analysis, since the work was formed with real properties sold in the Juiz market.

Keywords: Properties; intelligent algorithms; neural networks; TLBO; kNN; SVM

Resumo

Os corretores imobiliários avaliam os imóveis, sugerindo um preço com base em sua experiência e dados de mercado, enquanto os engenheiros avaliam os imóveis, realizando cálculos, avaliações estruturais e conservação do imóvel. No entanto, pode-se criar tecnologia especializada combinando o conhecimento técnico de ambos os profissionais para definir uma faixa de avaliação para um determinado imóvel. Este estudo apresenta a aplicação de algoritmos de classificação inteligente, a saber, k-neest neighbor (kNN), rede neural artificial multilayer perceptron (MLP) e algoritmos de máquina de vetor de suporte (SVM), e otimização baseada em ensino-aprendizagem (TLBO) como o seletor de entrada. O algoritmo desenvolvido neste estudo pode ser utilizado em três bairros da cidade de Juiz de Fora, Minas Gerais, Brasil, para classificar o preço de um imóvel em faixas que variam de A a F com velocidade de classificação de 75 milissegundos e 82% precisão. O sistema pode servir para orientar e auxiliar agentes imobiliários e engenheiros em sua avaliação, facilitando assim o trabalho desses profissionais ao incorporar as técnicas de ambos em sua estrutura, já que o MLP foi formado com imóveis vendidos no mercado de Juiz de Fora.

Palavras-Chave: Propriedades, algoritmos inteligentes, redes neurais, TLBO, ANN, kNN, SVM

1 Introduction

Estate agents play an important role in property markets. They use their local market knowledge to link information between potential buyers and sellers. Hence, estate agents use their information as leverage to reduce research costs and to find the best possible prices for their clients (Agarwal et al., 2019).

Thus, estate agents are able to buy at lower prices or even increase their sales profitability. An explanation of these effects is related to information asymmetries in the property market. In other words, estate agents have information advantages over less informed stakeholders.

In this context, data can become information, in turn, can represent knowledge. Currently, various types of data are generated and captured very quickly, but converting these data into information and knowledge is not an easy task. For this reason, new fields are emerging and gaining market share, such as data science.

Previous studies have produced various models to automatically estimate property value. These models include techniques based on statistics, machine learning, and artificial neural networks. More complex models are based on vision using internal, external, and satellite images (Nouriani and Lemke, 2022). Soltani et al. (2022) incorporated spatiotemporal dependency to define property prices using a regression model. Baur et al. (2023) used nonparametric property description-based machine learning models aimed at creating covariates for the learning model.

While each of these methods has its own advantages and disadvantages, artificial neural networks have been shown to be effective in predicting property value as they can handle large amounts of data and learn complex, nonlinear patterns Borba et al. (2022); Domingos et al. (2022).

However, artificial neural networks can be sensitive to the choice of input parameters, which can significantly affect the performance of the model. To overcome this problem, the TLBO algorithm was used in this study to define input parameters in classifiers Ziyad Sami et al. (2022).

Thus, the input vectors were different for each individual, and the solution adopted to evaluate each individual was to vary the size of the input layer while keeping the number of neurons in the hidden and output layers constant. This way, the study was able to evaluate each individual accurately and effectively, allowing for more precise results in predicting property value.

Therefore, the aim of the present study is to apply intelligent algorithms, along with an optimization algorithm, in the property market to convert data into information and knowledge and to verify which of these algritomes suited a small bank in order to be implemented in a model of short term memory. Given the constant oscillation of the real estate market in the region of Juiz de Fora, Minas Gerais, Brazil. By implementing TLBO, it becomes possible to adapt the algorithm during training, which in turn eliminates the requirement to use long-term memory. For this purpose, a small database extracted from the internet was created and promptly converted into a price range for a specific property using three main techniques: an artificial neural network multilayer perceptron (MLP), a support vector machine (SVM), and a k-nearest neighbor (kNN) algorithm, as well as an algorithm based on learning behaviors as an input selector called teaching-learning-based optimization (TLBO).

2 Theoretical background

2.1 Property prices

Property pricing dynamics stand out for several reasons. Owner-occupied houses account for most private sector wealth. In addition, housing prices can have a major impact on the distribution of economic well-being and are important in explaining household savings and consumption. From a theoretical standpoint, these relationships are not so clear because the increase in asset prices also implies an increase in renting costs, which tends to offset the impact on real wealth using an adequate cost-of-living index (Englund and Ioannides, 1997).

A set of economic and socio-cultural factors affect the home buying decision, which is made as a function of a local and specific property market (Levy et al., 2008). In the literature on residential mobility, price setting is directly linked to cost of living, demographic effects, supply/demand in a given region, total square footage of a property, number of bedrooms, rental price and regional businesses, which determine the value in the purchase decision (Hoesli and Oikarinen, 2019) (Bourassa et al., 2019; Iacoviello, 2005).

Residential estate agents use certain business tactics to set the price of a specific property. The list price is frequently discussed in a narrative that describes the seller preferences or the expected price of a specific property. If the price is lower than the market value, the property is more quickly sold; in contrast, if the price is higher than the market value, the sale may be delayed or even prevented (Hungria-Gunnelin et al., 2019).

2.2 Teaching-learning-based optimisation (TLBO)

TLBO is a population-based method that simulates classroom behaviours to find a global solution to a problem. This algorithm is divided into two parts. The first consists of the 'teacher phase', and the second consists of the 'learner phase'. In the teacher phase, learning occurs through the teacher, which is the best set of parameters, and in the learner phase, learning occurs through interactions between learners (Rao et al., 2011). The advantage of this algorithm is the need for little input data for its configuration and its faster convergence than other well-regarded optimisation algorithms, such as genetic algorithms (GAs) and particle swarm optimisation (PSO) (Magalhaes et al., 2017).

The Fig. 1 illustrates the basic flow chart of TLBO implementation.

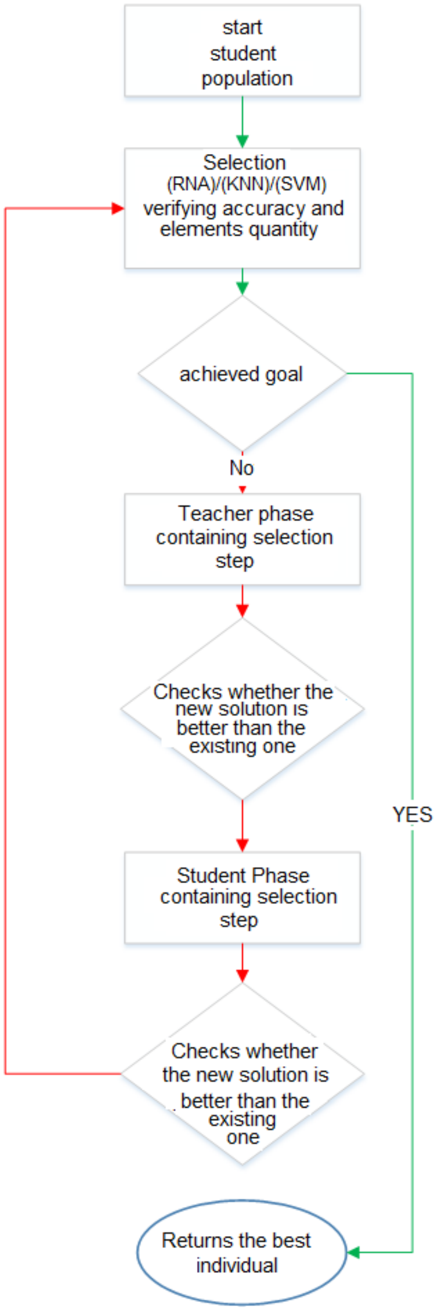

Figure 1: Flow chart for a basic TLBO.

2.3 K-nearest neighbour (kNN)

The kNN algorithm is the simplest and most common classifier of all machine learning algorithms. Its training method consists of storing the properties of a given datum and classifying it as previously specified. Thus, when analysing a new input parameter, the algorithm analyses k results near the parameter to classify it (Kim et al., 2012).

Within a given distance and k number of neighbours, the most frequent classification is the classification of the parameter analysed by the algorithm. The parameter classification process is similar to a voting process, whereby the parameter under analysis is classified based on the most representative class (Tahernezhad-Javazm et al., 2018).

Two factors determine the success of the kNN, namely, the number of neighbours and the distance calculation method. At low k values, the model may suffer from overfitting, impairing the generalisation of the method. Another factor that may reduce the performance of the method is the dimensionality of the parameters. As input dimensionality increases, efficiency decreases (Chen et al., 2015). To improve the functionality of the algorithm, TLBO reduces the dimensionality of the problem. The Fig. 2 illustrates the impact of K parameter on the classifier performance.

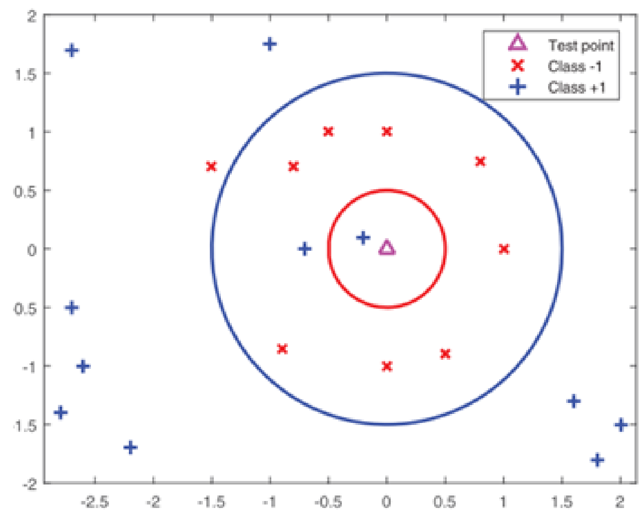

Figure 2: Impact of parameter k on the type classifier KNN, (Tahernezhad-Javazm et al., 2018)

2.4 Artificial neural network multilayer perceptron (MLP)

An MLP is a classifier inspired by brain connections between neurons. An MLP can be divided into three sectors, namely, an input layer, where input data are entered into the network; a hidden layer, which extracts characteristics of the problem under study; and an output layer, which presents the final results from the network. Such a network has a feedforward architecture in which information flows through the network in a single direction, from the input layer to the output layer. Each layer consists of nodes, which represent neurons, with weights assigned to each connection between two nodes. When calibrated, the weight provides the necessary information on the problem. The connections between nodes represent synapses (Naik et al., 2016).

Because an MLP has nonlinear characteristics and the network topology can be easily changed, this classifier is highly flexible in analysing the most diverse problems. However, this method requires an appropriately choice for the number of neurons in the hidden layer (Tahernezhad-

Javazm et al., 2018).

The MLP architecture consists of an input layer, one or more hidden layers (which process data and transform nonlinearly separable problems into separable), and an output layer.

Several algorithms are available for calibrating the weights of the MLP network. In this study, the back-propagation algorithm or generalised delta rule was used because this technique has already been widely applied and commonly implemented in neural network libraries.

The activation function used for the neurons was a hyperbolic tangent for the input and hidden layers and a linear function for the output layer. Its architecture is composed by input layer, one or more hidden layers, where occurs data processing and transform non-linearly separable problems in separable problems and output layer as illustrated in Fig. 3.

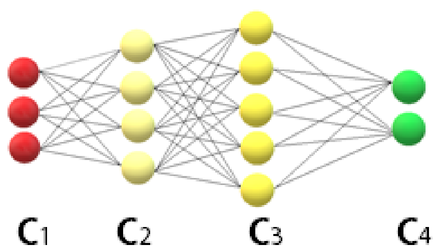

Figure 3: MLP network with two hidden layers, (Marques et al., 2019)

As the study aimed to predict property prices based on a small database extracted from the internet, multilayer perceptron (MLP) was chosen as a suitable algorithm due to its ability to learn complex non-linear relationships in data and handle large datasets. The study used the TLBO algorithm to define input parameters for the classifiers, and the solution adopted to evaluate each individual was to vary the size of the input layer while keeping the number of neurons in the hidden and output layers constant. This approach allowed for different input vectors for each individual, ensuring accurate evaluation.

3 Support vector machine (SVM)

An SVM algorithm classifies information based on the separation of data from different classes by defining a narrow margin and thus establishing a hyperplane that separates the two classes of data. The search criterion is to find the hyperplane that is the longest marginal distance from the data close to the hyperplane. These data or points in a coordinate system are known as support vectors. Assigning kernels facilitates increasing the complexity of the method and can be used to classify nonlinearly separable data (Paik and Kumari, 2017). The assignment of kernels makes possible to increase the complexity of the method, being possible to use in the classification of non-linearly separable data as illustrated in the Fig. 4

The main advantage of the SVM algorithm is its robustness against overfitting and the possibility

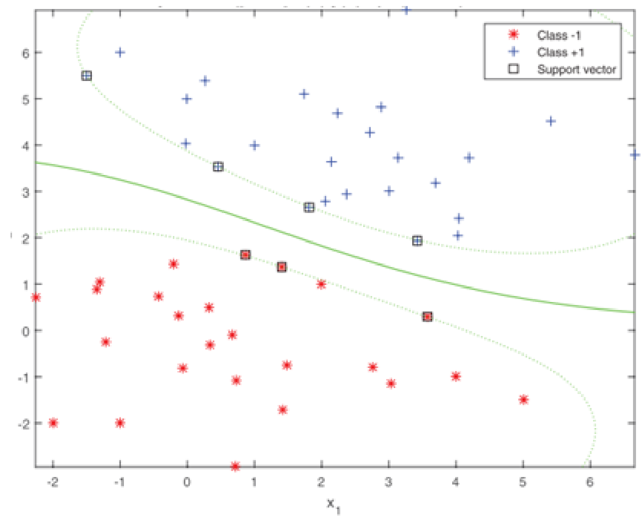

Figure 4: Classifier SVM with nonlinear hyperplane, adapted from (Tahernezhad-Javazm et al., 2018).

of analysing data with high dimensionality without decreasing the performance of the algorithm. Additionally, it requires few samples for training compared to the dimensionality of the input data (Tahernezhad-Javazm et al., 2018).

To configure the separation criterion through the hyperplane, the SVM has a control parameter that enhances the performance of the classifier. Parameter C determines the fitting condition of the problem. High C values may lead to model overfitting. When the C values are low, the model may violate marginal distances and fail to classify some data, resulting in underfitting (Tahernezhad-Javazm et al., 2018).

3.1 Confusion matrix and receiver operating characteristic (ROC) curve

The confusion matrix is a matrix used to store the results, calculating the sensitivity of the quality achieved by each predictor (Lopes et al., 2014). Confusion matrix analysis according to Eq. (1).

$$Sm = \frac{VP}{VP + CFN} \tag{1}$$

where VP is the number of true positive cases, CFN is the number of false negative cases and Sm represents the proportion of positives that were correctly identified.

Through concepts of quality sensitivity and specificity, the accuracy of a diagnostic test can be graphically described by plotting the ROC curve. The specificity consists of the probability of the test finding negative results that are truly negative, and this probability is calculated using Eq. (2) and its representation in matrix form can be illustrated in the Fig. 5 and its representation in graphic form can be illustrated in the Fig. 6.

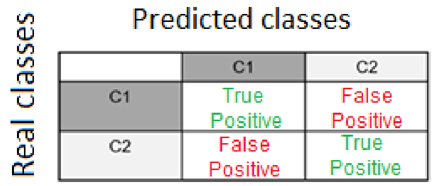

Figure 5: Confusion matrix analysis, adapted (Marques et al., 2019)

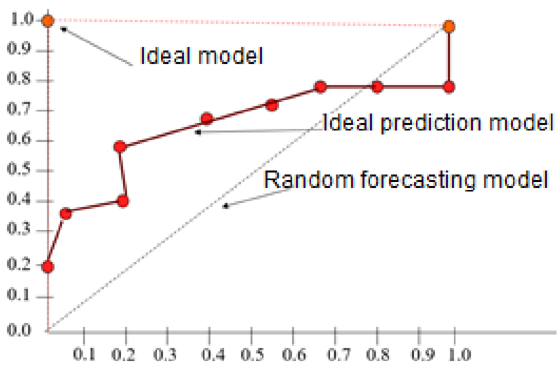

Figure 6: Analysis curve ROC, adapted (Marques et al., 2019)

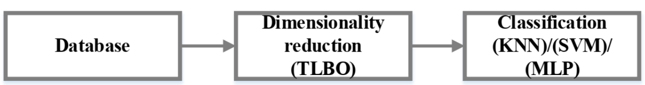

Figure 7: Data classification steps.

$$Sr = \frac{TN}{FPN + TN} \tag{2}$$

where FPN is the number of false positive cases, TN is the number of true negative cases and Sr is the percentage of the area under the curve.

4 Method

The method consisted of searching for properties in the city of Juiz de Fora, Minas Gerais, Brazil. The search was filtered for professional sellers available for consultation on the OLX (OLX, 07/18/2020 to 07/22/2020) platform and designed for a database of ten different neighbourhoods (50 properties in each neighbourhood).

The classification steps are presented in Fig. 7. Next, each step is individually discussed.

4.1 Database

The following data were collected: neighbourhood, number of bedrooms, garage, number of bathrooms, total square footage, whether the home is a house or an apartment, and whether there is a condominium fee, in addition to an external search for the number of large markets within 2 km of the property and the value that determines the property price. The database totalled 150 properties, of which 70% were used for training and 30% for network validation. All data were collected from 18/07/2020 to 22/07/2020 to avoid abrupt changes in the market and were available in a range of up to 30 days. To prevent data distortion, the property values were divided into six range categories, which were analysed and defined as the solution to the algorithm.

4.2 Dimensionality reduction

After the database was collected and simple characteristics were extracted, a dimensionality reduction method was applied to assess whether all entries were actually important for the classification process and thus reduce the dimensionality if necessary.

The TLBO algorithm was used for this purpose, identifying the choice with the best performance. Fig. 8 shows the basic flowchart of the TLBO algorithm implemented for the present problem.

In this study, each of the 8 TLBO algorithm entries indicates whether a datum is chosen as an entry in the classifier or not. However, the TLBO algorithm, unlike the binary GA, does not make it possible to analyse the characteristics of the individuals with constant values of 0 or 1. Such an analysis requires adjusting the individuals of the TLBO algorithm to the selection or not of characteristics.

This selection was defined as follows: individuals were created with random values of 0 or 1 in the initial population, but after the teacher and learner phases were complete, negative values outside the range of 0 to 1 started to appear. Thus, the individual was redefined to return it to the initial range by obtaining the absolute value of the negative values and by adjusting values greater than 1, including the absolute values, to 0.9, to increase the variability in the TLBO processing.

After this step, selection could be performed using a characteristic selection threshold value, and if an individual position was greater than this threshold, this input characteristic was chosen.

The creation of the initial population was also modified: individuals were created with an increasing number of selected characteristics, from a number of characteristics of practically zero to a high number of selected entries. Despite increasing in number, the characteristics were still selected at random. This helped the TLBO search for good solutions with low numbers of entries, broadened the search space, and accelerated the identification of good candidate solutions. The classifications (Class) were given through Eq. (3):

$$Class = AC - 0,1 * N_i \tag{3}$$

where AC corresponds to the accuracy and Ni is the number of individuals. The variable minimised the number of accumulators used; that is, if two individuals had the same accuracy rating, the individual with fewer

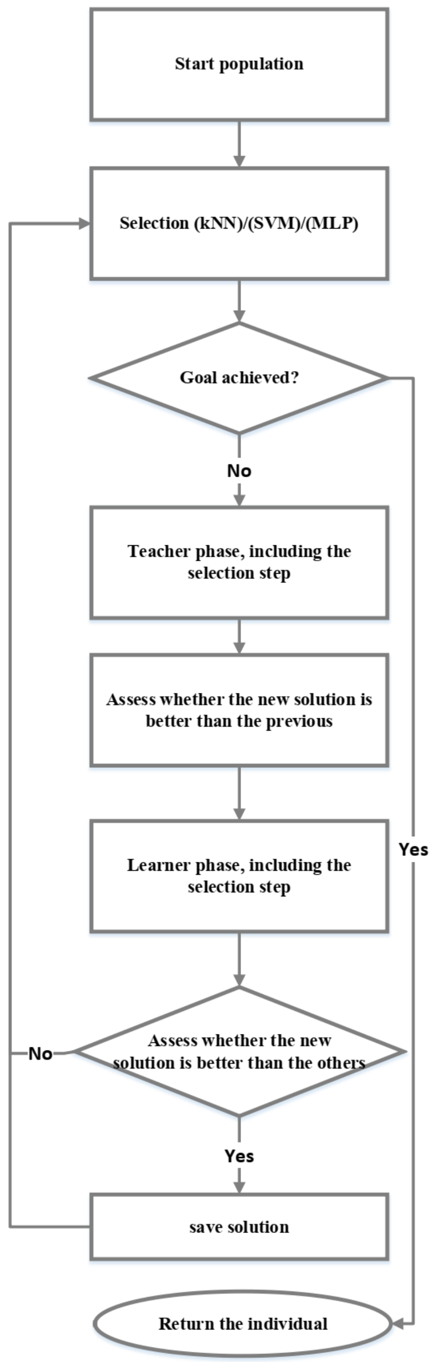

Figure 8: Data classification steps.

elements had a higher rating.

4.3 Classification method

Many studies have used different algorithms, such as ANNs, SVMs, and fuzzy logic, to solve similar problems. In this study, three main algorithms were used, SVM, MLP, and kNN, under different configurations to compare the efficiency of each choice with regard to the number of correctly classified data and the number of individuals.

4.4 Validation of the classification models

The TLBO cost function was defined as the accuracy of each classifier minus an applied penalty, defined in Eq. (3). This accuracy is the percentage of hits of the model for a set of test data. Although the division of the database into a training set and test set is indicated for model validation, only one division does not ensure that the classifier is effective in generalising new data because there is a chance that a data division and data mixtures are obtained that may favour high percentages of hits in the model for the configuration used. In addition, classifiers such as neural networks initialize their synaptic weights at random and are able to derive a model considered ideal only for the particular set of training and test data used.

These optimal random classifiers with high percentages of hits could disrupt the dynamics of the TLBO, influencing the convergence of individuals in the population towards the set of entries used in the classifier.

To avoid this type of random classifier, the stratified k-fold cross-validation technique was used.

This technique consists of dividing the database into k parts, with the classifier using one of the K parts as a test set and the remaining parts as training sets. This process is repeated until all k parts have been used as a test set and the rest as training sets, thus returning an average of the percentages of hits of the test and training sets. In addition, the stratification process ensures that all possible classification classes are present in the test sets in the same proportion, effectively testing the classification capacity of the classifier.

In this study, the database was divided into three parts. Each classifier was trained three times with 66.6% of the data and validated with 33.3% of the data, returning the average percentage of hits of the training and test data to the TLBO algorithm, thus helping to choose a good set of entries and a good classifier configuration.

The algorithm output was bounded by six ranges denoted A through F, where A includes properties priced from 80,000 to 125,000; B properties from 130,000 to 185,000; C 190,000 to 245,000; D 260,000 to 315,000; E 320,000 to 375,000; and F above 380,000 Brazilian reais.

4.5 Comparison method

Two main methods of visual comparison were used to individually analyse the results, namely, the confusion matrix and the ROC curve, which standardised the MLP, SVM and kNN evaluations.

5 Results and discussion

The TLBO algorithm and the classifiers were configured as follows: the TLBO used, in all cases, 20 individuals and 20 iterations for convergence. The kNN classifier used 10 neighbours for classification. The MLP neural network had a learning rate of 0.001, a moment of 0.9, and 10 neurons in the hidden layer. The SVM used the linear

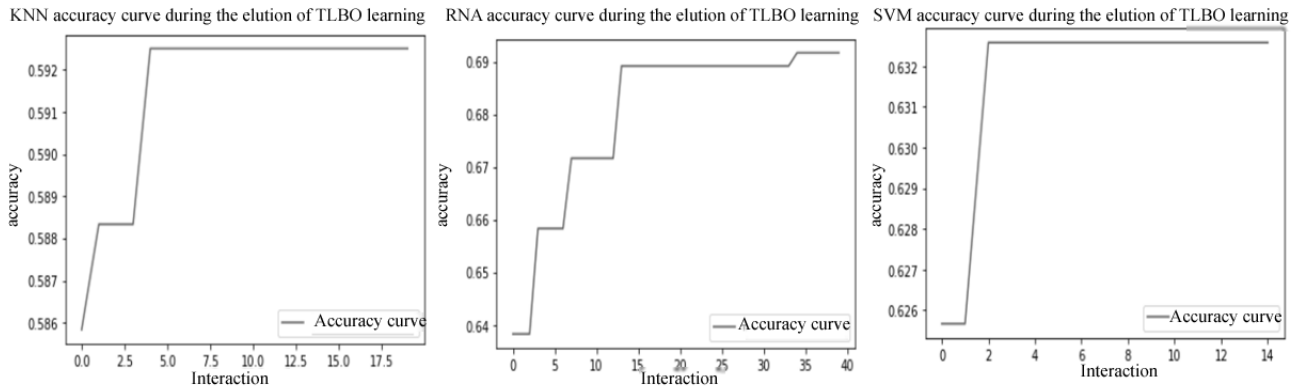

Figure 9: Comparison of the evolution of the classifier over time during training.

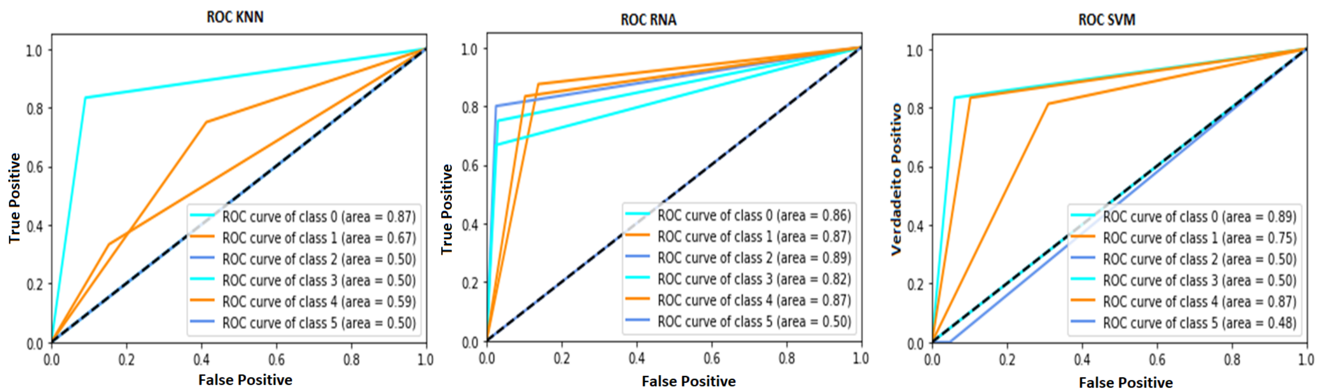

Figure 10: ROC curve comparison.

kernel and penalty parameter C equal to 1.

Analysis of the evolution graph shows that the MLP had the highest percentages of hits of all classifiers during training, reaching 70% during training, according to the graph in Figure 8. In the test using unknown data, the MLP neural network performed much better than the other models, reaching an average accuracy of 82%, much higher than the accuracy of the other models, as shown in Section 5.

For all tests performed (with all properties), the accuracy curves were constructed by iterating the TLBO and ROC, in addition to the confusion matrix. The curves and confusion matrix are shown in Section 5, Section 5, and Section 5, respectively.

When comparing this work with existing works, considerable advantages can be observed over other models previously applied to solve the proposed problem. It was observed that the best results were obtained through the RNA/TLBO algorithm, so the comparison will be made using this algorithm in relation to existing works so far.

The first strong point of the proposed methodology is the elimination of the need for long-term memory, compared to other methodologies presented so far, as described in Baur et al. (2023) and Gao et al. (2022).

Furthermore, it was observed that (Gao et al., 2022) divides its sample space into smaller quantities of properties to support its research, validating the observation of a smaller number of individuals within a database.

When comparing the results, we found that the proposed algorithms obtained errors around 17%. Although the algorithm proposed in this work has higher errors, with an average of 22%, it is important to note that it is better suited for groups of properties worth more than R$ 190.000, where it presented errors lower than 17%, decreasing further for clusters of higher value, reaching 14%, due to the clustering of the system. This reduction in error can be attributed to the presence of a larger number of relevant factors for houses worth more than R$190,000.

6 Conclusion

Among the classifiers used in this study, the MLP neural network, used together with the TLBO algorithm to reduce the input dimensionality, showed the best results, reaching extremely low numbers of inputs and high percentages of hits for the studied individuals. Despite the good results, the computational cost to find such results, approximately 6 hours in a 3.6 GHz i7-7700 computer with 8 Gb of memory, hinders its use as a classifier. However, after training, the model was found to be very agile in classifying an unknown property, with an average response time of approximately 75 milliseconds and an

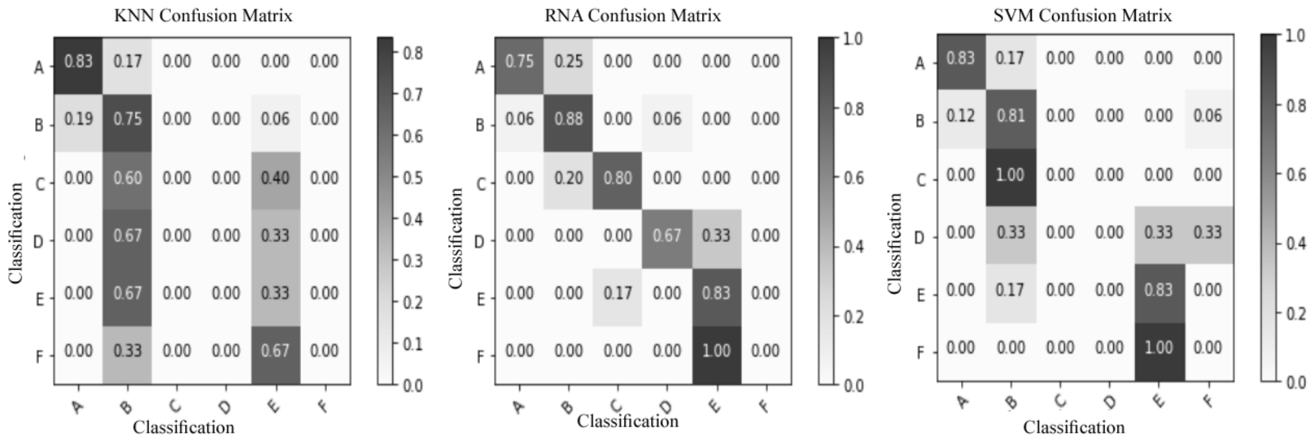

Figure 11: Confusion matrix.

accuracy of 82%.

The system can direct and assist an engineer and estate agent in their evaluation, thus facilitating the work of both professionals by linking the two techniques to its structure because the system was trained with real properties sold in the Juiz de Fora market.

Future studies should conduct tests with different neural network architectures and parameters since the MLP showed the best results and perhaps use a classifier decision committee, considering the same set of input parameters for all properties.

Some adaptations can be made to the proposed algorithm in order to improve its performance. One of these adaptations is the reduction of the cluster range, which would allow for a better segmentation of properties according to their characteristics. This would enable an increase in the number of clusters, resulting in greater precision in predictions.

Another important adaptation is the increase in the database and automatic collection of pricing data. This would allow the algorithm to be fed with a greater amount of information, making predictions more precise and reliable.

To make the algorithm functional and applicable in the real estate market, it is suggested that it be capable of constantly updating its training. One way to achieve this would be to use an LSTM (Long Short Term Memory) type algorithm, which is capable of learning from data sequences over time. This way, it would be possible to use TLBO in the input of predictor parameters, ensuring better prediction accuracy.

Acknowledgments

This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) – Finance Code 001.

The authors also thank the Federal University of Lavras (UFLA) and the Postgraduate Program in Agricultural Engineering (PPGEA), the Postgraduate Program in Systems Engineering and Automation (PPGESISA) and the Department of Physics (DFI) for enabling the development of research.

References

Agarwal, S., He, J., Sing, T. F. and Song, C. (2019). Do real estate agents have information advantages in housing markets?, *Journal of Financial Economics* **134**(3): 715–735. https://doi.org/10.1016/j.jfineco.2019.05.008.

Baur, K., Rosenfelder, M. and Lutz, B. (2023). Automated real estate valuation with machine learning models using property descriptions, *Expert Systems with Applications* **213**: 119147. https://doi.org/10.1016/j.eswa.2022.119147.

Borba, D. et al. (2022). Aplicação de redes neurais artificiais para classificar padrões de ruído eletromagnético conduzido nas linhas de alimentação de circuitos integrados. https://repositorio.pucrs.br/dspace/bitstream/10923/24354/1/000502706-Texto%2Bcompleto-0.pdf.

Bourassa, S. C., Hoesli, M. and Oikarinen, E. (2019). Measuring house price bubbles, *Real Estate Economics* **47**(2): 534–563. https://doi.org/10.1111/1540-6229.12154.

Chen, T., Hong, Z., Deng, F.-a., Yang, X., Wei, J. and Cui, M. (2015). A novel selective ensemble classification of microarray data based on teaching-learning-based optimization, *International journal of multimedia and ubiquitous engineering* **10**(6): 203–218. http://dx.doi.org/10.14257/ijmue.2015.10.6.21.

Domingos, A. C. et al. (2022). Comparação entre as redes lvq e mlp na previsão da atenuação provocada pela chuva.

Englund, P. and Ioannides, Y. M. (1997). House price dynamics: an international empirical perspective, *Journal of housing economics* **6**(2): 119–136. https://doi.org/10.1006/jhec.1997.0210.

Gao, Q., Shi, V., Pettit, C. and Han, H. (2022). Property valuation using machine learning algorithms on statistical areas in greater sydney, australia, *Land Use Policy* **123**: 106409. https://doi.org/10.1016/j.landusepol.2022.106409.

Hungria-Gunnelin, R., Kopsch, F. and Enegren, C. J. (2019). Real estate agents' list price strategies: Higher sales price, shorter time on the market or asymmetric information? http://urn.kb.se/resolve?urn=urn:nbn:se:kth:diva-245041.

Iacoviello, M. (2005). House prices, borrowing constraints, and monetary policy in the business cycle, *American economic review* **95**(3): 739–764. https://doi.org/10.1257/0002828054201477.

Kim, J., Kim, B.-S. and Savarese, S. (2012). Comparing image classification methods: K-nearest-neighbor and support-vector-machines, *Proceedings of the 6th WSEAS international conference on Computer Engineering and Applications, and Proceedings of the 2012 American conference on Applied Mathematics*, pp. 133–138. https://dl.acm.org/doi/abs/10.5555/2209654.2209684#sec-cit.

Levy, D., Murphy, L. and Lee, C. K. (2008). Influences and emotions: exploring family decision-making processes when buying a house, *Housing Studies* **23**(2): 271–289. https://doi.org/10.1080/02673030801893164.

Lopes, B., Ramos, I. C. d. O., Ribeiro, G., Correa, R., Valbon, B. d. F., Luz, A. C. d., Salomão, M., Lyra, J. M. and Ambrósio Junior, R. (2014). Biostatistics: fundamental concepts and practical applications, *Revista Brasileira de Oftalmologia* **73**: 16–22. https://doi.org/10.5935/0034-7280.20140004.

Marques, L. S., Magalhães, R. R. and Ferreira, D. D. (2019). Inteligência computacional aplicada à detecção de câncer de mama, *Revista Brasileira de Computação Aplicada* **11**(1): 28–35. https://doi.org/10.5335/rbca.v11i1.8727.

Naik, B., Nayak, J., Behera, H. S. and Abraham, A. (2016). A self adaptive harmony search based functional link higher order ann for non-linear data classification, *Neurocomputing* **179**: 69–87. https://doi.org/10.1016/j.neucom.2015.11.051.

Nouriani, A. and Lemke, L. (2022). Vision-based housing price estimation using interior, exterior & satellite images, *Intelligent Systems with Applications* **14**: 200081. https://doi.org/10.1016/j.iswa.2022.200081.

OLX (07/18/2020 to 07/22/2020). Pesquisa de imóveis. https://www.olx.com.br/.

Paik, P. and Kumari, B. (2017). Stock market prediction using ann, svm, elm: A review, *International Journal of Emerging Trends & Technology in Computer Science (IJETTCS)* pp. 88–94.

Rao, R. V., Savsani, V. J. and Vakharia, D. (2011). Teaching–learning-based optimization: a novel method for constrained mechanical design optimization problems, *Computer-aided design* **43**(3): 303–315. https://doi.org/10.1016/j.cad.2010.12.015.

Soltani, A., Heydari, M., Aghaei, F. and Pettit, C. J. (2022). Housing price prediction incorporating spatio-temporal dependency into machine learning algorithms, *Cities* **131**: 103941. https://doi.org/10.1016/j.cities.2022.103941.

Tahernezhad-Javazm, F., Azimirad, V. and Shoaran, M. (2018). A review and experimental study on the application of classifiers and evolutionary algorithms in eeg-based brain–machine interface systems, *Journal of neural engineering* **15**(2): 021007. https://doi.org/10.1088/1741-2552/aa8063.

Ziyad Sami, B. F., Latif, S. D., Ahmed, A. N., Chow, M. F., Murti, M. A., Suhendi, A., Ziyad Sami, B. H., Wong, J. K., Birima, A. H. and El-Shafie, A. (2022). Machine learning algorithm as a sustainable tool for dissolved oxygen prediction: a case study of feitsui reservoir, taiwan, *Scientific Reports* **12**(1): 3649. https://doi.org/10.1038/s41598-022-06969-z.