Award Price Estimator for Public Procurement Auctions Using Machine Learning Algorithms: Case Study with Tenders from Spain

Manuel J. GARCIA RODRIGUEZ[1], Vicente RODRIGUEZ MONTEQUIN[1]*,
Andoni ARANGUREN UBIERNA[2], Roberto SANTANA HERMIDA[2],
Basilio SIERRA ARAUJO[2], Ana ZELAIA JAUREGI[2]

[1] Project Engineering Area, University of Oviedo, Oviedo, 33004, Spain
 manueljgarciar@gmail.com, montequi@uniovi.es (*Corresponding author*)

[2] Department of Computer Sciences and Artificial Intelligence, University of the Basque Country,
San Sebastián, 20018, Spain
andoni.aranguren@gmail.com, roberto.santana@ehu.eus, b.sierra@ehu.eus, ana.zelaia@ehu.eus

Abstract: The public procurement process plays an important role in the efficient use of public resources. In this context, the evaluation of machine learning techniques that are able to predict the award price is a relevant research topic. In this paper, the suitability of a representative set of machine learning algorithms is evaluated for this problem. The traditional regression methods, such as linear regression and random forest, are compared with the less investigated paradigms, such as isotonic regression and popular artificial neural network models. Extensive experiments are conducted based on the Spanish public procurement announcements (tenders) dataset and employ diverse error metrics and implementations in WEKA and Tensorflow 2.

Keywords: Machine learning, Neural networks, Public procurement, Spanish tender.

1. Introduction

The importance of the public procurement is well known. In terms of projects and cost, the largest adjudicators of a country are the public procurement agencies. For example, the public authorities of the European Union (EU) spent around 14% of their GDP (around €2 trillion) on public procurement (purchase of services, works and supplies) in 2017 (European Commission, 2017). Therefore, improving public procurement can yield enormous savings: even a 1% efficiency gain could save €20 billion per year. It is crucial to analyse the public procurement notice (also called auctions, requests for tender or simply tenders) in order to understand its behaviour in terms of prices. Through the use of new technologies, like machine learning (ML), among others, new tools can be created to improve these public procurement processes.

ML involves computer algorithms that are used for knowledge discovery from large amounts of data. It is considered to be a type of artificial intelligence (AI), and it is regarded as one of the most disruptive innovations and a strong enabler of competitive advantages. While ML has been around for more than 60 years, it has only recently showed significant potential for disrupting economies and societies (Lee & Shin, 2020). Mirroring a trend that has increased pace in the last 5 years in the private sector worldwide, the adoption of AI within public administration processes has the potential to provide enormous benefits. It improves the efficiency and effectiveness of policy making and service delivery to businesses and citizens, ultimately enhancing their level of satisfaction and trust in the quality of public service (Kuziemski & Misuraca, 2020).

The award price estimator is a regression problem. The tender has x known input features (e.g., date, tender price, type of contract, and public procurement agency) and a y unknown output feature (award price). The tender price, which is calculated by the public procurement agency, is the key input parameter to the award price estimator. The tender price is the theoretical price and the estimator adjusts it regarding the real and changing market conditions to predict the award price by the winning bidder.

The aim of this article is to improve the accuracy of the award price estimator studied previously in (García Rodríguez et al., 2019a). That article applied only one algorithm (random forest) to predict the award price, and it was validated over two tender datasets from Spain and Europe. Further, this article increases the prediction accuracy, and it compares four algorithms: linear

regression, isotonic regression, random forest and artificial neural network. The last two algorithms are ML methods, particularly supervised learning.

An award price estimator would produce significant benefits. It would be an excellent tool for the cost planning of public tendering agencies by allowing them to have more realistic budgets. Additionally, such a price estimator would provide support to small- and medium-sized enterprises (SMEs) that play a crucial role in most economies. For example, SMEs represent 50% of the GDP in the EU (European Commission, 2020). However, they have difficulty when competing on equal terms with big suppliers in the public procurement space. Other benefits could be the reduction of fraud between bidders, which would improve the transparency of the process and lead to better quantification of the product quality.

The paper begins with reviewing the literature and identifying the research gap to be examined (Section 2). Then, the dataset of public procurement auctions, the ML algorithms being compared (random forest, linear regression, isotonic regression and artificial neural networks (ANNs)) and the error metrics that are used are described (Section 3). Next, the major quantitative results of the experimental analysis are summarized for identifying the best ML algorithm to predict the award price (Section 4). Finally, some concluding remarks, limitations, and avenues for future research are presented (Section 5).

2. Literature Review

While an increasing number of studies in public procurement is being published every year, an overview of the field is missing. In the literature on public procurement, an ambiguous wording is usually used, and a consensus on the terminology and concepts involved has not been reached yet (Obwegeser & Müller, 2018). Technological and organizational challenges faced during public electronic procurement processes are not well understood despite past studies focusing on these topics (Mohungoo, Brown & Kabanda, 2020). The data analysis of public tenders can provide valuable information for different stakeholders: public tendering agencies, public

procurement managers, project managers, executives, politicians and, indirectly, citizens. In the particular case of Spain, an initial analysis published in (García Rodríguez et al., 2019b) explains the Spanish public tendering system and the potential applications and benefits of employing massive data processing.

The past decades have seen the rapid development of the computer hardware, communication technologies and computer sciences (artificial intelligence and big data). These new technologies make it possible to implement the informatization of conventional public procurement tendering processes. Public procurement has the typical objectives of the private sector: to acquire the right goods or services from the right supplier, at the right price, at the highest service level, and considering laws and norms requirements. But it also requires strict compliance with the principles of non-discrimination, free competition, and transparency of the awarding procedures (Dotoli, Epicoco & Falagario, 2020).

There is extensive literature about prediction techniques (forecasting) and data analysis in public tendering. There are mainly two approaches: statistical models (e.g., mathematical algorithms) and statistical learning (e.g., ML algorithms). There is not a clear demarcation or boundary between both approaches because research on ML also covers the conception of mathematical algorithms. Thus, statistics and ML are closely related fields in terms of methods, but distinct with regard to their principal goal: statistics draws population inferences from a sample, while ML finds generalizable predictive patterns (Bzdok, Altman & Krzywinski, 2018).

Statistical models are the traditional or conventional approach used to analyse and validate hypotheses. For example, there are models for statistical relationships for tender forecasting in capped tender (Ballesteros-Pérez, González-Cruz & Cañavate-Grimal, 2012), scoring probability graphs (Ballesteros-Pérez, González-Cruz & Cañavate-Grimal, 2013), multicriteria decision making (Dotoli, Epicoco & Falagario, 2020), the probability of bidder participation (Ballesteros-Pérez et al., 2015; Ballesteros-Pérez et al., 2016), and the optimal

bidder participation to achieve the lowest procurement prices (Onur & Tas, 2019). There is also a mathematical model where the bidders are evaluated on the basis of price and quality through a score function (Lorentziadis, 2020), the detection of groups of bidders in collusive auctions (also called not competitive tenders or bid-rigging cartels) (Conley & Decarolis, 2016) or discriminatory competitive procedures in public procurement with unverifiable quality (Albano, Cesi & Iozzi, 2017).

On the other hand, a variety of ML techniques has also been successfully applied to public procurement and created empirical models. For example, among the particular problems addressed by this type of algorithm are those related to the behaviour of bidders: the estimation of the number of bidders in tenders (KNN) (Gorgun, Kutlu & Onur Tas, 2020), the identification of the optimal bidder (fuzzy logic) (Wang et al., 2014), creating a search engine of suppliers to recommend potential bidders for a characterized tender (random forest) (García Rodríguez et al., 2020) the detection of collusive auctions (ensemble method) (Huber & Imhof, 2019), or the proposal of an objective system (key performance indicators) for supporting the estimators (benchmarking) during the tender evaluation process (ANNs) (Bilal & Oyedele, 2020).

However, there are almost no studies about award price forecasting, so there is a research gap. The first holistic approach that considers all kinds of tenders (multi-sectorial) and a large volume of tenders is (García Rodríguez et al., 2019a) whose dataset is used in this article. Previously, two articles created award price estimators with ML algorithms, but they were applied only to construction auctions: bridge projects (Chou et al., 2015) and highway procurement (Kim & Jung, 2019). It is typical to find literature focused only on public procurement for construction or civil engineering projects; this is mainly because they are the biggest and most important projects in public procurement (García Rodríguez et al., 2019a). This paper is the first attempt to compare different algorithms in order to improve the accuracy of award price forecasting in multi-sectorial tenders.

In conclusion, this article is a true reflection of the applicability of ML in public procurement. The fundamental insight behind this breakthrough is as much statistical as computational. Artificial intelligence became possible once researchers stopped approaching intelligence tasks procedurally and began tackling them empirically (Mullainathan & Spiess, 2017). ML algorithms produce a powerful, flexible way of making quality predictions, but they have a weakness: they do not contain strong assumptions and instead contain mostly unverifiable assumptions due to the fact that ML approaches do not generally produce stable estimates of the underlying parameters (Mullainathan & Spiess, 2017).

3. Experimental Procedures

The main objective of this work is to analyze different ML paradigms for predicting the award (winning) price of Spanish public tenders. In this section, the dataset and the learning models are presented, and details about the error metrics and validation method are given.

3.1 Dataset

The original data were extracted from the information files published by the Spanish Ministry of Finance (see Data Availability). It contained information about tenders published between 2012 and 2018. The data were preprocessed for a preliminary study published in (García Rodríguez et al., 2019a) and a dataset of 58,337 Spanish tenders was obtained. To compare the results, the same dataset was used in the experiments presented in this article.

Tenders in the dataset were defined by 14 input variables that provided the following information: the name of the public procurement agency that made the tender, geographical information about the agency (municipality, province, region, and wider region code), the tender price (the amount of budgeted bidding), the duration (days to execute the contract), the type of work according to the common procurement vocabulary (CPV) in two levels of detail, the type of contract defined by legislation (in two levels of detail), the procedure by which the contract was awarded, the urgency level

and the date of agreement in the award of the contract. Note that during preprocessing, all this information was converted to integer values to make it suitable for the learning methods being evaluated. The output variable was the award price, which is the amount offered by the winning bidder of the contract.

3.2 Machine Learning Algorithms

The random forest for regression ML model was selected to create an award price predictor in the preliminary study (García Rodríguez et al., 2019b). The research presented here aimed to investigate a wider range of ML paradigms, compare them and select the most suitable one for the task. Models for regression need to be selected, since the output variable to predict is the award price. Very widely used supervised ML algorithms were considered: random forest, linear regression, isotonic regression and artificial neural networks (ANNs). A brief description of them is presented here.

A random forest algorithm (Breiman, 2001) is a combination of tree predictors, where each tree depends on the values of a random vector independently sampled and with the same distribution for all the trees in the forest. The prediction of the ensemble is computed by averaging the predictions of the individual models. It is a typical example of an ensemble method that reduces the bias of individual models and provides a more flexible predictor that is less prone to overfitting.

While the random forest model is robust, there are situations where simpler algorithms, like linear regression, could produce better results. This explains the convenience of evaluating the performance of linear regression for award price estimation.

Linear regression (equation 1) is a machine learning technique used to model the linear relationship between the input variables x_i and the output variable y:

$$y = \sum_{i=1}^{n} (\beta_i x_i + \varepsilon) \qquad (1)$$

where β_i are the parameters that measure the influence of the input variables, and ε is a constant value.

Another technique that is increasingly applied to regression problems is isotonic regression (equation 2). This method tries to find a line as close to the observations as possible:

$$\min_{g} \sum_{i=1}^{m} w_i (g(x_i) - f(x_i))^2 \qquad (2)$$

where x_i are the input variables, g is the isotonic estimator, f is a function, w_i are the weights and m is the number of observations. This method produced a series of predictions for the training data that were the closest to the targets in terms of the mean square error (MSE). These predictions were interpolated to predict unseen data. The predictions from the isotonic regression thus formed a function that was piecewise linear (Chakravarti, 1989).

For the three ML algorithms presented in this section, the implementations available in WEKA (Hall et al., 2009; Witten et al., 2011) were used in the experiments. WEKA is a machine learning platform developed by Waikako University, that supports a large number of learning algorithms (Waikako University, 2021).

3.3 ANNs

Recently, ANNs have re-emerged as a powerful tool to deal with a variety of ML problems. In particular, they have been applied to regression problems where the input data can be noisy or not fully observed. An ANN is a computational model inspired by biological neural networks. It consists of a collection of units or nodes (artificial neurons) organized in connected layers. The parameters of the model are the weights and biases associated to the connections. Information is processed from the input layer to the output layer.

The learning process is based on minimizing a cost function (also known as loss function) that evaluates the performance of the network for the given task. Backpropagation is used to learn the weights associated to the connections. One of the ANNs used in this work is a multi-layer perceptron (MLP) (Hastie, Tibshirani & Friedman, 2009) implemented in WEKA (Hall et al., 2009; Witten et al., 2011).

3.4 ANN Optimization (Deep Learning)

In addition to using the MLP implementation in WEKA, a set of ANN architectures that represented a different number of layers (to evaluate the impact of the depth) and a different number of neurons in each layer was selected. The particular choice of the number of neurons is arbitrary and was intended to keep a balance between the goals of increasing the capacity of the model and keeping a manageable complexity.

The selected ANN architectures were the following: two architectures of one hidden layer with 16 nodes and 32 nodes; five architectures of three hidden layers of 16 nodes in each (16-16-16), 32 nodes in each (32-32-32), and a different number of nodes in each (16-8-16, 32-8-32, 32-16-32); and an architecture of five hidden layers with 32-16-8-16-32 nodes (see Figure 1). For each of these ANN designs, different activation functions, loss functions and gradient descent optimization algorithms were evaluated.

The activation function determines the type of non-linear transformation made to the linear combination of the weights and input neurons. In most cases, the rectified linear unit (ReLU) general activation function is used. Recently, the scaled exponential linear unit (SeLU) activation function (Klambauer et al., 2017) has been reported to produce promising results. This is an activation function that induces self-normalizing properties.

Regarding the choice of ReLU and SeLU, preliminary experiments were made with other activation functions including a sigmoid. Due to the page number restrictions and the poor results achieved with these functions, it was decided to include only the results for the ReLU and SeLU. It is emphasized that both functions are theoretically more sound since they address the vanishing and exploding gradient problems experienced by the sigmoid and hyperbolic tangent functions. They can be used for all the main neural network paradigms (i.e., MLPs, CNNs, and RNNs). In particular, SeLU, one of the newest activation functions proposed in the literature, was introduced with an eye on standard feed-forward neural networks and not envisioning CNNs.

Regarding the selection of the regression loss-functions, the common ones were used: MSE or the sum of squared distances, mean absolute error (MAE) or the sum of absolute differences (see subsection 3.5). Among the available gradient descent optimization algorithms commonly used, Adagrad (Duchi, Bartlett & Wainwright, 2012) and Adam and Adamax (Kingma & Ba, 2014) were selected for the experiments (Keras, 2021). For the optimization process, the maximum number of epochs (times the learning algorithm iterated through the training dataset) was set to 50,000.

The eight ANN structures combined with two activation functions, two loss functions and three optimizers provided 96 different ANN designs. Only the training dataset (46,670 tenders) was used for the optimization process. Two different validation frameworks were evaluated: a train/test
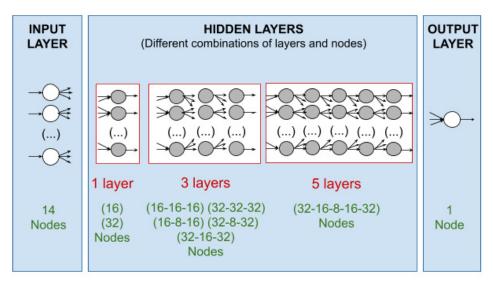

Figure 1. ANN structures

division (Hold-out 80/20) and a K-fold cross-validation with K=10. Figure 2 and Figure 3 show the results obtained for the four error metrics (MAE, root mean square error (RMSE), relative absolute error (RAE) and root relative square error (RRSE)). The rows of the tables correspond to the eight different ANN architectures, the columns correspond to the two activation functions (RELU, SELU), two loss functions (MAE, MSE) and three optimizers (Adam, Adamax, Adagrad). The best results (minimum error) are coloured in green, the intermediate ones are in orange and the worst are in red.

A set of ANN configurations that performed well during the optimization phase was selected for the final test:

- ANN1: Three hidden layers with 16-8-16 nodes, SeLU activation function, MAE loss function and Adam optimizer (see Figure 2);

- ANN2: Three hidden layers with 32-8-32 nodes, SeLU activation function, MAE loss function and Adagrad optimizer (see Figure 3);

- ANN3: Three hidden layers with 16-8-16 nodes, ReLU activation function, MAE loss function and Adagrad optimizer (see Figure 2);

- ANN4: Three hidden layers with 16 nodes in each, ReLU activation function, MSE loss function and Adam optimizer (see Figure 3).

The optimization for the ANN architectures was performed using Tensorflow (Abadi et al., 2016; Tensorflow, 2021), which is an open source software library for ML. It is a very appropriate platform to evaluate with different ANN architectures. Keras is an API designed to simplify the use of Tensorflow.

3.5 Error Metrics

In this subsection, the error metrics used to measure the deviation of the predicted values compared to the real ones are presented.

The MAE (equation 3) and RMSE (equation 4) are two of the most common metrics used to measure accuracy in absolute terms for continuous variables. The MAE measures the average magnitude of the errors in a set of predictions without considering their direction. It was also used as the loss function for the ANNs in the present experiments. The RMSE is a quadratic

scoring rule that also measures the average magnitude of the error:

$$MAE = \frac{1}{m} \sum_{i=1}^{m} |r_i - p_i| \qquad (3)$$

$$RMSE = \sqrt{\frac{1}{m} \sum_{i=1}^{m} (r_i - p_i)^2} \qquad (4)$$

where r_i are the actual observations (the true values), P_i are the predicted values and m is the number of observations.

In relative terms, the RAE (equation 5) and RRSE (equation 6) calculate the error values as a ratio (percentage):

$$RAE = \frac{\sum_{i=1}^{m} |r_i - p_i|}{\sum_{i=1}^{m} |r_i - \overline{r}|} \qquad (5)$$

$$RRSE = \sqrt{\frac{\sum_{i=1}^{m} (r_i - p_i)^2}{\sum_{i=1}^{m} (r_i - \overline{r})^2}} \qquad (6)$$

where $\overline{r}$ is the mean of the actual observations. Values over 100% appear when the absolute or quadratic difference between the predicted values and the actual observations are bigger than the differences between the actual observations and their means.

Finally, the MSE (equation 7) is used as a loss function in the present experiments with the ANNs:

$$MSE = \frac{1}{m} \sum_{i=1}^{m} (r_i - p_i)^2 \qquad (7)$$

3.6 Validation

To evaluate the performance of the different ML paradigms, models were trained with 80% of the tenders (46,670). The remaining 20% (11,667) were used as a test. The same validation framework and train/test division of the dataset were used in the preliminary study.

4. Experimental Results

The validation was performed for the 11,667 tenders in the test dataset. The random forest model was considered to be the baseline because it was used in the preliminary study and therefore enabled the comparison of the behaviour of different ML paradigms.

Error metrics	ANN configuration	Experimental results												Colour legend	
	Activation function	RELU						SELU							
	Regression loss function	Mean Absolute Error			Mean Squared Error			Mean Absolute Error			Mean Squared Error			Percentile value	Percentile
	Optimizer / Layer structure	Adam	Adamax	Adagrad	Adam	Adamax	Adagrad	Adam	Adamax	Adagrad	Adam	Adamax	Adagrad		
MAE (M€)	16x1	0.94M€	0.77M€	0.88M€	1.04M€	0.89M€	0.73M€	0.80M€	0.75M€	0.86M€	0.89M€	0.83M€	0.75M€	1.27M€	100
	32x1	0.91M€	0.76M€	0.94M€	0.92M€	0.86M€	0.75M€	0.82M€	0.75M€	1.03M€	0.94M€	0.88M€	0.74M€	0.87M€	75
	16x3	0.78M€	0.75M€	0.76M€	0.73M€	0.80M€	0.80M€	0.77M€	0.75M€	0.79M€	0.80M€	0.86M€	0.80M€	0.82M€	59
	32x3	0.78M€	0.74M€	0.76M€	0.94M€	0.77M€	0.74M€	0.78M€	0.74M€	0.79M€	0.80M€	0.82M€	0.73M€	0.79M€	41
	16-8-16	0.78M€	0.76M€	**0.67M€**	0.73M€	0.79M€	0.75M€	**0.55M€**	0.76M€	0.80M€	0.81M€	0.77M€	0.73M€	0.76M€	25
	32-8-32	1.11M€	1.03M€	0.86M€	0.80M€	0.79M€	0.74M€	0.95M€	1.25M€	0.86M€	0.77M€	0.77M€	0.85M€	0.75M€	16
	32-16-32	0.90M€	1.01M€	0.82M€	0.79M€	0.85M€	0.73M€	1.06M€	1.27M€	0.80M€	0.83M€	0.83M€	0.82M€	0.73M€	8
	32-16-8-16-32	1.00M€	1.23M€	0.84M€	0.83M€	0.83M€	0.81M€	0.98M€	1.22M€	0.81M€	0.76M€	0.78M€	0.76M€	0.55M€	0
RMSE (M€)	16x1	10.80M€	10.86M€	10.78M€	10.62M€	10.62M€	10.74M€	10.84M€	10.68M€	10.45M€	10.59M€	10.63M€	10.64M€	10.99M€	100
	32x1	10.68M€	10.76M€	10.63M€	10.69M€	10.61M€	10.65M€	10.71M€	10.71M€	10.68M€	10.70M€	10.62M€	10.56M€	10.78M€	75
	16x3	10.72M€	10.73M€	10.61M€	10.68M€	10.66M€	10.57M€	10.81M€	10.47M€	10.54M€	10.66M€	10.71M€	10.56M€	10.71M€	59
	32x3	10.95M€	10.53M€	10.60M€	10.60M€	10.51M€	10.66M€	10.82M€	10.50M€	10.63M€	10.62M€	10.62M€	10.62M€	10.66M€	41
	16-8-16	10.77M€	10.59M€	**10.36M€**	10.64M€	10.57M€	10.58M€	**10.11M€**	10.68M€	10.60M€	10.68M€	10.56M€	10.58M€	10.62M€	25
	32-8-32	10.91M€	10.79M€	10.86M€	10.70M€	10.71M€	10.67M€	10.86M€	10.86M€	10.93M€	10.73M€	10.64M€	10.64M€	10.58M€	16
	32-16-32	10.91M€	10.99M€	10.92M€	10.70M€	10.65M€	10.58M€	10.89M€	10.92M€	10.91M€	10.71M€	10.71M€	10.71M€	10.55M€	8
	32-16-8-16-32	10.95M€	10.84M€	10.86M€	10.70M€	10.73M€	10.68M€	10.92M€	10.84M€	10.83M€	10.71M€	10.72M€	10.71M€	10.11M€	0
RAE (%)	16x1	145%	119%	137%	160%	137%	112%	123%	116%	133%	137%	127%	115%	195%	100
	32x1	140%	118%	145%	142%	133%	115%	127%	116%	159%	145%	136%	114%	134%	75
	16x3	120%	116%	118%	113%	123%	123%	119%	115%	122%	123%	132%	124%	127%	59
	32x3	121%	115%	118%	145%	118%	114%	120%	114%	122%	124%	127%	112%	121%	41
	16-8-16	120%	117%	**104%**	113%	121%	115%	**85%**	117%	123%	125%	119%	113%	118%	25
	32-8-32	171%	159%	133%	124%	122%	115%	147%	192%	133%	119%	118%	131%	115%	16
	32-16-32	139%	156%	126%	122%	131%	112%	164%	195%	124%	128%	128%	127%	113%	8
	32-16-8-16-32	155%	189%	129%	128%	129%	120%	151%	188%	125%	117%	120%	117%	85%	0
RRSE (%)	16x1	109.4%	110.0%	109.2%	107.5%	107.6%	108.8%	109.8%	108.2%	105.8%	107.2%	107.6%	107.8%	111.3%	100
	32x1	108.2%	109.0%	107.7%	108.3%	107.5%	107.8%	108.5%	108.5%	108.1%	108.4%	107.5%	107.0%	109.2%	75
	16x3	**108.5%**	108.6%	107.5%	108.2%	108.0%	107.1%	109.5%	106.1%	106.7%	108.0%	108.5%	107.0%	108.5%	59
	32x3	110.9%	106.7%	107.4%	107.4%	106.5%	107.9%	109.6%	106.3%	107.6%	107.5%	107.5%	107.6%	108.0%	41
	16-8-16	109.1%	107.3%	**104.9%**	107.8%	107.0%	107.1%	**102.4%**	108.2%	107.4%	108.2%	106.9%	107.2%	107.5%	25
	32-8-32	110.5%	109.2%	110.0%	108.4%	108.5%	108.1%	110.0%	110.0%	110.7%	108.6%	107.7%	107.8%	107.2%	16
	32-16-32	110.5%	111.3%	110.6%	108.4%	107.9%	107.2%	110.3%	110.6%	110.5%	108.5%	108.5%	108.4%	106.8%	8
	32-16-8-16-32	110.9%	109.8%	110.0%	108.3%	108.6%	108.2%	110.6%	109.8%	109.7%	108.4%	108.5%	108.5%	102.4%	0

Figure 2. Error metrics (MAE, RMSE, RAE and RRSE) for different ANN configurations with validation framework train/test division (hold-out 80/20)

Error metrics	ANN configuration	Experimental results												Colour legend	
	Activation function	RELU						SELU							
	Regression loss function	Mean Absolute Error			Mean Squared Error			Mean Absolute Error			Mean Squared Error			Percentile value	Percentile
	Optimizer / Layer structure	Adam	Adamax	Adagrad	Adam	Adamax	Adagrad	Adam	Adamax	Adagrad	Adam	Adamax	Adagrad		
MAE (M€)	16x1	0.99M€	0.90M€	0.96M€	0.97M€	1.01M€	0.89M€	0.97M€	0.89M€	0.94M€	1.08M€	1.02M€	0.88M€	1.25M€	100
	32x1	1.15M€	0.91M€	0.95M€	1.01M€	1.14M€	0.92M€	1.01M€	0.90M€	1.02M€	0.97M€	1.01M€	0.91M€	0.97M€	75
	16x3	0.89M€	0.89M€	0.88M€	**0.81M€**	0.88M€	0.89M€	0.90M€	0.87M€	0.91M€	0.91M€	0.99M€	0.89M€	0.91M€	59
	32x3	0.89M€	0.88M€	0.86M€	0.91M€	0.83M€	0.98M€	0.90M€	0.90M€	0.86M€	0.82M€	0.83M€	0.98M€	0.89M€	41
	16-8-16	0.84M€	0.90M€	0.91M€	0.95M€	1.04M€	0.89M€	**0.81M€**	0.90M€	0.91M€	0.87M€	1.02M€	0.91M€	0.88M€	25
	32-8-32	0.89M€	0.91M€	0.85M€	0.86M€	0.90M€	0.84M€	0.82M€	0.83M€	**0.77M€**	0.92M€	0.95M€	0.87M€	0.86M€	16
	32-16-32	0.89M€	0.90M€	0.89M€	0.88M€	0.99M€	1.09M€	0.88M€	0.89M€	0.89M€	0.96M€	1.15M€	1.01M€	0.83M€	8
	32-16-8-16-32	0.83M€	0.87M€	0.87M€	0.98M€	0.91M€	0.88M€	0.85M€	0.89M€	**0.80M€**	0.86M€	1.25M€	0.87M€	0.77M€	0
RMSE (M€)	16x1	14.30M€	14.29M€	14.25M€	14.26M€	14.29M€	14.24M€	14.26M€	14.26M€	14.26M€	14.30M€	14.27M€	14.25M€	14.32M€	100
	32x1	14.30M€	14.31M€	14.26M€	14.25M€	14.29M€	14.23M€	14.31M€	14.28M€	14.24M€	14.32M€	14.24M€	14.27M€	14.27M€	75
	16x3	14.26M€	14.29M€	14.27M€	**14.18M€**	14.22M€	14.23M€	14.32M€	14.24M€	14.27M€	14.19M€	14.21M€	14.23M€	14.26M€	59
	32x3	14.30M€	14.28M€	14.26M€	14.25M€	14.21M€	14.28M€	14.32M€	14.28M€	14.27M€	14.25M€	14.19M€	14.22M€	14.24M€	41
	16-8-16	14.24M€	14.24M€	14.25M€	14.23M€	14.24M€	14.23M€	**14.21M€**	14.27M€	14.27M€	14.18M€	14.23M€	14.22M€	14.23M€	25
	32-8-32	14.27M€	14.29M€	14.18M€	14.27M€	14.25M€	14.18M€	14.30M€	14.19M€	**14.08M€**	14.24M€	14.25M€	14.25M€	14.22M€	16
	32-16-32	14.25M€	14.28M€	14.24M€	14.29M€	14.21M€	14.25M€	14.25M€	14.23M€	14.24M€	14.23M€	14.21M€	14.23M€	14.19M€	8
	32-16-8-16-32	14.21M€	14.26M€	14.23M€	14.26M€	14.26M€	14.23M€	14.25M€	14.25M€	**14.16M€**	14.23M€	14.30M€	14.23M€	14.08M€	0
RAE (%)	16x1	124%	112%	121%	121%	127%	112%	122%	112%	118%	136%	128%	110%	156%	100
	32x1	144%	114%	119%	127%	143%	115%	127%	113%	128%	122%	126%	114%	121%	75
	16x3	112%	111%	110%	**102%**	111%	112%	113%	109%	113%	115%	124%	112%	114%	59
	32x3	112%	111%	108%	114%	104%	123%	112%	113%	108%	103%	103%	123%	112%	41
	16-8-16	106%	113%	114%	119%	130%	112%	**102%**	112%	114%	109%	128%	115%	110%	25
	32-8-32	111%	114%	107%	108%	113%	105%	103%	104%	**96%**	116%	119%	109%	108%	16
	32-16-32	111%	112%	112%	110%	124%	137%	111%	112%	111%	121%	144%	127%	104%	8
	32-16-8-16-32	104%	109%	108%	123%	115%	110%	107%	112%	**101%**	108%	156%	110%	96%	0
RRSE (%)	16x1	103.0%	102.9%	102.6%	102.7%	103.0%	102.6%	102.7%	102.7%	102.7%	103.0%	102.8%	102.7%	103.2%	100
	32x1	103.0%	103.1%	102.7%	102.6%	103.0%	102.5%	103.1%	102.9%	102.6%	103.2%	102.6%	102.8%	102.8%	75
	16x3	102.7%	102.9%	102.8%	**102.1%**	102.4%	102.4%	103.1%	102.9%	102.8%	102.3%	102.4%	102.4%	102.7%	59
	32x3	103.0%	102.8%	102.8%	102.6%	102.4%	102.8%	103.1%	102.9%	102.8%	102.7%	102.2%	102.4%	102.6%	41
	16-8-16	102.6%	102.6%	102.7%	102.5%	102.6%	102.5%	**102.3%**	102.8%	102.8%	102.1%	102.5%	102.5%	102.5%	25
	32-8-32	102.8%	103.0%	**102.1%**	102.8%	102.6%	102.2%	103.0%	102.2%	**101.4%**	102.6%	102.7%	102.6%	102.4%	16
	32-16-32	102.6%	102.9%	102.6%	102.9%	102.4%	102.6%	102.7%	102.5%	102.6%	102.5%	102.3%	102.5%	102.2%	8
	32-16-8-16-32	102.4%	102.7%	102.5%	102.8%	102.8%	102.5%	102.6%	102.7%	**102.0%**	102.5%	103.0%	102.5%	101.4%	0

Figure 3. Error metrics (MAE, RMSE, RAE and RRSE) for different ANN configurations with validation framework K-fold cross-validation (K=10)

Table 1 shows that the results obtained for the random forest model were improved for all the error metrics (the lowest errors are in bold). The linear regression model did not perform well because the results obtained are worse than the ones obtained with the random forest model for all the error metrics. Therefore, it was concluded that the model is not appropriate for the problem at hand. Isotonic regression and MLP performed better. In fact, both improved the results obtained with the random forest model for some of the error metrics. Isotonic regression improved all the error metrics. MLP substantially improved the results for the RMSE and RRSE error metrics (the values in bold) and are the best compared to the results obtained from the other models.

For all the error metrics, the ANNs are the models that obtained the best results (values in bold). The ANN2 architecture improved the simple MLP and had the best MAE and RAE errors. This comprised a network structure of only 3 hidden layers with 32-8-32 nodes, SeLU activation function, MAE loss function. It was trained using the Adagrad optimizer and appears to be a very promising configuration in terms of the MAE. The simplicity of this network design makes it very suitable in terms of generalization to other data.

Similarly, when considering the RMSE metrics, the MLP with parameters by default outperformed all the other configurations. Relative errors for the previous two ANN configurations were also very good. The ANN2 model had the best RAE, and the MLP model had the best RRSE. These results confirmed that these are the best ANN designs among the ones evaluated herein. Experts may select ANN or ANN2 depending on the risk they are taking: ANN2 minimizes the absolute error value, while ANN (MLP) obtains the minimum value for the square of the errors, which could be considered as a riskier bidding.

Summarizing, ANNs are very promising models for award price prediction. The quality of the final predictions is very good considering that only 96 ANN designs were tested.

5. Conclusion and Future Work

While the importance of using public datasets to make a more efficient use of public resources is generally acknowledged, the choice of the particular type of ML technique to apply to each problem is not straightforward. For award price prediction in public procurement auctions, it was previously reported that the random forest model is an efficient algorithm. The present paper investigates this question considering a larger set of ML models. Extensive experiments were conducted aiming to predict the award price of Spanish tenders.

The contributions of this study are the following. Using different metrics, it was demonstrated that ANNs and isotonic regression can improve the performance of random forests for the award price estimation of public procurement auctions. Furthermore, the influence of the neural network hyperparameters and gradient optimizers on the performance of the ANN was evaluated in detail and it was concluded that a careful choice of hyperparameters can further improve the predictions of the model.

These experiments used different error metrics, and the performance of different ML paradigms was evaluated. Upon analysing the obtained results, it was concluded that among the methods that are not based on ANNs, isotonic regression is the model that gives the best results. Using its implementation in WEKA, it was corroborated that it is a fast and efficient method for training and testing. However, according to all the error metrics considered, the ANN models can

Table 1. Results compared with the baseline model (random forest)

	MAE	RMSE	RAE	RRSE
Random Forest (baseline model)	179,247.80€	6,621,784.24€	31.11%	74.86%
Linear regression	228,491.36€	15,535,231.61€	39.66%	175.63%
Isotonic regression	136,971.39€	5,648,693.54€	23.76%	63.86%
ANN (MLP)	270,953.50€	**1,974,981.24€**	47.03%	**22.33%**
ANN1	140,763.27€	7,416,004.50€	23.03%	83.84%
ANN2	**123,570.91€**	5,110,687.50€	**20.22%**	57.78%
ANN3	157,181.00€	9,543,883.00€	25.71%	107.90%
ANN4	124,035.82€	3,304,259.20€	20.29%	37.36%

outperform the results from isotonic regression. It was proved that a hyperparameter optimization phase can contribute to improving the predictions made by the ANNs.

There are a number of ways in which this work could be extended. Procurement datasets are updated daily, so we can increase the size of the dataset. An update of the dataset in order to include tender information up to 2021 and a revaluation of the performance of the ML algorithms are planned. On the other hand, three interesting input variables that have not yet been used and that could improve the award price estimator in terms of accuracy were discovered during the analysis. These variables include the price criteria weighing variable and the number of bidders for each tender

and their economic offers. Unfortunately, this information has not been consistently collected in the Spanish public procurement datasets until now. When these values become available, they will be added to the input variables of this study.

Data Availability

The processed data used to support the findings of this study are available from the corresponding author upon request. The raw data from Spain are available from the Ministry of Finance, Spain. Open data of Spanish tenders are hosted in:

https://www.hacienda.gob.es/es-ES/GobiernoAbierto/Datos%20Abiertos/Paginas/licitaciones_plataforma_contratacion.aspx

REFERENCES

Abadi, M. et al. (2016). *TensorFlow: Large-Scale Machine Learning on Heterogeneous Distributed Systems*. Available at: <http://arxiv.org/abs/1603.04467>, last accessed: 20 May, 2021.

Albano, G. L., Cesi, B. & Iozzi, A. (2017). Public procurement with unverifiable quality: The case for discriminatory competitive procedures, *Journal of Public Economics*, *145*, 14-26. DOI: 10.1016/j.jpubeco.2016.11.004

Ballesteros-Pérez, P., Campo-Hitschfeld, M., Mora-Meliá, D. & Domínguez, D. (2015). Modeling bidding competitiveness and position performance in multi-attribute construction auctions, *Operations Research Perspectives*, *2*, 24–35. DOI: 10.1016/j.orp.2015.02.001

Ballesteros-Pérez, P., González-Cruz, M. C. & Cañavate-Grimal, A. (2012). Mathematical relationships between scoring parameters in capped tendering, *International Journal of Project Management*, *30*(7), 850-862. DOI: 10.1016/j.ijproman.2012.01.008

Ballesteros-Pérez, P., González-Cruz, M. C. & Cañavate-Grimal, A. (2013). On competitive bidding: Scoring and position probability graphs, *International Journal of Project Management*, *31*(3), 434-448. DOI: 10.1016/j.ijproman.2012.09.012

Ballesteros-Pérez, P., Skitmore, M., Pellicer, E. & Gutierrez, J. H. (2016). Improving the estimation of probability of bidder participation in procurement auctions, *International Journal of Project Management*, *34*(2), 158-172. DOI: 10.1016/j.ijproman.2015.11.001

Bilal, M. & Oyedele, L. O. (2020). Big Data with deep learning for benchmarking profitability performance

in project tendering, *Expert Systems with Applications*, *147*, 113194. DOI: 10.1016/j.eswa.2020.113194

Breiman, L. (2001). Random forests, *Machine Learning*, *45*(1), 5-32. DOI: 10.1023/A:1010933404324

Bzdok, D., Altman, N. & Krzywinski, M. (2018). Statistics versus machine learning, *Nature Methods*, *15*(4), 233-234. DOI: 10.1038/nmeth.4642

Chakravarti, N. (1989). Isotonic Median Regression: A Linear Programming Approach, *Mathematics of Operations Research*, *14*(2), 303-308. DOI: 10.1287/moor.14.2.303

Chou, J. S., Lin, C.-W., Pham, A.-D. & Shao, J.-Y. (2015). Optimized artificial intelligence models for predicting project award price, *Automation in Construction*, *54*, 106-115. DOI: 10.1016/j.autcon.2015.02.006

Conley, T. G. & Decarolis, F. (2016). Detecting Bidders Groups in Collusive Auctions, *American Economic Journal: Microeconomics*, *8*(2), 1-38. DOI: 10.1257/mic.20130254

Dotoli, M., Epicoco, N. & Falagario, M. (2020). Multi-Criteria Decision Making techniques for the management of public procurement tenders: A case study, *Applied Soft Computing*, *88*, 106064. DOI: 10.1016/j.asoc.2020.106064

Duchi, J. C., Bartlett, P. L. & Wainwright, M. J. (2012). Randomized smoothing for (parallel) stochastic optimization. In *2012 IEEE 51st IEEE Conference on Decision and Control (CDC)* (pp. 5442-5444). IEEE. DOI: 10.1109/CDC.2012.6426698

European Commission (2017). *Public procurement, European semester thematic factsheet*. Available at:

<https://ec.europa.eu/info/sites/info/files/file_import/european-semester_thematic-factsheet_public-procurement_en_0.pdf >, last accessed: 20 May, 2021.

European Commission (2020). *Unleashing the Full Potential of SMEs*, p. 3. DOI: 10.2775/296379

García Rodríguez, M. J., Montequín, V. R., Ortega-Fernández, F. & Balsera, J. (2019a). Public Procurement Announcements in Spain: Regulations, Data Analysis, and Award Price Estimator Using Machine Learning, *Complexity*, *2019*(v), 1-20. DOI: 10.1155/2019/2360610

García Rodríguez, M. J., Montequín, V. R., Ortega-Fernández, F. & Balsera, J. (2019b). Spanish Public Procurement: Legislation, open data source and extracting valuable information of procurement announcements, *Procedia Computer Science*, *164*, 441-448. DOI: 10.1016/j.procs.2019.12.204

García Rodríguez, M. J., Montequín, V. R., Ortega-Fernández, F. & Balsera, J. (2020). Bidders Recommender for Public Procurement Auctions Using Machine Learning: Data Analysis, Algorithm, and Case Study with Tenders from Spain, *Complexity*, *1*, 1-20. DOI: 10.1155/2020/8858258

Gorgun, M. K., Kutlu, M. & Onur Tas, B. K. (2020). Predicting The Number of Bidders in Public Procurement. In *2020 5th International Conference on Computer Science and Engineering (UBMK)*. (pp. 360-365). IEEE. DOI: 10.1109/UBMK50275.2020.9219404

Hall, M. Frank, E., Holmes, G., Bernhard Pfahringer, B., Reutemann, P. & Witten, I. H. (2009). The WEKA data mining software, *ACM SIGKDD Explorations Newsletter*, *11*(1), 10-18. DOI: 10.1145/1656274.1656278

Hastie, T., Tibshirani, R. & Friedman, J. (2009). *The Elements of Statistical Learning*, second edition. New York, NY: Springer New York. Part of the Springer Series in Statistics book series. DOI: 10.1007/978-0-387-84858-7

Huber, M. & Imhof, D. (2019). Machine learning with screens for detecting bid-rigging cartels, *International Journal of Industrial Organization*, *65*, 277-301. DOI: 10.1016/j.ijindorg.2019.04.002

Keras. (2021). *Optimizers*. Available at: <https://keras.io/api/optimizers/>, last accessed: 20 May, 2021.

Kim, J. M. & Jung, H. (2019). Predicting bid prices by using machine learning methods, *Applied Economics*, *51*(19), 2011-2018. DOI: 10.1080/00036846.2018.1537477

Kingma, D. P. & Ba, J. (2014). Adam: A Method for Stochastic Optimization. In *3rd International Conference on Learning Representations, ICLR 2015 - Conference Track Proceedings* (pp. 1–15). Available at: <http://arxiv.org/abs/1412.6980>, last accessed: 20 May, 2021.

Klambauer, G., Unterthiner, T., Mayr, A. & Hochreiter, S. (2017). Self-Normalizing Neural Networks, *Advances in Neural Information Processing Systems*, *2017*(Decem), 972-981. Available at: <http://arxiv.org/abs/1706.02515>, last accessed: 20 May, 2021.

Kuziemski, M. & Misuraca, G. (2020). AI governance in the public sector: Three tales from the frontiers of automated decision-making in democratic settings, *Telecommunications Policy*, *44*(6), 101976. DOI: 10.1016/j.telpol.2020.101976.

Lee, I. & Shin, Y. J. (2020). Machine learning for enterprises: Applications, algorithm selection, and challenges, *Business Horizons*, 63(2), 157-170. DOI: 10.1016/j.bushor.2019.10.005.

Lorentziadis, P. L. (2020). Competitive bidding in asymmetric multidimensional public procurement, *European Journal of Operational Research*, *282*(1), 211-220. DOI: 10.1016/j.ejor.2019.09.005

Mohungoo, I., Brown, I. & Kabanda, S. (2020) A Systematic Review of Implementation Challenges in Public E-Procurement, *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics) (2020) 12067 LNCS*, 46-58. Springer International Publishing. DOI: 10.1007/978-3-030-45002-1_5

Mullainathan, S. & Spiess, J. (2017). Machine Learning: An Applied Econometric Approach, *Journal of Economic Perspectives*, *31*(2), 87-106. DOI: 10.1257/jep.31.2.87

Obwegeser, N. & Müller, S. D. (2018). Innovation and public procurement: Terminology, concepts, and applications, *Technovation*, *74–75*(April 2016), 1-17. DOI: 10.1016/j.technovation.2018.02.015

Onur, I. & Tas, B. K. O. (2019). Optimal bidder participation in public procurement auctions, *International Tax and Public Finance*, *26*(3), 595-617. DOI: 10.1007/s10797-018-9515-2

TensorFlow (2021). *An end-to-end open source machine learning* platform. Available at:<https://www.tensorflow.org>, last accessed: 20 May, 2021.

Waikako University (2021). *Weka 3: Machine Learning Software in Java*. Available at: <http://old-www.cms.waikato.ac.nz/ml/weka/>, last accessed: 20 May, 2021.

Wang, Y. Xi, C., Zhang, S., Yu, D., Zhang, W. & Li, Y. (2014). A combination of extended fuzzy AHP and Fuzzy GRA for government e-tendering in hybrid fuzzy environment, *Scientific World Journal*, *2014*(1), 123675. DOI: 10.1155/2014/123675

Witten, I., Frank, E. & Hall, M. (2011). *Data Mining: Practical Machine Learning Tools and Techniques*. Third. Boston: Elsevier. DOI: 10.1016/C2009-0-19715-5