ISSN 1870-4069

Predicting Airline Customer Satisfaction using *k*-nn Ensemble Regression Models

V. García[1], R. Florencia-Juárez[1], J. P. Sánchez-Solís[1], G. Rivera-Zarate[1], R. Contreras-Masse[2]

[1] Universidad Autónoma de Ciudad Juárez, División Multidisciplinaria en Ciudad Universitaria
Ciudad Juárez, Chihuahua, Mexico
{vicente.jimenez,rogelio.florencia,julia.sanchez,gilberto.rivera}@uacj.mx
[2] Doctorado en Tecnología
Universidad Autónoma de Ciudad Juárez
Ciudad Juárez, Chihuahua, Mexico
rcontreras@itcj.edu.mx

Abstract. Customer satisfaction questionnaires are a rich and strong source of information for companies to seek loyalty, customer and client retention, optimize resources, and repurchase products. Several advanced machine learning and statistical models have been employed to estimate the customer satisfaction score; however, there is not a single model that can yield the best result in all situations. Ensembles of regression techniques have demonstrated their effectiveness for various applications, where the success of these models lies in the construction of a set of single models. We perform an experimental study using a real database of 129,890 samples from airline companies, in order verify the benefits of ensemble models for predicting customer satisfaction. Accordingly, the present paper evaluates the BAGGING ensemble model using the well-renowned *k*-nn algorithm as the base learner. The obtained results indicate that the BAGGING ensemble performs better than the single classifier in terms of RMSE and MAE.

Keywords: regression, customer satisfaction, ensemble, *k*-NN, BAGGING.

1 Introduction

Companies utilize customer satisfaction surveys and questionnaires to find out what a client think about a product, service, brand or the company. Therefore, the customers' satisfaction databases play a crucial role in the productive decision process. Exploiting datasets to get valuable information allows to make an overall marketing strategy that helps the companies retain existing customers and add new customers. Consequently, an effective customer satisfaction data analysis represents a challenge and provide opportunities in several areas as machine learning, data mining, and marketing [1,2,3].

Machine learning methods can predict the customer satisfaction when a service is provided. The research in this problem has been addressed from a classification and regression point of view. Park et al. [2] used four machine learning algorithms (naïve

Bayes, decision tree, logistic regression, and support vector machine) on a dataset acquired from a contact center. The prediction task was addressed as a classification problem, where the best model obtained an accuracy of 66%. Roy et al. [4] employed naïve Bayes, multiclass classifier, k-star, and IBK (Instance-Based learning with parameter K) as classifiers models for predicting customer satisfaction from a database constructed from a customer survey conducted by the San Francisco International Airport. Aktepe et al. [5] showed that using classifier algorithms combined with programming software and structural equation modeling is able to analyze the level of customer satisfaction and loyalty. Farhadloo et al. [6] analyzed the satisfaction of customers using reviews from different states parks in California. These reviews were left by real visitors on TripAdvisor.com. Grigoroudis and Politis [7] suggested that the customer satisfaction problem can be seemed as a multicriteria evaluation problem. Thus, they proposed the MUlticriteria Satisfaction Analysis system (MUSA) that uses ordinal regression techniques. Bockhorst et al. [8] developed a hybrid customer satisfaction system by integrating a linear ranking sub-model and a non-linear isotonic regression. The system was trained on a database constructed by using phone calls from five surveys. Experimental results revealed that the proposed model is better than standard regression techniques. Other algorithms that have also been employed in the customer satisfaction analysis are the CART (Classification And Regression Tree) algorithm [9], artificial neural network approaches [10,11,12], the principal component analysis [13], and the support vector machine algorithm [14].

Although previous studies conclude that data mining and machine learning techniques can be successfully used for prediction of customer satisfaction, there is not an overall best algorithm for dealing with customer satisfaction problems. Consequently, ensembles models have emerged to exploit the different behavior of individual techniques and reduce prediction errors. Several practical investigations have demonstrated that ensemble models perform better than single prediction methods in classification [15,16,17,18,19] and regression problems [20,21,22 23].

The focus of this paper is primarily on exploring the use of ensembles of regression models in the scope of customer satisfaction. In particular, we analyze the performance of the k-nearest neighbor (*k*-nn) model for regression as base classifier in the BAGGING (Bootstrap AGGregatING) ensemble model. This approach is the most common ensemble learning algorithm, where the diversity is achieved by the manipulation of the training samples [24]. BAGGING has showed a good performance on various real problems and provide practical algorithms for constructing ensemble models [16]. Therefore, the paper shows how *k*-nn ensemble regression models can result useful to estimate a customer satisfaction score from a real-database constructed from 129,889 surveys supplied by several airline companies.

Henceforth, the rest of the paper is organized as follows. Section 2 introduces the bases of the ensemble model and regression technique that will be explored in this study. Section 3 provides the experimental set-up and the description of the database used in our experiments. Next, the results are reported and discussed in Section 4. Finally, Section 5 summarizes, the main conclusions and points out some directions for future research.

2 Regression Ensemble Models

An ensemble of regressors consists of a set of individually trained models (the base learners) whose decision is integrated in some way when predicting the output of new examples. By combining individual regression models, the ensemble approaches aiming to minimize the error on problems where the output variable is continuous. The construction process of ensembles is performed by following three steps [25]: 1) generation, 2) pruning, and 3) integration. In the former a set of base learners is constructed; if the base learners are the same, then the construction is homogenous, on the otherwise is heterogeneous. Some redundant models can be generated; therefore, a pruning process can be performed. Finally, the prediction of the models is combined in different ways by fusion or selection.

The set of models is trained by using different subsamples of the training data. A popular resampling method is the bootstrap that, given a set D containing n examples, generates training sets by drawing n examples at random with replacement from D.

2.1 *k*-NN Regression Model

The k-nearest neighbor (k-NN) model is a non-parametric technique that works under the assumption that new samples share similar properties with the set of stored samples. In brief, given a set of n labeled examples (training set), say $D = \{(x_1, a_1), (x_2, a_2), \cdots, (x_n, a_n)\}$ and identically distributed (i.i.d.) random pairs (x_i, a_i), where $x_i \in \mathbb{R}^d$ and a_i denotes the target value associated it, the k-nn classifier consists of assigning a new input sample y to the class most frequently represented among the k closest instances in the training set, according to a certain similarity measure (generally, the Euclidean distance) in the d-dimensional feature space $\mathbb{R}^d$.

The straightforward implementation of the k-nearest neighbor regression model applied to a test sample y first calculates $\delta_i = d(x_i, y)$, for $i = 1, \cdots, n$. Then we get the indices $\{p_1, \cdots, p_k\}$ of the k smallest values such that $\delta_{p_b} \leq \delta_j$, $\forall j \notin \{p_1, \cdots, p_k\}$, $b = 1, \cdots, k$, and $\delta_{p_1} \leq \cdots \leq \delta_{p_k}$. We define x_{p_b} as the p_b −nearest neighbor sample and a_{p_b} is the corresponding target value.

Regression aims to learn a function $f: y \to a$ to predict the a value for a new sample $y = [y_1, y_2, \cdots, y_d]$. The k-nn regression model estimates the target value $f(y)$ of a new input sample y by averaging the estimated target values of its k-nearest neighbors [26]:

$$f(y) = \frac{1}{k} \sum_{b=1}^{k} a_{p_b}, \tag{1}$$

where a_{p_b} denotes the target value of the b-th nearest neighbor.

An illustrative example of the k-nn regression model is displayed in Fig. 1. Using $k = 3$, the output of a new example y is estimated computing the mean among the responses of the 3 nearest neighbors (instances A, B, and C). When $k = 5$, the output of instances A, B, C, D and E are averaged. If $k = 1$, then the k-nn regression model assigns to $f(y)$ the value a_i from x_i, that is the training sample nearest to y.

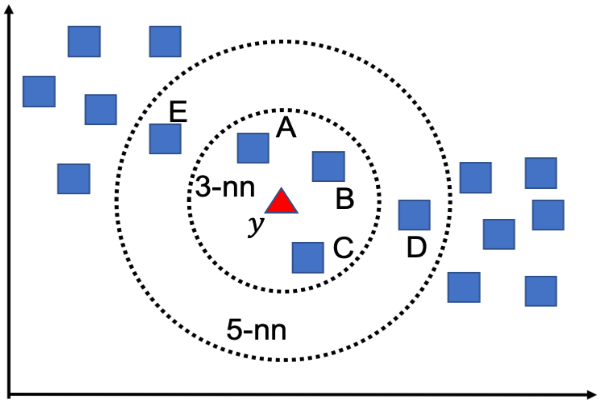

Fig. 1. An example of k-NN regression model with k = 3 and 5.

2.2 BAGGING Ensemble Model

Breiman [26] proposed the BAGGING algorithm that creates multiple models trained from different bootstrap subsamples $D_1, D_2, \cdots, D_M$, each one consisting of n samples drawn at random with replacement from the original D training set. Fig. 2 shows an example of a regression ensemble model based on BAGGING, where in the first step M bootstrap replicates of the training set D are generated. Afterwards, each base regressor R_i is trained on the bootstrap D_i. Thus, the ensemble is conformed by different models, where each one is not exposed to the same set of samples. This creates the diversity necessary to cover a wide range of situations [16]. By this way, the output of new observations will be predicted by taking the average of the ensemble R^* built from $R_1, R_2, \cdots, R_M$. The procedure of training and testing of BAGGING can be resumed in nine steps:

Training step

1. $R^* = \emptyset$, the ensemble
2. M number of regression models to train
3. $For\ i = 1, \cdots, M$
4. Create a boostrap sample of size n, D_i from D
5. Train a regression model R_i using D_i as the training set
6. Add the trained model to the current ensemble, $R^* = R^* \cup R_i$

Testing step

7. Run $R_i, \cdots, R_M$ on the new input y and compute $f_{R_i}(y)$ using Eq. (1)
8. Average the outputs of $R_i, \cdots, R_M$ using

$$f_{R^*}(y) = \frac{1}{M}\sum_{1=1}^{M} f_{R_i}, \quad (2)$$

9. f_{R^*} is the predicted target value.

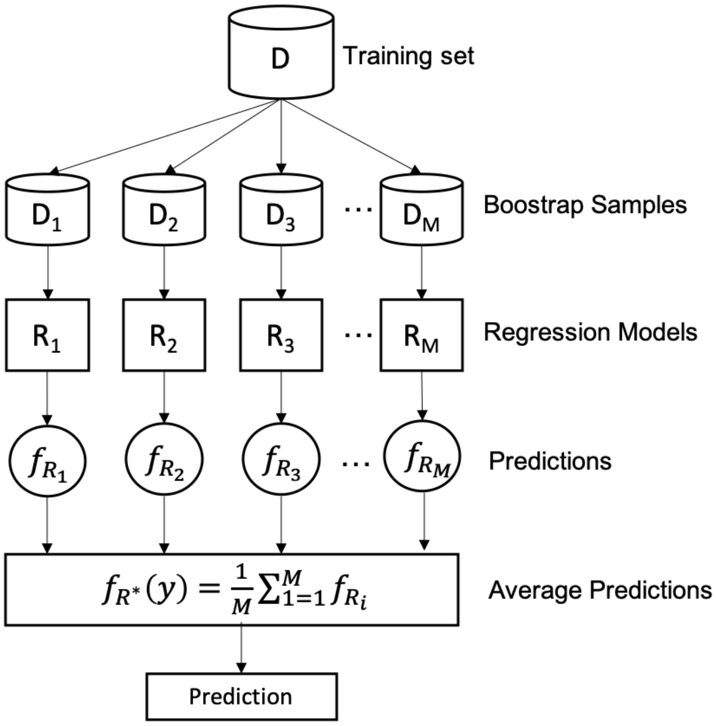

Fig. 2. An ensemble regression model based on BAGGING.

3 Database and Experimental Set-up

As already stated, this study aims to evaluate the performance of regression ensembles for airline customer satisfaction prediction. Thus, we conducted a pool of experiments on a real dataset taken from airline companies in the USA. This is a subset of samples collected on 2014 from customer satisfaction questionnaires. It consists of 129,889 instances, where each one is represented by 24 input variables: airline status, age, gender, price sensitivity, year of first flight, no. of flights, percentage of flights with other airlines, type of travel, no. of other loyalty cards, shopping amount at the airport, eating and drinking at airport, class, day of month, airline code, airline name, origin city, destination, schedule departure hour, departure delay in minutes, arrival delay in minutes, flight cancelled, flight time in minutes, flight distance, and arrival delay greater 5 minutes. The output variable is satisfaction that is a five-point score measurement (i.e., from 1 to 5). Since some variables were categorical, they were first converted into numeric value, and then these values were normalized into the range [0,1].

Table 1 sums up the main characteristics of the database used in the empirical analysis: the attribute number, the attribute description and some statistics, such as the minimum and maximum values of the attribute, the mean and the standard deviation.

We focused our study on the BAGGING ensemble model and the simple *k*-nn algorithm used as the base classifier. The *k*-value used for the regression algorithm was set to 3. Besides, we aim to analyze the performance of the BAGGING ensemble when varying the number of bootstrap subsamples; therefore, six different bootstrap values were tested: 5, 10, 15, 20, 25, and 30. All models were taken from the WEKA toolkit [28].

To prevent inaccurate performance estimates and following the standard strategy used in other works, we evaluate the performance of regression models by a 5-fold cross-validation model [29,30,31]. The original dataset was randomly divided into five disjoint parts. Each fold is used as a test set and the remaining four folds are used for training the regression models. Thus, we can get 5 different test set performances and therefore 5 different trained models. Note that the bootstrap samples are generated for each training set.

To assess the model performance in regression problems we used two of the most popular performance measures that estimate how much the prediction $(e, \cdots, e_n)$ deviate from the actual target values $(a_1, \cdots, a_n)$ [32]. These metrics are the Root Mean Square Error (RMSE) [33,35,36],

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(e_i - a_i)^2}, \tag{3}$$

and the Mean Absolute Error (MAE) [34, 35],

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|e_i - a_i|. \tag{4}$$

Both metrics are minimized when the predicted value for each test sample agrees with their true value. For each fold, we record the $RMSE^{(i)}$ and the $MAE^{(i)}$ (i=1,2, …,5) and compute the final estimate as the mean of all folds:

$$RMSE_{avg} = \frac{1}{5}\sum_{i=1}^{5} RMSE^i, \tag{5}$$

$$MAE_{avg} = \frac{1}{5}\sum_{i=1}^{5} MAE^i. \tag{6}$$

Table 1. Characteristics of the airline customer satisfaction dataset used in the experiments.

No.	Description	Minimum	Maximum	Mean	Std. Dev.
	Input variables				
1	Airline status	1	4	1.747	1.205
2	Age	15	85	46.196	17.321
3	Gender	1	2	1.435	0.496
4	Price sensitivity	0	5	1.276	0.546
5	Year of first flight	2003	2012	2007.209	2.977
6	No. of flights	0	100	20.091	14.362
7	Percentage of flights with other airlines	1	110	9.314	8.761
8	Type of travel	1	3	1.696	0.911
9	No. of other loyalty cards	0	12	0.884	1.142
10	Shopping amount at the airport	0	879	26.553	53.081
11	Eating and drinking at the airport	0	895	68.242	52.21
12	Class	1	3	2.024	0.431
13	Day of month	1	31	15.723	8.659
14	Airline code	1	14	8.134	4.441
15	Airline name	1	14	7.061	4.341
16	Origin city	1	295	128.852	83.059
17	Destination city	1	296	123.289	82.153
18	Scheduled departure hour	1	23	12.896	4.623
19	Departure delay in minutes	0	1592	14.713	38.07
20	Arrival delay in minutes	0	1584	15.045	38.415
21	Flight canceled	1	2	1.018	0.135
22	Flight time in minutes	0	669	109.164	72.8
23	Flight distance	31	4983	793.804	592.125
24	Arrival delay greater than 5 minutes	1	2	1.343	0.475
	Output variable				
25	Satisfaction	1	5	3.379	0.965

4 Results and Discussions

Fig. 3 and 4 display the RMSE and the MAE averaged across the five runs. For each prediction model, we have plotted the results for the base learner (3-nn) and also the results of the 3-nn ensemble regression model when varying the number of bootstrap subsamples from D. Note that the line parallel to X-axis corresponds to the case of base learner, which indicates that the results were achieved by learning directly from the training set D.

From Fig. 3 and 4 as expected, the individual regression model achieves the highest error values ($RMSE \approx 0.7985,\ MAE \approx 0.6365$), whereas, the ensemble regression models appear as the learning model with the lowest overall error ($RMSE \approx 0.7650$ and $MAE \approx 0.5664$ when number or regression models is set to 30). This is a difference between the two models $\approx$ 5% when $RMSE$ is used and $\approx$ 11% in the case of MAE.

On the other hand, when varying the number of bootstraps from D, both the RMSE and the MAE rates decrease as the number of bootstrap subsamples increases, although the configuration of the base classifier does not vary along the different bootstrap subsamples.

From the results here reported, it appears that the customer satisfaction prediction problem can be handled better using ensembles approaches than single models. Likewise, when the number of base models is increased, the error decrease.

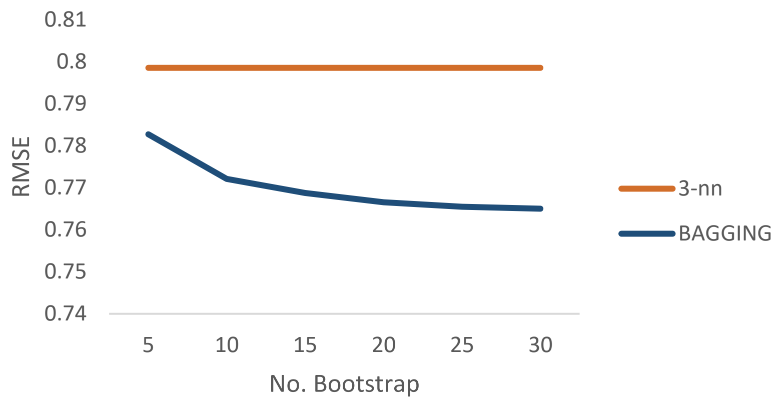

Fig. 3. Average RMSE for the *3*-nn classifier (single learner) and the BAGGING when varying the number of bootstrap subsamples.

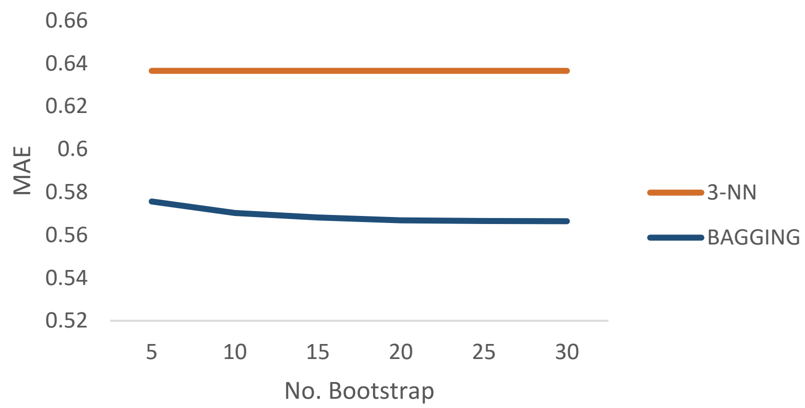

Fig. 4. Average MAE for the *3*-nn classifier (single learner) and the BAGGING when varying the number of bootstrap subsamples.

5 Conclusions and Future Work

The present paper has analyzed the performance of ensemble regression models for an airline customer satisfaction prediction problem. In particular, BAGGING has been taken as representative of ensemble models, whereas the *k*-nn has been employed as the base learner. To this end, a real-life airline customer satisfaction dataset was used to build all the models. From the experiments carried out, we have observed that regarding both RMSE and MAE the ensemble regression models have produced the best results. A final indication of the experiments is that using a significant number of bootstrap subsamples the error may decrease.

Several directions for further research have emerged from this study: (i) to extend the present analysis to other ensemble approaches; (ii) to incorporate a feature selection phase to remove any attribute that might be considered noisy or irrelevant; (iii) to use some prototype selection method to reduce the complexity of data (border and redundant points).

Acknowledgement. The authors would like to acknowledge the financial supports from the Mexican PRODEP [UACJ-PTC-373, UACJ-PTC-423].

References

1. 10 key marketing trends for 2017, https://www-01.ibm.com/common/ssi/cgi-bin/ssialias?htmlfid=WRL12345USEN, last accessed 2018/06/25
2. Park, Y., Gates, S.: Towards real-time measurement of customer satisfaction using automatically generated call transcripts. In: Proceedings of the 18th ACM Conference on Information and Knowledge Management, pp. 1387–296. ACM (2009)
3. Gurau, C., Ranchhod, A.J.: Measuring customer satisfaction: A platform for calculating, predicting and increasing customer profitability. Journal of Targeting, Measurement and Analysis for Marketing 10(3), 203–209 (2002)
4. Roy, S.S., Kaul, D., Barna, C., Mehta, S., Misra, A.: Prediction of customer satisfaction using naïve Bayes, multiclass classifier, k-star, and IBK. In: Balas V., Jain, L., Balas, M. (eds.) Soft computing, SOFA 2016, Advances in Intelligent Systems and Computing, vol 634, 153–161. Springer, cham, Arad, Romania (2016)
5. Aktepe, A., Ersoz, S., Toklu, B.: Customer satisfaction and loyalty analysis with classification algorithms and structural equation modeling. Computers & Industrial Engineering 86, 95–106 (2015)
6. Farhadloo, M., Patternson, R.A., Rolland, E.: Modeling customer satisfaction form unstructured data using a Bayesian approach. Decision Support Systems 90, 1–11 (2016)
7. Grigoroudis, E., Politis, Y.: Multiple criteria approaches for customer satisfaction measurement. In: Matsatsinis, N., Grigoroudis, E. (eds) Preference Disaggregation in Multiple Criteria Decision Analysis. Multiple Criteria Decision Making, pp. 95–123. Springer, cham (2018)
8. Bockhorst, J., Yu, S., Polania, L., Fung, G.: Prediction self-reported customer satisfaction of interactions with a corporate call center. In: Altun Y. et al. (eds) Machine Learning and Knowledge Discovery in Databases, ECML PKDD 2017, Lectures Notes in Computer Science, vol. 10536, pp. 179-190. Springer, Skopje, Macedonia (2017)

9. Dobrota, M., Bulajic, M., Radojicic, Z.: Data mining models for prediction of customer satisfaction: The CART analysis. In: Jaksic, M.L., Rakocevic, S.B. (eds) Innovative Management and Firm Performance, pp. 401-421. Palgrave (2014)
10. Mashinchi, R., Selamat, A., Ibrahim, S., Krejcar, O., Penhaker, M.: Evaluating customer satisfaction: Linguistic Reasoning by Fuzzy Artificial Neural Network. In: Barbucha, D., Nguyen, N., Bartubara, J. (eds), New Trends in Intelligent Information and Database Systems, Studies in Computational Intelligence, vol. 598, 91–100, Springer cham (2015)
11. Yau, H.K., Tang, H.Y.H.: Analyzing customer satisfaction in self-service technology adopted in airports. Journal of Marketing Analytics 6(1), 6–18, (2018)
12. Segura, C., Balcells, D., Umbert, M., Arias, J., Luque, J.: Automatic speech feature learning for continuous prediction of customer satisfaction in contact center. In: Abad, A. et al. (eds), Advances in speech and language technologies for Iberian languages, IberSPEECH 2016. Lecture Notes in Computer Science, vol. 10077, pp. 255–265, Springer, Cham (2016)
13. Hu, B.: Application of data mining in power customer satisfaction evaluation. In: Qian, Z., Cao, L., Su, W., Wang, T., Yang, H. (eds), Recent advances in computer science and information engineering, Lecture Notes in Electrical Engineering, vol. 124, 37-44, Springer, Berling (2015)
14. Jiang, Z., Zan, W., Liu, X.: Customer satisfaction analysis based on SVM. In: Zu, Q., Hu, B. (eds), Human centered computing, HCC 2016. Lecture Notes in Computer Science, vol. 9567, 683–688. Springer, Cham (2016)
15. Rodriguez, J.J., Kuncheva, L. I., Alonso, C.J.: Rotation forest: a new classifier ensemble method. IEEE Transactions on Pattern Analysis and Machine Intelligence 28(10), 1619–1630 (2006)
16. Kuncheva, L. I.: Combining pattern classifiers: Methods and algorithms. 2nd edn. Wiley & Sons, Hoboken (2014)
17. Osareh, A., Shadgar, B.: An efficient ensemble learning method for gene microarray classification. BioMed Research International 210, 1–10 (2013)
18. Rokach, L.: Ensemble-based classifiers. Artificial Intelligence Review 33(1-2), 1–39 (2010)
19. Cabrera-Hernández, L., Morales-Hernández, A., Casas-Cardoso, G.M.: Medidas de diversidad para la construcción de sistemas-multiclasificadores usando algoritmos genéticos. Computación y Sistemas 20(4), 729–747 (2016)
20. Mendes-Moreira, J., Soares, C., Jorge, A.M., De Sousa, J.F.: Ensembles approaches for regression: A Survey. ACM Computing Surveys 45(1), 10:1–10:40 (2012)
21. Frayman, Y., Rolfe, B.F., Webb, G.I.: Solving regression problems using competitive ensemble models. In: LNAI vol. 2557, 511–522 (2002)
22. Sun, Q., Pfahringer, B.: Bagging ensemble selection for regression. In: Thielscher, M., Zhang, D. (eds), AI 2012: Advances in Artificial Intelligence. Lecture Notes in Computer Science, vol 7691, 695–706, Springer, Berlin (2012)
23. Trejo, K., Angulo, C.: Single-camera automatic landmarking for people recognition with an ensemble of regression trees. Computación y Sistemas 20(1), 19–28 (2016)
24. Bonte, I., Rodríguez, A., García, M. M, Grau, R.: Combinación de clasificadores para bioinformática. Computación y Sistemas 16(2), 191–201 (2012)
25. Ren, Y., Zhang, L., Suganthan, P.N.: Ensemble classification and regression – Recent developments, applications and future directions. IEEE Computational Intelligence 11(1), 41–53 (2016)
26. Breiman, L.: Bagging Predictors. Machine Learning 24(2), 62–-635 (1996)
27. Guyader, A., Hengartner, N.: On the mutual nearest neighbors estimate in regression. Journal of Machine Learning Research 14, 2361–2376 (2014)

28. Witten, I., Frank, E., Hall, M., Pal, C.: Practical machine learning tools and techniques. 4th edn., Morgan Kaufmann (2016)
29. Molinaro, A. M., Simon, R., Pfeiffer, R.M.: Prediction error estimation: a comparison of resampling methods. Bioinformatics 21(15), 3301–3307 (2005)
30. Reader, T., Forman, G., Chawla, N. V.: Learning from imbalanced data: evaluation matters. In: Holmes, D.E., Jain, L.C. (eds), Data Mining: Foundations and Intelligent Paradigms. Intelligent System Reference Library, vol. 23., 315–331, Springer, Berling
31. Gacto, M.J., Galende, M., Alcará, R., Herrera, F.: METSK-HD^e: A multiobjective evolutionary algorithm to learn accurate TSK-fuzzy systems in high-dimensional and large-scale regression problems. Information Sciences 276, 63–79 (2014)
32. Caruana, R., Niculescu-Mizil, A.: Data mining in metric space: an empirical analysis of supervised learning performance criteria. In: 10th ACM SIGKDD international conference on knowledge discovery and data mining, pp. 69–78, New York (2004)
33. Madrigal Espinoza, S. G.: Modelos de regression para el pronóstico de series temporales con estacionalidad creciente. Computación y Sistemas 18(4), 821–831 (2014)
34. Luna Sandoval, M. R. O, Ruiz Ascencio, J.: MUREM: Un método multiplicativo de regresion para estimar el esfuerzo de desarrollo de software. Computación y Sistemas 20(4), 763–787 (2016)
35. Béjar Chacón, W. E., Valeriano Valdez, K. Y., Ilachoque Umasi, J. L., Sulla Torres, J.: Predicción de caudales medios diarios en la cuencia del Amazonas aplicando redes neuronales artificiales y el modelo neurodifuso ANFIS. Research in Computing Science 113, 23–25 (2016)
36. Pinzón Pineda, S. A., Hernández Aguilar, J. A., Arroyo-Figueroa, G.: Aplicación de modelos auto regresivos para la predicción de generación de energía eléctrica a partir de datos eólicos. Research in Computing Science 139, 59–70 (2017)