Supplier classification by applying AutoML

Daniela-Lizbeth Hernández-Sánchez
Artificial Intelligence Laboratory,
National Technological, Misantla Campus
Veracruz, Mexico

Eddy Sánchez-DelaCruz
Artificial Intelligence Laboratory
National Technological, Misantla Campus
Veracruz, Mexico
eddsacx@gmail.com (Corresponding author)

Cecilia-Irene Loeza-Mejía
Artificial Intelligence Laboratory
National Technological, Misantla Campus
Veracruz, Mexico

Arnulfo-Gamaliel Hernández-González
Artificial Intelligence Laboratory
National Technological, Misantla Campus
Veracruz, Mexico

***Abstract*—The interest of this research referred to the possibility of detecting patterns to identify the best supplier in a citrus exporter under a classification model that would provide a high degree of confidence. This work aims to verify the validity of Automated Machine Learning (AutoML) employing automatic model selection and hyperparameter optimization method applied in suppliers classification. These results can be used as support for decision-makers to evaluate and select the best of their alternatives. 786 historical records of suppliers were analyzed. The historical records were separated into datasets classified by season (high, low) and production area (north, center, and south). The sampling criteria implemented on the datasets were: cross-validation, percentage split (66%), and representative sample. The classification was evaluated by employing confusion matrix and performance indicators for each dataset according to the sampling criterion chosen. Through AutoML, the following algorithms: Vote, Random Forest, Attribute-Selected-Classifier, Bayes Network, Repeated Incremental Pruning to Produce Error Reduction (RIPPER), and Naive Bayes found the best percentage of sensitivity (89.44, 70.79, 87.50, 88.59, 76.62, 90.13) for each dataset respectively under the highest performance sampling criterion. The AutoML approach helped discriminate Persian lemon suppliers based on their sourcing history, providing support for decision-making under acceptable percentages in reliability indicators.**

***Index Terms*—classification algorithm, machine learning, supervised machine learning**

I. Introduction

Currently, achieving competitiveness in the market is a success factor for any company to maintain or increase profits and meet objectives established in its organizational philosophy (i.e. quality and customer satisfaction). The above, together with a desire to remain competitive, the availability of suppliers, and the different indicators that the decision-maker must evaluate to choose the appropriate suppliers may result in a complex selection process susceptible to errors that are reflected in the responsiveness of the suppliers, quality in the product requested and customer satisfaction [1].

The motivation for this work arises from the interest in addressing the problem of a citrus exporter when identifying the best supplier to address a specific order meeting the parameters established by the exporter, as the supplier selection process becomes complex. Due to the above, the use of computational techniques to convert historical data into useful information for the exporter when classifying its suppliers is proposed.

This work is established within the domain of experimental Machine Learning. Using information as a strategic resource will allow baler decision-makers to discriminate suppliers with a high confidence degree since currently no computational model is implemented in the selection process. In that sense, the most significant contribution lies in the implementation of Machine Learning algorithms in a field not previously studied given the needs of the region's citrus production sector (see Section III-A: *Dataset*). AutoML was implemented under the approach of Thornton *et al.* [2] to help solve the previous problem.

The rest of the article is structured in section II, where the related works are described from a computational perspective. In section III the dataset and the algorithms implemented are described. Section IV shows the results obtained, and finally, in section V, conclusions and future work are presented.

II. State-of-the-art

This section highlights contributions from previous works that address the method of classification algorithms and hyperparameterization and previous work on the supplier selection process that analyzes historical data.

A. Classification and hyperparameterization algorithms

Corso conducted a study to discover the confidence level of supervised rule-based classification algorithms, and to reveal if the confidence level increases or decreases with more training data. The author analyzed datasets with 300, 1420, and 2000 instances employing the Nearest Neighbor with generalization for the first and second datasets, Nearest Neighbor with generalization, and PART decision list for the third. The results showed that the values for the statistics Kappa=1, absolute error=0, and 100% of correctly classified instances are maintained, for all tests. It was found that the number of

978-1-6654-5508-4/22/$31.00 ©2022 IEEE

instances does not significantly influence the application of rule-based classification algorithms [3].

Lindauer *et al.* exposed an approach called AutoFolio to address simultaneously the selection of algorithms and the specification of parameters by the automatic configuration of algorithms. They analyzed the performance of AutoFolio (experiments based on claspfolio2) using the library sklearn and sequential model-based algorithm configuration (SMAC). AutoFolio was executed in 12 algorithm selection scenarios. AutoFolio was shown to significantly improve claspfolio2 throughput in 11 of the 12 scenarios in the Algorithm Selection Library, showing performance improvements of a factor between 1.3 and 12.4 in terms of PAR10 scores compared to the best single solution for the given set of algorithms [4].

Thornton *et al.* mentioned the intention of previous work on the choice of learning algorithms and, independently, in the optimization of hyperparameters, however, the proposal of their research was based on addressing both problems jointly. The approach they developed called Auto-WEKA was applied to the full range of classifiers implemented in Waikato Environment Knowledge Analysis (WEKA), which includes 3 assembled methods, 14 meta-methods, 30 base classifiers, and a wide range of hyperparameter settings for each of these, 10 datasets were used, and better classification performance was shown than that of full cross-validation on the default hyperparameter configuration of the 47 classification algorithms. Each dataset was divided into 70% for training and 30% for testing to evaluate the configurations found by the various optimization methods. Tested hyperparameter configurations based on cross-validation for exhaustive-default (Ex-Def) technique, Hoeffding races, tree-structured Parzen estimator (TPE), and SMAC. In all 10 datasets, TPE and SMAC gave better or the same performance than Hoeffding races and Ex-Def. Among Bayesian optimization methods, SMAC performed better in 8 of 10 datasets. The classifiers selected by each of the methods for each dataset, TPE, and SMAC showed a tendency to choose between the same set of classifiers for each dataset, the most common methods selected were the assembled methods and Random Forest [2].

Thornton *et al.* extended the previous study by considering the combined problem of algorithm selection and hyperparameter optimization (CASH). The Auto-WEKA was evaluated in 21 classification datasets, optimizing the classification error rate, with 100% corresponding to classifiers who did not make correct predictions. The search on the hyperparameters of all the base classifiers yielded better results than Ex-Def in 17/21 cases, where not only the correct algorithm is chosen but also the correct configuration of hyperparameters. The basic methods chosen were shown, AdaBoostM1 frequently chose support vector machine (SVM) in small datasets, while, Random Trees and REP were chosen for large datasets. In the multiclass classifier, Logistic regression and SVM were frequently selected. On the other hand, Logistic regression, as well as the random tree frequently selected by AdaBoostM1. SMAC often found the best hyperparameter configuration compared to TPE (in 10/13 cases) and Iterated F-Race (in 8/13 cases) [5].

Mohr *et al.* presented an algorithm for automated service, demonstrating that automated service composition can overcome the combination of manual software by maximizing classification accuracy. They compared the tool MLS-Plan with other AutoML tools like Auto-WEKA, auto-sklearn, and TPOT (scikit-learn, Python), the experiments were carried out by 20 runs on 21 datasets. The significance of an improvement or degradation was determined using the t-test with a threshold for a score of 2.086. Used 70% for training and 30% for testing, demonstrating that the performance of approaches varies greatly across datasets. TPOT dominated on small datasets, demonstrating that there is no one approach that is the best among most datasets. It was concluded that the combination of automated services is a relevant approach to solving the AutoML problem, so it should be on the list of solution techniques [6].

B. Evaluation and selection of suppliers based on artificial intelligence.

Lambert *et al.* proposed using an artificial neural network (ANN) to address the problem of identifying and selecting reliable producers-suppliers that ensure a supply and quality of fruit in response to the demand of their markets. They considered a diffuse logic (DL) model for the prediction of orchard yield and fruit quality from historical data on crop traceability in an orchard. DL considered the evaluation of the orchard yield and the quality of the fruit from the experience and knowledge of the expert. In addition, employing a genetic algorithm, they optimized the result of a multicriteria mathematical model maximizing the production in orchards and maximizing the quality of the fruit. Finally, they made a comparison between the DL model, the ANN, and the mathematical model to optimize uncertain parameters, where they exhibited previous research, they have made use of various techniques of artificial intelligence, however, among them are not reported works that address the problem of selection of suppliers in that sector [1].

Gerón Fernádez explained the problem of a citrus baler on orders that leave after the estimated time and do not reach the customer on the scheduled date. A vendor selection model was proposed using deep learning and the analytic hierarchy process (AHP) to ensure the supply of fruit. The dataset was built with historical supplier information. The criteria evaluated by AHP were described based on their importance, where each supplier was evaluated on a particular scale for each criterion that was defined by the expert. The hierarchical structure through AHP showed the decision alternatives (suppliers), and the evaluated criteria (gauges) to weight each of the alternatives according to the criteria Saaty scale was used. The author used software programs Microsoft Excel and WEKA. The author implemented Assembled Classifiers and Multilayer Perceptron. The classifier that obtained the lowest error was the Filtered Classifier, which was used for the second experiment with the combination of Filtered Classifier + Multilayer Perceptron. The sampling criteria used were cross-validation with a division of 10 folds and criterion 2/3-1/3. Results showed that the filtered classifier got the smallest error [7].

C. Analysis

Through the previous investigations, it is identified that there are no works related to the classification of Persian lemon suppliers based on their supply history by the automatic model selection and hyperparameter optimization method. Therefore, we have selected the AutoML method proposed in [2] to implement it in this case study. The criteria for selection were to address the problems of simultaneous selection of learning algorithms involving the configuration or adjustment automatically of various parameters [8], [9]. Likewise, the practice of this approach demonstrates the importance of its application in obtaining superior performance compared to the standard selection of algorithms.

III. Materials and methods

The study population consists of historical data on suppliers that caters to the requirement of Persian lemon in a citrus exporter. The classification technique was applied by implementing the Thornton *et al.* method [2] in the machine learning software[1]. The origin of each dataset was defined to implement the classifier (combined selection of classifiers and hyperparameterization), select the attribute (provider) as the target class, and set the sampling criterion. Subsequently, classification techniques were verified using the performance indicators, where data from the classifier and the confusion matrix obtained were analyzed. The sample was constructed employing cross-validation, percentage split, and representative sample [10], [11].

A. Dataset

Historical records made up of attributes collected during the years 2014-2018 from Veracruz and Tabasco states as the main producers of Persian lemon, were divided into files with comma-separated values (CSV) based on production season, considering that, the citrus harvest occurs in two chronological times during the year, high season production and low season production. Then, the records were grouped by production area (north, center, and south; considering that the citrus production of the State of Veracruz, Mexico, is divided into three regions: north, center, and south. The latter region also includes a supplier from Tabasco, Mexico. The historical records resulted in the generation of six datasets (high_north, high_center, high_south, low_north, low_center, and low_south) which correspond to the three production areas in the two seasons. Each dataset consists of supply records of the citrus exporter in which the relevant attributes are identified taking into account the expert's judgment supporting its decision in the historical data of supply. The inclusion criteria in the case study correspond to the attributes: year (year in which production was obtained), quality (categorization of the fruit based on the characteristic calibre i.e the size of a lemon), specification (categorization of fruit based on preselection characteristics), month (month of the year in which the production was obtained), customer (entities that source the product), supplier (23 entities supplying the company), expertise (weighting obtained after applying the AHP methodology).

[1]It was employed Auto-WEKA which is available as a WEKA package through the WEKA package manager. The Auto-WEKA source code is available at https://github.com/automl/autoweka

B. Automatic model selection and hyperparameter optimization method

The parameter optimization method is a model that focuses on a classification problem, where a learning algorithm maps a set of points as training data. In most learning algorithms, hyperparameters are exposed which changes the way the algorithm works on its own [12].

- Model selection: Given a set of learning algorithms A and training data $\mathcal{D} = \{(\mathrm{x}_1, y_1), \dots, (\mathrm{x}_n, y_n)\}$, the selection of the model aims to determine the algorithm $A^* \in A$ with optimal performance. Performance is estimated by dividing D into separate training datasets $\mathcal{D}^{(i)}_{train}$ and validation $\mathcal{D}^{(i)}_{valid}$, learning functions f_i through the implementation of A^* to $\mathcal{D}^{(i)}_{train}$, and evaluating the predictive performance of those functions on $\mathcal{D}^{(i)}_{valid}$. Cross-validation is used in *K-fold*, which divides training data into k partitions of equal size $\mathcal{D}^{(1)}_{valid}, \dots, \mathcal{D}^{(k)}_{valid}$, and establishes $\mathcal{D}^{(i)}_{train} = \mathcal{D} \diagup \mathcal{D}^{(i)}_{valid}$ for $i = 1, \dots, k$. This allows for the learning algorithm selection problem to be written as:
$$A^* \in \underset{A \in \mathcal{A}}{\text{argmin}} \ \frac{1}{k} \cdot \sum_{i=1}^{k} \mathcal{L} \ (A, \mathcal{D}^{(i)}_{train}, \mathcal{D}^{(i)}_{valid}),$$
where $\mathcal{L} \ (A, \mathcal{D}^{(i)}_{train}, \mathcal{D}^{(i)}_{valid})$ is the loss achieved by A when trained on $\mathcal{D}^{(i)}_{train}$ and evaluated on $\mathcal{D}^{(i)}_{valid}$. For classification problems, the loss is typically defined as the rate at which the predictions have different labels than the validation data, whereas for regression problems the loss is often expressed as the root mean squared error (RMSE).
- Combined algorithm selection and hyperparameter optimization: Is modeled the problem of selecting one of k learning methods $A_1, \dots, A_k$ with associated hyperparameter spaces $\Lambda^{(1)}, \dots, \Lambda^{(k)}$ as a single combined hyperparameter optimization problem with algorithm A and parameter space Λ. This combined problem features the union of the parameters and their domains in $\Lambda^{(1)}, \dots, \Lambda^{(k)}$, plus a new root-level hyperparameter $\lambda_r \in \{A_1, \dots, A_k\}$ that selects between the k methods. The root-level parameters of each subspace Λ^i are made conditional on λ_r being instantiated to A_i. The goal of hyperparameter optimization is to determine the hyperparameters λ^* optimizing generalization performance of $\mathcal{A}_{\lambda^*}$ based on a limited amount of training data $\mathcal{D} = \{(\mathrm{x}_1, y_1), \dots, (\mathrm{x}_n, y_n)\}$.

Finally, Fig. 1. graphically represents the supplier classification process considering the datasets grouped by production season and production area. Datasets were divided by different sampling criteria and used as input to the AutoML model where the process of selecting and configuring algorithms according to input data is carried out automatically and iteratively. The model chosen by AutoML is evaluated by different metrics until a maximum execution time limit is reached, in this case of study, the maximum defined time for training was 15 minutes.

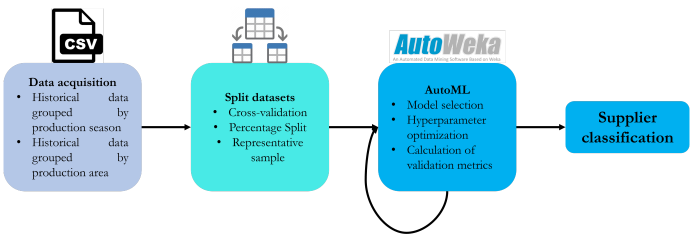

Fig. 1. Supplier classification Pipeline

IV. Results

The experiments were carried out on a personal computer on the platform Windows 8.1 PRO, Intel ®Core^{TM} processor i3 CPU M380 a @2.53 GHz, 4 GB of RAM, and operative system of 64 bits. The platform WEKA stable version 3.8.3 [13] was employed to implement the automatic model selection and hyperparameter optimization method for each dataset under the parameters set according to the sampling criterion (cross-validation, percentage split=66%, and representative sample).

Table I shows for each dataset the best classifier found and the overall performance of the rating performance indicators. The metrics calculated were sensitivity or the true positive rate (TP rate), false positive rate (FP rate), precision, F-Measure, Matthew's correlation coefficient (MCC), ROC area [14], PRC area, and Kappa statistic. Reliability percentage in terms of sensitivity for the sampling criterion where the highest performance was obtained. According to the percentage of sensitivity in each dataset can be determined that for the dataset high_north the highest percentage corresponds to the criterion representative sample. On the other hand, for the datasets, high_center, high_south, and low_center the highest percentages correspond to the criterion cross-validation. Whereas, for the datasets low_north and low_south the percentage split sampling criterion (66%) obtains the highest reliability percentage. Overall it can be observed that the classification is good considering that, in most cases, the values of TP rate, ROC area, and Kappa statistic are superior to the 80%.

Table II shows the confusion matrix using representative sample for dataset high_north. In the confusion matrix, it is noted that of the class "a" four instances are confused as a class "b" and an instance is confused as a class "c". Of the class "b", five instances confuse them as "a". While of the class "c" eight confuse them as "b" and one as "a".

Table III shows validation metrics for each of the classes which allows knowing the effectiveness of the classifier. From these statistics, the maximum value of the validation metric is taken as the reference TP rate of the class considered. In this case, for the dataset high_north, the maximum value of the TP rate establishes S-2 as the best class, this information matches the analysis of the confusion matrix generated for the same dataset (Table II) where it is also inferred to S-2 as the best class, obtaining the maximum value of correctly classified instances in the diagonal (56 instances).

Based on the experiments and the analysis of metrics of the confusion matrix generated for each dataset under the criterion of highest performance sampling, the class with the highest number of correctly classified instances was determined. Table IV shows the maximum value found on the main diagonal of the confusion matrix and the maximum value of TP rate for each dataset, in the same way, the name of the class with the highest performance in the classification is shown for both cases. Also, it is showed the supporting data to generate the ROC area, in which cut-off points are established for each dataset taking into account the best-classified class. The values of the ROC area depict a good overall performance in the classification process, with a value close to 1, this indicates the quality that exists in the classifier.

Below are the indicators of the degree of confidence (TP rate) that exists for the classes best classified in the two study seasons. Reliability percentages concentrate on suppliers with the best classification in the high season and low season are shown in Fig. 2 and Fig. 3, respectively.

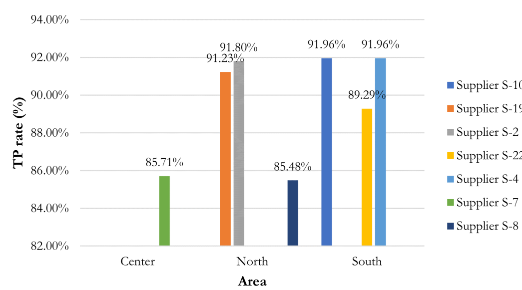

Fig. 2. Suppliers in high season

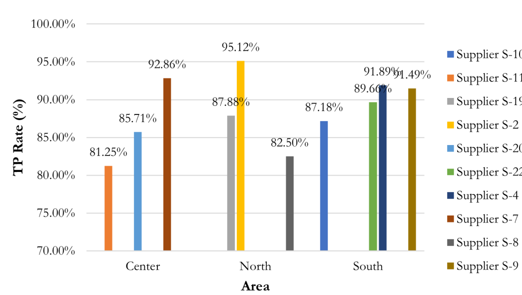

Fig. 3. Suppliers in low season

TABLE I
COMPARISON OF PERFORMANCE IN CLASSIFICATION

Dataset	Number of classes	Best classifier algorithm	Sampling criterion	TP rate	FP rate	Precision	F-Measure	MCC	ROC area	PRC area	Kappa statistic
high_north	3	Vote	Representative sample split: 46.3%	0.8944	0.053	0.901	0.895	0.845	0.977	0.947	0.8417
high_center	17	Random Forest	Cross-validation 10 folds	0.7079	0.018	0.763	0.722	0.711	0.899	0.689	0.6897
high_south	4	Attribute-Selected-Classifier	Cross-validation 10 folds	0.875	0.042	0.881	0.875	0.836	0.968	0.934	0.8333
low_north	3	Bayes Network	Percentage split: 66%	0.8859	0.052	0.895	0.887	0.836	0.967	0.927	0.8289
low_center	16	RIPPER	Cross-validation 10 folds	0.766	0.016	0.799	0.773	0.763	0.959	0.835	0.7506
low_south	4	Naive Bayes	Percentage split: 66%	0.9013	0.040	0.906	0.902	0.866	0.991	0.970	0.8666

TABLE II
CONFUSION MATRIX FOR DATASET HIGH_NORTH

a	b	c	classified as
52	4	1	a = S-19
5	56	0	b = S-2
1	8	53	c = S-8

TABLE III
DETAILED METRICS BY CLASS

	TP rate	FP rate	Precision	F-Measure	MCC	ROC Area	PRC Area	Class
	0.912	0.049	0.897	0.904	0.860	0.976	0.951	S-19
	0.918	0.101	0.824	0.868	0.798	0.980	0.945	S-2
	0.855	**0.008**	**0.981**	**0.914**	**0.878**	**0.991**	**0.975**	S-8
Weighted avg.	0.894	0.053	0.901	0.895	0.845	0.982	0.957	

In general, Fig. 2 and Fig. 3 allow for visualizing the performance of the best-ranked suppliers in the high and low seasons, in this way it is possible to determine that the best performance indicators are concentrated in the low season and correspond to the classes S-2 and S-7 with 95.12% and 92.86% TP rates respectively, this means that for such suppliers in the low season the supply capacity is higher than in the high season.

TABLE IV
COMPARED METRICS FOR THE BEST CLASS

Dataset	TP rate	FP rate	ROC area	Class	Classifies instances correctly
high_north	0.918	0.101	0.980	S-2	56
high_center	0.857	0.013	0.944	S-7	24
high_south	0.920	0.080	0.974	S-10	103
	0.920	0.039	0.988	S-4	103
low_north	0.951	0.041	0.974	S-2	39
low_center	0.929	0.054	0.987	S-7	104
low_south	0.919	0.026	0.992	S-4	43

A. *Discussion*

Through the experiments, it was found that it is possible to discriminate Persian lemon suppliers based on their supply history by automatic model selection and hyperparameter optimization method with high confidence. This method was implemented to perform the experiments by combining the results of the individual classifiers covering the full range of classifiers. The sampling criteria used in the study were cross-validation, percentage split (66%), and representative sample. The results varied between 70.70% and 90.13% in terms of sensitivity, while in the Kappa statistic metric the results varied between 0.6897 and 0.8666. In both metrics, the worst performance was obtained in the high_center dataset containing 17 classes using the cross-validation 10 folds, and the algorithm selected by AutoML was Random Forests, while the best performance was achieved in the low_south dataset containing 4 classes achieving better performance in the low_south dataset using the percentage split criterion and the algorithm selected by AutoML was Naive Bayes. The results are related to the study conducted in [1] where algorithms are applied to select suppliers based on their supply history in which classification models are addressed, but the work referred to differs from the present study in that it proposes the paradigm of combination of classifiers against individual classification models.

V. CONCLUSIONS AND FUTURE WORK

The results of the classification allowed finding the best classifying algorithm for each dataset in this study, determining the best-performing sampling criterion, discriminating classes, obtaining confidence indicators for classes in the two study seasons (high and low), and identifying the supplier with the best classification in a given supply area (north, center, and south). With the above, it can be inferred that for the dataset: high_north, the highest ranked class is S-2. While for the dataset high_center the highest ranked class is S-7. On the other hand, the highest performing classes for the dataset high_south are S-10 and S-4. While, in the dataset low_north, S-2 is the highest ranked class. In the dataset low_center, S-7 is the best classified. Finally, for the dataset low_south, the best class is S-9. Based on the percentage of sensitivity for suppliers in high season it was determined that the classes S-10 and S-4 corresponding to the southern area, have the best performance with a sensitivity percentage of 91.96%, and in low season, the class S-2 which corresponds to the northern zone presents the best result with a sensitivity percentage of 95.12%. The sensitivity indicators employing the automatic model selection and hyperparameter optimization method are acceptable and can serve as decision support for the best supplier. The following are considered for future directions:

- Develop a highly trusted software tool that allows supplier classification.
- Perform classification by selecting attributes by exploring different AutoML algorithms than the one proposed in this work.
- Identify other case studies related to supplier historical datasets, to study the behavior of the process performed in another scenario.

REFERENCES

[1] G. F. Lambert, A. A. A. Lasserre, B. O. I. Rivera, and N. Montalvo, "Una propuesta basada en técnicas de ia para seleccionar proveedores en una exportadora de cítricos."

[2] C. Thornton, F. Hutter, H. H. Hoos, K. Leyton-Brown *et al.*, "Auto-weka: Automated selection and hyper-parameter optimization of classification algorithms," *CoRR, abs/1208.3719*, 2012.

[3] C. L. Corso, "Aplicación de algoritmos de clasificación supervisada usando weka," *Córdoba: Universidad Tecnológica Nacional, Facultad Regional Córdoba*, 2009.

[4] M. Lindauer, H. H. Hoos, F. Hutter, and T. Schaub, "Autofolio: Algorithm configuration for algorithm selection," in *Workshops at the Twenty-Ninth AAAI Conference on Artificial Intelligence*, 2015.

[5] C. Thornton, F. Hutter, H. H. Hoos, and K. Leyton-Brown, "Auto-weka: Combined selection and hyperparameter optimization of classification algorithms," in *Proceedings of the 19th ACM SIGKDD international conference on Knowledge discovery and data mining*, 2013, pp. 847–855.

[6] F. Mohr, M. Wever, and E. Hüllermeier, "Automated machine learning service composition," *arXiv preprint arXiv:1809.00486*, 2018.

[7] M. Y. Gerón Fernéndez, "Evaluación de proveedores para la empacadora limones Mónica S.A. de C.V., basado en la técnica deep learning," Master's thesis, Instituto Tecnológico Superior de Misantla, 2018.

[8] L. Kotthoff, C. Thornton, H. H. Hoos, F. Hutter, and K. Leyton-Brown, "Auto-weka: Automatic model selection and hyperparameter optimization in weka," in *Automated machine learning*. Springer, Cham, 2019, pp. 81–95.

[9] L. Kotthoff, C. Thornton, and F. Hutter, "User guide for auto-weka version 2.6," *Dept. Comput. Sci., Univ. British Columbia, BETA lab, Vancouver, BC, Canada, Tech. Rep*, vol. 2, pp. 1–15, 2017.

[10] S. Aguilar-Barojas, "Fórmulas para el cálculo de la muestra en investigaciones de salud," *Salud en tabasco*, vol. 11, no. 1-2, pp. 333–338, 2005.

[11] E. Sánchez-DelaCruz, R. Weber, R. Biswal, J. Mejía, G. Hernández-Chan, and H. Gómez-Pozos, "Gait biomarkers classification by combining assembled algorithms and deep learning: results of a local study," *Computational and mathematical methods in medicine*, vol. 2019, 2019.

[12] C. Thornton, "Auto-weka: combined selection and hyperparameter optimization of supervised machine learning algorithms," Ph.D. dissertation, University of British Columbia, 2014.

[13] Weka Development Team, "Waikato environment knowledge analysis." [Online]. Available: https://www.cs.waikato.ac.nz/ml/weka/

[14] A. Rodrigo, A. Penas, and F. Verdejo, "Comparación de enfoques para evaluar la validación de respuestas," *Procesamiento del lenguaje natural*, vol. 43, pp. 277–285, 2009.