Volume 59, January-December 2025, ISSN: 2079-3480

ANIMAL SCIENCE

CUBAN JOURNAL OF AGRICULTURAL SCIENCE

Journal site: https://cjascience.com Cu-ID: https://cu-id.com/1996/v59e01

CLASSIFICATION OF SMALL-SCALE DAIRY PRODUCTION IN THE ECUADOR-COLOMBIA BORDER AREA. A COMPARATIVE STUDY OF AUTOMATIC LEARNING TECHNIQUES

CLASIFICACIÓN DE LA PRODUCCIÓN LECHERA A PEQUEÑA ESCALA EN LA ZONA FRONTERIZA ECUADOR-COLOMBIA. UN ESTUDIO COMPARATIVO DE TÉCNICAS DE APRENDIZAJE AUTOMÁTICO

L. CARVAJAL-PÉREZ[1]*, F. MONTENEGRO-ARELLANO[1], G. TERÁN-ROSERO[1], GLADYS URGILÉS-URGILÉS[1], NAYELI CHULDE-CHULDE[1], R. COBO-CUÑA[2], MAGALY HERRERA-VILLAFRANCA[2]

[1]*Universidad Politécnica Estatal del Carchi, Antisana y Universitaria, Tulcán, Ecuador*
[2]*Instituto de Ciencia Animal, C. Central, km 47 ½, San José de las Lajas, Mayabeque, Cuba*

**Email: luis.carvajal@upec.edu.ec*

The socioeconomic factors determining production in dairy farms were researched. The classification of small-scale farmers in the border area between Ecuador and Colombia was involved. A total of 532 farmers participated in the survey and the data collected was analyzed using automatic learning techniques. The data were subjected to an exhaustive preprocessing to remove errors and outliers related to socioeconomic factors in milk production in Carchi, Ecuador. Among the variables examined, economic income, the price per liter of milk and the quantity of liters used for cheese production emerged as the most influential factors. The results showed that automatic learning techniques can effectively classify small-scale dairy production, with accuracy above 96 %. The presence of a child who provides economic support to the house, the allocation of milk for the production and sale of cheese, together with its use for family consumption, significantly influenced 90 % of the surveyed participants.

Key words: *classification models, dairy productivity, economic well-being, small dairy farmers*

Se investigaron los factores socioeconómicos determinantes en la producción en granjas lecheras. Se involucró la clasificación de los productores a pequeña escala en la zona fronteriza entre Ecuador y Colombia. Un total de 532 agricultores participaron en la encuesta y los datos recopilados se analizaron mediante técnicas de aprendizaje automático. Los datos se sometieron a un preprocesamiento exhaustivo para eliminar errores y valores atípicos relacionados con los factores socioeconómicos en la producción de leche del Carchi, Ecuador. Entre las variables examinadas, el ingreso económico, el precio por litro de leche y la cantidad de litros utilizados para la producción de queso surgieron como los factores más influyentes. Los resultados mostraron que las técnicas de aprendizaje automático pueden clasificar eficazmente la producción láctea a pequeña escala, con precisión superior a 96 %. La presencia de un hijo que proporciona apoyo económico al hogar, la asignación de leche para la producción como para la venta de queso, junto con su utilización para el consumo familiar, influyeron significativamente en 90 % de los participantes encuestados.

Palabras clave: *bienestar económico, modelos de clasificación, pequeños productores lecheros, productividad lechera*

Received: August 10, 2024
Accepted: November 21, 2024

Conflict of interest: The authors declare that there is no conflict of interest between them.

CRediT Authorship Contribution Statement: L. Carvajal-Pérez: **Conceptualization, Investigation, Formal analysis, Writing- original draft.** F. Montenegro-Arellano: **Conceptualization, Investigation.** G. Terán-Rosero: **Methodology, Formal analysis.** Gladys Urgilés-Urgilés: **Funding acquisition, Resources.** Nayeli Chulde-Chulde: **Data curation.** R. Cobo-Cuña: **Validation.** Magaly Herrera-Villafranca: **Formal analysis, Writing- original draft**

This is an open access article distributed under the terms of the Creative Commons Attribution-NonCommercial (CC BY-NC 4.0). https://creativecommons.org/licenses/by-nc/4.0/

Introduction

Milk production is an important economic activity in the world. By 2023, milk production exceeded 950 million tons. In emerging economies, approximately 80 % of production comes from family farms with limited use of inputs, which translates into lower yields per animal. The 20 % of farms are medium and large, of which 4 % invest in technology to fulfill quality standards (FAO 2023a).

In 2022, the European Union (made up of 27 countries) was the world's largest producer with 144 million tons. It was followed by the United States with 103 million tons and India with 97 million tons (Orús 2022). In Ecuador, approximately 6.15 million liters of milk were produced per day, which generated income for 1.3 million inhabitants (Ionita 2022). Milk production contributes 4 % to the country's agro-industrial gross domestic product and shows growth of 10.92 % compared to 2020. The Sierra region contributes 73 % of production, the Coast 19 %, and the Amazonian 8 % (CIL Ecuador, 2023).

Milk production uses production factors including land, capital, labor, technology and, according to some authors, business management to transform them and contribute to improving the living conditions of farmers.

The social factors with the greatest impact are gender, level of education, training, experience or associativity (Zemarku *et al.* 2022). Likewise, economic factors such as income, costs, herd size, and production volume were identified (Vásquez *et al.* 2022); in addition, the availability of land, foods, and veterinary care is essential in the production process (Peña *et al.* 2018), without neglecting innovations in the rearing system and the use of automation equipment for quality production (Tangorra *et al.* 2022).

The dairy sector allows rural populations to produce and market their products, contributing to local economic development, food security, economic development and therefore a better quality of life for farmers (FAO 2022a). It is a sector that is always changing. It needs to invest in new technology to be efficient. This harms small farmers, who cannot afford to invest (Gil and Hernández 2019). In addition, the dairy value chain promotes small, micro and medium farmers by helping them process and sell dairy products (Gaudin and Padilla 2020).

The study area includes the Carchi province, located in northern Ecuador, on the border with Colombia. The 63 % of the territory is in the humid temperate zone. It is between 1,800 and 3,000 m o. s. l and between 12 and 18 °C. The temperature depends on if the weather is dry or rainy (Franco 2016). The other 37 % is in the subtemperate region, which is very humid. It is in the low moors, between 3,000 and 4,000 m o. s. l. The temperature is 6 to 12 °C. The rainfalls are from 1000 to 1500 mm per year, with no month of maximum rainfall (Requelme and Bonifaz 2012).

Carchi's dairy production ranks third in national production. It is based on families, has a strong presence in the informal market (Morocho *et al.* 2021), employs 36 % of the population (Terán and Cobo 2017). There are 8,957 livestock farms (Prefectura del Carchi 2023).

The main system is extensive, with traditional practices and the presence of a lot of native cattle. The cows produce an average of 9.4 L per day. This is higher than the national average of 5.9 L (Carvajal 2014). Farms with Holstein cattle achieve yields of 15 to 18 L per cow per day (Balarezo *et al.* 2016), but they are only 6 % of the total.

Agricultural production units (APU) have small milking facilities or stables, which reflects their limited economic capacity (Velasteguí 2019). In terms of land area, there is a large difference between farmer groups. Small farmers have an average of 3 ha. Medium farmers have 7 ha. Large farmers have 120 ha (Requelme and Bonifaz 2012).

The average age of producers is 50 years old. This shows few young people and little generational change (Moreno 2018). In terms of education, 60 % of farmers have primary education, 25 % have secondary education and 15 % have university education. The production chain is not competitive, harms production and limits the agricultural sector in the region.

Several tools are used around the world to evaluate socio-economic factors (SEF) and analyze strategies for sustainable agricultural and food development (FAO 2018). Today, the implementation of inclusive and sustainable artificial intelligence (AI) practices in agriculture provides solutions to achieve food and nutritional security. The AI is applied in agricultural robotics, soil and crop monitoring, as well as predictive analysis (FAO 2022b).

Machine Learning (ML) is the field of study known as a scientific method or art, where computers can learn from data through programming (Valdez 2019 and Kassahun *et al.* 2022). The data used for learning are called samples and are part of the training set. The part of the ML system that learns and makes predictions is called a model, which is commonly tested using the test set (Gaurav and Patel 2020 and Slob *et al.* 2021). Automatic learning is good, for example, in problems that require many rules, fluctuating environments, and in problems that require discovering insights in large amounts of data.

Géron (2019) proposes three main ML systems: those that are supervised during training, those that can learn incrementally on the course, and those that allow comparing new data points with known data points. Automatic learning systems can classify data based on the training data used to learn the model. This opens up several categories, but this study is driven by supervised learning, which requires the solutions in the training data, commonly called labels. An example of this learning is the classification of *spam* emails (Valdez 2019).

For Alwadi *et al.* (2024), the gradient boosting classifier (GBC) uses large data sets to develop models that forecast production and find relevant patterns. This method, used in a study in Jordan, where sensors were used to track 4,000 cows, showed great potential for increasing productivity. Similarly, Bai *et al.* (2022) showed that GBDT-AdaBoost achieved an average recognition accuracy of 98.0 %, exceeding other models such as the random forest and extremely random tree, which had accuracies of 79.9 % and 71.1 %, respectively.

Bovo *et al.* (2021) showed a random forest (RF) classifier with an average prediction error of 18 % for daily milk production of each cow, and only 2 % for total production. This shows that the random forest classifier is effective in calibrating models that help improve sustainability and efficiency in dairy livestock.

Piwczyński *et al.* (2020) used a decision tree (DT) classifier to identify factors that influence on high monthly milk production in Holstein-Friesian cows in 27 herds with milking robots. The results showed that the highest monthly production (47.24 kg) was recorded in multiparous cows, milked more than three times a day, in stables with deep bedding. In contrast, the lowest production (13.56 kg) was observed in cows milked less than twice a day, with an average of less than 3.97 quarters milked. This model allows breeders to fit factors to maximize milk production.

Finally, Fadillah *et al.* (2023) in a study with Indonesian dairy farmers on milk quality and factors associated with total plate count (TPC) and somatic cell count (SCC). Multinomial regression models and Firth-type logistic regression were used to identify factors related to the knowledge of TPC and SCC. They identified as significant variables belonging to cooperatives, distance from neighboring farmers and the adoption of technology to increase awareness about milk quality among small farmers. In general, such results provide evidence that these are models applicable to any region and facilitate decision-making based on results with effective measurements.

This research compared four different automatic learning techniques: gradient boosting classifier (GBC), random forest classifier (RF), decision tree classifier (DT), and logistic regression (LR). The results showed that GBC and RF were the most effective automatic learning techniques for classifying milk production.

Methodology

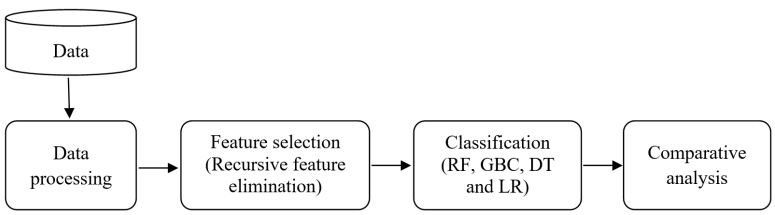
This study involves an experimental analysis consisting of four phases: data preprocessing, feature selection, classification, and comparative analysis of the classifiers. The workflow of the proposed methodology is shown in figure 1, which illustrates the relations between the different phases and the application of specific algorithms at each stage.

Data collection: The population of small and medium dairy farmers from Carchi province was surveyed, totaling 532 individuals. An applied research approach was used with an exploratory and correlational methodology (Hernández-Sampieri and Mendoza 2018). The questionnaire deal with a variety of factors, providing information on relevant aspects to the dairy farming community:

- **Social**: age, gender, educational level, family structure, training, access to technology, housing conditions, basic services, employment, associativity, governance and participation, government technical support
- **Economic**: livestock incomes, other incomes, production costs, income distribution, financing, marketing, farm size.
- **Productive**: land use, herd size and structure, number of heads of cattle, grasses, milk production per hectare (L ha^{-1}), adoption of technology and productive diversification. number of heads of cattle.

A total of 17 questions with quantitative information, 23 interval questions and 10 dichotomous questions were incorporated. The questionnaire was rigorously developed and its content and structure were validated. Field data collection was carried out in collaboration with Business Administration students from the Universidad Politécnica Estatal del Carchi (UPEC), Ecuador, during the second semester of 2022. Simple random sampling was applied.

Data preprocessing: The collected data were subjected to a rigorous preprocessing process, which included the removal of errors and outliers, as well as the treatment of missing values. Min-Max normalization was applied to ensure that all features had a common range and were comparable to each other (Treviño Cantú 2022). This allowed eliminating any bias due to the data scale, ensuring a more accurate and fairing analysis.



Source: Own elaboration

Figure 1. Workflow for predicting small-scale dairy production

Feature Selection: Function selection plays an important role in the data preprocessing phase before applying automatic learning techniques (Siddiqui and Amer 2024). It involves selecting the most relevant and informative features from the data set, while discarding irrelevant or redundant features. In this study, feature selection was used to improve the yield and interpretability of automatic learning models to classify small-scale dairy farmers in the border region between Ecuador and Colombia.

The dataset used in this research contains several socioeconomic and production-related variables that could potentially influence on milk production. However, not all of these variables are equally important for the prediction task. Some features may introduce noise, increase computational upload, or cause an overfitting, which make difficult the model's ability to generalize well unseen data.

To deal with these challenges and identify the most influential features, recursive feature elimination (RFE) technique was used. It is a popular and powerful feature selection method that works by recursively fitting the automatic learning model, removing the least significant features in each iteration. The process continues until the desired number of features is obtained. The importance of RFE lies in its ability to rank features based on their contribution to the model yield, allowing to focus on the most relevant attributes and discard the less informative ones (Mannepalli *et al*. 2024).

The initial database consisted of 134 items, including numerical, dichotomous and categorical variables. In order to reduce the dimensionality of the data and the computational cost during model training, feature selection was applied and finally the set was reduced to 10 variables. The type of house, access to drinking water and electricity, marketing of raw milk, sales of pasteurized cheese, use of milk for cheese production, customer relations, total annual income from primary activity, liters used for cheese production and price per liter were included.

Classification algorithm

Gradient Boosting Classifier (GBC): Is a classifier that highlights for its accuracy and prediction speed on large and complex data sets. It also minimizes the bias error of the model (Bentéjac *et al.* 2020). This method is used when there are only two classes in the target features, i.e. binary classes (positive and negative). The loss function as log-likelihood is used in the creation (training) of the model (Natekin and Knoll 2013). This loss is shown in equation (1):

$$L(\theta) = -\sum y_i \log(p(y_i|x_i;\theta)) \quad (1)$$

where y_i is the classification target, p is the predicted probability of class 1, and θ is the input.

The loss function finds the residuals after creating the decision tree with all the independent variables and the target. When the first tree is built, the final output is by the leaves (Saini 2021). The direct formula to calculate the final result is shown in equation (2):

$$\gamma = \frac{\sum_{i=1}^{n} Residual_i}{\sum_{i=1}^{n} [Previous\ \ probability_i \times (1 - Previous\ \ probability_i)]} \quad (2)$$

where Y is the objective function for the classification decisión.

Random Forest classifier (RF): It is called a decision tree forest. This method is based on the principle of bagging with random feature selection and the model uses voting to combine tree predictions. RF works well for most of the problems; it can manage noise and select only the most important features. However, the interpretability of the model is limited and its fitting requires some effort in data management (Gaurav and Patel 2020).

Decision Tree classifier (DT): It is a supervised automatic learning algorithm that can be used for categorization or prediction. The DTs are designed to mimic human thinking, making the results easy to understand and interpret. The six key components of a DT are the root node, split, decision node, leaf node, pruning and branch (Suthaharan 2016).

The DTs are used in problems which involve data and variables, both numerical and categorical.

They are effective for modeling problems with multiple results and for testing the reliability of trees. Another advantage of DTs is that they require less data cleaning compared to other data modeling techniques. However, it is important to recognize that DTs can be affected by noise and may not be ideal for larger datasets (Kliś *et al.* 2021).

Logistic regression (LR): Also called logit regression, is used to estimate the probability that an instance belongs to a given class. Typically, it is used for binary classification tasks where classes are labeled as 0 and 1, according to a probability threshold (Géron 2019). The estimated probability of LR is showed in equation (3):

$$\hat{p} = h_\theta(x) = \sigma\left(\theta^t \cdot x\right) \quad (3)$$

where σ (t) is a sigmoid function that produces a number between 0 and 1, given by the logistic function shown in equation (4):

$$\sigma(t) = \frac{1}{1 + e^{-t}} \quad (4)$$

where t is the time

The evaluation of automatic learning models is described below:

- *Accuracy or Proximity of results*: It uses the parameters true positive (TP), true negative (TN), false positive (FP), false negative (FN).

- *Area under the curve (AUC):* It measures the ability of the model to discriminate between two classes.
- *Recall or probability of classifying true positives:* It uses the parameters true positive (TP), false negative (FN).
- *Precision or dispersion of the set of values obtained:* Uses the parameters true positive (TP) and false positive (FP).
- *F1 (F-Score):* Combines precision and recall measures into a single value.
- *Kappa* quantifies the agreement between predictions made by a model and the true classes. It is used to evaluate the different predictive yield between classes.
- Training Time (TT Sec) measures the time it takes for a model to learn from the training dataset and fit its parameters to obtain accurate predictions.

Results and Discussion

Automatic learning algorithm preparation, including feature selection and model training, was performed using a combination of state-of-the-art data science tools. The code used for this purpose, based on the 'pycaret' and 'scikit-learn' libraries in Python, formed the cornerstone of the methodological approach.

Implementing the model using standard 'scikit-learn' functions provided a solid foundation for the training process. In this study, hyperparameter fitting was intentionally omitted, relying instead on the default parameters inherent to each model. This strategic choice was made to maintain methodological consistency and facilitate direct comparisons between models. The adoption of default settings inherent to each algorithm was intended to maintain a standardized framework across all analyses, ensuring transparency and reproducibility of the experiments.

The best model trained with the dataset discussed above was GBC, which achieved 96.77 % correct predictions in the testing phase. Additionally, the percentage of the predictive evaluation ability of the trained model was 96.9 %, and in the performance evaluation it reached 93.50 %. Other important metrics such as AUC, recall and precision were also measured, which scored 99.4, 97.90 and 96.10 % respectively. Also, metrics for models such as RF, DT and LR are showed in table 1.

In this study, the training time of the models was measured. In GBC, the training took approximately 0.9 seconds. RF, DT and LR achieved 1, 0.63 and 0.77 seconds in their training respectively. These results and the accuracy of each model are shown in figure 2.

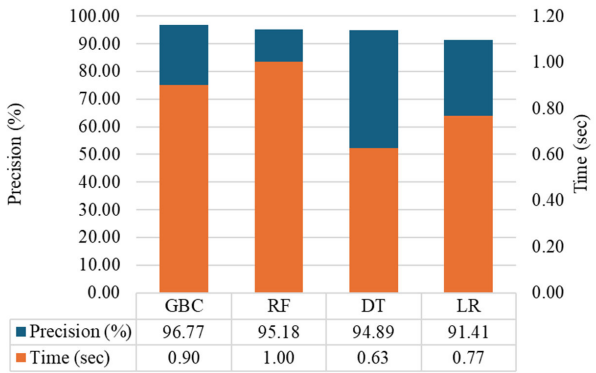


Figure 2. Accuracy and execution time of the top-rated automatic learning algorithms

An essential phase in forming the best model was feature importance. The GBC model, which is the best, found that the feature corresponding to "main income" had a metric of 80 %. The feature importances are showed in figure 3.

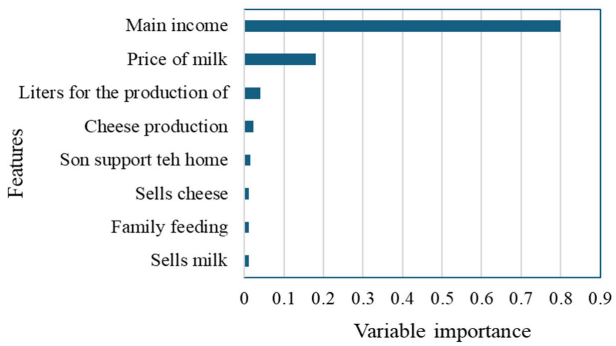


Figure 3. Important features of the GBC model

Figure 4 shows the prediction matrix and the top left and bottom right boxes correspond to correct predictions, while the top right and bottom left boxes contain incorrect predictions or false positives.

Table 1. Results of classification algorithms

Algorithm	Accuracy, %	AUC, %	Recall, %	Prec, %	F1, %	Kappa, %	TT, seg.
GBC	0.9677	0.994	0,979	0.961	0.969	0.935	0.90
RF	0.9518	0.984	0.964	0.946	0.954	0.903	1.00
DT	0.9489	0.956	0.943	0.96	0.95	0.898	0.63
LR	0.9141	0.977	0.948	0.894	0.919	0.828	0.77

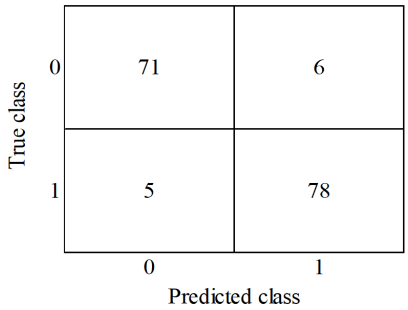


Figure 4. Confusion matrix of the best classification model.

Nyambo et al. (2023) applied automatic learning techniques (ML) in the dairy industry from Tanzania. Their study focused on three main issues: inadequate infrastructure, outdated technology and low productivity. They analyzed the data and found homogeneous production groups. Then they made recommendations to increase milk production. Similarly, Mwanga *et al.* (2020) used ML to identify groups of farmers. In their case, the classification was based on the farm location. It was also based on the system of feeding and caring of animals. This information facilitated better planning and resource management. It allowed for more precise interventions in each group to improve services.

Authors such as Abdukarimova *et al.* (2016) mention that estimating milk production helps to assess production performance and it is necessary for efficient resource management. However, there are several challenges associated with milk production prediction, especially in effective classification.

Ji *et al.* (2022) ran an automatic learning framework using five years of productivity and behavioral health data from 80 cows. They achieved an accuracy of over 80 %.

Other authors such as Radwan *et al.* (2020) have proposed a dynamic linear model (DLM) and an artificial neural network (ANN) in the prediction of milk production. The DLM achieved 95 % accuracy using a dataset consisting of 1,094,780 observations of sensor data provided by Lely Industries (Masslui, The Netherlands). The ANN achieved 79.5 % accuracy, exceeding milk production expectations.

Despite the challenges involved, this study compared different automatic learning models (GBC, RF, DT, LR) on a milk production dataset from Carchi, Ecuador province. The results showed significant classification accuracy: GBC achieved 96.77 % precision and 97.9 % recall. RF achieved 95.18 % accuracy and 95.4 % F1 score.

The abundance of data in the livestock sector requires innovative analytical approaches. This study researched the potential of deep learning models, specifically six neural network algorithms, as an alternative to traditional statistical methods. Compared to these traditional methods, deep learning models can achieve higher accuracy, making them valuable tools for identifying agricultural variables and developing safe dairy products and risk management practices (Suseendran and Duraisamy 2021).

The researchers used classification methods to identify relevant variables, and then used these variables to train several predictive models. These models included not only deep learning algorithms but also established ones such as logistic regression, k nearest neighbors, decision trees, and random forests. While most models achieved high predictive yield of 93 %, neural networks and Gaussian mixture models proved to be more sensitive to variations in the dataset. In response, researchers combined random forest and decision tree algorithms to improve factor selection (Mwanga *et al.* 2020).

The survey results showed that the main economic income derived from milk production (89 %), the price per liter of milk (46 %) and the amount of liters of milk used for cheese production (18 %) were the most important factors in the production. The presence of a child as the economic support of the house (5 %), the use of milk for the production and sale of cheese (21 %) and the use of milk and cheese production for domestic consumption (53 %) also had a significant impact, but to a lesser extent.

The study describes the key SEFs that shape family dynamics and agricultural production in the studied community. It is noted that 90 % of farmers who maintain adequate home conditions, the educational level does not show any influence on family welfare decisions. However, the university education level of some farmers shows the presence of higher incomes and better production rates. In addition, a patriarchal model of family breadwinner prevails, in which husbands assuming this role in 75 % of houses. Age also emerges as a factor. There was increase in cohabitation between the ages of 50 and 55. Also, the experience is intertwined with education, as both have a significant impact on production levels. These findings underscore the complex interplay between education, income, house structure and agricultural productivity and provide valuable information for developing socioeconomic models and development strategies.

The study suggests further exploration through an analysis of technical production efficiency, which would include variables such as infrastructure, labor, products management, milking processes, management, environmental practices and quality control. This type of analysis would allow optimizing production capacities in a production unit. This can lead to specific interventions to improve production efficiency, facilitate fair market access and rationalize value-added dairy processing activities.

Conclusions

This study has identified the factors that influence on production in small dairy farms in the border region between Ecuador and Colombia. The results of this study can be used to inform future researchers and decisions aimed at supporting the sustainability and development of the dairy sector in the

region. By shedding light on the key determinants of milk production and its impact on the economic well-being of rural families, this research provides a valuable guidance to stakeholders and policy makers in formulating targeted interventions and initiatives.

This study, in the unique context of the Ecuadorian border region, highlights the potential of automatic learning techniques to accurately classify small farmers' milk production. The successful application of automatic learning algorithms including Gradient Boosting Classifier and Random Forest has proven effective in classifying milk production with remarkable accuracy.

The results of this study have significant implications for the dairy industry in the Ecuador-Colombia border region, and beyond. The identified factors which influence on milk production provide a roadmap for improving productivity and livelihoods in small-scale dairy farming communities.

As the dairy sector continues to play an essential role in the region's economy, harnessing the power of automatic learning to identify relevant variables will be critical to shaping predictive models, promoting sustainable growth, and strengthening the sector's overall economic well-being.

References

Abdukarimova, M., Abdukarimov, A. & Abdukarimov, N. 2016. *Handbook of Industrial and Innovation Economics*, editado por Munisa, 466p. Uzbekistan: Independently. ISBN: 979-8412353852. Available at: https://www.researchgate.net/profile/Munisa-Abdukarimova/publication/344279960_Handbook_of_Industrial_and_innovation_economics/links/62493f3621077329f2ed6414/Handbook-of-Industrial-and-innovation-economics.pdf.

Alwadi, M., Alwadi, A., Chetty, G. & Alnaimi, J. 2024. Smart dairy farming for predicting milk production yield based on deep machine learning. *International Journal of Information Technology*, 16: 4181-4190, ISSN: 2511-2112. https://doi.org/10.1007/s41870-024-01998-5.

Bai, J., Xue, H., Jiang, X. & Zhou, Y. 2022. Recognition of bovine milk somatic cells based on multi-feature extraction and a GBDT-AdaBoost fusion model. *Mathematical Biosciences and Engineering: MBE*, 19(6): 5850-5866, ISSN: 1551-0018. https://doi.org/10.3934/mbe.2022274.

Balarezo, L., García-D, J., Hernández, M. & García-L, R. 2016. Metabolic and reproductive state of Holstein cattle in the Carchi region, Ecuador. *Cuban Journal of Agricultural Science*, 50(3): 381-392, ISSN: 2079-3480. https://cjascience.com/index.php/CJAS/article/view/632/699.

Bentéjac, C., Csörgő, A. & Martínez-Muñoz, G. 2020. A comparative analysis of gradient boosting algorithms. *Artificial Intelligence Review*, 54(3): 1937-1967, ISSN: 1573-7462. https://doi.org/10.1007/s10462-020-09896-5.

Bovo, M., Agrusti, M., Benni, S., Torreggiani, D, & Tassinari P. 2021. Random Forest Modelling of Milk Yield of Dairy Cows under Heat Stress Conditions. *Animals*, 11(5): 1305, ISSN: 2076-2615. https://doi.org/10.3390/ani11051305.

Carvajal, L.A. 2014. La asociatividad en el sector agropecuario del Carchi y su potencial de producir y comercializar semielaborados de papa y leche. *SATHIRI*, 7(7): 153-163, ISSN: 2631-2905. https://doi.org/10.32645/13906925.348.

CIL Ecuador. 2023. *La industria láctea fomenta la economía circular, a través de una producción sostenible,* Comprometidos con el Desarrollo de la Cadena Láctea. Available at: https://www.cil-ecuador.org/post/la-industria-láctea-fomenta-la-economía-circular-a-través-de-una-producción-sostenible. [Consulted: March 10, 2024].

Fadillah, A., van den Borne, B.H.P., Poetri, O.N., Hogeveen, H., Umberger, W., Hetherington, J., & Schukken, Y.H. 2023. Smallholder milk-quality awareness in Indonesian dairy farms. *Journal of Dairy Science*, 106(11): 7965-7973, ISSN: 0022-0302. https://doi.org/10.3168/JDS.2023-23267.

FAO. 2018. *Panorama de la pobreza rural en América Latina y el Caribe*. Roma, 114p. ISBN: 978-92-5-131085-4 Available at: https://openknowledge.fao.org/handle/20.500.14283/ca2275es. [Consulted: February 03, 2024].

FAO. 2022a. *The State of Food and Agriculture 2022.* Roma, 182p. ISBN: 978-92-5-136043-9. https://doi.org/10.4060/cb9479en.

FAO. 2022b. *La aplicación de las mejores prácticas de la inteligencia artificial en el contexto de la agricultura,* editado por Bishan Dong, 136. Roma: FAO Publications Catalogue 2022. ISBN: 78-92-5-136969-2.

FAO. 2023a. *FAO analiza fortalezas y brechas de la producción láctea en América Latina y el Caribe,* Más Allá de La Finca Lechera. Available at: https://www.fao.org/americas/noticias/ver/es/c/1617544/. [Consulted: July 18, 2024].

Franco, W. 2016. Propuestas para la innovación en los sistemas agroproductivos y el desarrollo sostenible del Valle Interandino en Carchi, Ecuador. *Tierra Infinita*, 2(1): 49-87, ISSN: 2631-2921. https://doi.org/10.32645/26028131.104.

Gaudin, Y. & Padilla, R. 2020. Los intermediarios en cadenas de valor agropecuarias: un análisis de la apropiación y generación de valor agregado (N° 186 (LC/TS.2020/77; LC/MEX/TS.2020/15). Serie Estudios y Perspectivas-Sede Subregional de La CEPAL en México. Available at: https://www.cepal.org/es/publicaciones/45796-intermediarios-cadenas- valor-agropecuarias-un-analisis-la-apropiacion-generacion. [Consulted: August 20, 2024].

Gaurav, K.A. & Patel, L. 2020. Machine Learning With R. In S. Khalid (Ed.), Applications of Artificial Intelligence in Electrical Engineering (pp.

291-331), ISBN: 9781799827184. IGI Global. https://doi.org/10.4018/978-1-7998-2718-4.ch015.

Géron, A. 2019. Hands-on machine learning with Scikit-Learn and TensorFlow: concepts, tools, and techniques to build intelligent systems (2nd ed.). O'Reilly Media. ISBN: 978-1-492-03264-9. Available at: https://books.google.com.ec/books?id=HnetDwAAQBAJ&printsec=frontcover&hl=es&source=gbs_book_other_versions#v=onepage&q&f=false. [Consulted: August 10, 2024].

Gil Montelongo, M. & Hernández Villa, X. 2019. Risk management as a tool in the internal control on organizations of the dairy sector. *Ekotemas*, 5(2): 51-66, ISSN: 2414-4681. https://www.ekotemas.cu/index.php/ekotemas/article/view/63/54.

Hernández-Sampieri, R., & Mendoza, C. 2018. Metodología de la investigación. Las rutas cuantitativa, cualitativa y mixta. In Interamericana (Ed.), McGRAW-HILL Interamericana Editores S.A. de C.V. Mc Graw Hill. ISBN: 978-1-4562-6096-5.

Ionita, E. 2022. La producción de leche en Ecuador, Veterinaria Digital. Available at: https://www.veterinariadigital.com/articulos/la-produccion-de-leche-en-ecuador/. [Consulted: January 20, 2024].

Ji, B., Banhazi, T., Phillips, C.J.C., Wang, C. & Li, B. 2022. A machine learning framework to predict the next month's daily milk yield, milk composition and milking frequency for cows in a robotic dairy farm. *Biosystems Engineering*, 216(9): 186-197, ISSN: 1537-5110. https://doi.org/10.1016/j.biosystemseng.2022.02.013.

Kassahun, A., Bloo, R., Catal, C. & Mishra, A. 2022. Dairy Farm Management Information Systems. *Electronics*, 11(2): 1-18, ISSN: 2079-9292. https://doi.org/10.3390/electronics11020239.

Kliś, P., Piwczyński, D., Sawa, A. & Sitkowska, B. 2021. Prediction of Lactational Milk Yield of Cows Based on Data Recorded by AMS during the Periparturient Period. *Animals*, 11(383): 1-11, ISSN: 2076-2615. https://doi.org/10.3390/ANI11020383.

Mannepalli, P.K., Kulurkar, P., Jangade, V., Khan, A., & Singh, P. 2024. An Enhanced Classification Model for Depression Detection Based on Machine Learning with Feature Selection Technique. En P. K. Jha, B. Tripathi, E. Natarajan, & H. Sharma (Eds.), *Proceedings of Congress on Control, Robotics, and Mechatronics* (Vol. 364, pp. 589-601). Springer Nature Singapore. https://doi.org/10.1007/978-981-99-5180-2_46

Moreno, F. 2018. Caracterización socioeconómica y productiva de la cadena de valor agroalimentaria de la leche en la provincia de Tungurahua. Tesis presentada en opción al Título de carrera de Ingeniería de los alimentos, Universidad Técnica de Ambato, Ecuador.

Morocho, B., Carvajal, H. & Vite, H. 2021. Análisis socioeconómico del agronegocio ganadero: Caso productores de la Aso Ganaderos del Altiplano Orense 5 de noviembre del cantón Atahualpa. *Revista Metropolitana de Ciencias Aplicadas*, 4(1): 26-32, ISSN: 2631-2662.

Mwanga, G., Lockwood, S., Mujibi, D., Yonah, Z. & Chagunda, M. 2020. Machine learning models for predicting the use of different animal breeding services in smallholder dairy farms in Sub-Saharan Africa.*Tropical Animal Health and Production,*52(3): 1081-1091, ISSN: 1573-7438. https://doi.org/10.1007/s11250-019-02097-5.

Natekin, A. & Knoll, A. 2013. Gradient boosting machines, a tutorial. *Frontiers in Neurorobotics*, 7(21): 1-21, ISSN: 1662-5218. https://doi.org/10.3389/fnbot.2013.00021.

Nyambo, D.G., Malamsha, G.C. & Mavura, F. 2023. Leveraging Machine Learning Techniques to Improve Learning and Recommendations Within Dairy Farms: Towards High Milk Yields for Small-Scale Farmers. In F. Mtenzi, G. Oreku, & D. Lupiana (Eds.), *Impact of Disruptive Technologies on the Socio-Economic Development of Emerging Countries* (pp. 172-188), ISBN: 9781668468739. IGI Global. https://doi.org/10.4018/978-1-6684-6873-9.ch011.

Orús, A. 2022. Leche de vaca: principales productores a nivel mundial en 2022. Estatista. Available at: https://es.statista.com/estadisticas/600241/principales-productores-de-leche-de-vaca-en-el-mundo-en/. [Consulted: April 30, 2024].

Piwczyński, D., Sitkowska, B., Kolenda, M., Brzozowski, M., Aerts, J. & Schork, P.M. 2020. Forecasting the milk yield of cows on farms equipped with automatic milking system with the use of decision trees. *Animal Science Journal,* 91(1): e13414, ISSN: 1740-0929. https://doi.org/10.1111/asj.13414.

Peña, Y., Benitez, D., Ray, J. & Fernández, Y. 2018. Factores determinantes de la producción ganadera en una comunidad campesina del suroeste de Holguín, Cuba. *Cuban Journal of Agricultural Science*, 52(2): 155-163, ISSN: 2079-3480. http://scielo.sld.cu/scielo.php?pid=S2079-34802018000200155&script=sci_arttext&tlng=es

Prefectura del Carchi. 2023. Datos informativos de la provincia. Available at: https://carchi.gob.ec/2016f/index.php/informacion-provincial.html. [Consulted: April 25, 2024].

Radwan, H., Qaliouby, H. & Elfadl, E. 2020. Classification and prediction of milk yield level for Holstein Friesian cattle using parametric and non-parametric statistical classification models. *Journal of Advanced Veterinary and Animal Research*, 7(3): 429-435, ISSN: 2311-7710. https://doi.org/10.5455/javar.2020.g438.

Requelme, N. & Bonifaz, N. 2012. Caracterización de sistemas de producción lechera de Ecuador. *La Granja*, 15(1): 56-69, ISSN: 1390-3799.

Saini, A. 2021. Gradient Boosting Algorithm: A Complete Guide for Beginners. Analytics Vidhya. Available at:

https://www.analyticsvidhya.com/blog/2021/09/gradient-boosting-algorithm-a-complete-guide-for-beginners/. [Consulted: March 21, 2024].

Siddiqui, T. & Amer, A.Y.A. 2024. A comprehensive review on text classification and text mining techniques using spam dataset detection. In *Mathematics and Computer Science*, vol. *2*, editado por Ghosh, S., Niranjanamurthy, M., Deyasi, K., Mallik, B. & Das, S., 1-17. Editorial Wiley, ISBN: 978-111989671-5. https://doi.org/10.1002/9781119896715.ch1.

Slob, N., Catal, C. & Kassahun, A. 2021. Application of machine learning to improve dairy farm management: A systematic literature review. *Preventive Veterinary Medicine*, 187: 105237, ISSN: 1873-1716. https://doi.org/10.1016/j.prevetmed.2020.105237.

Suthaharan, S. 2016. Decision Tree Learning, In *Machine Learning Models and Algorithms for Big Data Classification*, Integrated Series in Information Systems, vol 36. Springer, Boston, MA., 237-269, ISBN: 9781489976413. https://doi.org/10.1007/978-1-4899-7641-3_10.

Suseendran, G. & Duraisamy, B. 2021. Predication of Dairy Milk Production Using Machine Learning Techniques. In: Peng, SL., Hsieh, SY., Gopalakrishnan, S., Duraisamy, B. (eds) *Intelligent Computing and Innovation on Data Science. Lecture Notes in Networks and Systems, 248*: Springer, Singapore, ISBN: 978-981-16-3153-5. https://doi.org/10.1007/978-981-16-3153-5_60.

Tangorra, F. M., Calcante, A., Vigone, G., Assirelli, A. & Bisaglia, C. 2022. Assessment of technical-productive aspects in Italian dairy farms equipped with automatic milking systems: A multivariate statistical analysis approach. *Journal of Dairy Science*, 105(9): 7539-7549, ISSN: 0022-0302. https://doi.org/10.3168/jds.2021-20859.

Terán, G. & Cobo, R. 2017. Determining management factors in dairy farms in Carchi, Ecuador. *Cuban Journal of Agricultural Science*, 51(2): 175-182, ISSN: 2079-3480. http://cjascience.com/index.php/CJAS/article/view/724.

Treviño Cantú, J.A. 2022. Alternativas de estandarización para índices compuestos espacio-temporales. El caso del rezago educativo en los estados de México, 2000 a 2020. *Investigaciones Geográficas*, 109: 1-14, ISSN: 2448-7279. https://doi.org/10.14350/rig.60615.

Valdez, A. 2019. Machine Learning para todos. En *IV Congreso Nacional de Profesionales de Computación, Informática y Tecnologías*. pp. 60. Perú: Ministerio de Educación. https://doi.org/10.13140/RG.2.2.13786.70086.

Vásquez, H., Barrantes, C., Vigo, C. & Maicelo, J. 2022. Factores socioeconómicos que influyen en la adopción de tecnologías para mejoramiento genético de ganado vacuno en Perú. *Agricultura, Sociedad y Desarrollo*, 19(3): 312-330, ISSN: 2594-0244. https://doi.org/10.22231/asyd.v19i3.1358.

Velasteguí, N. 2019. Cadena productiva del sector lechero en la provincia de Tungurahua, cantón Píllaro: Un estudio socio-económico de la producción de la leche cruda. Tesis presentada en opción al Título de carrera de Economía, Universidad Técnica de Ambato, Ecuador.

Zemarku, Z., Senapathy, M. & Bojago, E. 2022. Determinants of Adoption of Improved Dairy Technologies: The Case of Offa Woreda, Wolaita Zone, Southern Ethiopia. *Advances in Agriculture*, 2022: 1-19, ISSN: 2314-7539. https://doi.org/10.1155/2022/3947794.