DOI: https://doi.org/10.1590/1809-4430-eng.agric.v42n5e20210239/2022

Engenharia Agrícola

ISSN: 1809-4430 (on-line)

www.engenhariaagricola.org.br

Scientific Paper
Doi: http://dx.doi.org/10.1590/1809-4430-Eng.Agric.v42n5e20210239/2022

CLASSIFICATION OF SUGARCANE YIELDS ACCORDING TO SOIL FERTILITY PROPERTIES USING SUPERVISED MACHINE LEARNING METHODS

**Jhonnatan Yepes[1*], Gian Oré[2], Marlon S. Alcântara[2],
Hugo E. Hernandez-Figueroa[2], Bárbara Teruel[1]**

[1*]Corresponding author. School of Agricultural Engineering, University of Campinas-UNICAMP/ Campinas - SP, Brasil.
E-mail: jayepesg@unal.edu.co | ORCID ID: https://orcid.org/0000-0001-9313-9834

KEYWORDS

random forest, naïve bayes, decision trees, multi-class classification, classification techniques.

ABSTRACT

Action planning and decision-making in the sugarcane management chain depend on yield estimates, which, in turn, vary with the soil. This study aimed to describe an applicable method of classifying sugarcane productivity into three categories, based on soil properties (medium, low, and high), determining which is most associated with biomass production. To this end, we applied the machine learning methods Naïve Bayes, Decision Trees, and Random Forest, as they proved to be useful tools for faster and more accurate results. Our results indicate that Random Forest is the most suitable for classifying all yield categories, and Naïve Bayes had good results for classification into "medium" and "low" and potential for solving multiclass problems in agriculture. Organic matter was the property most closely related to sugarcane biomass yield by the Random Forest and Decision Trees algorithms. The methods described can be used to obtain subsidies for sugarcane chain management, contributing to more sustainable decisions.

INTRODUCTION

Brazil is the largest sugarcane producer worldwide. This crop is important for the economy, which has led to the search for scientific and technological advances, including productivity and fertilization estimates. Early estimates allow saving time, work, and resources, serving as a subsidy for decision-making public policies (Everingham et al., 2016).

Several methods have been used to estimate sugarcane yield. However, productivity has not been classified based on soil fertility properties. Furthermore, knowing which soil properties are most closely related to sugarcane biomass allows for adequate fertilization of soils, preserving them for agricultural purposes. Organic fertilizers provide greater environmental control and relevant savings in sugarcane fertilization, contributing to sustainability (Xu et al., 2021).

In this sense, Machine Learning techniques have been used to find correlations and connect information or patterns between attributes that make up a data network (Liakos et al., 2018). Among the techniques applied to agricultural processes, the tools Decision Trees, Random Forest, and Naïve Bayes stand out because they are based on input attributes (independent variables), describing interactions with a target attribute.

Decision Trees are one of the most common and powerful data structures in Computer Science because they are easy to understand and interpret (Rajeswari & Suthendran, 2019). Random Forest is one of the most explored methods for crop yield prediction (van Klompenburg et al., 2020) to model and correlates it to soil fertility parameters in sugarcane (Charoen-Ung & Mittrapiyanuruk, 2019; Kouadio et al., 2018). Finally, the Naïve Bayes is an algorithm capable of performing multiclass classification, through a linear independence analysis of factors influencing dependent variables (Drury et al., 2017; Pham & Brabyn, 2017).

Given their potential to provide subsidies to the sugarcane production chain, this research had two objectives:

[1] School of Agricultural Engineering, University of Campinas-UNICAMP/Campinas - SP, Brasil.
[2] School of Electrical and Computer Engineering, University of Campinas-UNICAMP/Campinas - SP, Brasil.

Area Editor: Edna Maria Bonfim-Silva
Received in: 12-29-2021
Accepted in: 9-17-2022

1) to describe the method that can be applied for sugarcane yield classification in three categories (medium, low, and high) based on soil properties; 2) determine the soil fertility property mainly associated with biomass production.

MATERIAL AND METHODS

Study area

The field study was carried out between July 2019 and June 2020 in the city of Campinas, São Paulo State, Brazil. The experiment had a total area of three hundred and twenty square meters (320 m^2), with the geographical coordinates being 22°49'12"S, 47°03'41"W, 625-m altitude at the central point.

According to the USDA soil classification, the soil of the experimental area corresponds to a clayey Oxisol with a grain size distribution of 570, 250, and 180 g kg^{-1} of clay, sand, and silt, respectively. To eliminate weeds, the first soil layer was mixed to generate a stable and aerated layer, and a mower was used for root fixation. Then, two operations were carried out with a 0.66 m diameter disc plow.

In this study, 400 sugarcane plants were grown in rows spaced 1.50m apart and at 0.75m in a row. The plants used belonged to the variety IACSP97-4039, which is characterized by its short cycle (up to 13 months), high sucrose content, and resistance to water deficit and sudden temperature changes.

Soil sampling and fertility analysis

The experimental area was divided into eight 4.5x9.0m regular meshes (FIGURE 1). A 300-g composite sample was collected from each mesh at a depth of 0-20cm, with all sampling sites being geo-referenced. The samplings were performed monthly from December 2019 to June 2020 to measure spatial and temporal variability of soil fertility.

Calcium chloride (CaCl$_2$) was used for pH determination. Organic content was determined by the photometric method, with the resin method to measure essential macro-elements (P, K, Ca, and Mg). Pentetic acid (DTPA) was the active component used to determine micronutrients (Cu, Fe, Mn, and Zn). Finally, the hot water method was used to quantify available boron (B).

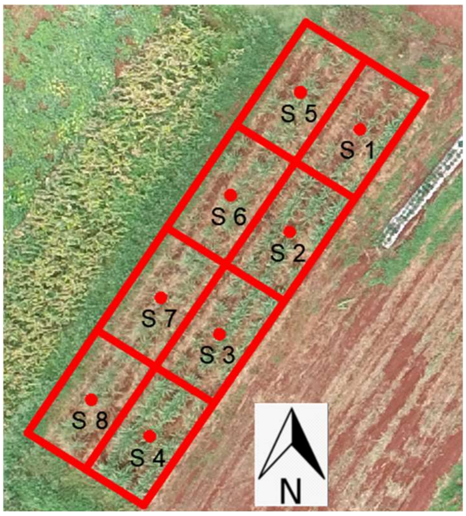


FIGURE 1. Distribution of sampling sites for determination of soil fertility properties in the experimental area.

Harvest

In mid-June 2020, sugarcane plants reached the optimum maturation point and were harvested manually. The mass of two contiguous plants within the same row was adopted as productivity per square meter. Thus, a regular mesh was generated over the experimental area to later generate a georeferenced productivity map.

Data processing and algorithms application

During monthly monitoring, undeveloped plants were detected and geo-referenced. Outliers were identified from yield histogram analysis. Afterward, soil chemical properties and sugarcane productivity were modeled with nearest-neighbor interpolation, using the free software QGIS™ version 3.10.12.

Python programming language from Visual Studio Code editor was chosen for exploratory data analysis (EDA), measures of central tendency, analysis of correlations between variables, and removal of soil fertility parameters not correlated with the target attribute, reducing dimensionality problems for the application of the machine learning algorithms.

The filtered database was divided into two sets. The first contained 70% of the total volume of data and allowed the training of each classification algorithm, while the remaining 30% had the function of validating the performance of generated models. Each classifier was treated in the selection of training set adjustment parameters to obtain the best results (Table 1). Data splitting, model training, results in visualization, and further evaluation of modeling metrics was performed with Pandas Libraries, Scikit-learn, and Matplotlib

TABLE 1. Model parameters used in each classifier.

Algorithm	Parameter
Naïve Bayes	Batch size= 100 Kernel estimator: Yes Normalized values: Yes
Decision Trees	Criterion: Entropy Minimum number of samples required to split an internal node= 2 Minimum number of samples required to be at a leaf node= 5 Maximum leaf nodes= 5 Class weight: Balanced Normalized values: Yes
Random Forest	Criterion: Entropy Number of trees in the forest= 50 Minimum number of samples required to split an internal node= 2 Minimum number of samples required to be at a leaf node= 5 Maximum leaf nodes= 5 Class weight: Balanced Normalized values: Yes

Evaluation metrics (Models performance evaluation)

Receiver Operating Characteristics (ROC) analysis was used to evaluate, compare, and select the best classifier based on performance. The first phase consisted of the generation of a confusion matrix (Majnik & Bosnić, 2013).

Based on the confusion matrix values, recall is the true positive rate and indicates the probability of detecting a positive sample correctly. On the other hand, accuracy refers to the dispersion of a set of values from repeated measurements of a magnitude, and the less dispersed the values, the more accurate they are. The metric is represented by the ratio between the number of correctly classified instances (positive and negative) and the total number of instances.

Then, a two-dimensional ROC plot was built considering the probability of a false positive on the abscissa versus the recall on the ordinate. Based on this projection, the area under the ROC curve (AUC) was determined. The AUC of a classifier is equivalent to the probability that the classifier will rank a randomly chosen positive instance higher than a randomly chosen negative instance.

Cohen's Kappa coefficient (k, equation [1]) represents the proportion of agreements observed beyond chance. It generally varies between 0 and 1 although negative numbers can occur. The closer the coefficient is to 1, the greater the reliability, and values close to zero, or below, denote purely random agreement:

$$k = \frac{P_0 - P_e}{1 - P_e} \qquad (1)$$

Where:

P_0 (equation 2) represents the proportion of agreements observed,

P_e corresponds to the proportion of agreements expected in the hypothesis of independence among classifiers, that is, agreements by chance, and n is the number of samples.

$$P_0 = \frac{TP + TN}{n}, \qquad (2)$$

$$P_e = \frac{[(TP + FP) \times (TP + TN)] \times [(FP + TN) \times (FN + TN)]}{n^2}. \qquad (3)$$

Where:

TP corresponds to positive data correctly classified by the algorithm;

FN is the number of incorrectly classified positive data;

TN is when the negative instance is classified as such;

FP is in the case of positive data classified incorrectly.

Mean absolute error (MAE, equation [4]), root mean square error (RMSE, equation [5]), and relative absolute error (RAE, equation [6]) were determined for reference, evaluation, and comparison purposes.

$$Mean\ Absolute\ Error\ (MAE) = \frac{1}{n}\sum_{i=1}^{n}|y_i - x_i|, \qquad (4)$$

$$Root\ Mean\ Square\ Error\ (RMSE) = \sqrt{\frac{1}{n}\sum_{i=1}^{n}|y_i - x_i|}, \qquad (5)$$

$$Relative\ Absolute\ Error\ (RAE) = \sum_{i=1}^{n} \frac{|y_i - x_i|}{|\bar{y}_i - x_i|}. \qquad (6)$$

Where:

y_i is the response of the predicted variable, and

x_i is the observed variable response.

RESULTS AND DISCUSSION

Climatic factors

From crop sowing to harvest, the average temperature was 22.7°C, average relative humidity 64.2%, and accumulated rainfall 1,213.1 mm, with 25% of rains concentrated between January and May 2020. FIGURE 2 shows the historical series of climatic factors measured during the experimental period.

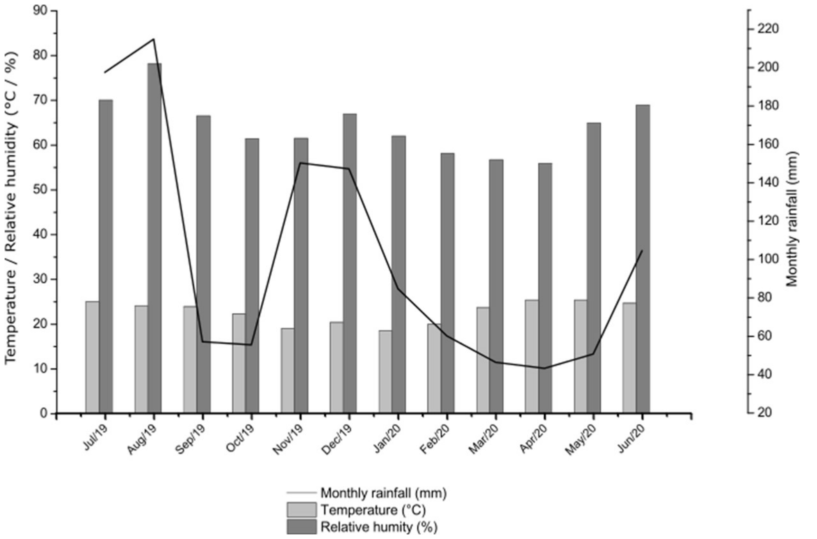


FIGURE 2. Monthly averages of temperature (°C), relative humidity (%), and accumulated rainfall (mm) between July 2019 and June 2020.

Soil fertility analysis

Soil analyses were performed from sugarcane maturation to harvest to monitor the uniformity and availability of chemical and nutritional properties. Eight sampling regions were defined, in which 56 sampling points were geo-referenced. Table 2 summarizes the results of soil analysis. In terms of temporal variability, the first layer of the soil profile under investigation showed uniform behavior.

That behavior is not satisfied only in the case of calcium, whose standard deviation was 30 mg dm^{-3}. This effect is because Ca is mainly absorbed during sugarcane rooting. Moreover, the cationic substitution of exchangeable Ca, in the Ca^{2+} form, improves soil structure, permeability, and water infiltration, helping plants withstand saline stress (Rahman et al., 2018).

TABLE 2. Summary of the descriptive statistical analysis of soil fertility factors expressed as mg dm^{-3}, OM expressed as g dm^{-3}, and pH non-dimensional.

	OM	pH	P	K	Ca	Mg	B	Cu	Fe	Mn	Zn
Mean	25.7	4.53	11.3	12.9	90.8	6.1	0.30	15.1	18.6	22.6	4.1
Standard deviation	1.7	0.18	2.7	2.5	30.4	1.4	0.006	3.5	3.3	0.4	2.5
Standard error	0.14	0.02	0.23	0.21	2.53	0.12	0.0005	0.29	0.28	0.03	0.21
Lower value	22.3	4.2	8.2	9.0	44.0	3.6	0.29	9.0	13.2	21.3	0.8
25%	24.4	4.3	8.7	10.8	62.0	4.8	0.296	11.7	16.1	22.5	1.7
50%	25.8	4.5	10.7	12.9	86.4	6.0	0.3	15.0	18.9	22.7	3.2
75%	27.5	4.7	13.9	15.2	119.0	7.3	0.30	18.1	20.6	22.8	6.3
Maximum value	28.2	4.8	16.5	16.8	144.0	8.6	0.31	21.0	26.5	23.3	9.8

OM - Organic matter, P- Phosphorus, K- Potassium, Ca- Calcium, Mg- Magnesium, B-Boron, Cu- Copper, Fe-Iron, Mn- Manganese, and Zn-Zinc.

Variation in intracellular free Ca^{2+} concentration is one of the earliest events following plant perception of environmental changes (Aldon et al., 2018). Table 3 highlights this variation that occurred between December 2019 and March 2020 when monthly accumulated rainfall was reduced by one-third (from 147 to 47 mm). The monthly average of relative humidity also declined by 10.2%, while temperature increased by 14% (from 20.4 to 23.7°C). During this time interval, there was an increase in the consumption of Ca^{2+} available in the soil.

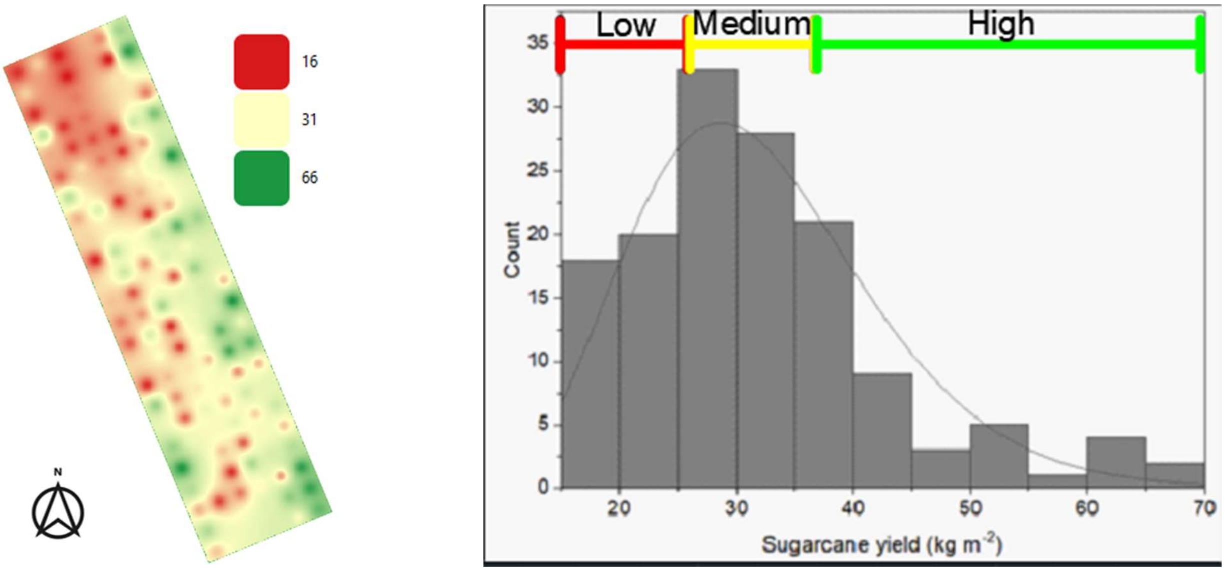
TABLE 3. Averages of available calcium in the first layer of the soil profile.

Ca^{2+} (mg dm^{-3})					
Dec/19	Jan/20	Feb/20	Mar/20	Apr/20	May/20
98.5	90.0	85.0	86.5	88.5	81.5

Harvest data

The rigorous monthly monitoring of the sugarcane field, together with other parallel investigations, allowed the detection of 20 undeveloped plants, which generated a set of missing values in the global dataset (Luebeck et al., 2020; Oré et al., 2020). These values were removed from the dataset and modeled by inverse distance weighting (IDW), using the QGIS™ software version 3.10.12 (FIGURE 3a).

The changes made in the main database provided: a) a decrease in standard deviation by 26%; b) a reduction in the range of values between maximum and minimum crop productivity by 36.5%; c) an increase in yield per square meter from 29.14 to 32 kg m^{-2} for the estimated productivity. As consequence, the overall value of fresh mass increased from 4196.2 to 4616.5 kg m^{-2}.



(a)

(b)

FIGURE 3. Sugarcane harvest results: (a) yield map after removal of outliers (kg m^{-2}), (b) yield histogram.

Frequency histogram analysis and interpretation, in parallel with a mean ($\bar{x}$) and standard deviation (σ), allowed us to categorize productivity into three classes: low (equation [7]), medium (equation [8]), and high (equation [9]). As a result, the range corresponding to each dataset was divided based on the mean and standard deviation of the dependent variable. The category "medium" contained 36% of the data, while "high" and "low" contained 33 and 31% of the data, respectively.

$$Low \leq \bar{x}_{prod} - \frac{\sigma_{prod}}{2}, \tag{7}$$

$$\bar{x}_{prod} - \frac{\sigma_{prod}}{2} \leq Medium < \bar{x} + \frac{\sigma_{prod}}{3}, \tag{8}$$

$$High \geq \bar{x}_{prod} + \frac{\sigma_{prod}}{3}. \tag{9}$$

Consecutively, exploratory data analysis was executed to reduce the dimensionality of model input variables, and the data correlation matrix was calculated. The micronutrients boron (B) and manganese (Mn), which were used as input variables, were not correlated according to FIGURE 4. For this reason, these micronutrients were discarded from the input group of the model training sets.

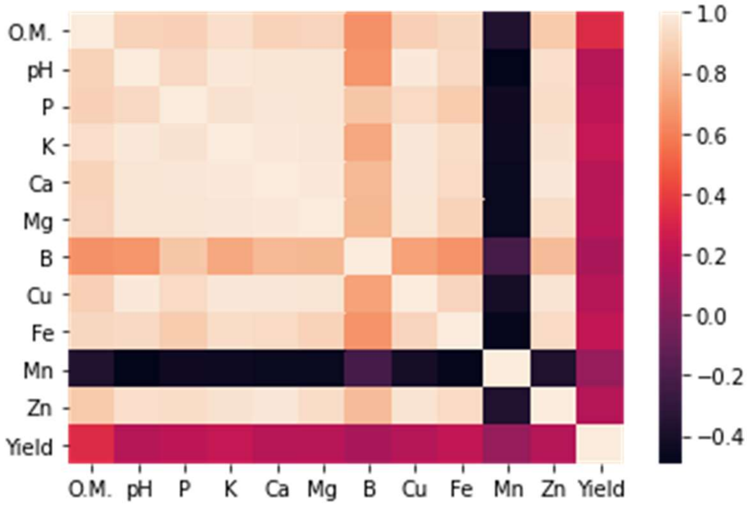


FIGURE 4. Correlation matrix of fertility properties and sugarcane yield.

Performance of models

Three classification algorithms were trained and evaluated: Naïve Bayes, Decision Trees, and Random Forest. Tables 4, 5 and 6 show the performance of each algorithm corresponding to the ROC curve from the machine learning models used. Additionally, table 7 shows the statistical analysis results of each model. The summary of results demonstrates that the Random Forest algorithm had a better performance in classifying the proposed labels for the development of this research.

However, a more detailed analysis of true positive and true negative rates opened a debate about the choice of the best classifier. Although the Naïve Bayes algorithm had a superior performance for "medium" and "low" classes, its low performance in the "high" class reduced its overall performance.

When analyzing values in the area under ROC (AUC), which is interpreted as the probability of a classifier considering a positive instance superior to a negative instance under random conditions, the Naïve Bayes algorithm had the best results, demonstrating its ability to solve multi-class problems due to its conditional independence analysis setup.

TABLE 4. Summary of the ROC characteristics of the Naïve Bayes algorithm.

Naïve Bayes					
Class	True positive rate	False positive rate	Accuracy	Recall	AUC
High	0.36	0.18	0.40	0.36	0.70
Medium	0.40	0.22	0.62	0.40	0.70
Low	0.83	0.32	0.50	0.83	0.84
Weighted average	0.51	0.24	0.53	0.51	0.74

TABLE 5. Summary of the ROC characteristics of the Decision Trees algorithm.

Decisions Tree					
Class	True positive rate	False positive rate	Accuracy	Recall	AUC
High	1.00	0.78	0.30	1.00	0.47
Medium	0.00	0.00	--	--	--
Low	0.50	0.03	0.86	0.50	0.63
Weighted average	0.40	0.21	--	0.40	--

TABLE 6. Summary of the ROC characteristics of Random Forest algorithm.

Random Forest					
Class	True positive rate	False positive rate	Accuracy	Recall	AUC
High	0.64	0.28	0.44	0.64	0.70
Medium	0.35	0.26	0.54	0.35	0.52
Low	0.75	0.16	0.64	0.75	0.79
Weighted average	0.58	0.24	0.54	0.54	0.64

TABLE 7. Performance metrics for classification algorithms.

Metrics	Algorithm		
	Naïve Bayes	Decision Trees	Random Forest
Correctly classified instances (%)	51.16	39.53	53.48
Incorrectly classified instances (%)	48.84	60.46	46.51
Kappa coefficient	0.27	0.18	0.30
MAE	0.34	0.40	0.38
RMSE	0.47	0.44	0.47
RAE	0.77	0.89	0.85

The recall is responsible for evaluating how well a model predicts a positive class when the actual result is positive. It showed that the three models promoted a better interpretation of the range of values corresponding to high productivity, while the two other classes showed discontinuous behavior, with the "medium" class being the most susceptible to misclassification despite having more observations to categorize productivity.

The Decision Trees was the least successful method for label classification. According to the graphical result in FIGURE 5, the root node corresponds to organic matter, and a rule reaches its terminal nodes to classify productivity as low or medium.

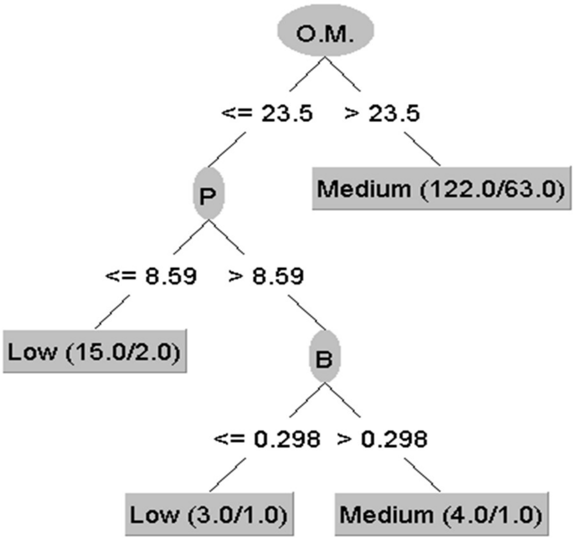


FIGURE 5. Productivity classification by the decision trees method (Organic matter [OM]).

Although the distribution and availability of soil fertility properties were homogeneous in the experimental field, productivity showed a dispersed range of values. Thereby, the Kappa concordance indexes obtained were unsatisfactory for all methods used. The maximum Kappa index achieved was 0.30 for Random Forest followed by 0.27 for Naïve Bayes.

The situations discussed are based on the internal structure of each algorithm. For instance, Decision Trees and Random Forest are methods that automate interaction, with mostly non-linear effects. Furthermore, predictors or classifiers are split to imply a more solid forest. In this process, overfitting problems of the models are generated so that Random Forest is less prone than Decision Trees (Jiang et al., 2020).

Despite its limitations, Random Forest could improve performance in predicting unbalanced categorical variables, due to its ability to divide trees of the same size in parallel in the learning process (Zhou et al., 2020). As previously shown, data imbalance is reflected in accuracy and recall metrics. For this reason, Random Forest performed better than the Decision Trees in terms of labeling.

The main constraint of Decision Trees was their instability, as a small change in data may build different trees (Deepa & Ganesan, 2018). Furthermore, this algorithm is sensitive to noisy and irrelevant attributes and may cause the absence of global functions of various types. Therefore, it loses representativeness in complex multi-class classification problems.

By contrast, for being based on a conditional independence test, the Naïve Bayes algorithm is suitable for solving multiclass problems in the case of several categorical inputs. Its main limitation, however, is that it implicitly assumes that all attributes are independent of each other, limiting their applicability in real-life cases since completely independent predictors are difficult to obtain (Poroikov et al., 2019). Also, if a categorical variable has a label in the test dataset that was not observed during model training, a value of zero will be assigned, and classification will not be performed. To solve this problem, the training data must be cleaned and verified or smoothed with Laplace functions.

Soil fertility parameters associated with sugarcane productivity

Organic matter (OM) had the highest importance factor among the chemical properties mostly related to biomass production when the Random Forest algorithm was used (FIGURE 6). The same result was obtained by the Decision Trees algorithm, which was supported by the correlation matrix in the exploratory data analysis, which, in turn, described the low association of manganese and boron with the targeted variable.

OM contents must stabilize so that organic carbon deposits can be maintained. As a result, positive effects are generated on soil structure, water retention, aeration, fertility, plant rooting, fauna development, as well as microbial biomass, and diversity (Moreno-Barriga et al., 2017). The soil is a dynamic habitat for countless living beings and is where biotic interactions occur for the ecosystem's functioning. In this context, macrofauna and microfauna are crucial for OM crushing and transforming. In turn, Ca^{2+} promotes a bond between OM and mineral particles, forming aggregates in the soil, which, in turn, control the dynamics of OM in the soil and directly increase its ability to capture and stabilize organic carbon (Juriga et al., 2018).

In the relationship between water, soil, and environment, OM showed to be closely related to soil water retention, facilitating root growth, soil nutrient absorption, and crop establishment (Ankenbauer & Loheide, 2017; Minasny & McBratney, 2018).

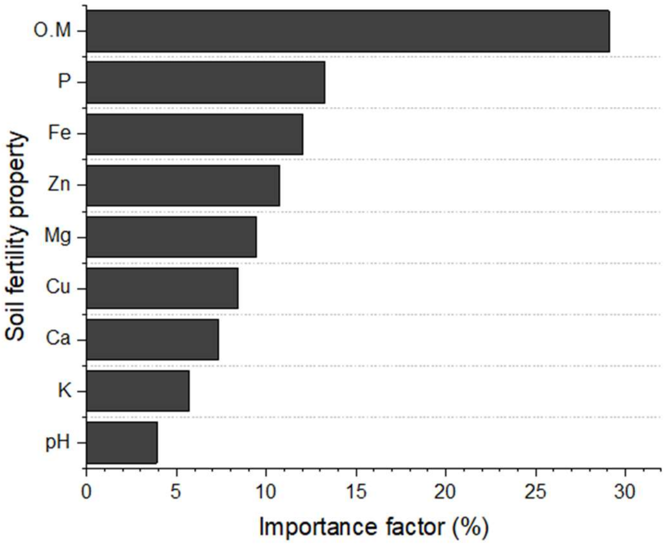


FIGURE 6. Ranking of the importance of soil fertility properties according to sugarcane productivity using the Random Forest algorithm.

Given the close relationship between crop yields and climate changes, especially in sugarcane fields (Jin et al., 2018), future studies should evaluate such factors in a time series. Therefore, we recommend replicating our study in other regions of the country to increase the volume of data and then build robust and reliable models. Accordingly, sugarcane production volume can be predicted, ensuring that national and international demands are met, rigorous monitoring of water and soil management, land-use planning, and support in decision-making regarding storage, transport, and logistics in the sector.

Besides the widely used algorithms for solving regression and classification problems in agriculture (Random Forest and Decision Trees), we included the Naïve Bayes. And, although it has been little used to classify agricultural productivity, it had a good performance and potential due to its simpler internal structure. We could also detail the procedures required for database exploration, in addition to identifying, removing, and filling in missing data, dividing classes taking as a starting point the two typical measures of central tendency.

CONCLUSIONS

The application of Random Forest is recommended to classify all categories of crop productivity. The Naïve Bayes is a good technique for the classification of productivity at "medium" and "low" levels, with the potential for solving multiclass problems in agriculture.

Organic matter has a considerable relationship with sugarcane biomass production when using the Random Forest and Decision Trees algorithms. This fact reinforces the convenience of using organic fertilizers to reduce impacts on the environment.

For future studies, we recommend replicating the experiments under different conditions, including climatic factors within the time series analyzed, including hybrid machine-learning systems to mitigate limitations and improve analysis.

ACKNOWLEDGMENTS

This study was funded by the Coordination for Improvement of Higher Education Personnel (CAPES) and São Paulo State Research Foundation (FAPESP). The agencies financed the research through projects (PITE 2017/19416-3, PIPE 2018/00601-8, 88882.434658/2019-01, and 88882.434694/2019-01).

REFERENCES

Aldon D, Mbengue M, Mazars C, Galaud JP (2018) Calcium signalling in plant biotic interactions. International Journal of Molecular Sciences 19(3):665. DOI: http://dx.doi.org/10.3390/ijms19030665

Ankenbauer KJ, Loheide SP (2017) The effects of soil organic matter on soil water retention and plant water use in a meadow of the Sierra Nevada, CA. Hydrological Processes 31(4):891–901. DOI: http://dx.doi.org/10.1002/hyp.11070

Charoen-Ung P, Mittrapiyanuruk P (2019) Sugarcane yield grade prediction using random forest with forward feature selection and hyper-parameter tuning. In Advances in Intelligent Systems and Computing 769:33–42). DOI: http://dx.doi.org/10.1007/978-3-319-93692-5_4

Deepa N, Ganesan K (2018) Multi-class classification using hybrid soft decision model for agriculture crop selection. Neural Computing and Applications 30(4):1025–1038. DOI: http://dx.doi.org/10.1007/s00521-016-2749-y

Drury B, Valverde-Rebaza J, Moura MF, Andrade Lopes A de (2017) A survey of the applications of Bayesian networks in agriculture. Engineering Applications of Artificial Intelligence 65:29-42. DOI: http://dx.doi.org/10.1016/j.engappai.2017.07.003

Everingham Y, Sexton J, Skocaj D, Inman-Bamber G (2016) Accurate prediction of sugarcane yield using a random forest algorithm. Agronomy for Sustainable Development 36(2). DOI: http://dx.doi.org/10.1007/s13593-016-0364-z

Jiang T, Gradus JL, Rosellini AJ (2020) Supervised machine learning: A Brief Primer. Behavior Therapy 51(5):675–687. DOI: http://dx.doi.org/10.1016/j.beth.2020.05.002

Jin X, Kumar L, Li Z, Feng H, Xu X, Yang G, Wang J (2018) A review of data assimilation of remote sensing and crop models. European Journal of Agronomy 92:141–152. DOI: http://dx.doi.org/10.1016/j.eja.2017.11.002

Juriga M, Šimanský V, Horák J, Kondrlová E, Igaz D, Polláková N, Buchkina N, Balashov E (2018) The effect of different rates of biochar and biochar in combination with N fertilizer on the parameters of soil organic matter and soil structure. Journal of Ecological Engineering 19(6):153–161. DOI: http://dx.doi.org/10.12911/22998993/92894

Kouadio L, Deo RC, Byrareddy V, Adamowski JF, Mushtaq S, Phuong Nguyen V (2018) Artificial intelligence approach for the prediction of Robusta coffee yield using soil fertility properties. Computers and Electronics in Agriculture 155:324–338. DOI: http://dx.doi.org/10.1016/j.compag.2018.10.014

Liakos KG, Busato P, Moshou D, Pearson S, Bochtis D (2018) Machine learning in agriculture: A review. Sensors18(8):1–29. DOI: http://dx.doi.org/10.3390/s18082674

Luebeck D, Wimmer C, Moreira LF, Alcântara M, Oré G, Góes JA, Oliveira LP, Teruel B, Bins LS, Gabrielli LH, Hernandez-Figueroa HE (2020) Drone-borne differential SAR interferometry. Remote Sensing 12(5):778. DOI: http://dx.doi.org/10.3390/rs12050778

Majnik M, Bosnić Z (2013) ROC analysis of classifiers in machine learning: A survey. Intelligent Data Analysis 17(3):531–558. DOI: http://dx.doi.org/10.3233/IDA-130592

Minasny B, McBratney AB (2018) Limited effect of organic matter on soil available water capacity. European Journal of Soil Science 69(1):39–47. DOI: http://dx.doi.org/10.1111/ejss.12475

Moreno-Barriga F, Díaz V, Acosta JA, Muñoz MÁ, Faz Á, Zornoza R (2017) Organic matter dynamics, soil aggregation and microbial biomass and activity in Technosols created with metalliferous mine residues, biochar and marble waste. Geoderma 301:19–29. DOI: http://dx.doi.org/10.1016/j.geoderma.2017.04.017

Oré G, Alcântara MS, Góes JA, Oliveira LP, Yepes J, Teruel B, Castro V, Bins LS, Castro F, Luebeck D, Moreira LF, Gabrielli LH, Hernandez-Figueroa HE (2020) Crop growth monitoring with drone-borne DInSAR. Remote Sensing 12(4): 615. DOI: http://dx.doi.org/10.3390/rs12040615

Pham LTH, Brabyn L (2017) Monitoring mangrove biomass change in Vietnam using SPOT images and an object-based approach combined with machine learning algorithms. Journal of Photogrammetry and Remote Sensing 128:86–97. DOI: http://dx.doi.org/10.1016/j.isprsjprs.2017.03.013

Poroikov VV, Filimonov DA, Gloriozova TA, Lagunin AA, Druzhilovskiy DS, Rudik AV, Stolbov LA, Dmitriev AV, Tarasova OA, Ivanov SM, Pogodin PV (2019) Computer-aided prediction of biological activity spectra for organic compounds: the possibilities and limitations. Russian Chemical Bulletin 68(12):2143–2154. DOI: http://dx.doi.org/10.1007/s11172-019-2683-0

Rahman MA, Lee SH, Ji HC, Kabir AH, Jones CS, Lee KW (2018) Importance of mineral nutrition for mitigating aluminum toxicity in plants on acidic soils: Current status and opportunities. International Journal of Molecular Sciences 19(10). DOI: http://dx.doi.org/10.3390/ijms19103073

Rajeswari S, Suthendran K (2019) C5.0: Advanced Decision Tree (ADT)classification model for agricultural data analysis on cloud. Computers and Electronics in Agriculture 156:530–539. DOI: http://dx.doi.org/10.1016/j.compag.2018.12.013

van Klompenburg T, Kassahun A, Catal C (2020) Crop yield prediction using machine learning: A systematic literature review. Computers and Electronics in Agriculture 177: 105709. DOI: http://dx.doi.org/10.1016/j.compag.2020.105709

Xu N, Bhadha JH, Rabbany A, Swanson S, McCray JM, Li YC, Strauss SL, Mylavarapu R (2021) Crop nutrition and yield response of bagasse application on sugarcane grown on a mineral soil. Agronomy 11(8):1–15. DOI: http://dx.doi.org/10.3390/agronomy11081526

Zhou X, Lu P, Zheng Z, Tolliver D, Keramati A (2020) Accident prediction accuracy assessment for highway-rail grade crossings using random forest algorithm compared with decision tree. Reliability Engineering and System Safety 200:106931. DOI: http://dx.doi.org/10.1016/j.ress.2020.106931