Machine Learning for Cup Coffee Quality Prediction from Green and Roasted Coffee Beans Features

Chapter · October 2020

DOI: 10.1007/978-3-030-61834-6_5

Machine Learning for Cup Coffee Quality Prediction from Green and Roasted Coffee Beans Features

Javier Andrés Suarez-Peña[1], Hugo Fabián Lobaton-García[2], Jose Ignacio Rodríguez-Molano[1], and William Camilo Rodriguez-Vazquez[2](✉)

[1] Research Group GICOECOL, Universidad Distrital Francisco José de Caldas, Carrera 7 No. 40B–53, Bogotá D.C., Colombia
[2] Research Group GEICOS, Universitaria Agustiniana, Ak. 86 #11b-95, Bogotá D.C., Colombia
William.rodriguez.v@gmail.com

Abstract. Coffee is one of the main exported products of Colombia. It is grown in different regions throughout the territory and is recognized worldwide for its flavor and freshness. Its quality is evaluated by professional tasters, who taste the coffee drink obtained from roasted coffee beans. They qualify it according with the platform or method requested by customers. This study proposes the use of different Machine Learning (ML) algorithms for the prediction of cup coffee quality, based on a set of measurements made to almond and roasted coffee beans. The data was obtained with the support of Almacafé, a company belonging to the National Federation of Coffee Growers (FNC) of Colombia. The classification results with the validation set, showed a higher accuracy with the Neural Network algorithm, with an average score of 81% for a 10-fold stratified cross validation. This work demonstrates the possibility of qualifying cup coffee quality with ML algorithms.

Keywords: Cup coffee quality · Machine Learning · Neural network · Classification · SVM

1 Introduction

Colombia is the third-largest producer of coffee in the world and the main producer of Arabian coffee [1], recognized by its coffee quality, which goes through rigorous selection processes and a final qualification that allows its definition for exportation [2]. Coffee exports in Colombia for 2017 were 710,440 Metric tons and 710,836 for 2018, representing 7% and 5% of total exports for their respective years [3]. Colombian coffee is produced by more than 560,000 small coffee producers, who are grouped in the National Federation of Coffee Growers (FNC) and follow standards to guarantee coffee quality [2]. To assess the quality of coffee, professional tasters are required, specialized in recognizing the different cup coffee features, who transfer their sensory experience to a score. His knowledge is based on empirical models of interpretation of coffee quality [1], however, human errors can occur in some of the measurements made,

J. C. Figueroa-García et al. (Eds.): WEA 2020, CCIS 1274, pp. 48–59, 2020.
https://doi.org/10.1007/978-3-030-61834-6_5

which requires standardized statistical methods, forms, and analyzes, in addition to the experience of the tasters in interpreting the results [2]. It is necessary to find statistical methods based on numerical measurements of physicochemical properties that can be performed by a wider range of professionals, with reproducible results and with less variation. Alternatively, ML techniques emerge as a novel alternative given their ability to emulate human activities and operations.

Among the examples of ML applications related to coffee, we found works such as the one proposed by [3], who proposed a model capable of classifying green coffee beans by differences in their color, using neural networks first to convert RGB space colors to CIELab. Later they use a Bayesian classifier, to classify the images of coffee in four colors white, green, blue green and cane green. This study was justified in the importance of color for coffee quality definition, associated with a greater market value. [4] made a classifier using a neural network algorithm for green coffee samples containing several grains, corresponding to Indonesian standards. [5], performs a segmentation of coffee fruits to classify them according to their maturation state. To do this, the author uses a connectivity algorithm, which uses homogeneity criteria to group pixels.

Another example is an electronic selection system for specialty coffee (“excelso”) based on color through image processing from [6], where they trained a neural network to recognize a coffee with defects (bad) of one without defects. [7] presented a method to classify coffee based on its color. A similar approach is addressed by [8], who, using a sphere spectrophotometer, quantitatively determined the color of the fruit during its different stages of development, which, although they do not use ML algorithms, their results can be used for the application of classifiers.

According to the articles reviewed, it can be inferred that most authors have aimed their studies towards the classification of coffee after harvest, working with fruits or grains due to their importance in quality definition, either by building mobile identification systems, models or algorithms, for segmentation or segregation of defects, such as immature or overripe fruits. In addition, no articles were found that use ML to predict cup coffee quality by analyzing the characteristics of fruits, almonds, green beans or roasted coffee beans. Therefore, the objective of this article is to propose a ML model that allows to predict cup coffee quality as high or low, based on variables measured in green coffee beans in Colombia.

2 Materials and Methods

2.1 Coffee Samples and Collected Data

The experimental data were provided by Almacafé. The data correspond to measurements made on samples from different parts of Colombia of almond coffee beans. It obtained by wet process and analyzed as they arrived at the facilities of the Almacafé laboratory, located in the city of Bogotá during the period from May to August 2019. Each sample is received in packages of 500 gr of parchment coffee beans according to FNC definitions. From this sample 250 gr has taken, to the wrap or parchment of the grain. This it is removed by a process called threshing. The process leaves the green almond coffee an average weight of 210 gr per sample. Table 1 lists the measurements made. In total 56 samples were analyzed.

Regarding the humidity measurement, it has done in equipment calibrated according to ISO 6673 method. The particle size distribution gives an idea of superior quality if its size is above the 15/64-inch mesh. According to the experience of Almacafé, a small grain would likely to be a not fully developed grain, causing its attributes being not optimal. The color is measured with a spectrophotometer on the CIELAB and CIEXYZ scale defined by the Commission Internationale de lÉclairage (CIE), as to [9]. The spectrophotometers collect the entire spectrum of each wavelength and by means of an algorithm they transform it into values that resemble human perception. For this work the color of green coffee beans and roasted beans was measured. For each of these two grains, the color space coordinates L *, a *, b *, X, Y and Z are recorded in the database. As for sensory evaluation, 10 coffee quality descriptors are measured on an ordinal scale from 1 to 10. These are aroma, taste, residual taste, acidity, body, balance, uniformity, cup cleaning, sweetness and overall cup score (overall). To make this measurement, the roasting of the green coffee beans of each sample is carried out, under a standardized method and with controlled conditions of Almacafé. Then, the roasted coffee is ground and 5 cups of coffee are prepared, which are evaluated by professional tasters. An important precision regarding the tasters is that they perform the evaluation, without any prior knowledge of the physical analyzes results carried out on the coffee beans, or on the traceability of the crop, allowing them to have no bias at the time of making the qualifications to the respective cups of coffee.

Considering the empirical evidence and the experience of Almacafé, a high correlation between the measured variables and the results of the cupping would be expected. According to Siswoputrano (1993) cited by [4], the main aspects for determining the quality of coffee are size, shape, color, defects and other materials.

Table 1. Summary of measurements made to coffee samples

Variable	Unit of measurement	Measurement equipment	Measurement method	Sample size
Altitude	meters	Google Earth	Approximate location according to the sample place location	500 g taken from several sacks
Moisture	percentage	MOISTURE KAPPA AK-60-B	AL-CC-EC-P-0003 Parchment coffee - Excelso coffee: Moisture determination (kappa)	400 g
Initial parchment weight	grams	Digital Scales	Direct weight on digital scale	250 g

(*continued*)

Table 1. (*continued*)

Variable	Unit of measurement	Measurement equipment	Measurement method	Sample size
Green almond weight	grams	Digital Scales	Direct weight on digital scale	210 g approx.
Waste (peel)	percentage	Manual	It is the percentage resulting from dividing the difference between the weight of the parchment and the almond, divided by the weight of the parchment	210 g approx.
Defective grains group 1	Units and grams	Manual	The beans are spread on a table, where is proceeded to identify, separate, count and weigh the beans with color defects	210 g approx.
Defective grains group 2	Units and grams	Manual	The beans are spread on a table, where is proceeded to identify, separate, count and weigh the beans with physical defects other than color	210 g approx.
Granulometric distribution	Mesh retention percentage	ZARANDA MECANICA	AL-CC-EC-P-0008 Green Coffee: Granulometric Distribution. Mesh sizes	210 g approx.
Color	Cielab	spectrophotometer LABSCAN EX	Direct scale measurement CIELAB y CIEXYZ	210 g approx.
Sensory assessment	Ordinal scale	Professional taster	AL-CC-EC-I-0002 SENSORY ANALYSIS PREPARATION OF THE TEST	5 coffee cups

2.2 Support Vector Machine

SMV are supervised learning algorithms (MathWorks, s.f.), proposed by [10], used for binary classification or regression, in pattern recognition problems. This classification is based on the basis of optimal separation between classes, so that if the classes are subject to separation, the result is chosen to segregate the classes as much as possible. [11]. The SVM builds a hyperplane in a space of very high dimensionality (which can become infinite) that separates the classes that are held in a data set. A good dissociation between classes will generate a correct categorization of the new sample, in other words, it is required to find the maximum separation to the points closest to the generated hyperplane [12]. Because SVMs are a known method of ML implemented in various fields of knowledge, libraries and tools have emerged to facilitate their use, one of them is LIBSVM (SVM library), used within the Python-based Scikit-Learn library. The LIBSVM allows several SVM formulations for classification, regression and distribution estimation [13]. One of these formulations is the Classification of support vectors C, in which, there are two training vectors Xi that belongs to the set Rn, where i = 1, …, l, in two classes, and an indicator vector and that belongs to the set Rl such that yi belongs to {1,−1}, C-SVM solve the primary optimization problem, according with Boser et al.,1992; Cortes & Vapnik, 1995; cited by [13], as follows,

$$\min_{w,b,\xi} \frac{1}{2}w^T w + C \sum\nolimits_{i=1}^{l} \xi_i \tag{1}$$

Subject to,

$$y_i\left(w^T \emptyset(x_i) + b\right) \geq 1 - \xi_i, \xi_i \geq 0, i = 1, \ldots, l \tag{2}$$

where $\emptyset(x_i)$ Assign xi to a space of larger dimensions and C > 0 is the regularization parameter. In view of the possible high dimensionality of the variable W, the dual problem is solved with the following equation [13],

$$\min_{\alpha} \frac{1}{2}\alpha^T Q\alpha - e^T \alpha \tag{3}$$

Subject to,

$$0 \leq \alpha_i \leq C \quad i = 1, \ldots, l \tag{4}$$

Where e = [1, …,1] T is the vector of each one of these, Q is a positive semi-defined matrix l by l and α is Lagrange multiplier that acts as an inverse of the C term, which solves the problem of dual optimization. After the application of the previous equation, the optimal satisfaction equation (w) is applied [13],

$$w = \sum\nolimits_{i=1}^{l} y_i \alpha_i \emptyset(x_i) \tag{5}$$

And finally the decision function [13],

$$sgn\left(w^T \emptyset(x) + b\right) = sgn\left(\sum\nolimits_{i=1}^{l} y_i \alpha_i K(x_i, x) + b\right) \tag{6}$$

2.3 Neural Networks

In ML one of the most used models is neural networks (NN). An artificial neural network, is a computational model consisting of a number of basic computational units called neurons, that are connected to each other forming a network of communication, allowing complex computations [14]. Where each node is a neuron and the arcs represent the connections between them. Neurons are grouped in layers. An input layer that receives the input data of the model, which passes through the next group, called hidden layers. In this group, we can find as many layers as the complexity of the problem or the data requires. Finally, an output layer, where the values to the functions resulting from the network processing will be obtained. The way in which networks acquire their knowledge is contained in the connection weights that each neuron (node) possesses within the network. They compute their values (connection weights) during the training phase [15].

The way in which neurons interconnect allows two types of architecture to be distinguished [15]: The feedback architecture, in which there are connections between neurons of the same or the previous layer; The feedforward architecture [16], without feedback connections, that is, the signals go only to the neurons of the next layer. Neural networks can be configured in multiple ways (the possibilities become infinite), so the choice of the optimal configuration for the data or problem handled, must be primarily an objective function of the application. The most common neural networks are Perceptron and Multilayer Perceptron.

Perceptron: t is the simplest network, which is made up of a single neuron, with n inputs and a single output Rosenblatt (1958) Minsky and Papert (1969) cited by [15].

Multilayer Perceptron (MLP): networks with an input layer, intermediate layers (which may range from one to those required) and an output layer are called multilayer perceptron [16].

2.4 Programming Tools

The development of this work is achieved using the free execution environment of Google Colaboratory web books. It allows users to write and execute the Python programming language, access to hosted execution environments, providing free use of up to 25 GB of RAM, and large capacity GPU processors, which reduce the execution time of the algorithms. The link to the Colab book is shared in Appendix A. The Keras API is used for the implementation of the neural network [17], and the Scikit-Learn module of SVM for the SVC algorithm [18].

2.5 Coffee Quality Classification

The authors [1], mention that there are different methods to evaluate coffee quality, most of them are based on a score scale of 0 to 100, similar to the wine ratings. Score augmentations are obtained from the ratings assigned to non-standardized descriptors qualified on a smaller scale, which together give a result relative to 100 according to their quality score. Among the methods mentioned by these authors, there is a use of similar descriptors to those used by Almacafé for this work. The first step to define the classification categories is to add the scores selected in the different descriptors, grouping them into a variable called "Sum". Then a separation of the data is performed looking for a balance between the data. Although there are scores of great interests than others, like scores greater than 80 points for the qualification of a coffee as excellent, in this work scores were obtained between the range of 32.5 to 63.5, because the sample selection was random. Therefore, it is decided to divide the results into two groups, samples with low quality scores and samples with high quality scores, with the differentiating value being a score of 51.5. In this way, 28 data were obtained in each group. Subsequently, the variables "X" and "y" are created to store in "X" the input data and in "y" the respective labels for each sample. With the "train_test_split" function of Scikit-Learn, the data is separated in training and validation, using a validation set size of 20% of the total samples and a seed of 4 to obtain a balanced amount of the categories in the validation data.

2.6 Model Validation

Following [19, 20], a stratified cross validation is applied with 10 folds for model validation. This method divides the whole dataset in 10 groups, where it uses 9 for training and one for test, interchanging the test set with each one of the 10 groups, until it evaluates all of them. This method has the advantage that considers the elements of each subset of categories defined and is designed to have the same number of examples from each class, on both training and test sets.

3 Results and Discussion

For data processing through defined ML algorithms, a statistical analysis of the data was carried out in first place, the categorical variables were transformed into numerical. With the purpose of validating if the normalized data would improve the prediction, a normalization of the data was performed as suggested by [21]. A correlation analysis between the variables was also conducted.

In this work, the color coordinates on green coffee beans and roasted coffee beans were measured. Additionally, this information was used to predict the quality of the coffee drink in the defined groups. Similarly [3], measured the coordinates of the CIELAB space in green coffee beans to classify them with the help of a neural network and a Bayesian classifier. [22] used an electronic tongue (ET) for coffee quality prediction, trained a neural network with the measurements obtained with the ET and the quality scores assigned by professional tasters, to classify in the defined groups. In this work the quality scores assigned by professional tasters were also used, with the difference that the variables or input measurements, were not made on the coffee drink, but on the beans, resembling the filter and quality evaluation performed by Almacafé and secondly because it could generate savings by reducing the need for beverage preparation. [6] they developed a machine capable of selecting defective coffee beans from good ones, from the implementation of the multilevel thresholding segmentation method, applied to images of good and defective beans, which allowed to determine the ranges of each color of the RGB space for the categories defined. In contrast, in this work the CIELAB space was used, where the values of each component of the color space, together with the other variables, contribute a weight to each cupping result obtained.

3.1 Programming Tools

The neural network was made using the Keras API [17]. An input layer of 36 nodes was defined, corresponding to each of the measurements made per sample. Then the size was reduced by half to match the defined classes. Table 2 lists the parameters used in the neural network. The dataset used was the normalized version.

Table 2. Neural network parameters

PARAMETERS	VALUE
Input size	36
Number of layers	4
NN architecture	Input layer 36 nodes – dense layer 18 nodes – Dropout layer 50% - dense output layer with 2 nodes and activation function “Softmax” [23]
Activation function (Dense layer)	ReLu [24]
Kernel initializer	he-normal [25]
Optimizer	Adam [26]
Learning rate	1e−2
Learning rate reduction	50% after 4 epochs without improvement in validation accuracy (val_acc)
Loss function	“categorical_crossentropy”
Early stopping (callback)	Patience de 20 epochs monitoring validation accuracy with a maximum number of 100 epochs
Reduce (callback)	Factor = 0,5, patience = 3, min = 1e−8
Batch size	1
Epochs	26

The initial results of this network are shown in Fig. 1. The accuracy obtained was 83%. It is observed that this network was better classifying the low score samples, correcting in 100% of the cases, without modifying, that improvement in category 0 implied only 67% of successes in class 1, as shown in the Confusion matrix of Fig. 1. The classification metrics, confusion matrix and classification report were used. This lats one, groups the following metrics: accuracy, precision, recall and F1 score.

After evaluating the initial results and tuning the hyper parameters of the neural network, the next step was to run the stratified cross validation. A for loop was defined to visualize the elements of each fold and to let the use of the callbacks early and reduce, as well as some auxiliary functions to plot each epoch and the confusion matrix. On Table 3 are resumed the results of each fold in terms of accuracy and F1 scores.

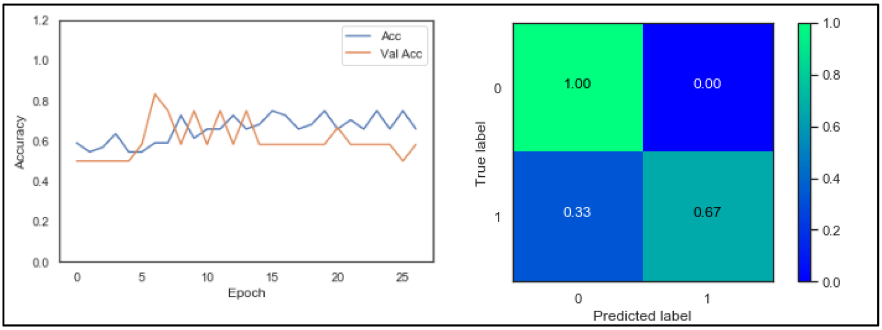

Fig. 1. Neural network results. Left, training precision curve and validation. Right, confusion matrix. Source: self-made.

Table 3. Results of the stratified cross validation

Iteration	Accuracy	F1-0	F1-1	Test samples selected	Class of test samples
1	1	1	1	[0 1 2 3 4 8]	[1. 1. 0. 1. 0. 0.]
2	0,67	0,75	0,5	[5 6 7 11 12 13]	[1. 1. 1. 0. 0. 0.]
3	0,83	0,8	0,86	[9 10 14 15 16 17]	[1. 1. 0. 0. 0. 1.]
4	0,67	0,75	0,5	[18 19 20 21 22 24]	[0. 1. 1. 0. 1. 0.]
5	0,67	0,75	0,5	[23 25 26 27 31 32]	[1. 0. 0. 0. 1. 1.]
6	0,83	0,8	0,86	[28 29 30 33 39 42]	[0. 0. 0. 1. 1. 1.]
7	0,83	0,86	0,8	[34 35 36 43 45 47]	[0. 0. 0. 1. 1. 1.]
8	0,83	0,86	0,8	[37 38 40 48 49 50]	[0. 0. 0. 1. 1. 1.]
9	0,75	0,8	0,67	[41 44 51 52]	[0. 0. 1. 1.]
10	1	1	1	[46 53 54 55]	[0. 1. 1. 0.]

The average accuracy was 81%, showing a better F1 score for low quality samples in comparison with high quality coffee. From the results was of interest to analyze the reasons of the folds (or iterations) where the high-quality samples were misclassified, founding that those samples were close of the partition point defined which may confuse the neural network. This also shows that more classes could be defined with the corresponding incremental number of samples per class. In contrast the folds with 100% accuracy come from sets where the difference on the score of the test samples were at least 3 points or more away from the partition point of class definition.

3.2 SVM (SVC) Results

The SVC algorithm was implemented using the Scikit-Learn Libsvm library, additional "GridSearchCV" function was used to define the penalty parameter C and gamma, using a Kernel = "rbf" (radial basis function), obtaining a value of C = 1000, a gamma = 0.01. The other parameters of this algorithm were left as they come by default in the "SVC" function.

From Fig. 2 the resulting accuracy, was the same achieved with the neural network, however the F1 scores show more balance on the performance of the model for both classes. Alike the previous algorithm a stratified cross validation is conducted to check the performance of the model with different training and test sets. In Table 4, are shown the results of the cross validation using the SVC.

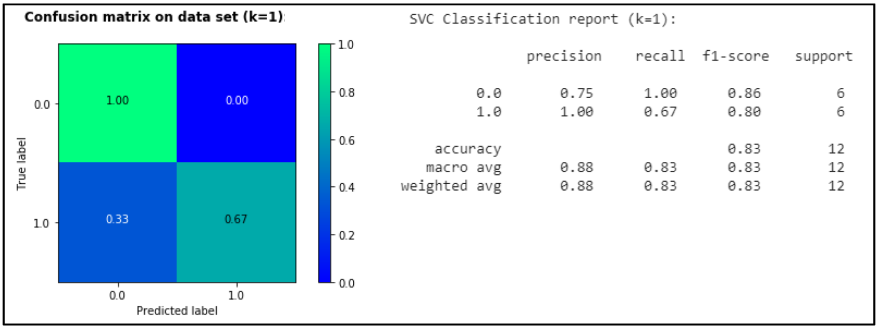

Fig. 2. Confusion matrix and classification report for SVC. Source: self-made.

Table 4. Cross validation results using SVC

Iteration	Accuracy	F1-0	F1-1	Test samples selected	Class of test samples
1	0,5	0,57	0,4	[0 1 2 3 4 8]	[1. 1. 0. 1. 0. 0.]
2	0,67	0,75	0,5	[5 6 7 11 12 13]	[1. 1. 1. 0. 0. 0.]
3	0,5	0,57	0,4	[9 10 14 15 16 17]	[1. 1. 0. 0. 0. 1.]
4	0,67	0,75	0,5	[18 19 20 21 22 24]	[0. 1. 1. 0. 1. 0.]
5	0,83	0,86	0,8	[23 25 26 27 31 32]	[1. 0. 0. 0. 1. 1.]
6	0,83	0,86	0,8	[28 29 30 33 39 42]	[0. 0. 0. 1. 1. 1.]
7	1	1	1	[34 35 36 43 45 47]	[0. 0. 0. 1. 1. 1.]
8	0,17	0	0,29	[37 38 40 48 49 50]	[0. 0. 0. 1. 1. 1.]
9	0,5	0,67	0	[41 44 51 52]	[0. 0. 1. 1.]
10	0,75	0,67	0,8	[46 53 54 55]	[0. 1. 1. 0.]

A 64% average accuracy resulted from the cross validation run. With this algorithm there was minimal difference between using the original dataset or the normalized version. It was also applied a Principal Component Analysis (PCA) to reduce dimensionality, however there was not any improvement on the results. One of most significant results was the computational time with 1 s (or less) for SVC and 20 to 60 s for NN algorithm.

4 Conclusions

This research shows that cup coffee quality can be predicted as high or low from the analysis of green coffee beans features. The neural network model defined can be an alternative to expert cup coffee qualification, reducing the cost of the evaluation, saving time in analysis and cup coffee preparation and providing an additional tool for coffee taster and coffee producers. Neural Network algorithm had a better performance in comparison with support vector classification. The early stopping callback was determinant to stop the NN training when the model was not improving the validation accuracy score. Despite the callbacks defined, the computational time was higher for the NN that the required for SVC algorithm. Future work may consider predicting coffee quality by using convolutional neural networks on green coffee beans pictures or a functional model to mix numerical data with images.

Acknowledgement. Special thanks are given to the Office of coffee quality Almacafé, for its interest in this work and for providing samples, measuring input attributes and cupping them.

Appendix A

Link to the notebook used in Colab and the corresponding database:
https://github.com/Javiersuing/GitHub/blob/master/AlmacafeDataBase_CrossV_v4.ipynb.

References

1. Thomas, E., Puget, S., Valentin, D., Songer, P.: Sensory Evaluation-Profiling and Preferences. Elsevier Inc., London (2017)
2. Salamanca, C.: Métodos Estadísticos Para Evaluar La Calidad Del Café. Universitat de Girona (2015)
3. de Oliveira, E.M., Leme, D.S., Barbosa, B.H.G., Rodarte, M.P.: A computer vision system for coffee beans classification based on computational intelligence techniques. J. Food Eng. **171**, 22–27 (2015). https://doi.org/10.1016/j.jfoodeng.2015.10.009
4. Faridah, F., Parikesit, G.O.F., Ferdiansjah, F.: Coffee bean grade determination based on image parameter. TELKOMNIKA (Telecommun. Comput. Electron. Control) **9**, 547–554 (2011). https://doi.org/10.12928/telkomnika.v9i3.747
5. Montes, N.: Segmentación De Imágenes De Frutos De Café En El Proceso De Beneficio. Universidad Nacional de Colombia, sede Manizales (2003)
6. Ruge, I.A., Pinzon, A.S., Moreno, D.E.: Sistema de selección electrónico de café excelso basado en el color mediante procesamiento de imágenes. Rev Tecnura **16**, 84–93 (2012). http://dx.doi.org/10.14483/udistrital.jour.tecnura.2012.4.a06
7. Ramos Giraldo, P.J., Sanz Uribe, J.R., Oliveros Tascón, C.E.: Identificación y clasificación de frutos de café en tiempo real, a través de la medición de color. Cenicafé **61**, 315–326 (2010)
8. Carvajal, J.J., Aristizábal, I.D., Oliveros, C.E., Mejía, J.W.: Colorimetría del Fruto de Café (Coffea arabica L.) Durante su Desarrollo y Maduración. Rev Fac Nac Agron Medellín **1**, 37–48 (2006)

9. Tobijaszewska, B., Mills, R., Jons, J.: El uso de la espectrometría para la medición simultánea del color y la composición en muestras de alimentos. In: FOSS (2018). https://www.fossanalytics.com/-/media/files/documents/papers/meat-segment/using-spectrometry-for-simultaneous-measurement_es.pdf
10. Cortes, C., Vapnik, V.: Support-vector networks. In: Saitta, L. (ed.) Machine Learning, pp. 273–297. KlugerAcademic Publishers, Boston (1995)
11. Nascimento, R.F.F., Alcântara, E.H., Kampel, M., et al.: O algoritmo SVM: avaliação da separação ótima de classes em imagens CCD-CBERS-2. XIV Simpósio Bras Sensoriamento Remoto 2079–2086 (2009)
12. Gala, Y.: Algoritmos SVM para problemas sobre big data. Universidad Autonoma de Madrid (2013)
13. Chang, C.C., Lin, C.J.: LIBSVM: a library for support vector machines. ACM Trans. Intell. Syst. Technol. **2** (2011). https://doi.org/10.1145/1961189.1961199
14. Shalev-Shwartz, S., Ben-David, S.: Understanding Machine Learning: From Theory to Algorithms. Cambridge University Press, New York (2014)
15. Gallo, C.: Artificial Neural Networks: tutorial (2015)
16. Hornik, K., Stinchcombe, M., White, H.: Multilayer feedforward networks are universal approximators. Neural Netw. (1989). https://doi.org/10.1016/0893-6080(89)90020-8
17. Chollet, F.: Keras (2015) (2017). http://keras.io/
18. Pedregosa, F., Varoquaux, G., Gramfort, A., et al.: Scikit-learn: machine learning in Python. J. Mach. Learn. Res. **12**, 2825–2830 (2011)
19. Forman, G., Forman, G., Scholz, M., Scholz, M.: Apples-to-apples in cross-validation studies: pitfalls in classifier performance measurement. HP Labs **12**, 49–57 (2009). https://doi.org/10.1145/1882471.1882479
20. Zhang, Y.D., Yang, Z.J., Lu, H.M., et al.: Facial emotion recognition based on biorthogonal wavelet entropy, fuzzy support vector machine, and stratified cross validation. IEEE Access **4**, 8375–8385 (2016). https://doi.org/10.1109/ACCESS.2016.2628407
21. Kotsiantis, S.B., Kanellopoulos, D.: Data preprocessing for supervised leaning. Int. J. Comput. Sci. **1**, 1–7 (2006). https://doi.org/10.1080/02331931003692557
22. Ferreira, E.J., Pereira, R.C.T., Delbem, A.C.B., et al.: Random subspace method for analysing coffee with electronic tongue. Electron. Lett. (2017) https://doi.org/10.1049/el:20071182
23. Dunne, R., Campbell, N.: On the pairing of the Softmax activation and cross-entropy penalty functions and the derivation of the Softmax activation function. In: Proceedings of 8th Australian Conference Neural Networks (1997). https://doi.org/10.1.1.49.6403
24. Agarap, A.F.: Deep learning using rectified linear units (ReLU), pp. 2–8 (2018)
25. He, K., Zhang, X., Ren, S., Sun, J.: Delving deep into rectifiers: surpassing human-level performance on imagenet classification. In: Proceedings of IEEE International Conference Computer Vision 2015, pp. 1026–1034 (2015). https://doi.org/10.1109/ICCV.2015.123
26. Kingma, D.P., Ba, J.: Adam: a method for stochastic optimization, pp. 1–15 (2014)