Artículo original

A new dataset for coffee rust detection in Colombian crops base on classifiers

Un nuevo conjunto de datos para la detección de roya en cultivos de café Colombianos basado en clasificadores

David Camilo Corrales
dcorrales@unicauca.edu.co
Grupo Ingenieripia Telemática
Grupo Estudios Ambientales
Universidad del Cauca, Popayán-Colombia

Agapito Ledezma
ledezma@inf.uc3m.es
Gr. de Control, Aprendizaje y Optimización de Sistemas
Universidad Carlos III, Madrid-España

Andrés J. Peña Q.
andres.pena@cafedecolombia.com
Centro Nacional de Investigaciones de Café, Chinchiná, Colombia

Javier Hoyos
javier.hoyos@supracafe.com
Supracafé S.A., Popayán, Colombia

Apolinar Figueroa
apolinar@unicauca.edu.co
Grupo Estudios Ambientales
Universidad del Cauca, Popayán-Colombia

Juan Carlos Corrales
jcorral@unicauca.edu.co
Grupo Ingeniería Telemática
Universidad del Cauca, Popayán-Colombia

Fecha de recepción: Marzo 11 de 2014
Fecha de aceptación: Abril 29 de 2014

Keywords
Coffee Rust, Classifier, SVR, BPNN, M5

Palabras clave
Roya en el Café, Clasificador, SVR, BPNN, M5

Abstract

Coffee production is the main agricultural activity in Colombia. More than 350.000 Colombian families depend on coffee harvest. Since coffee rust disease was first reported in the country in 1983, these families have had to face severe consequences. Recently, machine learning approaches have built a dataset for monitoring coffee rust incidence that involves weather conditions and physic crop properties. This background encouraged us to build a dataset for coffee rust detection in Colombian crops through data mining process as Cross Industry Standard Process for Data Mining (CRISP-DM). In this paper we define a proper data to generate accurate models; once the dataset is built, this is tested using classifiers as: Support Vector Regression, Backpropagation Neural Networks and Regression Trees.

Resumen

La producción de café es la principal actividad agrícola en Colombia. Más de 350.000 familias colombianas dependen de la cosecha de café. En este sentido, la roya fue reportada por primera vez en el país en 1983, y desde entonces estas familias han tenido que enfrentar graves consecuencias. Recientemente, diversos enfoques basados en aprendizaje automático han construido un conjunto de datos para el monitoreo de la incidencia de la roya del café, teniendo en cuenta las condiciones climáticas y las propiedades físicas de los cultivos. Estas investigaciones motivaron la creación de un conjunto de datos para la detección de la roya en cultivos Colombianos a través del proceso de minería de datos CRISP-DM. En este trabajo se definió un conjunto de datos con el objetivo de generar clasificadores precisos; una vez construido el conjunto de datos, fue probado mediante tres clasificadores: Maquinas de vector de regresión, Redes neuronales con propagación hacia atrás y Árboles de regresión.

Colciencias tipo 1

I. Introduction

Coffee production is the main agricultural activity in Colombia. More than 350.000 Colombian families depend on coffee harvest for their sole income. Diseases, pests and even low prices cause a big impact on the economic and social aspects of the main coffee-growing regions. Coffee rust, first reported in 1983 (Shieber & Zentmyer, 1984), is the most important and severe disease currently affecting the production of Colombian coffee. Varieties of coffee, that could resist this disease, have been developed through improvement with genes of Timor Hybrid (plant that features natural resistance to the disease) as a solution to the rust problem (Zapata & Ruíz, 1988). However, more than 50 percent of the country's coffee crop is still susceptible in the productive phase. Studies on coffee rust have concluded that the spores carrying the infection are spread by climatic elements such as wind and rainfall (Becker, 1979). Wind being the vector for long distance spore transport, while precipitation droplets are responsible for vertical propagation from infected leaves or soil (Becker, 1979). Once spores make contact with a susceptible leaf, the infection process is increased by high shadow index, high humidity (atmosphere and leaf), soil acidity, high coffee tree density and low soil fertility. The dataset proposed herein, joins each of the favorable conditions that coffee rust requires to infect the crop, by taking prophylactic measures (biological, chemical and weather), in order to allow the prevention of the onset of the disease.

Brazilian machine learning researchers, have, in recent times, built a dataset for monitoring the coffee rust incidence at the Experimental Farm of the Procafé Foundation, in Varginha, Minas Gerais, Brazil (21°34'0"S, 45°24'22"W), during 8 years (October, 1998 – October, 2006), which includes 182 examples and 23 attributes that involves weather conditions and physic crop properties (Cintra, Meira, Monard, Camargo, & Rodrigues, 2011; Luaces, Rodrigues, Alves Meira, & Bahamonde, 2011; Luaces et al., 2010; Meira, Rodrigues, & Moraes, 2008; Meira & Rodrigues, 2009; Meira, Rodrigues, & Moraes, 2009; Pérez-Ariza, Nicholson, & Flores, 2012). This dataset is tested through such classifiers as: decision trees, regression Support Vector Machines, non-deterministic classifiers and Bayesian Networks.

Nowadays, there are not Colombian approaches reported about coffee rust detection based on machine learning techniques and coffee producers lack technological systems to detect coffee rust incidence, in order to improve coffee quality and reduce investment costs. Therefore, the Cross Industry Standard Process for Data Mining (CRISP-DM) propose (Wirth, 2000) six phases (Business Understanding, Data Understanding, Data Preparation, Modeling, Evaluation, Deployment) to build a Data Mining solution. Our approach aims to define a proper data to generate accurate models (the dataset proposed herein, joins each of the favorable conditions that coffee rust requires to

infect the crop, by taking prophylactic measures); once the dataset is built, this is tested using classifiers as: Support Vector Regression, Backpropagation Neural Networks and Regression Trees.

The rest of the paper is organized as follows: data collection and classifiers are introduced in Section 2. Section 3 presents results and discussion, while Section 4 reports the conclusions.

II. Material and Methods

This section describes the data collection process and the generation of datasets used in experiments, and introduces three classifiers for prediction: Support Vector Regression, Backpropagation Neural Network, and Regression Tree M5.

a) Data Collection

The data used in this work was collected every three months for 18 plots (1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 37, in the Figure 1), closest to weather station (In Figure 1 represented as WS) at the Experimental Farm (Naranjos) of the Supracafé, in Cajibio, Cauca, Colombia (21°35'08"N, 76°32'53"W), during the last 3 years (2011-2013). The dataset includes 147 examples from the total of 162 available ones. The remaining 15 samples were discarded due to problems in the collection process.

The dataset is composed for 21 attributes which are divided in 4 categories: Weather conditions (6 attributes), Soil fertility properties (5 attributes), Physic crop properties (6 attributes), and Crop management (4 attributes). Table 1 describes the 21 attributes and its data type: Numerical (Nu) or Nominal (No).

Figure 1. Experimental Farm: Los Naranjos - Cajibio (Cauca)

Table 1. Attributes of data training for coffee rust detection at the Experimental Farm: Los Naranjos

Category	Attributes		Type
Weather conditions	1	Relative humidity average in the last 2 months.	Nu
	2	Hours of relative humidity > 90% in the last months.	Nu
	3	Temperature variation average in the last month.	Nu
	4	Rainy days in the last month.	Nu
	5	Accumulated precipitation in the last 2 months.	Nu
	6	Nightly accumulated precipitation in the last month.	Nu
Soil fertility properties	7	pH.	Nu
	8	Organic material.	Nu
	9	K.	Nu
	10	Ca.	Nu
	11	Clay.	Nu
Physic crop properties	12	Coffee Variety.	No
	13	Density of plants per hectare.	Nu
	14	Plant spacing.	Nu
	15	Furrow spacing.	Nu
	16	Crop age.	Nu
	17	Percentage of shade.	Nu
Crop management	18	Coffee rust control in the last month.	No
	19	Coffee rust control in the last 3 months.	No
	20	Fertilization in the last 4 months.	No
		Accumulated coffee production in the last 2 months.	Nu

In this sense, the class (variable to predict) was defined as, the Incidence Rate of Rust (IRR). IRR is calculated following the methodology developed by Cenicafé (Rivillas-Osorio, Serna-Giraldo, Cristancho-Ardila, & Gaitán-Bustamante, 2011) for a plot with area lower or equal of one hectare. The steps of the methodology are presented next:

1. The farmer must be standing in the middle of the first furrow and he has to choose one coffee tree and pick out the branch with greater foliage for each level (high, medium, low); leaves of selected branches are counted as well as infected ones for rust.
2. The farmer must repeat the step 1 for every tree in the plot until 60 trees are selected. It must be taken in consideration that the same number of trees must be selected in every furrow (e.g. if plot has 30 furrows, the farmer selects two coffee trees for each furrow).

Finished the step 1 and 2, the leaves of coffee trees selected (60) are added as well as the infected leaves of rust. Later it must be calculated the Incidence Rate of Rust (IRR) through Equation 1.

$$IRR = \frac{Infected\ leaves\ of\ 60\ coffee\ trees\ selected}{Total\ leaves\ of\ 60\ coffee\ trees\ selected}\ x\ 100$$

Equation 1

b) Classifiers

Support Vector Regression (SVR)

SVR is a supervised learning algorithm based on statistical learning theory and structural risk minimization principle (Vapnik, 1999; 2000). It can be expressed as the following equation:

$$f(x) = w^t\varphi(x) + b$$

Equation 2

Where $\varphi(.)$ is a non-linear mapping which takes the input data points into a higher dimensional feature space, w is a vector in the feature space and b is a scalar threshold (Balasundaram & Gupta, 2014). On the other hand, the unknowns w and b are solved as the solution of the constrained quadratic programming problems (Smola & Schölkopf, 2004) given below (Equation 3):

$$\min_{w,b,\varepsilon_1,\varepsilon_2} \frac{1}{2}w^tw + C(e^t\varepsilon_1 + e^t\varepsilon_2)$$

subject to:

$$\begin{aligned} y_i - w^t\varphi(x_i) - b &\le \varepsilon + \varepsilon_{1i}\ ;\ \varepsilon_{1i} \ge 0\ \ \forall_i = 1, \dots, m \\ w^t\varphi(x_i) + b - y_i &\le \varepsilon + \varepsilon_{2i}\ ;\ \varepsilon_{2i} \ge 0\ \ \forall_i = 1, \dots, m \end{aligned}$$

Equation 3

Where $\varepsilon_1 = (\varepsilon_{11}, \dots, \varepsilon_{1m})^t, \varepsilon_2 = (\varepsilon_{21}, \dots, \varepsilon_{2m})^t$, are vectors of slack variables and $C > 0,\ \varepsilon > 0$ are input parameters.

Rather than solving the primal problem considered above, it is introduced Lagrange multipliers $U_1 = (U_{11}, \dots, U_{1m})^t$ and $U_2 = (U_{21}, \dots, U_{2m})^t$ in R^m and it is applied the kernel function $k(.,.)$ (Cristianini & Shawe-Taylor, 2000; Vapnik, 2000) (Equation 4):

$$\min_{u_1,u_2} \frac{1}{2}\sum_{ij=1}^{m}(u_{1i} - u_{2i})\,k(x_i, x_j)(u_{1j} - u_{2j}) + \varepsilon\sum_{i=1}^{m}(u_{1i} + u_{2i}) - \sum_{i=1}^{m} y_i(u_{1i} - u_{2i})$$

subject to:

$$\sum_{i=1}^{m}(u_{1i} - u_{2i}) = 0\ ;\ \ 0 \le u_1, u_2 \le c_e$$

Equation 4

Lagrange multipliers in Equation 4 satisfy the equality $u_{1i}\,u_{2i} = 0$. The Lagrange multipliers, u_{1i} and u_{2i}, are calculated and an optimal desired weight vector of the regression hyperplane is given as next:

$$f(x) = \sum_{i=1}^{m} (u_{1i} - u_{2i})\, k(x, x_i) + b$$

Equation 5

Backpropagation Neural Network (BPNN)

Backpropagation neural network is a feed forward neural network used to capture the relationship between the inputs and outputs (Poh, 1991). The neural network is trained using backpropagation algorithm (Haykin, 2003). The backpropagation training algorithm the error in the output neuron q is given by Equation 6:

$$\delta_q = O_q(1 - O_q)(t_q - O_q)$$

Equation 6

Where O_q and t_q are the actual and desired outputs of neuron q in the output layer, respectively. The weight from neuron p in the hidden layer to neuron q in the output layer is adjusted using Equation 7:

$$W_{pq}(n+1) = W_{pq}(n) + \tau \delta_q O_p$$

Equation 7

Where τ is the learning rate coefficient, $0 < \tau < 1$, and $W_{pq}(n)$ and $W_{pq}(n+1)$ are the weights before and after adjustment, respectively (the criterion for choosing the value of the parameters is a trial and error). The error in the output layer is propagated backwards to adjust the weights in the hidden layers. The error in neuron p in the hidden layer is obtained using Equation 8:

$$\delta_p = O_p(1 - O_p) \sum_q \delta_q W_{pq}(n+1)$$

Equation 8

The error δ_p is used to adjust the weights connecting to neuron p in the hidden layer. This process is repeated for all the hidden layers. Application of all inputs once to the network and adjusting the weights is called an epoch. In the backpropagation training algorithm the network weights are adjusted for certain number of epochs to map the relationship between inputs and outputs (Suhasini, Palanivel, & Ramalingam, 2011).

Regression Tree (M5)

M5 is the most commonly used algorithm of regression trees family. Structurally, a model tree takes the form of a decision tree with linear regression functions instead of terminal class values at its leaves (Bonakdar & Etemad-Shahidi, 2011). The M5 model tree is a numerical prediction algorithm and the nodes of the tree are chosen over the attribute that maximizes the expected error reduction as function of the standard deviation of output parameter (Zhang & Tsai, 2007).

At first, the M5 algorithm constructs a regression tree by recursively splitting the instance space. The splitting condition is used to minimize the intra-subset variability in the values down from the root through the branch to the node. The variability is measured by the standard deviation of the values that reach that node from the root

through the branch, with calculating the expected reduction in error as a result of testing each attribute at that node (Bonakdar & Etemad-Shahidi, 2011). In this way, the attribute that maximizes the expected error reduction is chosen. The splitting process would be done if either the output values of all the instances that reach the node vary slightly or only a few instances remain. The standard deviation reduction (SDR) is calculated as follows (Equation 9):

$$SDR = sd(T) - \sum_i \frac{|T_i|}{|T|} x\ sd(T_i)$$

Equation 9

Where T, is the set of examples that reach the node, T_i are the sets that are resulted from splitting the node according to the chosen attribute and sd is the standard deviation (Wang, Witten, & Science, 1996). After the tree has been grown, a linear multiple regression model is built for every inner node, using the data associated with that node and all the attributes that participate in test in the sub-tree rooted at that node. In consequence, linear regression models are simplified by dropping attributes if it results in a lower expected error on future data. After this simplification, every sub-tree is considered for pruning. Pruning occurs if the estimated error for linear model at the root of a sub-tree is smaller or equal to the expected error for the sub-tree (Bonakdar & Etemad-Shahidi, 2011).

III. Results and Discussion

In this part of the paper are reported the results obtained by the classifiers described in the section 2.2. First, it is shown the performance evaluation methods (section 3.1) and after that are presented the results obtained by the classifiers (section 3.2). In all cases, the scores presented in Tables were estimated using a 10-fold cross-validation (Refaeilzadeh, Tang, & Liu, 2009).

a) Performance Evaluation Methods

Pearson correlation coefficient (PCC)

In statistics, Pearson correlation coefficient (r) is a measure of how well a linear equation describes the relation between two variables x and y measured on the same object or organism. The result of the calculus of this coefficient is a numeric value that runs from -1 to 1 (Monedero et al., 2012). This coefficient (r) is calculated by means of the following equation:

$$r = \frac{Cov(x,y)}{s_x\ s_y}$$

Equation 10

Where $Cov(x,y)$ is the covariance between x and y. s_x, s_y is the product of the standard deviations for x and y.

A value of 1 indicates that a linear equation describes the relationship perfectly and positively, with all data points lying on the same line and with Y increasing with X. A

score of -1 shows that all data points lie on a single line but Y increases as X decreases. At last, a value of 0 shows that a linear model is inappropriate – there is no linear relationship between the variables (Huitema, 1980).

Mean absolute error (MAE)

Measures the closeness among a prediction and the actual value of a data set (Hyndman & Koehler, 2006), and is defined by equation:

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|p_i - v_i|$$

Equation 11

Where p_i is the prediction, v_i the real value, and n the number of observations.

Root mean squared error (RMSE)

It represents the difference between the predicted value and the observed value by the mean square (Armstrong & Collopy, 1992), and is defined as (Equation 12):

$$ECM = \frac{\sqrt{\sum_{i=1}^{n}(p_i - v_i)^2}}{n}$$

Equation 12

Where p_i is the prediction, v_i the real value, and n the number of observations.

Relative absolute error (RAE)

It calculates the prediction error rate of a classifier (Armstrong & Collopy, 1992; Hall et al., 2009), using the following equation:

$$RAE = \frac{\sum_{i=1}^{n}|p_i - v_i|}{\sum_{i=1}^{n}|v_i - \bar{v}|}$$

Equation 13

Where p_i is the prediction, v_i the real value, and $\sum_{i=1}^{n}|v_i - \bar{v}|$ is the total absolute error.

b) Experimental Results

Experiments were carried out with SMOreg, Multilayer Perceptron (uses backpropagation to classify instances) , and M5P which are implementations of SVR, BPNN and M5 in the WEKA (Hall et al., 2009) framework.

Table 1 describes general characteristics of the datasets used, presenting number of attributes. The datasets were defined in order to avoid redundant attributes.

The first option (DS_1) includes all 21 attributes. Option DS_2 excludes attributes from Soil fertility properties (7-11) and three of Physic crop properties (13-15). While DS_3 replaces attributes 7-11 and 13-15 by plot number (plot_number). The evaluations of the three dataset options (DS_1, DS_2 and DS_3) are presented in Table 3, 4, and 5.

The dataset DS_1 was tested with performance evaluation methods: Pearson Correlation Coefficient (PCC), Mean Absolute Error (MAE), Root Mean Squared Error (RMSE) and Relative Absolute Error (RAE) using the classifiers: Support Vector Regression

Table 2. Three distinct subsets of attributes

Dataset	Attributes
DS1	1-21
DS2	1,2,3,4,5,6,12,16,17,18,19,20,21
DS3	1,2,3,4,5,6,plot_number,16,17,18,19,20,21

Table 3. Performance Evaluation of Dataset: DS_1

Classifier	Measures			
	PCC	MAE	RMSE	RAE
SVR	0.3087	2.2651	3.3564	92.27 %
BPNN	0.1796	3.655	5.383	148.90 %
M5	0.2938	2.4646	3.3609	100.40 %

(SVR), Backpropagation Neural Network (BPNN), and Regression Tree (M5). The results are showed in Table 3.

The classifiers used for dataset DS_1 (Table 3) present a weak positive correlation among IRR-predicted and IRR-real, where SVR ($r = 0.3087$) and M5 ($r = 0.2938$) present best results. Moreover, the outcomes obtained by measures MAE and RMSE have a minimum difference among IRR-predicted mean, IRR-real mean with 2.2651 (MAE) and 3.3564 (RMSE) percentage score of IRR using SVR; whereas, M5 obtains the values 2.46 (MAE) and 3.3609 (RMSE). Furthermore, the minimum error rates, on the prediction of classifiers, are 92.27% (SVR) and 100.40% (BPNN).

The dataset DS_2 was tested in the same way as DS_1 (Table 4), but DS_2 excludes attributes from Soil fertility properties (7-11) and three of Physic crop properties (13-15), in order to remove redundant attributes.

The DS_2 outcomes (Table 4) are mildly better than those for DS_1. This is because of the positive correlation of $r = 0.3536$ (BPNN) and $r = 0.2992$ (SVR). MAE and RMSE evaluations for DS_2 are also better in comparison with those for DS_1. This is due to the difference that exists between IRR-predicted mean and IRR-real mean, since the number is smaller for the classifiers SVR (MAE = 2.2872, RMSE = 3.3897) and BPNN (MAE = 2.3499, RMSE = 3.3115). However, the minimum error rate prediction

Table 4. Performance Evaluation of Dataset: DS_2

Classifier	Measures			
	PCC	MAE	RMSE	RAE
SVR	0.2992	2.2872	3.3897	93.18 %
BPNN	0.3536	2.3499	3.3115	95.73 %
M5	0.2218	2.5556	3.5028	104.11%

Table 5. Performance Evaluation of Dataset: DS_3

Classifier	Measures			
	PCC	MAE	RMSE	RAE
SVR	0.4705	2.0512	2.8	89.72 %
BPNN	0.4549	2.2947	3.0015	100.37 %
M5	0.4532	2.1723	2.8679	95.01 %

obtained by classifiers in DS_1 did not improve on DS_2. The best results are: 93.18% (SVR) and 95.73% (BPNN).

Finally the dataset DS_3, replaces attributes 7-11 and 13-15 by the number of plot (plot_number), which clusters the values of removed attributes. The evaluation of DS_3, is presented in Table 5.

When the attribute plot_number is added, the outcomes improve (Table 5) regarding DS_1 and DS_2. In this sense, the evaluation of classifiers used for DS_3 is better. This is because it presents a positive correlation closer to 1, in respect to the values obtained for DS_1 and DS_2, where SVR ($r = 0.4705$) is the best result. Whereas BPNN ($r = 0.4549$) and M5 ($r = 0.4532$) present similar results. Furthermore, the outcomes acquired by measures MAE and RMSE show a high closeness between IRR-predicted mean and IRR-real mean with 2.0512 (MAE) and 2.8 (RMSE) percentage scores of IRR using SVR; while M5 gets the values 2.1723 (MAE) and 2.8679 (RMSE). Similarly, the prediction error rate decreases by SVR (89.72%) regarding the outcomes obtained by the classifiers in DS_1 and DS_2.

Conclusions

We have built a novel dataset for coffee rust detection in Colombian crops given weather conditions, soil fertility properties, physic crop properties, and crop management. This approach required a significant effort to analyze data and preprocessing it. We consider that understanding the nature of the problem is the first step towards solving it.

The proposed dataset was used to form three distinct subsets of selected attributes (DS_1, DS_2 and DS_3) which were evaluated with three classifiers: Support Vector Regression, Backpropagation Neural Network, and M5 Regression Tree. Dataset DS3 presents the best outcomes regarding DS_1 and DS_2, and Support Vector Regression obtains the best performance evaluation for each dataset (DS_1, DS_2 and DS_3). Nevertheless, few instances to train a classifier limit his performance, since the classifier cannot take the right decision if the dataset training does not have cases that support the expected decision.

For future work, we intend to use techniques that combine classifier results,

known commonly as ensemble methods (Ghosh, 2002; Ranawana, & Palade, 2006). It has been demonstrated theoretically and empirically that using combinations of multiple classifiers can substantially improve upon the performance of constituent members (Alfaro, García, Gámez, & Elizondo, 2008; Dietterich, 2000; Kim & Street, 2004; Li, Zou, Hu, Wu, & Yu, 2013; Mannino, Yang, & Ryu, 2009; Opitz & Maclin, 1999; Wei, Chen, & Cheng, 2008; Zhu, 2010).

Acknowledgments

The authors are grateful to the Environmental Study Group (GEA), the Telematics Engineering Group (GIT) of the University of Cauca, Control Learning and Systems Optimization Group (CAOS) of the Carlos III University of Madrid and Supracafé for technical support. S&T

References

Alfaro, E., García, N., Gámez, M., & Elizondo, D. (2008). Bankruptcy forecasting: An empirical comparison of AdaBoost and neural networks. *Decision Support Systems*, *45*(1), 110-122. doi: http://dx.doi.org/10.1016/j.dss.2007.12.002

Armstrong, J.S. & Collopy, F. (1992). Error measures for generalizing about forecasting methods: Empirical comparisons. *International Journal of Forecasting*, *8*(1), 69-80. doi: http://dx.doi.org/10.1016/0169-2070(92)90008-W

Balasundaram, S. & Gupta, D. (2014). Training Lagrangian twin support vector regression via unconstrained convex minimization. *Knowledge-Based Systems*, *59*(0), 85-96. doi: http://dx.doi.org/10.1016/j.knosys.2014.01.018

Becker, S. (1979) *La propagación de la roya del cafeto*: Eschborn, Alemania GTZ.

Bonakdar, L. & Etemad-Shahidi, A. (2011). Predicting wave run-up on rubble-mound structures using M5 model tree. *Ocean Engineering*, *38*(1), 111-118. doi: http://dx.doi.org/10.1016/j.oceaneng.2010.09.015

Cintra, M.E., Meira, C.A.A., Monard, M.C., Camargo, H.A., & Rodrigues, L.H.A. (2011, 22-24 Nov. 2011). *The use of fuzzy decision trees for coffee rust warning in Brazilian crops*. Paper presented at the Intelligent Systems Design and Applications (ISDA), 2011 11th International Conference on.

Cristianini, N. & Shawe-Taylor, J. (2000). *An Introduction to Support Vector Machines and Other Kernel-based Learning Methods*: Cambridge, UK: Cambridge University Press.

Dietterich, T.G. (2000). An Experimental Comparison of Three Methods for Constructing Ensembles of Decision

Trees: Bagging, Boosting, and Randomization. *Mach.Learn.*, *40*(2), 139-157. doi: 10.1023/a:1007607513941

Ghosh, J. (2002). Multiclassifier systems: back to the future. *Lecture Notes in Computer Sciences* [Third International Workshop, MCS 2002 Cagliari, Italy, June 24-26, 2002 Proceedings], *2364*, 1-15

Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P., & Witten, I. H. (2009). The WEKA data mining software: an update. *SIGKDD Explor. Newsl.*, *11*(1), 10-18. doi: 10.1145/1656274.1656278

Haykin, S.S. (2003). *Neural networks: a comprehensive foundation*: Prentice Hall.

Huitema, B.E. (1980). *The Analysis of Covariance and Alternatives*: John Wiley & Sons.

Hyndman, R.J. & Koehler, A.B. (2006). Another look at measures of forecast accuracy. *International Journal of Forecasting*, *22*(4), 679-688. doi: http://dx.doi.org/10.1016/j.ijforecast.2006.03.001

Kim, Y. & Street, W.N. (2004). An intelligent system for customer targeting: a data mining approach. *Decision Support Systems*, *37*(2), 215-228. doi: http://dx.doi.org/10.1016/S0167-9236(03)00008-3

Li, L., Zou, B., Hu, Q., Wu, X., & Yu, D. (2013). Dynamic classifier ensemble using classification confidence. *Neurocomputing*, *99*(0), 581-591. doi: http://dx.doi.org/10.1016/j.neucom.2012.07.026

Luaces, O., Rodrigues, L.H.A., Alves-Meira, C.A., & Bahamonde, A. (2011). Using nondeterministic learners to alert on coffee rust disease. *Expert Systems with Applications*, *38*(11), 14276-14283. doi: http://dx.doi.org/10.1016/j.eswa.2011.05.003

Luaces, O., Rodrigues, L.H.A., Meira, C.A.A., Jos, #233, Quevedo, R., & Bahamonde, A. (2010). Viability of an alarm predictor for coffee rust disease using interval regression. In Proceedings of the 23rd international conference on Industrial engineering and other applications of applied intelligent systems, Cordoba, Spain, [Vol. 2] (pp.337-346]. Berlin, Alemania: Springer-Varlag

Mannino, M., Yang, Y., & Ryu, Y. (2009). Classification algorithm sensitivity to training data with non representative attribute noise. *Decision Support Systems*, *46*(3), 743-751. doi: http://dx.doi.org/10.1016/j.dss.2008.11.021

Meira, C., Rodrigues, L., & Moraes, S. (2008). Análise da epidemia da ferrugem do cafeeiro com árvore de decisão. *Tropical Plant Pathology*, *33*(2), 114-124.

Meira, C.A.A., & Rodrigues, L.H.A. (2009). *Árvore de decisão na análise de epidemias da ferrugem do cafeeiro* [Paper - VI Simpósio de Pesquisa dos Cafés do Brasil]. Retrieved from: http://www.sbicafe.ufv.br/bitstream/handle/10820/3466/56.pdf?sequence=2

Meira, C.A.A., Rodrigues, L.H.A., & Moraes, S.A.d. (2009). Modelos de alerta para o controle da ferrugem-do-cafeeiro em lavouras com alta carga pendente. *Pesquisa Agropecuária Brasileira*, *44*, 233-242.

Monedero, I., Biscarri, F., León, C., Guerrero, J. I., Biscarri, J., & Millán, R. (2012). Detection of frauds and other non-technical losses in a power utility using Pearson coefficient, Bayesian networks and decision trees. *International Journal of Electrical Power & Energy Systems*, *34*(1), 90-98. doi: http://dx.doi.org/10.1016/j.ijepes.2011.09.009

Opitz, D. & Maclin, R. (1999). Popular ensemble methods: An empirical study. *Journal of Artificial Intelligence Research*, *11*, 169-198.

Pérez-Ariza, C.B., Nicholson, A.E., & Flores, M.J. (2012). *Prediction of Coffee Rust Disease Using Bayesian Networks*, Proceedings of the Sixth European Workshop on Probabilistic Graphical Models, (pp.259-266). Available at http://arrow.monash.edu.au/hdl/1959.1/821316

Poh, H.L. (1991). *A neural network approach for marketing strategies research and decision support* [Ph.D Thesis], Stanford University

Ranawana, R. & Palade, V. (2006). Multi-Classifier systems: Review and a roadmap for developers. *Int. J. Hybrid Intell. Syst.*, *3*(1), 35-61

Refaeilzadeh, P., Tang, L., & Liu, H. (2009). Cross-Validation. In L. Liu & T. Özsu [Eds.], *Encyclopedia of Database Systems* (pp. 532-538): Springer

Rivillas-Osorio, C., Serna-Giraldo, C., Cristancho-Ardila, M., & Gaitán-Bustamante, A. (2011). La roya del cafeto en Colombia, impacto, manejo y costos de control. In S. Marín [Ed.], *Avances Tecnicos Cenicafe*. Chinchiná, Colombia: Cenicafé

Shieber, E. & Zentmyer, G. A. (1984). Coffee rust in the western hemisphere *Plant disease*, *68*, 89-93

Smola, A. & Schölkopf, B. (2004). A tutorial on support vector regression. *Statistics and Computing*, *14*(3), 199-222. doi: 10.1023/b:stco.0000035301.49549.88

Suhasini, A., Palanivel, S., & Ramalingam, V. (2011). Multimodel decision support system for psychiatry problem. *Expert Systems with Applications*, *38*(5), 4990-4997. doi: http://dx.doi.org/10.1016/j.eswa.2010.09.152

Vapnik, V.N. (2000). *The nature of statistical learning theory*. New York, NY: Springer.

Vapnik, V.N. (1999). An overview of statistical learning theory. *Neural Networks, IEEE Transactions on*, *10*(5), 988-999. doi: 10.1109/72.788640

Wang, Y., & Witten, I.H. (1996). Induction of model trees for predicting continuous classes. *Working Paper Series*, *96*(23). Retrieved from de http://www.cs.waikato.ac.nz/pubs/wp/1996/uow-cs-wp-1996-23.pdf

Wei, C.-P., Chen, H.-C., & Cheng, T.-H. (2008). Effective spam filtering: A single-class learning and ensemble approach. *Decision Support Systems*, *45*(3), 491-503. doi: http://dx.doi.org/10.1016/j.dss.2007.06.010

Wirth, R. (2000). CRISP-DM: Towards a standard process model for data mining. In *Proceedings of the Fourth International Conference on the Practical Application of Knowledge Discovery and Data Mining*, Manchester, UK, (pp29-39).

Zapata, J.C. & Ruíz, G.M. (1988). *La variedad Colombia: selección de un cultivar compuesto resistente a la roya del cafeto* [Premio Nacional de Ciencias, Fundación Alejandro Angel Escobar, 1986]. Chinchiná, Colombia: Cenicafé

Zhang, D., & Tsai, J. J. P. (2007). *Advances in MacHine learning applications in software engineering*: Hershey, PA: Idea

Zhu, D. (2010). A hybrid approach for efficient ensembles. *Decision Support Systems*, *48*(3), 480-487. doi: http://dx.doi.org/10.1016/j.dss.2009.06.007

Currículum vitae

David Camilo Corrales
M.Sc., in Telematics Engineering and researcher of Telematics Engineering Group and Environmental Study Group at University of Cauca, Colombia.

Agapito Ledezma
Ph.D., in Sciences, Speciality Computer Engineering and Full Professor at University Carlos III of Madrid.

Andrés Peña
M.Sc., in Meteorology and researcher at National Coffee Research Center (Colombia).

Javier Hoyos
Agronomic Engineer and Farmer Manager of Los Naranjos (Supracafé - Colombia).

Juan Carlos Corrales
Doctor of Philosophy in Sciences, Speciality Computer Science, and Full Professor and Leader of the Telematics Engineering Group at University of Cauca, Colombia.

Apolinar Figueroa
Doctor of Biological Sciences, and Full Professor and Leader of the Environmental Study Group at University of Cauca, Colombia.