CHARACTERIZATION OF THE COMPOSTING PROCESS USING MACHINE LEARNING ALGORITHMS

Molina Monteleón C.M., Saviñon Flores M.F., Vidal Robles E., Hernández Santiago A.A., Arzola Flores J.A.

Meritorious Autonomous University of Puebla

av. San Claudio, S/N Col. San Manuel, Puebla, Pue. C.P. 71590, Mexico; e-mail: ximikad09@mail.ru

Received: 16.06.2019

Abstract. The compost is a biological process of degradation of organic matter that has different applications in agriculture and the remediation of soils. Use of the package for data mining Orange was possible the development of an artificial intelligence algorithm, which was carried out through the treatment of images and their classification in the middle of the methods Logistic Regression, Neural Network, Random Forest, Support Vector Machine (SVM) and k-Nearest-Neighbors. With the algorithm, the stage of the compost process is identified by comparing the images of compost under controlled conditions.It is possible to create a supervised learning algorithm to be able to predict the stages of the composting process using only photographic images of the compost. Because of this, the algorithm that best performs the classification is the multilayer perceptron neural network. This result will allow the development of a portable device that allows identifying the quality of the soil.

Key words: *machine learning algorithms, composting.*

INTRODUCTION

The composting process is known to be an excellent alternative as a final disposal of organic waste: microorganisms degrade waste under aerobic conditions, transforming them into a homogenous and assimilable material for plants. Different applications for compost are known as organic fertilizer in agriculture, horticulture, landscaping, erosion control and soil recovery, depending on the stage of the process. According to the variation of temperature, there are three main stages in a composting: mesophilic phase, thermophilic or hygienic phase and cooling or mesophilic phase, in addition to a maturing stage of variable duration [1].

In the previous project, "Monitoring and control of a compost using Arduino", the monitoring and control of a composting system and the collection of images for 2 months was carried out [2]. Now the use of these images is proposed to generate an artificial intelligence algorithm to identify each stage of the composting process. An artificial intelligence algorithm is defined as the simulation of human intelligence processes by machines, especially computer systems. These processes include learning (the acquisition of information and rules for the use of information), reasoning (using the rules to reach approximate or final conclusions) and self-correction [3].

For the development of the project the Orange package was used, a computer program to perform data mining and predictive analysis developed in the computer faculty of the University of Ljubljana. It consists of a series of components developed in C ++ that implement data mining algorithms, as well as preprocessing operations and graphical representation of data. It is an open source tool for data analysis and visualization where data mining is done through visual programming or by Python code. It has components for Machine Learning, additions for bioinformatics and text mining [4].

The development of the algorithm aims to streamline and facilitate the identification of the stage of the composition process, the use of compost can be used in other applications.

METHODOLOGY

For the training of the algorithm the images of the compost under controlled conditions of the previous project were used. The treatment that was given to them has the following order:

1. Data acquisition
2. Embedded images
3. Matrix construction
4. Multilayer Perceptron Neural Network
5. Training
6. Testing
7. Prediction

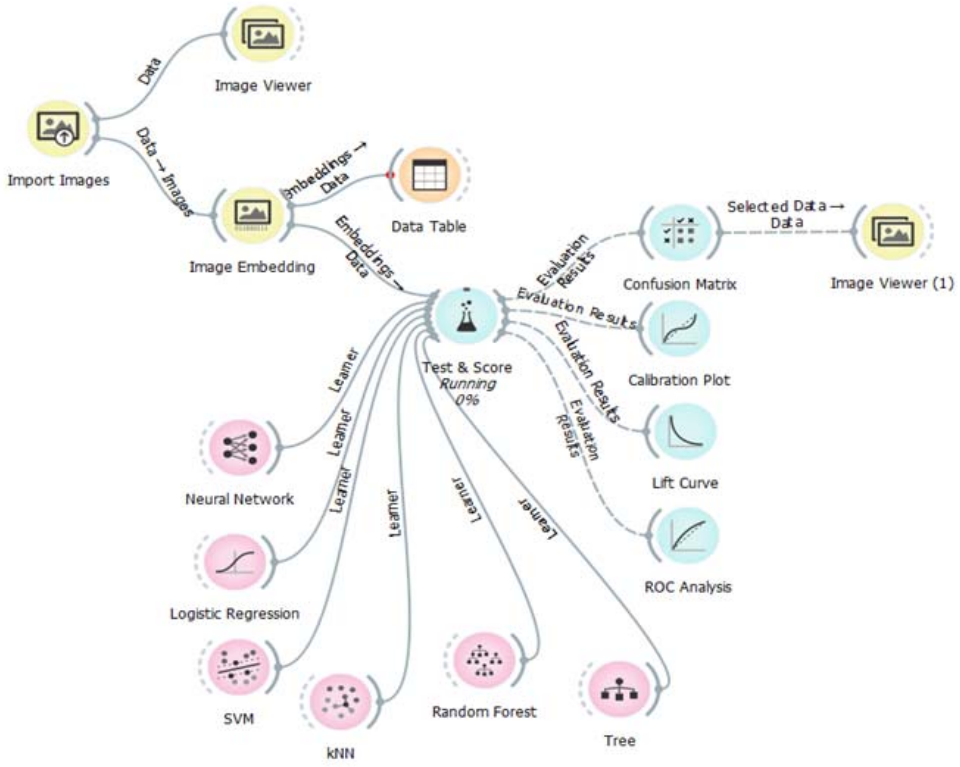

Figure 1. Models of supervised learning for the classification of images obtained from the composting process

The images were imported, and the embedding was performed, which reads the images and uploads them to a remote server or evaluates them locally. Machine learning models are used to calculate a feature vector for each image and return an improved data table with additional columns (image descriptors).

Then, 6 models were used to make the classification:

1. Logistic Regression: Learn a logistic regression model from the data. It only works for classification tasks [5].
2. Neural Network: Uses sklearn's (Python) multilayer perceptron algorithm that can learn nonlinear and linear models [6].
3. Random Forest: Build a set of decision trees. Each tree is developed from a bootstrap sample from the training data. When developing individual trees, an arbitrary subset of attributes is drawn (hence the term "Random"), from which the best attribute for division is selected. The final model is based on the majority vote of individually developed trees in the forest [7].

Table 1. Accuracy of the different supervised learning algorithms

Test & Score — **Tue May 28 19, 23:12:16**

Settings

Sampling type: Stratified 5-fold Cross validation
Target class: Average over classes

Scores

Method	AUC	CA	F1	Precision	Recall
kNN	0.690	0.243	0.233	0.234	0.243
Tree	0.566	0.158	0.156	0.166	0.158
SVM	0.526	0.174	0.185	0.353	0.174
Random Forest	0.775	0.239	0.228	0.227	0.239
Neural Network	0.846	0.332	0.321	0.315	0.332
Logistic Regression	0.828	0.377	0.379	0.394	0.377

1. SVM: Support vector support machine is a machine learning technique that separates the attribute space with a hyperplane, thus maximizing the margin between the instances of different classes or class values [8].
2. Tree: It is a simple algorithm that divides data into nodes by class purity. It is a precursor of the random forest. Tree in Orange is designed internally and can handle both discrete and continuous datasets [9].
3. k-Nearest-Neighbors: Use an algorithm that searches for "k" closest training examples in the feature space and uses its average as a prediction [10].

For the training and testing of the algorithms the cross-validation with 5 folds was used. The above, to ensure that the training models do not show overfitting [11].

RESULTS AND DISCUSSION

The 247 images divided into 23 categories were imported, corresponding to each one of the days in which the images were acquired. Subsequently, the cross-validation with 5-folds was applied for the training of the algorithms. Table 1 shows the accuracy (AUC) obtained by applying each of the 6 supervised learning algorithms.

The multilayer perceptron neural network has the highest accuracy. That is, the neuronal network with a hidden layer of 1000 neurons can perform the best classification of the images (AUC = 0.846), followed by the logistic regression (AUC = 0.828). Therefore, it is possible to make the prediction of the day in which the composting process is only using a photograph of the compost.

CONCLUSIONS

It is possible to create a supervised learning algorithm to be able to predict the stages of the composting process using only photographic images of the compost. Because of this, the algorithm that best performs the classification is the multilayer perceptron neural network. This result will allow the development of a portable device that allows to identify the quality of the soil.

References:

1. Oviedo-Ocaña E.R., Dominguez I., Komilis D., Sánchez, A. Co-composting of green waste mixed with unprocessed and processed food waste: influence on the composting process and product quality. *Waste and Biomass Valorization*, 219, vol. 10, no. 1, pp. 63-74.

2. Saviñon-Flores M.F. Diseño y construcción de un sistema de compostaje empleando las plataformas Arduino y Python (Tesis de pregrado). Benemérita Universidad Autónoma de Puebla, Puebla, México, 2019. [Saviñon-Flores M. F. *Design and construction of a composting system using the Arduino and Python platforms*. Undergraduate Thesis, Benemérita Autonomous University of Puebla, Puebla, Mexico, 2019. (In Span.)]

3. Al-Turjman, F. Artificial Intelligence in IoT. *Springer*, 2019.

4. Pejovic V., Majhen I., Janez M., Zupan B. *RICERCANDO: Data Mining Toolkit for Mobile Broadband Measurements*. 2019. arXiv preprint arXiv:1901.07287.

5. Jie M.A., Collins G.S., Steyerberg E.W., Verbakel J.Y., van Calster B. A systematic review shows no performance benefit of machine learning over logistic regression for clinical prediction models. *Journal of clinical epidemiology*, 2019.

6. Heidari A.A., Faris H., Mirjalili S., Aljarah I., Mafarja M. Ant Lion Optimizer: Theory, Literature Review, and Application in Multi-layer Perceptron Neural Networks. In Nature-Inspired Optimizers. *Springer, Cham*., 2019, pp. 23-46

7. Tyralis H., Papacharalampous G., Langousis A. A brief review of random forests for water scientists and practitioners and their recent history in water resources. Water, 2019, vol. 11, no. 5, p. 910.

8. Guo H., Wang W. Granular support vector machine: a review. *Artificial Intelligence Review*, vol. 51, no. 1, pp. 19-32.

9. Paez A., López F., Ruiz M., Camacho M. Inducing non-orthogonal and non-linear decision boundaries in decision trees via interactive basis functions. *Expert Systems with Applications*, vol. 122, pp. 183-206.

10. Jiang J., Chen Y., Meng X., Wang L., Li K. A novel density peaks clustering algorithm based on k nearest neighbors for improving assignment process. *Physica A: Statistical Mechanics and its Applications*, vol. 523, pp. 702-713.

11. Rabinowicz A., Rosset S. *Cross-Validation for Correlated Data*. 2019. arXiv preprint arXiv:1904.02438.