Original papers

Analysis of green manure decomposition parameters in northeast Brazil using association rule networks

Dario Brito Calçada[a,*], Solange Oliveira Rezende[a], Mauro Sergio Teodoro[b]

[a] ICMC, University of São Paulo, Av. Trabalhador São-carlense, 400, Centro, São Carlos, SP 13566-590, Brazil
[b] EMBRAPA Meio-Norte, Unidade de Execução de Pesquisa de Parnaíba (UEP), BR-343, 35, Parnaíba, PI, Brazil

ABSTRACT

The food sector is one of the most critical areas of the economy, and consumers are seeking safer, more readily available, more affordable, and better quality food. Therefore, organic agriculture has become a possible approach for optimizing the characteristics of processed foods. Vegetables have essential uses as green manure, but the greatest difficulty encountered when using these species is related to the time required for their residues to decompose. Computational intelligence and data mining techniques are used widely in agricultural studies and for process improvement purposes. Association rules are used as data mining techniques to identify patterns in large databases, where the patterns identified are analyzed and transformed into a graph to facilitate further analyses. In this study, we employed the patterns identified by association rule networks (ARNs) to determine directly the key parameters related to the half-life decomposition times of legumes used as green manure in northeast Brazil. We validated this method by comparing the results obtained by the ARNs and a decision tree. The proposed approach obtained promising results, thereby demonstrating its capacity to describe a set of objective items and facilitate the development of more integrated hypotheses. Thus, it was possible to optimize the discovery of the key parameters related to green fertilizers in order to identify the best species according to its culture requirements and to enhance productivity.

1. Introduction

Environmental degradation has been exacerbated by the need for rapid food production due to the inappropriate occupation of arable areas and the pursuit of economic interests to enhance profitability in the agricultural sector. Thus, environmental degradation has led to changes in the physical, chemical, and biological attributes of the soil, with consequent decreases in the potential productivity (Calegari, 2014).

The food sector is one of the most critical areas of the economy, where it encompasses agriculture, the food industry, and food retailers; hence, it affects virtually all members of society. Due to the need to provide consumers with safe, readily available, affordable, and high quality food, the food industry must be efficient with organized control processes in order to ensure the safety and quality of its products. In particular, organic farming has the potential to enhance the control of production in plantations (Lehmann et al., 2012).

Organic agriculture aims to substitute the inputs used in conventional agriculture. Thus, sustainable production is possible in ecologically-based agriculture, where the natural functions of ecosystems are promoted (Ambrosano et al., 2014). The use of vegetable species as manure can address the problem of ensuring the ecological output while maintaining the productivity. These vegetable species are known as green fertilizers and they are capable of improving the production environment because chemical additives are not allowed in ecologically-based agriculture (Martins and Barros, 2015).

Members of the family Leguminosae have important applications as green manure, but the main problem related to the use of these species is the time required for their residues to decompose, which directly affects the productivity of crops. Thus, the choice of species is related to the desired degradation time, which can make the selection process very complicated (Garcia, 2002).

Computational intelligence and data mining techniques have been employed widely in agriculture to assist with the resolution of complex problems such as disease detection (Phadikar et al., 2013) or even for identifying the locations of plantations to achieve higher productivity (Lucas and Chhajed, 2004).

At present, many companies and organizations manipulate and store large amounts of data in the course of their daily activities. Understanding these data, i.e., identifying the implicit information in

these data, is increasingly important for supporting decision making processes (Zaki and Meira, 2013).

Association rules are used as data mining techniques to identify specific patterns in the data found in large databases, where the interpretation of these patterns can allow the extraction of specific knowledge about the problem under consideration (Le and Vo, 2016; Kuo et al., 2014).

The mining of association rules can be divided into the following three main steps (Aggarwal et al., 2002, 1994; Nguyen and Nguyen, 2015):

- **Preprocessing:** The data set is prepared for data extraction where the removal of non-interesting items may occur.
- **Extraction of patterns:** Measurements are calculated, before constructing frequent item sets and obtaining the association rules.
- **Post-processing:** The rules of interest are identified to decrease the number of rules that need to be explored.

Definition 1. Let $I = \{i_1, i_2, ..., i_n\}$ be a set of objects called items that can assume a binary value of 0 or 1 (false or true), which represent the presence or absence of a particular object, respectively. Let T be a set of transactions, where each transaction D corresponds to a set of items such as $D \subseteq I$. It is also considered that a set of elements A is contained in a transaction D if all items in the set have a "true" value in the transaction. An expression of the form $A \Rightarrow B$ can represent an association rule R, with $A \subseteq I, B \subseteq I \ A \cap B = \emptyset$. It is also possible to treat quantitative or qualitative variables in order to create ranges of values for subsequent use as binary values. A is called the antecedent or left-hand side (LHS) of the rule and B is the consequent or right-hand side (RHS).

For each rule ($LHS \Rightarrow RHS$) extracted from a set of T transactions (propositional model after data transformation), a support value (sup) is calculated (Eq. (1)) to verify the strength of the association between the LHS and RHS (probability of transaction occurrence $LHS \cup RHS$), and a confidence value ($conf$) (Eq. (2)) to measure the strength of the logical implication of the rule (conditional probability of RHS given LHS) (Agrawal et al., 1994). $n(T)$ is the number of transactions.

$$sup(LHS \Rightarrow RHS) = P(LHS \cup RHS) \tag{1}$$

$$conf(LHS \Rightarrow RHS) = P(RHS|LHS) \tag{2}$$

The support can be described as the probability that any transaction satisfies both the LHS and RHS, whereas the confidence is the probability that a transaction satisfies the RHS because it satisfies the LHS.

Given the size of current databases, the number of rules discovered can be extremely high, and thus their interpretation almost becomes a new mining problem. Therefore, it is essential to understand the association rules and to develop better methods for interpreting the networks to facilitate the visualization of the rules (Pandey et al., 2009).

The use of networks facilitates the knowledge identification process and the use of decision support systems, and thus this technique can be applied to agricultural research. In the present study, we conducted pattern extraction based on association rule networks (ARNs) to determine the parameters related directly to the half-life decomposition time for legume plants used as green manure in northeast Brazil.

We compared the results obtained with those produced by a decision tree algorithm by using the relationships in the data set in order to validate the method. This comparison was conducted to verify the results and to determine the method that could describe the data set most appropriately according to the objective items.

The remainder of this paper is organized as follows. Related research is described in Section 2. In Section 3, we present the association rule mining process and the method employed for constructing ARNs. In Section 4, we explain the methods used for generating the data as well as the mining steps and the construction of the ARN. The results are presented and discussed in Section 5. We give our conclusions in Section 6.

2. Related research

The application of machine learning in agriculture is currently increasing among the academic community and industry professionals, which has led to the development of several different approaches based on various machine learning frameworks (Drury et al., 2017).

The generation of agricultural data related to productivity has increased in recent years. In order to evaluate the impact of various parameters, Hira and Deshpande (2015) constructed a multi-dimensional model of data by employing multivariate analysis, statistical analysis, and association rule mining techniques to extract knowledge. They presented an approach based on a case study to analyze agricultural productivity using several related parameters.

Drury et al. (2017) discussed the use of Bayesian networks in agricultural research. Bayesian networks can help with the extraction of knowledge because of their capacity to deal with incomplete information. Agricultural studies can aggregate new details using a network structure. The construction of Bayesian networks begins with a set of variables that describe the problem domain. Drury et al. (2017) survey the use of Bayesian networks but no previous studies were focused on objective items.

Recently, recommendation systems have become very popular in several areas, including agriculture. Kumar and Balakrishnan (2019) developed a recommendation system that uses the Apriori algorithm to forecast the consumption of agricultural items (vegetables/fruits), where data preprocessing was conducted by reducing the dimensionality to make the data more refined. The association rules were used to facilitate the analysis of the consumption of agricultural items. The use of association rules automated the process for generating hypotheses regarding farm product waste.

Phadikar et al. (2013) used a classifier rule-based approach in their study. The rules were fundamental for the results obtained but associations are not made in their study of rice disease parameters. Several characteristics of plants were used in the direct analysis of plantations and they used parameters extracted from diseased objects (plants) in a tree structure decision system. The trees were used to configure the classifier and generate the rules, but the best characteristics were not analyzed to optimize the classification.

Complex agricultural ecosystems need to be understood better to address the increasing challenges related to agricultural production. The search for organic resources to enhance productivity has been based on advances in agricultural science. Data analysis can allow farmers and companies to enhance value by improving their productivity. Kamilaris et al. (2017) reviewed previous studies in agrarian research based on big data analysis to solve various problems, where they considered 34 studies. The conclusions of these studies were consistent and demonstrated the great potential of applying big data analysis in the development of "smarter" agriculture.

Data related to agriculture have high complexity. Batarseh and Yang (2018) analyzed the production life cycle for agricultural systems using association rules, which assisted the testing process in the system building phase by adapting to the types of data used in agriculture.

Nourani and Molajou (2017) employed two data mining techniques (decision tree and association rules) to formulate a hybrid approach for identifying the associations between climatic situations in different regions. The two main steps in their proposed model comprised the classification and selection of data, and the extraction of hidden information. Decision tree techniques were used to classify and select the most useful data groups, before association rules were employed to extract predictive information to generate the candidate hypotheses. The results confirmed the reliable performance of the hybrid data mining method. Decision trees and association rule techniques have been used as complementary techniques, but they can be used to extract knowledge.

In all of the previous studies mentioned above, machine learning solutions were developed to facilitate decision making in agricultural

production, which is directly related to the current demand for enhanced productivity and sustainability. The use of network structures can facilitate the knowledge extraction process by allowing results to be visualized and the optimization of the descriptive process.

3. ARN

Data mining techniques, particularly mining with association rules, may help to identify the parameters related to agroecological production (de Barros et al., 2013). The identification of association rules is a data mining technique that aims to identify specific data patterns in large databases, which can be interpreted to obtain specific knowledge regarding the problem under consideration (Le and Vo, 2016).

An association rule quantitatively characterizes how the presence of a set of items in the records of a database indicate the presence of some other distinct set of elements in the same records (Agrawal and Srikant, 1994). Thus, the aim of association rules is to identify trends that can be used to understand and explore patterns of behavior in data. For example, if we consider supermarket sales data, we may find that 80% of the customers who buy product Q also buy product W, and thus 80% corresponds to the reliability of this rule (Domingues and Rezende, 2005).

ARNs have a structure that allows the synthesis, pruning, and analysis of a set of association rules to construct candidate hypotheses. From the perspective of knowledge discovery, an ARN allows context-oriented analysis to enhance our understanding of the relationships and the usefulness of the information obtained with association rule mining algorithms. From a mathematical viewpoint, ARNs are instances of directed hypergraphs (Pandey et al., 2009).

The central idea of ARN is that the association rules discovered by a mining algorithm can be synthesized, pruned, and integrated into the context of specific research objectives. In particular, if we consider a variable of interest, a network can be produced with the variables that are most strongly related to the aim, before obtaining a structure that can be tested using statistical methods. In practice, association rule mining involves hundreds of variables and the pruning strategy for the ARNs can be used to remove local inconsistencies between variables, such as cycles. ARNs have the following characteristics.

- Pruning in context: ARNs are used to prune rules in the context of a specific goal.
- Network structure: ARNs provide a method for determining the relationships between relevant variables and the goal by constructing a network. This technique can assist with analyzing the effects of direct and indirect changes in the mining of association rules.
- Generation and evaluation of hypotheses: ARN can provide a bridge between the outputs generated by association rule mining and their assessment.

According to Chawla (2010), ARNs are represented as a backward-directed hypergraph (B-graph), which can transform the ARN according to the selected target item in a data set after pruning. The following four steps are performed to create an ARN.

(1) Step A: Given a database *D*, minimum support, and minimum confidence, we must extract all of the association rules using a standard algorithm, such as *Apriori* (Agrawal and Srikant, 1994), *Aptitori-Tid* (Agrawal and Shafer, 1996), or *FP-Growth* (Grahne and Zhu, 2005), and a data structure to implement the algorithm (Yan et al., 2007).
(2) Step B: Choose a regular item *g*, which is represented as the target node in the rule set, and construct a *B-graph* that flows recursively to *g*.
(3) Step C: Perform pruning for the generated B-graph by removing hypercycles and reverse hyper-edges. The resulting B-graph obtained is called an ARN.
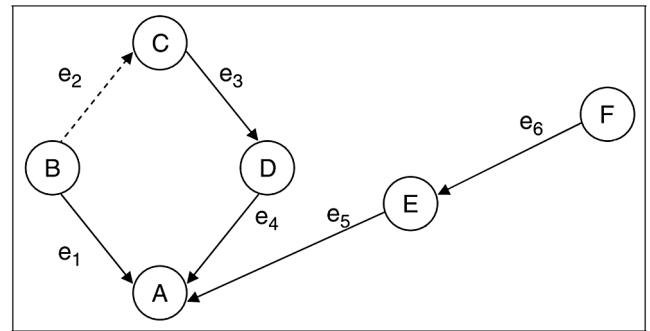

Fig. 1. Example of an ARN with a reverse hyper-edge (e_2). Adapted from Pandey et al. (2009).

(4) Step D: Find shorter paths between the target node and the other higher-level nodes (a variant of the distance between ends) in the ARN. The set of these paths represents the exploratory network for the target node.

Fig. 1 shows an example of an ARN where item "A" is selected as the target. All of the rules with "A" as a result are then selected, i.e., only the rules (B ⇒ A) and (D ⇒ A) in this case. Thus, the items are modeled at ARN level 1. "B" and "D" are then the targets, and the algorithm runs recursively until no more items are left as goals at the highest levels. In this case, the rules with "B" as their goal are modeled, before the rules with "D" and then "E," "C," and "F." In this example, no rules have "F" as a result. The highlighted hyper-edge "e_2" is one of those eliminated in the pruning process because although it has the "C" item as a consequence, the "B" item is already inserted into the ARN in the level below, thereby making the use of this rule impossible.

ARNs are tools that provide a context for understanding the relationships between rules while they are obtained. According to Pandey et al. (2009), ARNs have the following benefits.

- Organizing rules in a context: The main use of an ARN is for organizing a potentially large set of association rules so that an objective line of reasoning can be explored using the most relevant rules in the collection. This explanation can extend beyond the immediate rules, which have the objective item as a consequent, thereby allowing these rules to be organized in a context and then employing the set of association rules to generate statistically viable hypotheses.
- Local pruning: ARNs provide a method for pruning rules by associating redundant rules with hypercycles and reverse hyper-edges. Pruning occurs in the context of an objective node. Thus, a rule that is redundant for a particular target node may become relevant to another target, so this approach is more flexible than pruning based on statistical measures of interest.

If we consider the ARN presented in Fig. 1, when the target node is "A", the hyper-edge "e_2" is a redundant rule that can be eliminated. However, when the target node is "D" (Fig. 2), the same hyper-edge is relevant. Thus, rule pruning is dependent on the context of the target node, which is called local pruning.

This approach facilitates the exploration of computationally efficient research techniques for large databases in conjunction with traditional statistical methods. ARN provides a method for synthesizing association rules in a structured manner.

ARNs can also be used to systematically select attributes (Chawla, 2010) by adapting the processes to obtain association rules for the target parameter. A hypergraph grouping algorithm (Han et al., 1997) is applied to the ARN to extract the relevant characteristics (within the target domain). The first level nodes have immediate effects on the target node, whereas the top-level nodes have indirect effects.

Studying the parameters for various crops can allow us to optimize
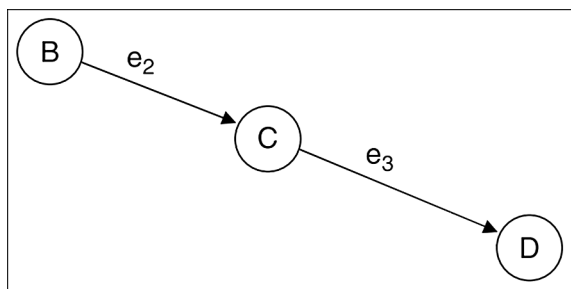
Fig. 2. ARN with target node "D." The hyper-edge "e_2" is part of the ARN. This proves the adaptive nature of local pruning. Adapted from Pandey et al. (2009).

Fig. 4. Average precipitation levels in Parnaíba/PI (2015).

Table 1
Spacing and sowing density of plants used as green manure.

Common Name	Depth (cm)	Spacing (cm)	Seeds/m (linear)
Crotalária breviflora	2–3	0.50	30–35
Crotalária juncea	2–3	0.50	22–27
Crotalária mucronata	2–3	0.50	40–45
Feijão de porco	2–5	0.50	4–5
Feijão Guandu Fava Larga	2–3	0.50	20–25
Feijão Guandu Anão IAPAR 43	2–3	0.50	20–25
Tefrósia	2–3	1.00	5

the selection of new cultivars, especially for plants that can be used as green manure in ecologically-based agriculture. Methods based on pattern mining (association rules) are frequently employed to obtain this information, mainly by using graphical elements such as ARNs.

In the present study, we collected data for vegetable species used as green manure and we analyzed their related parameters by using ARNs to identify the actual conditions that might directly interfere with the productivity of these plants in ecologically-based farming.

4. Materials and methods

We performed the entire data mining process comprising the preparation and pre-processing of the data set, before analyzing the extracted association rules and constructing the ARNs to obtain insights into vegetable species used as green manure (Fig. 3).

4.1. Agricultural experiment – data set generation

We conducted the study during the second semester of 2015 at the Embrapa Meio-Norte/Unidade de Execução de Pesquisa de Parnaíba (UEP) Parnaíba, Brazil ($3°05'S;41°46'W$ and 46.8 m altitude). The climate in this region belongs to the C1dA'a' type (Thornthwaite and Mather, 1955), which is characterized as dry, megathermal subhumid, with a small water surplus and a potential evapotranspiration concentration of 29.7% in the quarter comprising October, November, and December, and a mean relative air humidity of 77.5%, mean precipitation of 1107 mm, and average temperature of 27.6 °C (Bastos et al., 2012).

The soil in the experimental area was a medium-textured Yellow Latosol in the caatinga littoral phase (Melo et al., 2004). Soil samples were collected at different points in the experimental area to obtain highly uniform samples of the substrate. A composite sample was obtained using the collected samples, which was analyzed at the Embrapa Meio-Norte water and soil laboratory, UEP Parnaíba to determine the chemical characteristics of the soil.

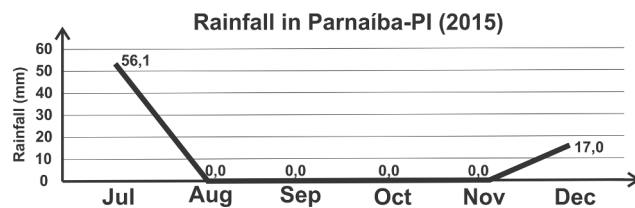Rainfall data for the experimental period were recorded at the

INMET (National Institute of Meteorology) climatological station located at the Embrapa Meio-Norte Experimental Field in Parnaíba, Piauí (Fig. 4).

The experiment was conducted in an experimental area that received no chemical fertilization in the previous five years. The soil was prepared by plowing and surface harvesting, with subsequent leveling followed by manual leveling.

The experimental design was completely randomized with seven treatments and four replicates. The plots measuring 2.0 m× 2.0 m contained five rows of plants with spacings of 0.50 m. We used a high seed density and the plant population was subsequently thinned at 20 days after sowing (DAS) (Table 1). Weeds were controlled manually. The irrigation system was a conventional sprinkler with a two-day irrigation shift and an average duration of 1 h.

The annual tropical climate species used in the treatments were: T1, Feijão guandu-anão (*Cajanus cajan*) cultivar IAPAR 43; T2, *Crotalaria mucronata*; T3, *Crotalaria juncea*; T4, Feijão de porco (*Canavalia ensiformis L.*); T5, *Crotalaria breviflora*; T6, Feijão guandu (*Cajanus cajan*) cultivar Caqui/Fava larga; and the perennial tropical climate species T7, Tefrósia (*Tephrosia candida*).

At 100 days after the start of the test, we performed the final cuts for all of the treatments at 0.05 m above the surface of the soil using pruning shears. The central lines were treated as the plot and we discarded the borders, thereby leaving a total area of 0.5 m² (1.00 m line with 0.5 m between the rows). The roots of the plants were also harvested manually with a hoe at a depth of 0.20 m to obtain the fresh and dry biomasses for the species tested. We determined the plant height at flowering (AP), new shoot mass (MFPA), dry shoot mass (MSPA), new
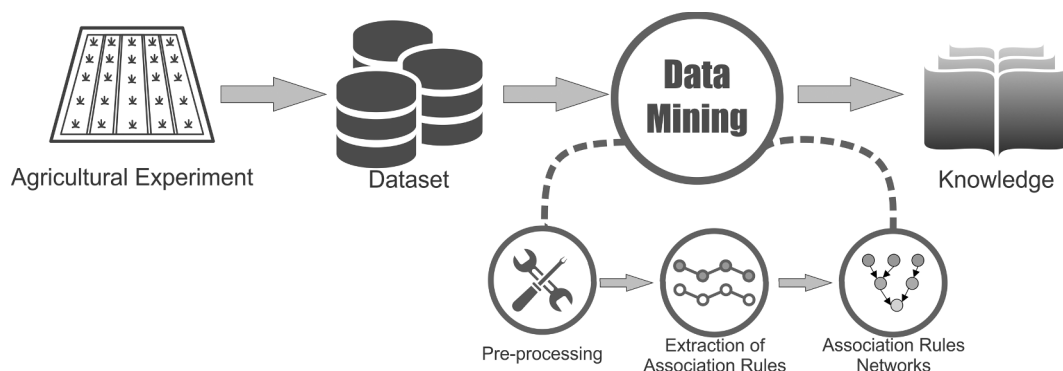

Fig. 3. Steps followed in this study.

root mass (MFR), and root dry mass (MSR). Germination (G) and flowering (F) were evaluated as well as pod formation (VF). Flowering by 50% of the plants was treated as full flowering.

All of the botanical materials, including the stems and roots, were weighed to obtain the fresh biomass values (MFPA and MFR) using precision scales. These materials were then placed in paper bags and dried in a forced ventilation oven at 65 °C for 48 h until constant weight to determine the dry biomass values (MSPA and MSR).

In addition to the measures mentioned above, we determined the diameters of the stems (DC) and calculated the mean values for emergence (E), flowering (F), AP, and pod ripening (V). A table was prepared containing all of the values to pre-process the data.

4.2. Data preprocessing

The data collected in the planting stage were categorized according to the values obtained. We distributed all of the parameters among the six classes according to Table 2. The classes were established using the statistical categorization method described by Barbetta et al. (2010) where we calculated the interval (I) of the values for each variable (Eq. (3)), and determined *Max* and *Min* as the maximum and minimum values, respectively.

$$I = \frac{Max - Min}{6} \tag{3}$$

In the interval set, the first class ranged from *Min* to *Min* + *I* and the last class ended at *Max*.

Using the data distributed into the six categories, we obtained array containing the parameters and classes, where each row represented a class and each column was an attribute. The data set was processed so each value was defined in the format of: "attribute = value." For example, in the case of the "MFPA" attribute, the value was changed to "MFPA = 1" when the value was within the limits of class "1" in the MFPA attribute and "MFPA = 2" for values in the second class, etc., where we executed the same procedure for all of the attributes and classes.

After obtaining all of the parameters, we calculated the half-life decomposition time for each plant sample at intervals of 30, 60, 90, and 120 days. The decomposition of the residues followed the simple exponential model described by Rezende et al. (1999) (Eq. (4)), where X represents the amount of dry matter after a period of time t, X_0 is the initial dry matter quantity, k is the decomposition constant, and t is the time in days.

$$X = X_0 . e^{-k.t} \tag{4}$$

By rearranging the terms in Eq. (4), it is possible to calculate the decomposition constant or value k (Eq. (5)).

$$k = \frac{ln\frac{X}{X_0}}{t} \tag{5}$$

The half-life decomposition time ($t_{\frac{1}{2}}$) can be used to evaluate the time required in days for half of the material to decompose. According to Rezende et al. (1999), it is possible to calculate the half-life decomposition time by using the decomposition constant k (Eq. (6)) as follows:

$$t_{\frac{1}{2}} = \frac{ln(2)}{k} \tag{6}$$

The six classes of the decomposition half-life times obtained in each closed interval were categorized. The period of 120 days was selected to ensure the consistency of the analysis because this was the longest time interval and it allowed better measurements to be obtained in large plantations.

4.3. Extraction of association rules

The *Apriori-TID* algorithm was used to extract the association rules. The data set comprised six classes related to the half-life decomposition time, 11 attributes, and 28 instances for the legume plants.

A minimum support value was selected as 0.3 in order to obtain a relevance of 30% for the values of each parameter. Minimum support values less than 0.3 generated an unusually high number of rules, whereas values above 0.3 produced far less rules than the number of parameters. The minimum confidence was set at 0.50 to consider an LHS influence of 50%, and we distinguished between plants from the same genus (Crotalária). In addition, the size of the rule was corrected in two elements, where we considered one item in the LHS and one item in the RHS. We obtained 154 association rules using these settings.

4.4. ARN construction

According to the method described by Pandey et al. (2009), we constructed an ARN with the "Half-life" attribute for six classes as the objective item. This target item was selected according to the decomposition criterion in order to determine the characteristics related to a longer half-life for the residues used as green manure, with the aims of obtaining greater protection for the soil and better resources for planting by promoting greater decomposition. We plotted the ARN with "Half-life = 6.0" as the target.

4.5. Decision tree

The decision tree method is generally employed for classification and regression tasks in data mining. The decision tree method continually decomposes a data set into smaller subsets according to resource priority until it reaches an appropriate level of disassembly (Nourani and Molajou, 2017).

The final graphical result is a tree with decision nodes and leaf nodes. A decision node comprises non-class attributes and the sub-branches represent all possible values for the attribute. A leaf node represents a specific class considered in a study.

In this study, we only analyzed the final graphical result regardless of the accuracy indexes for the classification of the data set. It was possible that several different configurations could improve the classification accuracy but all methods generated the same final graphic result, and thus we selected the default settings for the search process.

We generated the decision trees using the J48 algorithm available from Weka[1] with the default configuration.

5. Results and discussion

According to the method described above, we calculated the half-life times ($t_{\frac{1}{2}}$) by choosing intervals of 30, 60, 90, and 120 DAS. These intervals were selected to standardize the decomposition behavior of the plants.

The half-life times obtained at intervals of 30, 60, 90, and 120 DAS are shown in Table 3.

The parameter $t_{\frac{1}{2}}$ = 6.0 or "[Half-Life] = 6.0" (Fig. 5) was selected as the ARN target, which represented the parameter with the most extended half-life, thereby resulting in a longer decomposition time for the green manure.

After verifying the nodes with a level of 1 (one), i.e., directly connected to the target, we identified seven appropriate conditions related a longer decomposition time. Thus, in order to visualize this information, these rules (edges) were highlighted so the observation of the patterns was optimized. The complete network can be accessed online. [2]

---

[1] http://www.cs.waikato.ac.nz/ml/weka/.
[2] http://sites.labic.icmc.usp.br/dariobcalcada/tese/dariobcalcada_files/greenmanure/.

Table 2
Distribution of parameters analyzed in the six classes.

Variable	Class = 1	Class = 2	Class = 3	Class = 4	Class = 5	Class = 6
MFPA (t/ha)	9.60–22.47	22.47–35.34	35.34–48.20	48.20–61.07	61.07–73.93	73.93–86.80
MSPA (t/ha)	5.00–9.27	9.27–13.53	13.53–17.8	17.80–22.07	22.07–26.33	26.33–30.60
MFR (t/ha)	0.40–1.53	1.53–2.67	2.67–3.80	3.80–4.93	4.93–6.07	6.07–7.20
MSR (t/ha)	0.24–0.63	0.63–1.03	1.03–1.42	1.42–1.81	1.81–2.21	2.21–2.60
E (DAS)	2.00–2.50	2.50–3.00	3.00–3.50	3.50–4.00	4.00–4.50	4.50–5.00
F (DAS)	57.00–75.33	75.33–93.67	93.67–112.00	112.00–130.33	130.33–148.67	148.67–167.00
AP (m)	0.65–0.87	0.87–1.08	1.08–1.30	1.30–1.51	1.51–1.72	1.72–1.94
V (DAS)	73.00–91.00	91.00–109.00	109.00–127.00	127.00–145.00	145.00–163.00	163.00–181.00
DC (cm)	0.10–0.22	0.22–0.33	0.33–0.45	0.45–0.57	0.57–0.68	0.68–0.80
R/plant	1.00–4.50	4.50–8.00	8.00–11.50	11.50–15.00	15.00–18.50	18.50–22.00

Table 3
Categorization of the half-life times for samples collected at 30, 60, 90, and 120 DAS.

| Class	$t_{\frac{1}{2}}30$ (days)	$t_{\frac{1}{2}}60$ (days)	$t_{\frac{1}{2}}90$ (days)	$t_{\frac{1}{2}}120$ (days)
1	94.44–877.37	40.57–404.13	71.70–1017.70	88.98–264.16
2	877.37–1660.30	404.13–767.70	1017.70–1963.69	264.16–439.35
3	1660.30–2443.22	767.70–1131.26	1963.69–2909.68	439.35–614.53
4	2443.22–3226.14	1131.26–1494.82	2909.68–3855.67	614.53–789.71
5	3226.14–4009.07	1494.82–1858.38	3855.67–4801.66	789.71–964.89
6	4009.07–4792.00	1858.38–2221.94	4801.66–5747.65	964.89–1140,08
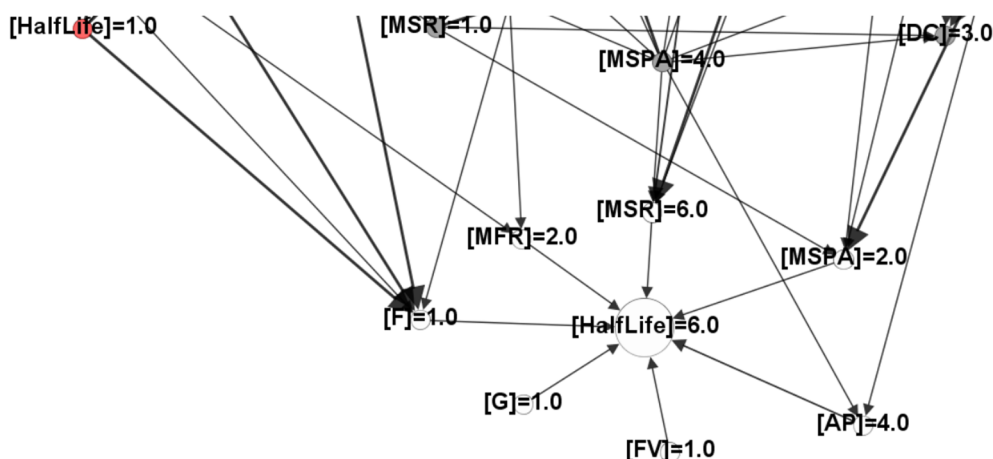
Fig. 5. ARN with the target "Half-Life = 6.0" and level one node (L = 1).

Knowledge exploration was performed strictly with level one nodes (L = 1) because they were the variables that directly affected the target situation, i.e., rules formed by predecessors that are more likely to interfere with the outcome in terms of the half-life in the long term.

After analyzing the ARN, we first determined the nodes without predecessors, i.e., "[G] = 1.0" and "[FV] = 1.0." Thus, we inferred that plants with faster times for germination and pod formation tended to decompose more slowly, thereby suggesting that these are important characteristics for the evaluation of new compounds. Ferreira et al. (2008) stated that plants with early germination also tend to decompose more slowly, which supports the results that we obtained by knowledge extraction with the ARN technique.

For the MFR parameter, we found that the "[MFR] = 2.0" node indicated a low rate for this index in plants with a longer half-life time. The "[MSR] = 6.0" node was connected to high values for all other mass items (MFR, MSR, and MSPA), which suggests that more suitable species should have high dry masses in terms of their root and aerial compositions. Thus, we hypothesize that plants with higher dry mass production have greater potential for use as green manure because they have a longer half-life decomposition time.

The nodes comprising "[MSPA] = 2.0" and "[F] = 1.0" were

influenced by lower half-life times, i.e., "[Half-Life] = 3.0" and "[Half-Life] = 1.0," respectively, which indicates that they merit further study. The decomposition rate was also related to category 4.0 (four) for the AP parameter.

We compared the ARN results with the decision tree results. The decision tree generated by the J48 algorithm is presented in Fig. 6. The J48 algorithm obtained a tree with 16 leaves.

After comparing the ARN output with the graphical results obtained by the decision tree (Fig. 6), we observed differences in the explanations of the objective items. For the "[HalfLife] = 6.0" objective, only the direct possibility of the parameter [F] referring to the flowering of plants was observed in the tree, and the values "[F] = 1.0" were obtained when "[MSPA] = 1.0" and "[F] = 1.0, 2.0, 4.0, or 5.0" when "[MSPA] = 2.0." The J48 algorithm completely ignored the other parametric conditions for obtaining higher half-lives.

Seven hypotheses with different parameters were formulated based on the ARN visualization. By contrast, only two suggestions were possible when we viewed the decision tree. A specialist evaluated the hypotheses obtained based on the ARN and confirmed by experimental evidence (Teodoro et al., 2018). The recommendations made based on the decision tree visualization were not confirmed by the expert.
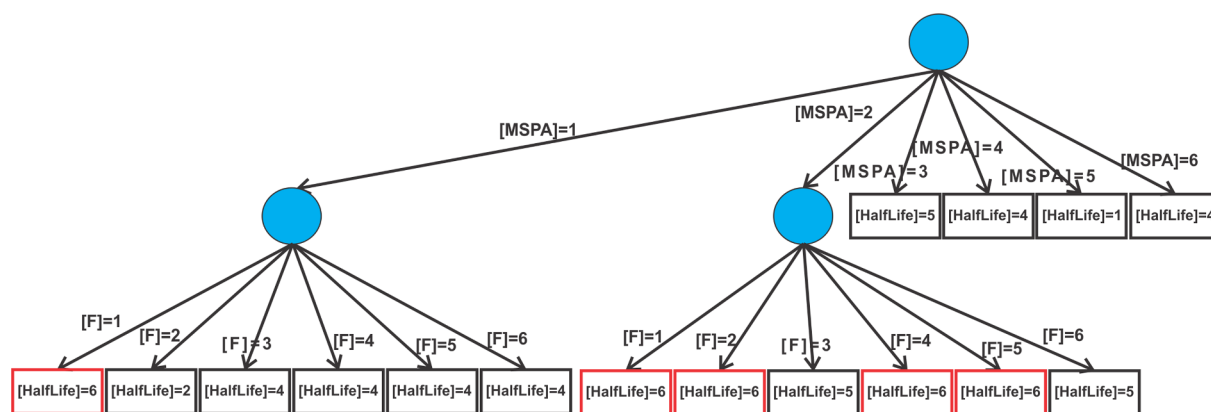
Fig. 6. Decision tree constructed with the J48 algorithm.

Therefore, the application of ARN allowed the consideration of more parameters to optimize the extraction of knowledge based on more relevant hypotheses.

6. Conclusion

In this study, we showed that ARNs could be used to optimize the discovery of knowledge related directly to green fertilizer by generating hypotheses connected directly to the decomposition characteristics of plants.

The results obtained using the association rule mining technique with ARNs were confirmed by similar conclusions reported in experimental agricultural studies (Teodoro et al., 2018), thereby supporting the credibility of the association rule mining technique. The ARN also allowed us to generate new hypotheses that might be confirmed by further experiments. The ARN facilitated the simple graphical analysis of the content of the data considered.

We also compared the ARN results with the decision tree generated by the J48 algorithm, which is a decision tree algorithm that is available in Weka. The decision tree only classified the elements; hence, it did not perform well at describing the relationships. The ARN obtained excellent results by describing the data using the extracted association rules, and thus we could select the rules that were most likely to be responsible for the occurrence of the target item.

The hypotheses generated in this study can be investigated in future agricultural experiments to assess the patterns identified based on the results of this research.

In future research, we will conduct association rule mining for other species that can be used as green manure and compare them with the productivity of crops that are commonly used, as well as applying different techniques for probabilistic analysis to investigate the effects of the parameters on the desired objective item.

Acknowledgments

References

Aggarwal, C.C., Procopiuc, C., Yu, P.S., 2002. Finding localized associations in market basket data. IEEE Trans. Knowl. Data Eng. 14 (1), 51–62.

Agrawal, R., Imielinski, T., Swami, A., 1994. Mining association rules between sets of items in large databases. Special Interest Group Manage. Data 22 (2 22 2), 207–216.

Agrawal, R., Shafer, J.C., 1996. Parallel mining of association rules. IEEE Trans. Knowl. Data Eng. 8 (6), 962–969.

Agrawal, R., Srikant., R., 1994. Fast algorithms for mining association rules. In: Bocca, Jorge B., Jarke, Matthias, Zaniolo, Carlo (Eds.), Proceedings of Twentieth International Conference on Very Large Data Bases, VLDB, pp. 487–499.

Ambrosano, E.J., Rossi, F., Guirado, N., Schammass, E.A., Muraoka, T., Trivelin, P.C.O., Ambrosano, G.M.B., 2014. Adubação verde na agricultura orgânica. In: Filho, O.F. de L, Ambrosano, E.J., Rossi, F., Carlos, J.A.D. Adubação Verde e plantas de cobertura no Brasil: fundamentos e prática., first ed., EMBRAPA, Brasília, pp. 45–80 (Ch. 15).

Barbetta, P.A., Reis, M.M., Bornia, A.C., 2010. Estatística para cursos de engenharia e informática, third ed. Atlas, São Paulo.

Bastos, E.A., de Andrade Junior, A.S., Rodrigues, B.H.N., 2012. Boletim agrometeorológico de 2011 para o município de Parnaíba, Piauí. Tech. rep., Empresa Brasileira de Pesquisa Agropecuária (Embrapa Meio Norte) (Documents, 221), Teresina.

Batarseh, F.A., Yang, R., 2018. A Deployment Life Cycle Model for Agricultural Data Systems Using Kansei Engineering and Association Rules. In: Federal Data Science. Elsevier Inc, pp. 141–159 Ch. 9.

Calegari, A., 2014. Perspectivas e estratégias para a sustentabilidade e o aumento da biodiversidade dos sistemas agrícolas com o uso de adubos verdes. In: Filho, O.F. de L., Ambrosano, E.J., Rossi, F., Carlos, J.A.D. (Eds.), Adubação Verde e plantas de cobertura no Brasil: fundamentos e prática, first ed. EMBRAPA, Brasília, pp. 21–36 Ch. 1.

Chawla, S., 2010. Feature Selection, Association Rules Network and Theory Building. In: JMLR: Workshop and Conference Proceedings - The Fourth Workshop on Feature Selection in Data Mining 10, pp. 14–21.

de Barros, F.M.M., Oliveira, S.R. d. M., de Oliveira, L.H.M., 2013. Desenvolvimento e validação de um sistema de recomendação de informações tecnológicas sobre cana-de-açúcar. Bragantia 72 (4), 387–395.

Domingues, M.A., Rezende, S.O., 2005. Post-processing of Association Rules using Taxonomies. In: Proceedings of the 12th Portuguese Conference on Artificial Intelligence (EPIA 2005), pp. 192–197.

Drury, B., Valverde-Rebaza, J., Moura, M.F., de Andrade Lopes, A., 2017. A survey of the applications of Bayesian networks in agriculture. Eng. Appl. Artif. Intell. 65 (July), 29–42.

Ferreira, N.R., Medeiros, R.B., Soares, G.L.G., 2008. Potencial Alelopático do Capim-Annoni-2 (Eragrostis plana Nees) na germinação de sementes de gramíneas perennes estivais. Revista Brasileira de Sementes 30 (2), 043–050.

Garcia, L.F., 2002. Introdução e avaliação de leguminosas para adubação verde em solos arenosos de Tabuleiros Costeiros do Piauí. Rev. Fac. Agron. (Maracay) 28, 93–103.

Grahne, G., Zhu, J., 2005. Fast algorithms for frequent itemset mining using FP-trees. IEEE Trans. Knowl. Data Eng. 17, 1347–1362.

Han, E.-H., Karypis, G., Kumar, V., Mobasher, B., 1997. Clustering based on association rule hypergraphs. SIGMOD Workshop on Research Issues on Data Mining and Knowledge Discovery, (SIGMOD-DMKD'97).

Hira, S., Deshpande, P.S., 2015. Data analysis using multidimensional modeling, statistical analysis and data mining on agriculture parameters. Proc. Comput. Sci. 54, 431–439.

Kamilaris, A., Kartakoullis, A., Prenafeta-Boldú, F.X., 2017. A review on the practice of big data analysis in agriculture. Comput. Electron. Agric. 143 (September), 23–37.

Kumar, M.B.S., Balakrishnan, K., 2019. Development of a Model Recommender System for Agriculture Using Apriori Algorithm. In: Springer, Singapore, pp. 153–163.

Kuo, Y.T., Lonie, A., Pearce, A.R., Sonenberg, L., 2014. Mining surprising patterns and their explanations in clinical data. Appl. Artif. Intell. 28 (2), 111–138.

Le, T., Vo, B., 2016. The lattice-based approaches for mining association rules: a review. Wiley Interdiscip. Rev.: Data Min. Discov. 6 (4), 140–151.

Lehmann, R.J., Reiche, R., Schiefer, G., 2012. Future internet and the agri-food sector: state-of-the-art in literature and research. Comput. Electron. Agric. 89, 158–174.

Lucas, M.T., Chhajed, D., 2004. Applications of location analysis in agriculture: a survey. J. Oper. Res. Soc. 55 (6), 561–578.

Martins, C.R., Barros, I., 2015. Intensificação ecológica da fruticultura: sistema de produção ecologicamente intensivo de coco e citros, na Região Norte e Nordeste do Brasil. In: IX Congresso Brasileiro de Agroecologia, vol. 10. Belém-PA, pp. 1–5.

Melo, F.d.B., Cavalcante, A.C., de Andrade Junior, A.S., Bastos, E.A., 2004. Levantamento Detalhado dos Solos da Área da Embrapa Meio-Norte/UEP de Parnaíba et al. Tech. rep., Empresa Brasileira de Pesquisa Agropecuária (Embrapa Meio Norte) (Documents, 89), Teresina.

Nguyen, L.T.T., Nguyen, N.T., 2015. Updating mined class association rules for record

insertion. Appl. Intell. 42 (4), 707–721.

Nourani, V., Molajou, A., 2017. Application of a hybrid association rules/decision tree model for drought monitoring. Global Planet. Change 159, 37–45.

Pandey, G., Chawla, S., Poon, S., Arunasalam, B., Davis, J.G., 2009. Association rules network: definition and applications Gaurav. Stat. Anal. Data Min. 1 (4), 179–260.

Phadikar, S., Sil, J., Das, A.K., 2013. Rice diseases classification using feature selection and rule generation techniques. Comput. Electron. Agric. 90, 76–85.

Rezende, C.d.P., Cantarutti, R.B., Braga, J.M., Gomide, J.A., Pereira, J.M., Ferreira, E., Tarré, R., Macedo, R., Alves, B.J.R., Urquiaga, S., Cadisch, G., Giller, K.E., Boddey, R.M., 1999. Litter deposition and disappearence in Brachiaria pastures in the Atlantic Forest region of the south of Bahia, Brazil. Nutr. Cycl. Agroecosyst. 54, 99–112.

Teodoro, M.S., Castro, K.N. d. C., Magalhães, J.A., 2018. Assessment of legumes with potential use as green manure in the Coastal Tablelands of Piauí States, Brazil. Caatinga.

Thornthwaite, C.W., Mather, J.R., 1955. The water balance. In: of Technology, C.D.I. (Ed.), Publication in climatology, 1st ed. Drexel Institute of technology, pp. 104.

Yan, X., Zhang, S., Zhang, C., 2007. On data structures for association rule discovery. Appl. Artif. Intell. 21 (2), 57–79.

Zaki, M.J., Meira, M.J., 2013. Data Mining and Analysis: Fundamental Concepts and Algorithms. Cambridge University Press.