



Using Decision Trees to Extract Patterns for Dairy Culling Management

M. Lopez-Suarez¹, E. Armengol²(✉), S. Calsamiglia¹,
and L. Castillejos¹

¹ Animal Nutrition and Welfare Service, Department of Animal and Food Sciences, Universitat Autònoma de Barcelona, Bellaterra, Spain
{Montserrat.Lopez.Suarez, Sergio.Calsamiglia, Lorena.Castillejos}@uab.cat

² Artificial Intelligence Research Institute (IIIA - CSIC),
Campus UAB, 08193 Bellaterra, Catalonia, Spain
eva@iiia.csic.es

Abstract. The management of a dairy farm involves taking difficult technical and economic decisions such as the replacement of some cows to either maintain or increase the productivity of the farm. However, there is not a standard method supporting the selection procedure of which animals need to be culled. In the present study we used decision trees to develop a model able to classify a cow according to the average herd productivity. This model, obtained from a data base around 98000 cows, predicts the average milk production of the first lactation of a cow based on the monthly milk controls corresponding to the lactation peak. Our goal is to identify poor productive cows during her first lactation in order to make more accurate selections of which cows should be culled.

Keywords: Veterinary · Dairy farms · Milk production · Voluntary culling
Artificial intelligence · Machine learning · Decision trees

1 Introduction

To be highly productive a dairy farm needs good management. There are many complex technical and economic decisions that have to be taken in order to maintain or increase the productivity. Direct benefits of a dairy are those coming from either milk or animal sales. However, raising a cow has also many costs that should be discounted from these benefits (feeding, labor, veterinary care, depreciation of facilities, utilities, etc.).

Culling is the departure of cows from the herd due to sale, slaughter or death. The main reasons to cull a cow are infertility, mastitis and poor milk production. Commonly, culling reasons have been classified as voluntary or involuntary or, as suggested by Fetrow et al. [4], economic or biologic. Biological culls are those cows for which no possible productive future exists due to disease, injury or infertility. Thus, this class of culls are mainly involuntary as most of the times are “forced” decisions. Economic culls mean that a cow is removed because a replacement is expected to produce greater

profit. In this case, farmer has freedom of choice over which cows are removed from the herd, although they are healthy [1, 4]. Hence, the farmer can do a voluntary selection of cows to cull based in the herd size and herd production level. Therefore, herd profitability can be improved by minimizing the proportion of the herd culled for involuntary or biological reasons and maximizing the proportion culled for voluntary or economic reasons [1]. Because it is important to know as soon as possible if a cow will be poorly productive, we propose to analyze first lactation production data to identify those animals in a herd which are candidates to be culled following milk yield improvement criteria.

For such a purpose, in the present paper, we used Decision Trees (DT) [9], an Artificial Intelligence technique that allows the classification of objects, to extract patterns for making decision support during the culling process. This is our main contribution since the use of these patterns is an easy and understandable way to support the farmer in culling decision making. A DT represents a model of a given domain that can be used for classification tasks. Thus, given the description of a cow, a DT could classify her as *Good* or *Bad*, supporting in this way the culling process. In Sect. 2 we briefly revise some works that use Artificial Intelligence technologies for dairy management. Section 3 explains how we have modelled the culling task, and the data base we used in our experiments (Sect. 3.1). Section 4 explains our results, and in Sect. 5 these results are discussed. Finally, Sect. 6 is devoted to conclusions.

2 Related Work

Most of the work focused on modelling the culling task or aspects related to it, try to construct statistical models based on the analysis of past cases of a dairy [2]. The use of artificial intelligence techniques is still not widely used for managing the culling although they have been used for other purposes. For instance, Cavero et al. [3] developed a fuzzy logic model for mastitis detection; [13, 15] are some examples of using neural networks for classification and control of mastitis in cows milked using an automatic milking system; Shainfar et al. [10] used fuzzy neural networks to predict breeding values for dairy cattle; Grzesiak et al. [6] also used neural networks to predict milk production; Sugiono et al. [12] built an adaptive system (BPNN) to predict performance of dairy cattle based on environmental and physical data; and Sitkowska et al. [11] used decision trees to predict the increment of the levels of somatic cells in milk. Thus, to our knowledge, Artificial Intelligence techniques and, particularly, decision trees, have not been used to support the decision of culling the herd.

Also the multi-agent technology has been used in dairy farms to support the process of decision making in several aspects of a dairy farm. For instance, Parrot et al. [8] investigated the feasibility of using a multi-agent system for heifer management; Thangaraj et al. [14] used a multi-agent system to make integrated decisions taking into account pasture availability, nutrients, and herd economic aspects. Goel et al. [5] also developed a multi-agent system to implement electronic contracting of food grains integrating various millers and producers in a food supply chain for better negotiations to reach a mutually acceptable price.

The approach we propose uses decision trees, a technique that, as far as we know, has not been used before in dairy management although Kamphuis et al. [7] used decision trees to improve the detection of clinical mastitis with sensor data from automatic milking systems.

Most of the mentioned studies using Artificial Intelligence technologies employed neural networks and fuzzy logic. This is mainly due to real time needs, since they deal with problems that need to be solved in a short time. However, we considered that, for the objective of this study, the use of inductive learning methods to construct a domain model could be a better approach. Inductive models, differently than neural networks, can explain the results they provide since the descriptions used to classify an object can be interpreted as an explanation of the result. This is an interesting characteristic as it enables the final user to completely understand how an automatic system has reached the solution it proposes. For this reason, our approach can contribute to support the farmer in taking more informed decisions when he needs to cull his herd.

3 Modelling the Culling Task

In this paper we propose the use of inductive learning methods to construct a model able to predict and explain why a cow should be culled during her first lactation. The goal of inductive learning algorithms is the construction of a domain theory from the known data. Commonly this domain theory is further used to predict the classification of unseen objects.

To characterize each one of the classes of the domain by means of discriminatory descriptions we have to solve the so-called *discrimination task*. Given a solution class C_i , the *discrimination task* for inductive learning methods is defined as follows:

- Given: a set E of positive E^+ and negative E^- examples of a class c_i .
- Find: a description D_i such that it is satisfied by elements in E^+ and it is not satisfied by any of the elements of E^- .

The discrimination task produces discriminatory descriptions, i.e., descriptions satisfied only by objects of one of the classes. A class C_i can be described by more than one discriminatory description D_i . To build a model of a domain we have to perform the discrimination task over each one of the classes. In that way, the model is composed of descriptions that can classify univocally an object as belonging to a class. One of the widely used inductive learning methods are the *Decision Trees (DT)*.

The goal of using DT is to create a domain model predictive enough to classify future unseen domain objects. The leaves of a tree determine a partition of the original set of examples, since each domain object only can be classified following one of the paths of the tree. The construction of a decision tree is performed by splitting the source set of examples into subsets based on an attribute-value test. This process is repeated on each derived subset in a recursive manner. Figure 1 shows the ID3 algorithm [9] commonly used to construct decision trees. The path from the root to each one of the leaves of a decision tree can be seen as a description of a class. When all the examples of a leaf belong to the same class the description is *discriminatory*, otherwise it is *non-discriminatory*. In our experiments we implemented our own version of the algorithm

in Fig. 1 in order to control overfitting and also to construct the patterns from the tree paths.

Decision trees can be useful for our purpose because their paths give us patterns describing classes of objects (cows in our approach) in a user-friendly manner. One shortcoming of decision trees is *overfitting*, meaning that there are few objects in most of the leaves of the tree. In other words, paths are actually descriptions that poorly represent the domain. The main procedure to either avoid or reduce overfitting is by pruning the tree, i.e., under some conditions, a node is no longer expanded. However, this means that leaves can contain objects belonging to several classes and, therefore, paths do not represent discriminatory descriptions of classes, i.e., these descriptions are satisfied by objects of more than one class. In our approach, we managed overfitting by controlling the percentage of elements of each class. Let S_N be the set of objects associated with an internal node N , the stopping condition in expanding N (the *if* of the ID3 algorithm) holds when the percentage of objects in S_N that belong to the majority class decreases in one of the children nodes. In such a situation, the node N is considered as a leaf.

3.1 The Data Base

We used a data base containing 97987 objects. These objects are descriptions of Holstein-Frisian cows which lived from 2006 to 2016, belonging to dairy farms within the CONAFE register system¹. Nowadays, most of farms have automatic systems to collect data, so there is a lot of information about each cow (genetics, production, morphology, reproductive indexes, disease control, etc.). In addition, most of farms pass a monthly control in which, for every lactating cow, the day milk yield is registered and a milk sample for analyses is taken. Therefore, the first step we carried out was to select which pieces of this information could be useful to detect as soon as possible poorly productive cows. To reach this goal, we decided to use only information relative to the first lactation. The attributes we considered for every cow were the following one:

- **BirthMonth:** Month (season) in which the cow was born.
- **Month1Calving:** Month (season) of the first calving of a cow.
- **KI:** Milk production genetic index.
- **ICO:** Official cattle breeding index in Spain.
- **Morpho:** Morphologic qualification of a cow.
- **KgMilkPeak:** Average test-day milk yield (kg/day) of the second and third control of the first lactation (lactation peak).
- **Fat:** Fat average percentage from the second and third controls of the first lactation.
- **Protein:** Protein average percentage from the second and third control of the first lactation.
- **SCC:** Somatic cell count in the milk. It is an indicator of the quality of milk as it expresses the likeliness to contain harmful bacteria.

¹ The Confederacion Nacional de la Raza Frisona (CONAFE) is a Spanish entity whose goal is to develop programs oriented to the improvement and selection of the Holstein-Frisian herd.

```

ID3 (examples, attributes)
  create a node
  if all examples belong to the same class return class as the label for the node
  otherwise
    A ← best attribute
    for each possible value  $v_i$  of A
      add a new tree branch below node
      examples $_{v_i}$  ← subset of examples such that A =  $v_i$ 
      ID3(examples $_{v_i}$ , attributes - {A})
  return node

```

Fig. 1. ID3 algorithm for growing a decision tree.

- **OpenDays:** Days from calving to conception.
- **Calving1stAI:** Interval of days between the first calving and the first insemination after it.
- **AI:** Number of artificial insemination attempts to conceive after the first calving.
- **Production/DIM:** Average daily milk production of the first lactation (kg/day) calculated dividing total amount of milk produced by a cow during the whole lactation by the total days in milk (DIM).

For the attributes **BirthMonth** and **Month1Calving** we divided the months according to seasons. All the remaining attributes have numerical values and we have discretised them. For all of them, we calculated the quartiles and divided the whole interval of values in four according to these quartiles. We associated to each of the 4 quartile interval the labels: *VeryLow (VL)*, *Low (L)*, *High (H)*, and *VeryHigh (VH)*. Table 1 shows the quartiles of the attributes **KgMilkPeak** and **Production/DIM**. For instance, for the attribute **KgMilkPeak**, the interval from 5 to 28 of milk average during the lactation peak, means that the cow has had very low milk production (*VL*). We considered **Production/DIM** as the solution class, i.e., the decision tree will model and predict the first lactation milk production performance of a cow (kg/day).

4 Results

We used the whole data base as input for a decision tree algorithm to obtain patterns (the tree paths) describing the classes of cows according to the values of the attribute **Production/DIM**. The resulting tree was formed by patterns that are discriminatory and by patterns that are not (due to the stopping condition explained in previous section). Figure 2 shows an example of discriminatory pattern formed by 6 attributes (**KgMilkPeak**, **KI**, **Morpho**, **Month1Calving**, **AI**, and **Fat**). This pattern is satisfied by 104 cows, where the 93.27% of them have a very low **Production/DIM** and the 6.73% of them have a low **Production/DIM**. None of the cows satisfying this pattern has **Production/DIM** high or very high. For model evaluation purposes, we decided to consider two final solution classes: *Good* formed by cows with **Production/DIM** high or very high; and *Bad* formed by cows with **Production/DIM** low or very low. This is

Table 1. Intervals corresponding to each quartile of the attributes **KgMilkPeak** and **Production/DIM**

Attribute	<i>VL</i>	<i>L</i>	<i>H</i>	<i>VH</i>
KgMilkPeak	(5, 28]	(28, 32]	(32, 36]	(36, 65]
Production/DIM	(6, 25]	(25, 29]	(29, 32]	(32, 60]

because the actual goal is to identify the less productive cows of the herd, Therefore, the pattern in Fig. 2 classifies all the cows that satisfy it as *Bad* with a 100% of predictability. Once the good cows have been separated, a more fine procedure could be used to select which of the bad cows are the worst. In fact, this procedure could depend on factors different for each dairy.

We evaluated the model using 10-fold cross-validation. Therefore, the whole data base has been divided in 10 parts. In each experiment we taken 9 of these parts to generate the patterns and the remaining part has been used as test on which we estimated the accuracy of the model. After the 10 experiments we obtained a mean accuracy of 84.10% in classifying a cow as *Bad* or *Good*.

We also tested the developed model using a different data base to check if the set of patterns shown in Table 2 could be used to classify the cows, using the same criteria, on any farm. For this test, we took a data base with information about 5474 cows different from the ones used during the construction of the model. In fact, these cows were alive from 1999 to 2006. We divided the data base in subsets of around 500 cows, and each subset was discretised according to the quartiles of the own subset. Then we used the patterns in Table 2 to classify the cows of each subset simulating, in that way, the difference between dairies. The mean accuracy of classification in this test was 82.72%. Comparing both this accuracy with the one from the 10-fold cross-validation (84.10%) we can conclude that it is feasible to use the model obtained from the data base of around 98000 since it provides good patterns for the classification of cows coming from different farms.

5 Discussion

All the patterns composing the model in Table 2 explain that milk production (kg/day) in the lactation peak of the first lactation (**KgMilkPeak**) of a cow is directly related to the average milk production (kg/day) of the whole first lactation (**Production/DIM**). This result is very interesting as the performance of a cow in her first lactation can be predicted in early lactation stages instead of having to wait until the end of the lactation (after 9 to 14 months). This is especially interesting to take early decisions about culling since, with around a 97% of predictability, poorly productive cows of a herd can be early detected. Therefore, the model composed by patterns that take into account only the attribute **KgMilkPeak** is a good predictor of the first lactation production of a cow.

We also performed a statistical analysis to compare the results of the DT model with the statistical one. Table 3 shows that all the variables included in the model were significant ($p < 0.05$). The most relevant variable is **KgMilkPeak** as in the DT model.

Pattern : [[KgMilkPeak, VL], [Kl, VL], [Morpho, VL],
 [Month1Calving, FirstTrim], [AI, VH], [Fat, VH]]

Classes of the cows satisfying the pattern: [[VL, 93.27], [L, 6.73]]
 Final class: Bad (100 %)
 Satisfied by 104 cows

Fig. 2. Discriminatory pattern satisfied by 104 cows having **Production/DIM** low or very low. According to this pattern, all these cows will be classified as Bad.

Table 2. Model formed by non-discriminatory patterns. For each pattern it is shown the number of cows that satisfy it, and the percentage of them that are *VeryLow* (VL), *Low* (L), *High* (H), and *VeryHigh* (VH). In addition, column *Bad* shows the sum of both bad and very bad cows, and column *Good* shows the sum of both good and very good cows.

Pattern	Cows	VL	L	H	VH	Bad	Good
[[KgMilkPeak, VL]]	27179	75.16	22.02	2.36	0.46	97.18	2.82
[[KgMilkPeak, L]]	25571	22.04	55.21	19.16	3.58	77.25	22.74
[[KgMilkPeak, H]]	23651	4.57	31.71	41.18	22.54	36.28	63.72
[[KgMilkPeak, VH]]	21586	0.65	7.66	21.55	70.14	8.31	91.69

Other variables, although significant, are much less important in the statistical model. This result can be interpreted as that the production of the first lactation can be predicted from the records of the milk control from the 3–4 months of lactation with a determination coefficient of 0.7176, i.e., the variable **KgMilkPeak** explains the 71.76% of the variability of the variable **Production/DIM**. By adding **Kl**, the model explains the 74.36%.

Therefore, the statistical model confirms the results we obtained with the DT model. However, whereas the statistical model only shows the high correlation between both **KgMilkPeak** and **Production/DIM**, the patterns of the DT model show explicitly the relation between certain values of both attributes. For instance, a Low value of **KgMilkPeak** corresponds to a Low value of **Production/DIM**.

The model in Table 2 supports to split the herd in two groups corresponding to *Good* and *Bad* cows, regarding first lactation production. Once the *Good* cows have been separated, a next step to take could be the addition of more specific patterns to select which of the *Bad* cows are the worst. We could use, for instance, patterns including the next more relevant attribute which, as in the statistical model, is **Kl**. An example of such a pattern is [[KgMilkPeak, L], [Kl, VL]] satisfied by 7145 cows, the 89.85% of them are *Bad* and the remaining 10.15% are *Good*. This example shows the advantage of the DT model compared to the statistical model since the latter only shows that both **KgMilkPeak** and **Kl** explain the 74.36% of **Production/DIM**, whereas by using the patterns from the DT model, we know the range of values in both **KgMilkPeak** and **Kl** (L and VL respectively) that support the decision of culling a cow.

Table 3. Statistical model obtained using Multiple Regression Model using Stepwise Selection Method (SAS 9.4).

Step	Variable	Partial R ²	Model R ²	C(p)	F-value	Pr > F
1	KgMilkPeak	0.7176	0.7176	12984.4	248933	<.0001
2	Kl	0.0260	0.7436	2757.89	9948.78	<.0001
3	Morpho	0.0045	0.7481	995.217	1747.00	<.0001
4	Month1Calving	0.0019	0.7500	239.494	755.91	<.0001
5	Protein	0.0003	0.7503	111.662	129.69	<.0013
6	Age1Calving	0.0002	0.7506	20.8978	92.75	<.0026
7	SCC	0.0000	0.7506	13.8943	9.00	<.0001
8	BirthMonth	0.0000	0.7506	11.1981	4.70	<.0001
9	ICO	0.0000	0.7506	0.7506	4.17	<.0001

6 Conclusions

In the present work we used decision trees to obtain patterns supporting the culling decision process of a dairy farm. From a data base with around 98000 cows, we obtained a model formed by 4 patterns that predicts the milk production of a cow on her first lactation (**Production/DIM**). By using this model, **Production/DIM** can be predicted based on milk yield records from first lactation peak (**KgMilkPeak**) with an 84% of accuracy after one trial of 10-fold cross-validation, and 83% of accuracy on different dairy farms. This result is consistent with the statistical model constructed from the same data base, which shows a high correlation between both variables (**Production/DIM** and **KgMilkPeak**). Thus, the model could be a helpful tool for the decision of culling a cow in early stages of her lactation since, poorly productive cows can be identified using a pattern that has a predictability of around 97%.

Acknowledgments. The authors acknowledge data support from CONAFE (Confederacion Nacional de la Raza Frisona). This research is partially funded by the projects (Project AGL2015-67409-C2-01-R) from the Spanish Ministry of Economy and Competitive-ness; RPREF (CSIC Intramural 201650E044); and the grant 2014-SGR-118 from the Generalitat de Catalunya. Authors also thank to Àngel García-Cerdaña his helpful comments.

References

1. Ansari-Lari, M., Mohebbi-Fani, M., Rowshan-Ghasrodashti, A.: Causes of culling in dairy cows and its relation to age at culling and interval from calving in Shiraz, Southern Iran. *Vet. Res. Forum* **3**, 233–237 (2012)
2. Calsamiglia, S., Castillejos, L., Astiz, S., Lopez-DeToro, C., Baucells, J.: A dairy farm simulation model as a tool to explore the technical and economical consequences of management decisions. In: Buiatrics, W.A. (ed.) *Proceedings of the World Buiatrics Congress 2016*, p. 406 (2016)
3. Cavero, D., Tölle, K., Buxadé, C., Krieter, J.: Mastitis detection in dairy cows by application of fuzzy logic. *Livestock Sci.* **105**, 207–213 (2006)

4. Fetrow, J., Nordlund, K.V., Norman, H.D.: Culling: nomenclature, definitions, and recommendations. *J. Dairy Sci.* **89**, 1896–1905 (2006)
5. Goel, A., Zobel, C.W., Jones, E.C.: A multi-agent system for supporting the electronic contracting of food grains. *Comput. Electr. Agric.* **48**(2), 123–137 (2005). <http://www.sciencedirect.com/science/article/pii/S016816990500058X>
6. Grzesiak, W., Blaszczyk, P., Lacroix, R.: Methods of predicting milk yield in dairy cows: Predictive capabilities of wood's lactation curve and artificial neural networks (ANNs). *Comput. Electron. Agric.* **54**(2), 69–83 (2006)
7. Kamphuis, C., Mollenhorst, H., Heesterbeek, J., Hogeveen, H.: Detection of clinical mastitis with sensor data from automatic milking systems is improved by using decision-tree induction. *J. Dairy Sci.* **93**(8), 3616–3627 (2010)
8. Parrott, L., Lacroix, R., Wade, K.M.: Design considerations for the implementation of multi-agent systems in the dairy industry. *Comput. Electr. Agric.* **38**(2), 79–98 (2003). <http://www.sciencedirect.com/science/article/pii/S0168169902001394>
9. Quinlan, J.R.: Induction of decision trees. *Mach. Learn.* **1**(1), 81–106 (1986)
10. Shahinfar, S., Mehrabani-Yeganeh, H., Lucas, C., Kalhor, A., Kazemian, M., Weigel, K.A.: Prediction of breeding values for dairy cattle using artificial neural networks and neuro-fuzzy systems. *Comput. Math. Methods Med.* (ID 127130) (2012)
11. Sitkowska, B., Piwczyński, D., Aerts, J., Kolenda, M., Özkaya, S.: Detection of high levels of somatic cells in milk on farms equipped with an automatic milking system by decision trees technique. *Turkish J. Vet. Anim. Sci.* **41**, 532–540 (2009)
12. Sugiono, S., Soenoko, R., Riawati, L.: Investigating the impact of physiological aspect on cow milk production using artificial intelligence. *Int. Rev. Mech. Eng.* **11**(1), 30–36 (2017)
13. Sun, Z., Samarasinghe, S., Jago, J.: Detection of mastitis and its stage of progression by automatic milking systems using artificial neural networks. *J. Dairy Res.* **77**, 168–175 (2009)
14. Thangaraj, A., Patricia, A., Samarasingh, S.: Modelling a multi agent system for dairy farms for integrated decision making. In: 22nd International Congress on Modelling and Simulation, Hobart, Tasmania, Australia (2017)
15. Wang, E., Samarasinghe, S.: On-line detection of mastitis in dairy herds using artificial neural networks. In: Proceedings of the International Congress on Modelling and Simulation (MODSIM 2005), Melbourne, Australia (2005)