

# Towards modelling beef cattle management with Genetic Programming

Francesca Abbona<sup>a,d</sup> Leonardo Vanneschi<sup>b,c</sup> Marco Bona<sup>d</sup> Mario Giacobini<sup>a</sup>

<sup>a</sup> University of Torino, Largo Paolo Braccini 2, 10095 Turin, Grugliasco, Italy

<sup>b</sup> NOVA Information Management School (NOVA IMS), Universidade Nova de Lisboa, Campus de Campolide, 1070-312 Lisboa, Portugal

<sup>c</sup> LASIGE, Departamento de Informàtica, Faculdade de Ciências, Universidade de Lisboa, 1749-016 Lisboa, Portugal

<sup>d</sup> Associazione Nazionale Allevatori Bovini Razza Piemontese, Carrù, Italy

***This is the accepted author manuscript of the following article published by Elsevier:***

Abbona, F., Vanneschi, L., Bona, M., & Giacobini, M. (2020). Towards modelling beef cattle management with Genetic Programming. *Livestock Science*, 241, 1-12. [104205]. <https://doi.org/10.1016/j.livsci.2020.104205>



***This work is licensed under a [Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License](https://creativecommons.org/licenses/by-nc-nd/4.0/).***

---

1 *Type of the Paper (Research Article)*

## 2 **Towards Modelling Beef Cattle Management with Genetic** 3 **Programming**

4 **Francesca Abbona<sup>1,4,\*</sup>, Leonardo Vanneschi<sup>2,3</sup>, Marco Bona<sup>4</sup> and Mario Giacobini<sup>1</sup>**

5 <sup>1</sup> University of Torino, Turin, Italy;

6 <sup>2</sup> NOVA Information Management School (NOVA IMS), Universidade Nova de Lisboa, Campus de Campolide, 1070-312 Lisboa,  
7 Portugal;

8 <sup>3</sup> LASIGE, Departamento de Informática, Faculdade de Ciências, Universidade de Lisboa, 1749-016 Lisboa, Portugal

9 <sup>4</sup> Associazione Nazionale Allevatori Bovini Razza Piemontese, Carrù, Italy;

10

11 \* Corresponding author.

12 E-mail addresses:

13 [francesca.abbona@unito.it](mailto:francesca.abbona@unito.it) (F. Abbona), [lvanneschi@novaims.unl.pt](mailto:lvanneschi@novaims.unl.pt) (L.Vanneschi), [marco.bona@anaborapi.it](mailto:marco.bona@anaborapi.it) (M. Bona),

14 [mario.giacobini@unito.it](mailto:mario.giacobini@unito.it) (M. Giacobini).

15

16 **Abstract:** Among the Italian Piedmontese Beef Breedings, the yearly production of calves weaned per cow, that is the  
17 calves that survive during the period of 60 days following birth, is identified as the main target expressing the  
18 performance of a farm. Modelling farm dynamics in order to predict the value of this parameter is a possible solution  
19 to investigate and highlight breeding strengths, and to find alternatives to penalizing factors. The identification of such  
20 variables is a complex but solvable task, since the amount of recorded data among livestock is nowadays huge and  
21 manageable through Machine Learning techniques. Besides, the evaluation of the effectiveness of the type of  
22 management allows the breeder to consolidate the ongoing processes or, on the contrary, to adopt new management  
23 strategies. To solve this problem, we propose a Genetic Programming approach, a white-box technique suitable for  
24 big data management, and with an intrinsic ability to select important variables, providing simple models. The most  
25 frequent variables encapsulated in the models built by Genetic Programming are highlighted, and their zoological  
26 significance is investigated *a posteriori*, evaluating the performance of the prediction models. Moreover, two of the  
27 final expressions selected only three variables among the 48 given in input, one of which is the best performing among

---

28 GP models. The expressions were then analyzed in order to propose a zootechnical interpretation of the equations.  
29 Comparisons with other common techniques, including also black-box methods, are performed, in order to evaluate  
30 the performance of different type of methods in terms of accuracy and generalization ability. The different approaches  
31 of these algorithms provide more accurate models. However, the techniques do not allow feature selection and the  
32 final expressions result complex and unintelligible, often not even accessible. The structure of the Genetic  
33 Programming algorithm ensured instead a selection of the possibly most informative variables, as well as readable  
34 final expressions. The approach entailed constructive and helpful considerations to the addressed task, confirming its  
35 key-role in zootechnical field, especially in the beef breeding management.

36 **Keywords:** Precision Livestock Farming; Evolutionary Algorithms; Machine Learning; Cattle Breeding; Piedmontese Bovines.

37

## 38 1. Introduction

39

40 The management of livestock by continuous automated real-time monitoring of production, reproduction, health  
41 and welfare of the herd, and its environmental impact is defined as Precision Livestock Farming (PLF) (Berckmans, 2017;  
42 Berckmans and Guarino, 2017). PLF supplements the skills of the farmer, the veterinarian, and the technician by a  
43 continuous collection of livestock information, with the support of information technologies. It can play a crucial role in  
44 the early detection of diseases and it objectively assesses animal condition and welfare in modern livestock production,  
45 representing a tool that supports many farmers as decision-makers. Despite the biological process is too complex to  
46 replace farmers by technology, it still offers more possibilities to save money and to change farmers' lives, as a more  
47 accurate management system can be achieved, leading to better approach of the genetic potential of today's livestock  
48 species. The breeder must generally deal with animals' problems, like their health conditions and social behavior, that  
49 affect the quality of the product, the life of the animal, and the performance of the farm. Indeed, PLF is an emerging  
50 field in which the main aspect involves the development of proper tools and novel technologies, suitable to monitor  
51 each animal and the whole breeding. The resulting increased knowledge, elaborated through mathematical models,  
52 may provide the offset of overall incurred costs of the farm, as these issues are identified in advance, allowing decisions  
53 to be made in time.

54 In a recent study conducted by CREA of Lodi, a Research Centre for animal production and aquaculture, a survey  
55 was addressed to cattle farmers in one of the most intense dairy breeding provinces in Italy, to investigate the diffusion

---

56 of precision farming tools (Abeni et al., 2019). Most breeders reported the use of electronic equipment, and the owners  
57 of larger farms showed a greater propensity for PLF technology, stating that, considering costs and benefits, the biggest  
58 problem in purchasing monitoring systems is the time required to manage the generated data. The major consequence  
59 of continuous monitoring of animals is a huge amount of data, the so-called "Big Data" (Cole et al., 2012; Lokhorst et  
60 al., 2019). Given the numerosity and complexity of these data, the databases cannot be visually inspected. If on one side  
61 PLF approach aims for a greater "accuracy" on the quantity and quality of information, entailing the development of  
62 monitoring systems, on the other side it must deal with the transformation of big data into meaningful information. For  
63 this reason, the use of Machine Learning (ML) techniques is becoming increasingly common. ML is a subfield of artificial  
64 intelligence, addressed to the study of algorithms for prediction and inference. Learning from data is at the core of ML,  
65 and hence this field of research is suitable for the management of large data sets and used to predict livestock issues,  
66 such as time of disease events, risk factors for health conditions, failure to complete a production cycle, as well as the  
67 genome of complex traits (González-Recio et al., 2014; Morota et al., 2018).

68 Many are the research studies in the cattle sector based on the application of ML techniques in farm management,  
69 to model for example the individual intake of cow feed, to optimize health and fertility, and to identify potential disease  
70 predictors for several pathologies, such as Bovine Viral Diarrhoea Virus (BVDV), Infectious Bovine Rhinotracheitis (IBR),  
71 Bovine Tuberculosis (TB), lameness, and mastitis (Bovine Diseases and Resources; Yao et al., 2016; Williams et al., 2016;  
72 Amrine et al., 2014; Guzhva et al., 2016; Machado et al., 2015; Nasirahmadi et al., 2017; Ortiz-Pelaez and Pfeiffer, 2008;  
73 Rodero et al., 2012). Beef cattle management is usually an extensive breeding system: the bovines are bred completely  
74 on pastures and they are not subject to massive checks such as dairy cattle, since they are more resistant and less  
75 exposed to stress factors (Derner et al., 2017). Therefore, a precision farming approach is rarely used in these cases.  
76 However, nowadays the number of new intensive farms is also increasing also in the meat sector. The Piemontese,  
77 mostly concentrated in the Italian region of Piedmont, belongs to this category, mainly because available pastures are  
78 not sufficient for the total number of animals (Bona et al., 2005; Savoia et al., 2019). Consequently, in order to optimize  
79 the management of this type of breeding, it is necessary to constantly monitor the animals, introducing and adapting  
80 to beef cattle the necessary tools implemented for the dairy sector. The latter already offers a wide range of devices,  
81 because dairy bovines generally have a shorter average life compared to the lifespan of beef bovines. They are often  
82 affected by diseases and metabolic problems and are crossed often with beef cattle for better performance and higher  
83 yield (Rutten et al., 2013; Hesse et al., 2019). The breeding cycle is reduced compared to other income-producing  
84 species, and there is no daily movement of the animals (e.g. milking). The aspects of greatest interest are the

---

85 composition of the ration and the consumption of food and water, behavioral remarks, the quality of the structures that  
86 host the cattle (temperatures, humidity, lighting), growth, slaughter yield, and carcass quality. The unit of measurement  
87 is not individual in the case of beef cattle, but each animal affects the performance of the farm after all. The lower  
88 impact of critical points in the meat sector entails that the adoption of sensors, not yet specific for this type of animal  
89 and with a high cost, is probably not worth the economic investment. The lower impact of critical points in the meat  
90 sector entails that the adoption of sensors, not yet specific for this type of animal and with a high cost, is probably not  
91 worth the economic investment.

92       Among the major beef cattle breeds, Italian 'Piemontese' represents a characteristic element of the territory of  
93 Piedmont, a region in Northwestern Italy. Organoleptic and zootechnical remarkable qualities result above all in a  
94 greater tenderness of the meat and exceptional character skills, such as meekness, maternal attitude, resistance to  
95 diseases, little stress, and great adaptation to pasture. It, therefore, allows easy management and the development of  
96 the local area (ANABORAPI ; Bona et al., 2005). The Piemontese cattle derives its name from this region, its cradle of  
97 origin, even if today it is spreading in several foreign countries. The bovines are bred in beef intensive farms, which are  
98 therefore provided with the installation of stables to control the animals, a grazing for feeding purposes, the addition  
99 of different artificial fodder on feed and curative intents, and particular attention to the reproduction of the livestock.

100       Since the National Association of Piemontese Cattle Breeders (ANABORAPI) is responsible for promoting the  
101 breed through the study of all processes of the Piemontese breeding ("Lo svezzamento del vitello Piemontese", 2012;  
102 "Relazione tecnica", 2018), information is stored in a complex database, i.e. the Herd Book of the Race. Besides animals'  
103 pedigrees, morphological details, and genetic values, a wide section of statistics among the economic efficiency of the  
104 farm are available. In particular, a constant monitoring of the average situation of breedings due to the main fertility  
105 parameters are provided, summarized by the average number of calves per cow produced in the last year. This is then  
106 translated into a brief economic summary, which compares the gross revenue with the mortality losses, providing the  
107 farmer with an indicator of breeding performance. However, this index does not include the effects of the weaning  
108 period. The physiological development process of the animal reaches completion in 60 days after birth. Calf mortality is  
109 an important cause of economic damages in Piemontese cattle farms: it represents for the farmer the loss of the  
110 economic value of the calf, and the reduction of both the herd's genetic potential and the size of the breeding. It is  
111 straightforward that the gestational phase alone is not exhaustive: it is crucial to consider neonatal mortality, outlining  
112 the calf's ability to survive, and the source of stress such as congenital calf's defects (i.e. arthrogryposis and  
113 macroglossia), compromising eventually the immune response and the growth rate, environmental and food conditions,

---

114 that affect the quality of life of the newborn (Prince et al., 2003; Tao et al., 2018). Therefore, these zoological influential  
115 variables must be identified among the numerous parameters within the complex database. Given the size of the  
116 databases, it is extremely difficult to recognize many of the substantial factors, and to be able to hypothesize a  
117 prediction model for the number of calves weaned per cow per year, that is a more accurate measure for the yield of  
118 the farm.

119 In order to investigate the production of Piemontese calves and its modelling, this study aims to examine which  
120 variables available in the dataset influence the performance of a breeding. In contrast with previous studies conducted  
121 by ANABORAPI, in which models are based on traditional statistical identification approaches, a priori assumptions  
122 about data or the relationship between the response and independent variables are avoided.

123 In ML approaches, the choice of which techniques lend best to the problem depends mostly on the objective to be  
124 achieved. There are many methods that can produce excellent results, by building accurate prediction models. However,  
125 there are different characteristics, intrinsic to the techniques, capable of better address the question that arisen, and  
126 which tend to be privileged in the studies. Therefore, since the task and the data are full of zoological meaningful  
127 features, a Genetic Programming (GP) approach is adopted as a baseline, as models are resumed in simple and  
128 interpretable expressions (Abraham et al., 2006; Koza et al., 1994). Thanks to its structure, the algorithm, belonging to  
129 the white-box methods category, can automatically create models “learning” from the data, process accurate results,  
130 and even extract critical information. Among other ML techniques, some common methods are selected, including  
131 black-box ones as Neural Networks and Random Forest to compare the results obtained with GP. Black -box model are  
132 generally outperforming, since their structure is able to capture the high non-linearity lying on data. However, as their  
133 definition suggests, they can be very unclear and do not explain the links between input and output variables, as well  
134 as the internal mechanisms leading to the results (O. Loyola-González, 2019).

135  
136 The article is organized as follows: Section 2 is dedicated to the description of the background: the current applied  
137 method, modeling the breeding performance, is presented and the aim of the study is highlighted. The dataset is  
138 analyzed and the basic steps to prepare the benchmark are also described. Emphasis is placed on the division of the  
139 dataset into different partitions, illustrating the need for the techniques to learn on a portion of the dataset and to test  
140 the prediction models on new instances. Afterwards, the GP baseline and other ML methods are enlightened. Results  
141 are examined and discussed in Section 3, with regard to the readability of the expressions. The performance of GP is

142 illustrated, and the comparison with other techniques discussed. The research article concludes with some  
143 considerations and further developments, highlighted in Section 4.

144

## 145 **2. Materials and Methods**

### 146 *2.1 The reference model*

147 The model used to monitor the farm performance estimates the number of live calves produced per cow per year.  
148 It is a classic statistical model, formulated among zootechnical hypotheses, and it incorporates two variables extracted  
149 from the information of the single farm: the average calving interval (*intp*), that is the time span measured in days  
150 between a birth and the previous one, and the average calves mortality at birth:

$$151 \quad Y_a = \frac{365}{intp} \left( 1 - \frac{m}{100} \right). \quad (1)$$

152 It concerns the birth, taking into account the average values of the previous 12 months, without taking into account  
153 important factors related to the weaning of the calf. The effects of the two months following this event are significant,  
154 since in the immediate period following the birth, i.e. 60 days, the physiological development of the calf reaches  
155 completion. Multiple factors can affect the growth, such as its own ability to adapt, the genetic factors, congenital calf's  
156 defects, such as arthrogryposis or macroglossia (Lynch et al., 2019). Moreover, the environmental and food conditions  
157 can contribute positively or negatively to the quality of life of the newborn, becoming a remarkable source of stress,  
158 compromising the growth rate and the well-being of the animals. Among the farms considered in the study, (description  
159 in Section 2.2), we compared the reported number of calves that died at birth and the sixtieth day after. As  
160 straightforward from Figure 1, during birth almost all the farms did not report any deaths, while at the end of weaning  
161 the number of farms with zero deaths drops drastically. In fact, it is noted that the breeders reported a high number of  
162 dead calves at 60 days. For the farmer, the loss of the calf means the loss of economic value. The high mortality rate  
163 reduces then the number of young animals to be used to increase the farm size and the genetic potential of the herd.  
164 This makes it necessary to formulate a model that can predict the number of calves weaned per cow per year, shaped  
165 on data. Like equation (1), it should incorporate the influential variables affecting the output, and at the same time  
166 provide a not linear expression of simple interpretation, in order to be able to understand and explain zoologically the  
167 link with the output afterwards. On this porpouse, we approach the problem with GP, since it can produce accurate and

---

168 explainable models, uncovering relevant predictors, with similar performance to other well-established common  
169 methods, such as linear regression, k-nearest neighbour, neural networks, and random forest.

170

## 171 *2.2 Datasets and ML framework*

172 The available database provided by ANABORAPI is the event history for all farms registered in the Herd Book of  
173 the Race. For every farm the average values of the measurements, recorded by technicians during routine controls,  
174 veterinarians, and directly by farmers, are stored. Data are registered with a devoted device available to technicians (a  
175 mobile computer Workabout) or directly with the smartphone and the personal computer. In fact, ANABORAPI designed  
176 a web service, accessible to registered users, which provides the situation of the farm. The data entered, both from the  
177 PC and from the other devices, are sent in real time to the servers, stored and processed, to return the updated situation  
178 at last. There are several records for each farm, since it keeps track of every visit. The content of the database is  
179 processed by the system on the date of elaboration going back 365 days, starting from the last check. Statistics are  
180 finally provided, resumed in a web page that the farmers can consult on their own. In addition to ID data of the breeding  
181 farms, all information on the consistencies, information on the parturitions and births, the type of inseminations carried  
182 out (natural or artificial), controls' dates, Estimated Breeding Values (EBV), consanguinity of all registered bovines, and  
183 perinatal mortality rates are kept. Since the database contains data of the last twenty years of all farms, including those  
184 that are no longer active, we have processed it in order to create a representative data set on which to apply the ML  
185 techniques.

186 In ML, it is necessary to define a set of data intended for "learning" and a set on which to "test" the obtained  
187 models. In other words, the learning set is a list of records (instances) on which the technique will build a knowledge  
188 base. The algorithm analyses the data to find links between a series of input variables (i.e. consistencies, CI, mortality,  
189 EBVs, consanguinity, etc.) and a specific output variable (the target variable, i.e. the number of calves weaned per cow  
190 per year). This type of task is called supervised learning since the attribute to be predicted is known (target attribute).  
191 The test set must verify the effectiveness of the model just constructed, subjecting it to new situations, by checking the  
192 validity of its answers. Thus, the learning set is required to build the model, while the test set to measure its  
193 effectiveness. It is therefore strictly necessary for the two sets to be disjoint. For an optimal management, besides the  
194 contemporary situation of the farm, it is relevant for the breeder to know the prediction of the future trend. Therefore



195 the core of the issue is concentrated in the prediction of the number of viable calves after the weaning phase, which  
196 each cow will produce over the next 365 days. The conditions that allow the survival of the calf are partly related to the  
197 calf itself and its temperament. However, in large part they are due to its genetic characters and those of its ancestors,  
198 and even to the choices that the breeder adopts for the management of the animals and the environment. The variable  
199 which serves to supervise the construction of the model is the target value for the following year. Considering for each  
200 farm the corresponding values in 2018 for the number of the calves born alive and those unable to survive during  
201 weaning period and the number of cows, (i.e.  $N_{BALIVE}$ ,  $N_{ELIM}$ ,  $COWS$ ), the target attribute  $Y$  was obtained with the  
202 following:

$$203 \quad Y = \frac{N_{BALIVE} - N_{ELIM}}{COWS}.$$

204 In a preliminary study, we investigated the dataset and performed a genetic programming approach, in order to  
205 explore the possibility to address this task with GP (Abbona et al., 2020). The method performed well, entailing that ML  
206 horizon should be investigated further and that comparisons with other techniques should be carried out, even on larger  
207 datasets containing more features. In the previous experiment we extracted and processed 19 variables and we kept  
208 stricter filters on data: to perform GP method, we selected the farms, based on the date of checks recorded between  
209 2017 and 2018. In this study, only the farms that show constant visits between 2014 and 2019 were considered. In this  
210 way, the effects related to farm management are consolidated and just the breedings with substantial data were kept.  
211 Indeed, even if the investigation is based on farms with data from 2017 as input and from 2018 as target, as a change in  
212 the type of management stabilizes over time, we considered breedings with historical records updated between 2014-  
213 2019, in order to focus on farms with a solid management. A newly started company does not have completely  
214 representative data. Moreover, the summary produced by anaborapi elaborates the average values among recordings  
215 related to the previous 65 days from the reference date. To avoid data from farms not yet fully operational, with gaps in  
216 registrations or close to resigning at the end of 2018, we set the restriction to companies active in the previous 5 years.

217 As in our pilot study, filters were imposed on breedings located in Piedmont with at least 30 cows and a percentage  
218 of artificial insemination between 90% and 100%. For each breeding, data recorded in 2017 and 2018 were considered  
219 and the record for the last check in the corresponding year was considered. Subsequently, in correspondence with each  
220 farm, input and target variables were extracted, respectively from 2017 and 2018. However, two further conditions  
221 were added for this research study. As already stated, in order to keep in the pool of currently active breedings those  
222 with stable and consolidated situations, inspections had to be constantly carried out for at least more than two years.

223 Hence, as already explained, only farms exhibiting constant data recordings during the time interval 2014-2019 were  
 224 considered. Moreover, the fact that breeders carry out between 90% and 100% of artificial insemination, means that a  
 225 part of the considered farms own bulls and carry out also natural impregnations. Most of the time, instead of recording  
 226 the date on which the insemination took place, a period of several days followed by the diagnosis of the pregnancy is  
 227 set. These farms were therefore excluded from the analysis. A main group of 304 representative Piedmontese cattle  
 228 farms results from the selection. Since the performance of the farm mainly focuses on fertility, data concerning  
 229 multiparae cows were considered to elaborate the number of deliveries and the calving intervals. In the same way, data  
 230 on bulls used for artificial insemination were maintained (i.e. EBVs, that represent namely estimations of the additive  
 231 genetic effect of a subject). Parameters on heifers were included in the dataset, since these are bovines that did not  
 232 give birth but, in breeding farms, are mostly intended to the production of calves. Moreover, since many breeders carry  
 233 out also natural impregnation besides artificial insemination, related to the bulls used for natural insemination were  
 234 added to the analysis, as well as the levels of consanguinity of calves that will be born from ongoing pregnancies. The  
 235 only strictly environmental measure available in the dataset, that was hence kept, is the Livestock Unit (LU or LSU): it  
 236 has the purpose of synthetically expressing the zootechnical load, to easily compare the environmental impact of  
 237 different farms. Based on the age of the animals, appropriate coefficients are applied to the number of animals for each  
 238 age category in the breeding: cattle over 2 years old (1 \* LSU), cattle aged between 6 months and 2 years (0.6 \* LSU)  
 239 and cattle less than 4 months old (0.4 \* LSU) (Sistema Piemonte-UBA).

240 The final dataset counts 304 records, each one standing for a single farm, and a total of 48 input attributes  
 241 (referring to year 2017) and one target variable, that is the actual number for weaned calves recorded in 2018. All  
 242 variables represent positive quantities and are described in Table 1.

243

244 **Table 1.** Final attributes used in the studied dataset. The last line (variable Y) represents the dependent variable, target  
 245 attribute for the predictive models generated by ML techniques.

	<b>Attribute</b>	<b>Description</b>
1	<b>BOVINES</b>	Consistency of all animals in the farms, i.e. size of the cattle.
2	<b>COWS</b>	Consistency for cows, i.e. number of cows.
3	<b>HEIFERS</b>	Consistency for heifers, i.e. number of heifers.
4	<b>F_CALVES</b>	Consistency for female calves, i.e. number of female calves.
5	<b>BULLS</b>	Consistency for bulls, i.e. number of bulls.

---

6	<b>M_CALVES</b>	Consistency for male calves, i.e. number of male calves.
7	<b>PERCENT_FA</b>	Percentage of Artificial Insemination.
8	<b>C_AGE</b>	Mean value of age of cows expressed in days.
9	<b>C_PAR</b>	Mean number of parturitions per cow.
10	<b>N_PAR</b>	Number of occurred deliveries.
11	<b>SALXGRAV</b>	Mean number of necessary inseminations, which resulted in positive pregnancy diagnosis.
12	<b>N_CORRECT</b>	Percentage of calves born without birth defects, such as Macroglossia or Arthrogyphosys.
13	<b>H_EASE</b>	Number of easy parturitions for primiparae, that did not require human intervention and that did not cause stress to the cow nor the calf.
14	<b>H_DIFFICULT</b>	Number of difficult parturitions for primiparae, that partly required human intervention.
15	<b>H_CESAREAN</b>	Number of parturitions that required caesarean section for primiparae.
16	<b>C_EASE</b>	Number of easy parturitions for multiparae, that did not require human intervention and that did not cause stress to the cow nor the calf.
17	<b>C_DIFFICULT</b>	Number of difficult parturitions for multiparae, that partly required human intervention.
18	<b>C_CESAREAN</b>	Number of parturitions that required caesarean section for multiparae.
19	<b>C_N_IND</b>	Number of cows that gave birth in the last year.
20	<b>C_PARTIND</b>	EBV referred to ease of parturition of the cows.
21	<b>H_PARTIND</b>	EBV referred to ease of parturition of the heifer.
22	<b>N_TFA</b>	Number of AI bulls whose semen has been used during the last year.
23	<b>TFA_BIRTH</b>	Mean value of EBV referred to ease of birth of the bulls, which semen has been used on artificial inseminations
24	<b>TFA_PAR</b>	Mean value of EBV referred to ease of parturition with which the bulls, which semen has been used on artificial inseminations, were born.
25	<b>N_TFN</b>	Number of NI bulls used during the last year.
26	<b>TFN_BIRTH</b>	Mean value of EBV referred to ease of birth of the bulls, which is used for natural impregnation
27	<b>TFN_PAR</b>	Mean value of EBV referred to ease of parturition with which the bulls, which is used for natural impregnation, were born.
28	<b>C_GRAVID</b>	Number of pregnant cows at the end of 2017.
29	<b>C_INS</b>	Number of inseminated cows at the end of 2017.
30	<b>C_POSTPARTUM</b>	Number of post-partum cows at the end of 2017.
31	<b>C_EMPTY</b>	Number of cows in dry period.
32	<b>LSU</b>	Total Livestock Unit.
33	<b>LSU1</b>	Livestock Unity for bovines older than one year.
34	<b>LSU06</b>	Livestock Unity for bovines between 6 months and 1 year old.
35	<b>LSU04</b>	Livestock Unity for bovines between 4 and 6 months old.

---

---

36	<b>INTP</b>	Mean calving interval, i.e. the average number of days that elapse between a parturition and the following one.
37	<b>CONSANG_NEW</b>	Level of consanguinity of calves that will be born from ongoing pregnancies.
38	<b>N_CONSANG_NEW</b>	Number of ongoing pregnancies.
39	<b>BIRTHW_M</b>	Mean birth weight of male calves.
40	<b>BIRTHW_M</b>	Mean birth weight of female calves.
41	<b>MORT</b>	Mean value of neonatal mortality.
42	<b>ABORT</b>	Percentage of abortions.
43	<b>N_ABORT</b>	Number of abortions.
44	<b>N_ELIM</b>	Number of calves dead within 60 days after their birth.
45	<b>N_TOT</b>	Total number of newborns.
46	<b>N_BALIVE</b>	Total number of calves born alive.
47	<b>BORN_FA</b>	Total number of newborns with artificial insemination.
48	<b>BORN_FN</b>	Total number of newborns with natural impregnation.
49	<b>Y</b>	Number of expected weaned calves per cow per year.

---

246

247 *2.3. Application of ML techniques: basic steps*

248

249 As previously described, one of the basic steps for applying ML techniques is the subdivision of the dataset into  
250 two disjoint parts: the learning set and the test set. Therefore, the main dataset is split, the chosen algorithm builds a  
251 model, learning the hidden relations between the data on the "Learning set", and its generalization performances are  
252 finally evaluated on the "Test set". In order to compare the performance of different approaches, it is necessary to  
253 analyse the median behaviour of the models obtained. Therefore, we split 30 times the dataset in order to obtain 30  
254 different sets, each one with constant learning-test partitioning (75%-25%) and randomly selected instances. In this  
255 way, each experimental phase will determine one solution on the corresponding dataset, for a total of 30 prediction  
256 models for each technique. Regarding GP, the "Learning set" was further split, exactly in a similar way to the division  
257 into learning and test set. Each of the 30 learning sets was randomly divided, with a constant partitioning (75% -25%),  
258 into a Training Set and a Validation Set. The choice of this methodology, i.e. a 75%-25% split repeated for both partitions,  
259 is due to the size of the dataset. The division between learning and test entailed a learning set of size equal to 228  
260 instances. We initially considered partitioning the learning set in training-validation through a  $k$ -fold cross validation,  
261 but the reduced size did not allow us to find a suitable value of  $k$ : for example,  $k$  smaller than 10 leads to a restrained

262 number of training-validation subsets. On the contrary, a value of  $k$  greater than 10, led to a restrained number of  
263 records within the validation sets (i.e. less than 21 farms). Using a 30-fold would imply a validation of size 7, not  
264 representative at all. For this reason, we repeated further the subdivision 75%-25% to obtain disjoint training and  
265 validation sets, paying attention to avoid too much overlap between the 30 subsequent partitions. Finally, the sizes for  
266 learning and test were respectively 228 and 76, whereas 171 and 57 for training and validation.

267 Very often, during the construction and development of a model, two opposite problems may appear: underfitting  
268 or overfitting (Domingos, 2012; Bhattacharya, 2013). The first arises when the created model is too simple and fails to  
269 generalize neither on the learning set nor on the test set. In the opposite case, the model adapts extremely well to the  
270 learning set, but fails to generalize, which leads to small errors on the learning set and very large on the test. The noise,  
271 naturally present in the dataset, can result in a problem: the model learns the noise instead of the true hidden  
272 relationship among the variables. Setting correctly the regularization hyperparameters with which the technique learns  
273 and adding complexity to the model are crucial steps. Once the parameters are set and the model is obtained, the error  
274 is evaluated. When dealing with ML, the error is measured with a fitness function, that is an objective function that is  
275 used to evaluate how close a given solution is to achieving the experimental aims. In this case, since we are dealing with  
276 a regression problem, i.e. we want to calculate the continuous value of a dependent variable starting from the  
277 independent variables, we chose the Root Mean Square Error (RMSE) as fitness function:

278

279

$$RMSE = \sqrt{\sum_i \frac{(y_i - \phi(x_i))^2}{n}},$$

280

281 where  $i=1, \dots, n$  and  $n$  is the number of instances (depending whether it is calculated on the learning or the test set).

282 The predictor  $\phi$  is evaluated at  $x_i$  (values of input variables on 2017) and  $y_i$  are the target values (on 2018). A good  
283 fitness value means a small RMSE and viceversa. Moreover, RMSE is expressed in the response variable's unit and it is  
284 an absolute measure of accuracy. The choice of this fitness function was also determined by the comparison of  
285 different ML techniques, that build non-linear models. This issue excludes a discussion based on the coefficient of  
286 determination  $R^2$ , as its definition assumes linearly distributed data. When the assumption is violated,  $R^2$  can assume  
287 misleading values (Spiess and Neumeier, 2010).

288 Finally, GPLab package built in Matlab ("The package 'caret'"; Silva, 2007) was performed, and comparisons with  
289 other techniques were applied to the benchmark problem, with R software library caret (Table 2).

290

291

**Table 2.** ML techniques adopted, the corresponding belonging area and the respective used package.

<b>Method</b>	<b>Description</b>	<b>Package</b>
'GP'	Genetic Programming based algoritm (GP)	GPLab library built in Matlab
'knn'	k-Nearest Neighbour algorithm (kNN)	R software library caret
'nnet'	Neural Network algorithm (NN)	R software library caret
'lm'	Linear regression algorithm (LM)	R software library caret
'ranger'	Random Forest Tree-based algorithm (RF)	R software library caret

292

### 293 2.3 . Application of ML techniques: Genetic Programing

294

295 The algorithm based on GP creates a population of models, whose number is set by the user in the parameters'  
 296 settings (Silva, 2007; Poli et al., 2008). It is a tree-based algorithm, which, exactly as in an evolutionary process, with the  
 297 passing of generations lets the initial population evolve, through mechanisms of selection, mutation, and recombination  
 298 of individuals (i.e. mutation and crossover). GP evolves individuals, represented as tree structures, that can be  
 299 recursively evaluated. The tree nodes are operator functions and every terminal node is an operand. By selecting,  
 300 recombining, and mutating the best individuals at each generation, at each evolutionary step (i.e. new generation), the  
 301 members of the new population are on average fitter than previously generated individuals, i.e. show a smaller error.

302 The user sets the size of the population, initially constructed randomly by the algorithm, to find at the last  
 303 generation (parameter determined also by the user) a population of the same size but with evolved individuals. In our  
 304 implementation, initial population was generated with the Ramped half and half method: half the initial population is  
 305 constructed using the *full* method (generates trees where all the leaves, i.e. the variables, are at the same depth) and  
 306 half is constructed using *grow* method (creation of trees of different sizes and shapes). Among other parameters, it is  
 307 possible to set the conservation of the best individual at each run (Elitism) and the selection method: we decided to set  
 308 the lexicographic parsimony pressure, since this parameter optimize both fitness and tree size, as fitness is treated as  
 309 the primary objective and tree size as a secondary objective in a lexicographic ordering. This peculiarity leads to the  
 310 conservation of the most influential variables over generations. The algorithm performs, hence, an implicit feature  
 311 selection and among all the input variables, only the most relevant are encapsulated in the solutions.

312 At the end of the evolution process the population size consists of 500 members (population size), whereas a single  
 313 model should be extracted at the end of the run on the learning set. It is necessary to evaluate the 500 individuals,  
 314 obtained on the learning set, on the validation dataset: GP models evolve on the training set and finally the best ones  
 315 are selected among all the evaluations on the validation set. These models are hence evaluated on the test set, in order  
 316 to measure their generalization ability on unseen data. Parameters are summarized in Table 3.

317

318

**Table 3.** Parameters used to perform GP.

Parameter	Description
Maximum number of generations	40
Population size	500
Selection Method	Lexicographic Parsimony Pressure
Elitism	Keepbest
Initialization Method	Ramped half and half
Tournament Size	2
Subtree Crossover Rate	0.8
Subtree Mutation Rate	0.1
Subtree Shrinkmutation Rate	0.05
Subtree Swapmutation Rate	0.05

319

#### 320 *2.4 Application of ML techniques: Linear Model, k-Nearest Neighbour, Neural Network and Random Forest*

321

322 We compared GP performance with other classical black-box ML approaches used for regression tasks (Hastie et  
 323 al., 2009). Differently from GP, these methods do not carry out an automatic feature selection. By the end of the learning  
 324 process for each run, the final population already consists of one model. This model was built on the examples given in  
 325 the learning set and it only needs to be evaluated on the test set, that is a validation set is not necessary. The hyper-  
 326 parameters were obtained with a tuning process. A tune grid was manually specified for each algorithm, that is the  
 327 parameters to test were explained in a vector and the best results were then selected. Corresponding main parameters  
 328 for all ML approaches are listed in Table 4. The unmentioned parameters were kept as default values, since during tuning  
 329 no tangible improvements in terms of fitness were achieved.

330

331 The k-Nearest Neighbors algorithm (kNN) is an instance-based method, in which the input consists of the k closest  
 332 instances (neighbors) in the feature space, and the output is the mean of the output values of k nearest neighbors. To  
 333 predict the value of any new data point, the distance between the point and the k nearest ones are selected and the

---

334 mean value of their output is assigned as prediction. A small value of  $k$  leads to results highly influenced by noise and a  
335 large value would be computationally expensive. We configured  $k$  equal to 15, that is the square root of the size of the  
336 learning set (Lantz, 2015). Values greater larger than 15, generated overfitting models.

337

338 Neural Networks (NN) emulate the complex functions of the brain. A NN is a simplified model of the structure of  
339 the biological neural network and consists of interconnected processing units organized in a specific topology. A set of  
340 nodes is arranged in at least three layers, including an input layer, where the data enter the system, one or more hidden  
341 layers, where the learning takes place, and an output layer, where the prediction is given. Learning occurs by changing  
342 connections weights, based on the error affecting the output. At each update, the weights of the connection between  
343 nodes are multiplied by a factor in order to prevent the weights from growing too large and the model from getting too  
344 complex. In this experimental study, we set a size of 15 hidden units that fit a single hidden layer.

345

346 Linear regression (LM) is a linear model, that basically assumes a linear relationship between the input and the  
347 single output variables. More specifically, the representation is a linear equation that combines a specific set of input  
348 values, whose solution is the expected output. As such, both the input values and the output are numeric. Learning a  
349 linear regression model means estimating, with the available data, the values of the coefficients used in the  
350 representation.

351

352 Random forest (RF) is an ensemble learning method, which operates by constructing a multitude of decision trees  
353 (i.e. a forest) during learning phase and returning the mean prediction of the individual trees. Basically, a bootstrapping  
354 is first performed by the algorithm and a tree learns from a random sample of the training observations. The samples  
355 are drawn with replacement, which means that some samples will be used multiple times in a single tree. When dealing  
356 with a large number of features, it is hence common to reach greater bias. The choice of the optimal cut-point is, indeed,  
357 responsible for a large proportion of the variance of the induced tree. Instead of learning on bootstrap copies, it is  
358 possible to grow the trees by splitting nodes at fully randomly chosen cut-points. Extra-Tree parameter (extremely  
359 randomized trees) drops the attempt to find an optimal cut-point for each chosen variable at each node. The algorithm  
360 uses then the whole learning sample, instead of a bootstrapped dataset, and all the variables are selected at each split.  
361 Consequently, we set extraTrees as split rule and the number of variables available at each split was configured equal  
362 to the total number of features in the dataset.



363

364

**Table 4.** Parameters used to perform ML techniques with caret package in R.

ML technique	Parameters
kNN	k = 15
Nnet	size = 15; decay = 0.2
Lm	Intercept = TRUE
ranger	mtry = 48; splitrule = extratrees; min.node.size = 5

365

366 **3. Results and Discussion**

367

368 *3.1. Interpretability of GP models.*

369

370 The section is dedicated to the discussion among the models obtained with GP. The expressions obtained with this  
 371 approach are readable and interpretable: this is a crucial characteristic, since we aim to put the breeder in the conditions  
 372 of understanding the meaning of the model, to understand which aspects (i.e. the variables) of management are more  
 373 relevant in the definition and the measure of the performance of the farm. It is possible to make some interesting  
 374 considerations, in order to interpret the achieved models. First of all, it is possible to analyze the frequency with which  
 375 the variables are used by the 30 best models, that is those that show the best fitness (in other words, the lowest error)  
 376 on the validation set and that have been evaluated on the test set.

377 In order to highlight key role variables, it is useful to investigate the median frequencies among the best solutions  
 378 on all the runs. Parameters showing non-null median values were used in over 50% of the models, whereas negligible  
 379 features respectively correspond to null median values. Results are reported in Table 5.

380 Namely, the most frequent variable is the number of calves born from natural inseminations (BORN\_FN), followed  
 381 by the number of cows (COWS), the total number of born calves (N\_TOT) and the number of calves dead in the first 60  
 382 days after birth (N\_ELIM). In exactly half of the models the EBV referred to ease of parturition of the heifers was used  
 383 (H\_PARTIND). It is straightforward possible to infer that GP models detected the majority of information in the  
 384 aforementioned features. On the contrary, it should be noted that none of the final prediction models encapsulated th

385 number of parturitions that required caesarean section for multiparae (C\_CESAREAN) and the mean value of EBV  
 386 referred to ease of birth of the bulls, which semen has been used on artificial inseminations (TFA\_BIRTH).

387 The emphasis placed by GP among the listed features entails that the prediction of yearly weaned calves per cow  
 388 for 2018 depends above all on the quantity of natural inseminations in the farm is accomplished. It is also proportional  
 389 to the total number of newborns and calves not weaned during 2017. The result suggests that these variables could be  
 390 the main features involved in this kind of task, that is the prediction of weaned calves per cows per year. It does not  
 391 imply, however, that the other parameters are not important in the management of the farm.

392

393 **Table 5.** Median frequencies and percentage of use of each variable among the best 30 individuals found by GP.

<i>Variable</i>	<i>Median</i>	<i>% of use on 30 runs</i>	<i>Variable</i>	<i>Median</i>	<i>% of use on 30 runs</i>
X <sub>1</sub> BOVINES	0	27%	X <sub>25</sub> N_TFN	0	17%
X <sub>2</sub> COWS	1	57%	X <sub>26</sub> TFN_BIRTH	0	13%
X <sub>3</sub> HEIFERS	0	7%	X <sub>27</sub> TFN_PAR	0	20%
X <sub>4</sub> F_CALVES	0	3%	X <sub>28</sub> C_GRAVID	0	3%
X <sub>5</sub> BULLS	0	17%	X <sub>29</sub> C_INS	0	10%
X <sub>6</sub> M_CALVES	0	13%	X <sub>30</sub> C_POSTPARTUM	0	20%
X <sub>7</sub> PERCENT_FA	0	23%	X <sub>31</sub> C_EMPTY	0	17%
X <sub>8</sub> C_AGE	0	10%	X <sub>32</sub> LSU	0	7%
X <sub>9</sub> C_PAR	0	7%	X <sub>33</sub> LSU1	0	20%
X <sub>10</sub> N_PAR	0	43%	X <sub>34</sub> LSU06	0	7%
X <sub>11</sub> SALXGRAV	0	13%	X <sub>35</sub> LSU04	0	23%
X <sub>12</sub> N_CORRECT	0	33%	X <sub>36</sub> INTP	0	13%
X <sub>13</sub> H_EASE	0	10%	X <sub>37</sub> CONSANG_NEW	0	27%
X <sub>14</sub> H_DIFFICULT	0	7%	X <sub>38</sub> N_CONSANG_NEW	0	17%
X <sub>15</sub> H_CESAREAN	0	7%	X <sub>39</sub> BIRTHW_M	0	7%
X <sub>16</sub> C_EASE	0	33%	X <sub>40</sub> BIRTHW_M	0	27%
X <sub>17</sub> C_DIFFICULT	0	7%	X <sub>41</sub> MORT	0	17%
X <sub>18</sub> C_CESAREAN	0	0%	X <sub>42</sub> ABORT	0	7%
X <sub>19</sub> C_N_IND	0	40%	X <sub>43</sub> N_ABORT	0	10%
X <sub>20</sub> C_PARTIND	0	40%	X <sub>44</sub> N_ELIM	1	57%
X <sub>21</sub> H_PARTIND	0,5	50%	X <sub>45</sub> N_TOT	1	57%
X <sub>22</sub> N_TFA	0	30%	X <sub>46</sub> N_BALIVE	0	20%
X <sub>23</sub> TFA_BIRTH	0	0%	X <sub>47</sub> BORN_FA	0	17%

$X_{24}$	TFA_PAR	0	17%	$X_{48}$	BORN_FN	1	60%
----------	---------	---	-----	----------	---------	---	-----

394

395 We thereafter investigated the interpretability of the expressions, considering the fitness obtained in each of the  
 396 best final models, and taking into account also the number of variables involved in the formula. Considering the results  
 397 reported in Table 6, we can deduce that the model entailing the best predictions on the test set encapsulates only three  
 398 variables (*Model 13* in Table 6). The expression in infix notation is:

399

400

$$Y = \frac{X_{10} + \frac{X_2}{X_{45}}}{X_2 + \frac{X_{45}}{X_2} + \frac{X_{10}}{X_2} + \frac{X_{45}}{X_2 + \frac{X_{10}}{X_{45} + X_{10}}}}, \quad (2)$$

401 where  $X_2$  is the number of cows (COWS),  $X_{10}$  is the total number of deliveries occurred during the year in the farm  
 402 ( $N\_PAR$ ) and  $X_{45}$  is the total number of born calves ( $N\_TOT$ ). Since these quantities are always positive summed and  
 403 divided in Equation (2), the denominators never reach null values. The *mydivide* operator is actually a division and the  
 404 model can be reformulated as

405

406

$$Y = \left( \frac{X_2 + \frac{X_{45}}{X_2}}{X_{10} + \frac{X_2}{X_{45}}} + \frac{\frac{X_{10}}{X_2}}{X_{10} + \frac{X_2}{X_{45}}} + \frac{\frac{X_{45}(X_2 + X_{45} \cdot X_{10})}{X_2(X_2 + X_{45} \cdot X_{10}) + X_{10} \cdot X_{45}}}{X_{10} + \frac{X_2}{X_{45}}} \right)^{-1}. \quad (3)$$

407 **Table 6.** Fitness on the test set, number of involved variables and corresponding percentage are reported for each model  
 408 evolved by GP in each one of the 30 performed runs.

Model	Fitness on Test	N. of variables	% of variables	Model	Fitness on Test	N. of variables	% of variables
model 1	0,1274	9	19%	model 16	0,1946	18	38%
model 2	0,1361	7	15%	model 17	0,1097	10	21%
model 3	0,1480	9	19%	model 18	0,1238	8	17%
model 4	0,0999	13	27%	model 19	0,1373	6	13%
model 5	0,1262	9	19%	model 20	0,1263	3	6%
model 6	0,1263	7	15%	model 21	0,1404	9	19%
model 7	0,1088	6	13%	model 22	0,1242	4	8%
model 8	0,1309	11	23%	model 23	0,1130	8	17%
model 9	0,1330	8	17%	model 24	0,1390	7	15%
model 10	0,1617	12	25%	model 25	0,1385	10	21%

<i>model 11</i>	0,1325	10	21%	<i>model 26</i>	0,1391	6	13%
<i>model 12</i>	0,1370	12	25%	<i>model 27</i>	0,1177	5	10%
<i>model 13</i>	0,0974	3	6%	<i>model 28</i>	0,1222	13	27%
<i>model 14</i>	0,1025	7	15%	<i>model 29</i>	0,1075	10	21%
<i>model 15</i>	0,1328	20	42%	<i>model 30</i>	0,1502	10	21%

409

410 In Equation (3) it is possible to notice that the simplification led to an expression containing a sum of three terms.

411 Whenever such result is reached, the following considerations can be developed:

412

413 a) the obtained expression is in the form of

$$414 \quad y = (x_1 + x_2 + \dots + x_n)^{-1}$$

415 where  $y$  is the result (i.e. the prediction) obtained for the values  $x_i, i=1, \dots, n$  of input variables, that is equivalent

416 to

$$417 \quad \frac{1}{y} = x_1 + x_2 + \dots + x_n.$$

418

419 b) By multiplying each  $x_i$  on the right side in the previous expression, we complete the standardization process

420 and reach the final expression

$$421 \quad 1 = (y \cdot x_1 + y \cdot x_2 + \dots + y \cdot x_n)$$

422

423 or equivalently in a more compact expression

424

$$425 \quad 1 = (\tilde{x}_1 + \tilde{x}_2 + \dots + \tilde{x}_n).$$

426

427 c) The previous standardization process allows an analysis of the contribution of each component of the linear

428 combination. The boxplots of each component for  $i=1, \dots, n$  give a visual idea of the distribution of data in the

429 interval  $[0;1]$  and statistical tests highlight any difference between them and with respect to the range

430 boundaries.

431

432 We standardized hence Equation (3), in order to evaluate the contribution of each of the three components  
 433 isolated in the expression. Following the previous step and renaming the predictions  $Y$  obtained for all the instances  
 434  $j=1,\dots,304$  with the new label  $n_j$ , and with  $v_{i,j}$  the three components in parenthesis ( $i=1,2,3$ ), Equation (3) can be expressed  
 435 as

436

$$437 \quad 1 = n_j v_{1,j} + n_j v_{2,j} + n_j v_{3,j}, \quad (4)$$

438

439 or equivalently

$$440 \quad 1 = \tilde{v}_{1,j} + \tilde{v}_{2,j} + \tilde{v}_{3,j}. \quad (5)$$

441

442 whether referring to the rescaled values  $n_j v_{i,j}$  as  $\tilde{v}_{i,j}$ .

443

444 Since the distributions of  $\tilde{v}_{i,j}$  are not normal (Lilliefors test:  $p < 0.05$ ), the statistical significance was checked with  
 445 the non-parametric Wilcoxon test with Bonferroni correction ( $\alpha = 0.017$ ) for paired data: all components are significantly  
 446 different ( $p < 0.001$ ), that is the difference of the mean values is not zero, in particular comparing  $\tilde{v}_{2,j}$  and  $\tilde{v}_{3,j}$ . The  
 447 boxplots for each one of them (Figure 2) show that the predictions obtained with Equation (5) are mainly due to the  
 448 first addend, that is most of the information is contained in  $\tilde{v}_{1,j}$ . Stated otherwise, in Equation (3) the corresponding  
 449 value

450

$$451 \quad \frac{X_2 + \frac{X_{45}}{X_2}}{X_{10} + \frac{X_2}{X_{45}}} \quad (6)$$

452

453 is the part of the individual almost completely concurring in the prediction. The remaining components play a minor  
 454 role, with a minimal effect on the performance of the individual obtained, corresponding to a refinement of the value  
 455 gained with the main Component (6).

456

457 In order to further investigate the mentioned concept and the interpretability of GP models, we focused on a  
 458 second individual, namely Model 20 in Table 6. The model encapsulated 3 variables, showing a larger error. Despite this,  
 459 the model gains a great interpretability, since the expression released at the end of the run is given by

460

461

$$Y = \frac{X_{45}}{X_2 + X_{44}}, \quad (7)$$

462

where  $X_{44}$  is the number of calves that did not survive during the weaning period.

463

464

Because of the same reasons entailing the simplification of Equation (2) into (3), the previous expression leads to the

465

following:

466

467

$$Y = \left( \frac{1}{\frac{X_{45}}{X_2}} + \frac{X_{44}}{X_{45}} \right)^{-1}, \quad (8)$$

468

otherwise stated as

469

470

$$Y = (Calves^{-1} + DeadCalves)^{-1}, \quad (9)$$

471

472

where *Calves* is the yearly number of calves per cow and the number of calves per cow that do not survive during

473

weaning period is labelled as *DeadCalves*.

474

As in the previous case, we investigated how the prediction is distributed between the two variables *Calves* and

475

*DeadCalves*. We performed again the standardization procedure, supporting the analysis with an equivalent expression

476

of (8):

477

478

$$1 = \tilde{u}_{1,j} + \tilde{u}_{2,j}, \quad (10)$$

479

480

where, for  $k=1,2$ ,  $\tilde{u}_{k,j}$  are the rescaled quantities  $\tilde{u}_{k,j} = m_j \cdot u_{k,j}$ , the prediction  $Y$  obtained with Model (7) are renamed as  $m_j$ ,

481

the variables  $u_{k,j}$  are respectively  $Calves^{-1}$  and *DeadCalves*.

482

Performing once again the non-parametric single sample Wilcoxon test, we obtained extremely significant p-

483

values, supporting that the two components *Calves* and *DeadCalves* mean values are different respectively from the

484

range boundaries 0 and 1. Both variables are crucial in predicting the output, with more relevance given by *Calves*

485

(Figure 3). As we could entail for the first inspected model, the first component of the Expression (8) is crucial one in

486

predicting the output, since it assumes values close to the result. However, this second model is also interesting, as the

---

487 two plotted distributions assume the same complementary behavior. Boxplots in Figure 3 visually express the concept  
488 and in particular we inspected the corresponding instance. The lower outliers of the first distribution correspond to  
489 farms where cows produce a smaller number of calves. It seems reasonable that a higher portion of calves will not even  
490 survive during the weaning period, values corresponding hence to the upper outlier of the second distribution.

491

### 492 3.2. Comparison with other ML techniques.

493

494 In this section, the performance achieved with the five methods are compared. The 30 models obtained by each  
495 ML technique were first evaluated on the Test Set to measure capacity of generalization of each, analyzing the median  
496 Fitness. Finally, the best model (i.e. the one that presents the best fitness) was extracted for each technique.

497 We analyzed the fitness distribution among the thirty models, to assess the ability of the models to learn and to  
498 generalize. In particular, we first commented the results obtained on the learning set and thereafter on the test set.  
499 Figure 4 displays the boxplots of the fitness distribution for each technique.

500

501 For all statistical tests the significance level was set at  $\alpha = 0.05$ . The normality of the distributions among all sets  
502 was analyzed and Lilliefors test showed a deviation from the normal distribution for the results of the LM method ( $p =$   
503  $0.006$ ). Therefore, in order to compare the performance of the achieved models on the learning sets, a non-parametric  
504 test was performed, to asses wheter there is a significant difference between the samples' performace medians. The  
505 median values were compared with Kruskal-wallis test and the null hypothesis that all median values are equal was  
506 rejected ( $p < 0.001$ ). Indeed, all the distributions resulted significantly different to Wilcoxon signed-rank test with  
507 Bonferroni correction ( $\alpha = 0.005$ , since there are 10 comparisons), meaning that all performances differ one from the  
508 other on the learning set ( $p$ -values for all considered couples showed  $p < 0.001$ ). Among all the involved method, as  
509 straightforward from Figure 4, the models obtained with RF are indeed the best performing ones in the learning phase,  
510 whereas GP produced less accurate models.

511

512 The results on the test set were therefore investigated. Predicted values were plotted against the observed data  
513 to check their dispersion among the 30 test sets (Figure 5 (a)). In a supervised learning issue, a predictive model is more  
514 accurate as the predicted values are close to the observed ones. In order for the model to be very accurate, the  
515 regression line of the scatterplot should tend to overlap the bisector of the plane. For each technique we hence plotted

---

516 the regression line of all the predicted values versus the observed values on the test sets and compared the coefficients  
517 of the line: intercepts and slopes are reported in Figure 5(a). All the techniques overestimated target values smaller  
518 than  $\sim 0.85$  (i.e. the coordinates value of the intersection between the bisector and the regression lines). For values  
519 larger than  $\sim 0.85$ , the models underestimated the target. Indicating with  $\bar{x}$  the abscissa of the intersection, the observed  
520 values  $x < \bar{x}$  were estimated with greater prediction values. On the contrary, for  $x > \bar{x}$  the predicted values are lower  
521 than the observed data. The slope of the fitting line obtained with LM is the closest to 1 ( $\beta_1=0.613$ ): the predictions  
522 follow a linear distribution on each test set by construction and therefore the assumed value is expected. Among the  
523 other techniques, NNET, GP and kNN, reported slopes  $\beta_1=0.417$ ,  $\beta_1=0.391$  and  $\beta_1=0.248$  respectively. Finally RF showed  
524 a slope equal to 0.002 and a corresponding larger value for the intercept ( $\beta_0=0.856$ ). Although the latter showed a lower  
525 median RMSE compared to the other techniques, two almost symmetrical regions with respect to the bisector were  
526 identified, entailing that predictions vary into a fixed interval (0.63;1), also for values outside the previous range. It is  
527 clear that the models are not able to generalize. Regarding the corresponding achieved errors, all fitness samples  
528 showed normal distributions of the variables (conclusion supported also by the representation of q-q plot in Figure 5  
529 (b)), and parametric tests were performed. Since the Levène test did not show any difference between the variance of  
530 the distributions ( $p = 0.139$ ), we carried out the one-way ANOVA test: the result was extremely significant, entailing that  
531 at least one sample had a mean performance different from the others. Finally, the Tukey test with Bonferroni correction  
532 was performed, in order to highlight which samples' average performances are actually significantly different. As it is  
533 tangible from the previous boxplots, similar results were achieved on the Test set ( $p > 0.005$ ), that is the techniques that  
534 showed a lower median fitness on the learning set, revealed a lower median error also on the test set, compared to the  
535 other techniques. Moreover, the following pairs of techniques showed not significantly different fitness distributions  
536 among the 30 runs: GP-kNN, kNN-NNet and kNN-LM, NNNet-LM, stating that the pairs of considered methods performed  
537 likewise among the test. What is clear is that, once again, models obtained with RF are the best performing models also  
538 on the test, with respect to all other techniques.

539

540 It is a crucial step to assess the robustness of the model over unseen data with respect to its ability to generalize.  
541 On this purpose, we finally compared the fitness within each technique among learning and test sets respectively.  
542 Apart from LM results, analyzed with non-parametric Kruskal-Wallis and Wilcoxon signed-rank tests, all couples of  
543 results for each technique were tested with the Student's t-test. All techniques showed significant difference between



---

544 learning and test results, extremely remarkable among kNN, NNet, LM and RF ( $p < 0.001$ ). Regarding the results  
545 achieved with GP, high significance was detected comparing the learning and test results ( $p = 0.006$ ).

546

547 The statistical tests entail that all the models can achieve good results on unseen instances, in particular Random  
548 Forest algorithm, since it outperformed all other techniques on both learning and test sets. It is followed by LM, NNet,  
549 kNN, with similar results as stated in the previous paragraph, and finally GP. However, by analyzing the results on the  
550 test sets, their ability to generalize tends not to be as accurate as that obtained during the learning phase. In fact, only  
551 the application of RF led to significantly better results. It must be considered that, among all methods, GP is the  
552 technique that actually produced models that show a median error on the test set not too different from the one  
553 obtained on the learning set. The final models built with kNN, NNet, LM and RF involve all variables available in the  
554 datasets (in the study under consideration, the dataset contains 48 variables). The techniques can easily perform better  
555 and show a better fitness (Figure 4), since the predictions receive the contributions of all parameters. Feature selection  
556 is usually carried out previously in ML approaches. However, this is a not intrinsic operation in the indicated algorithms,  
557 unrelated to their structure. Considering the best model obtained with GP, i.e. showing the lowest RMSE, analyzed in  
558 the previous section (Model 13 in Table 6, i.e. Expression (3)), its performance is comparable to the median behavior  
559 obtained with RF models, even incorporating only three variables among the 48 in input, without imposing a priori  
560 hypotheses. We also managed to provide a zootechnical interpretation, which would not be possible with black-box  
561 techniques. This fact outlines that the different architecture of the evolutionary algorithm can be a good alternative,  
562 balancing overfitting issues, whereas other techniques could slightly be affected. The characteristics of GP outline  
563 models that combine few variables, leading to a great interpretability of the formula and allowing further speculations  
564 on influential parameters.

565

#### 566 **4. Conclusions**

567 In this study, we investigated the performance of Piedmontese cattle breedings, namely the number of weaned  
568 calves per cow produced per year. The sought prediction model should include relevant factors that describe the  
569 weaning period, that is the 60 days after the birth. Many calves do not survive during this time lapse, entailing great  
570 losses to the economic revenues of the breedings and affecting the performance. The expression was expected to

---

571 predict the value without imposing any kind of a priori assumption on its formulation, but it is shaped on the available  
572 data.

573 Medium to large farms located in Piedmont were considered. The dataset provided by the National Association of  
574 Piedmontese Cattle Breeders was accurately filtered, imposing some conditions: since the number of involved variables  
575 was much greater (we processed 19 variable in the previous study), we extracted records from biennium 2017-2018,  
576 among the most representative farms, i.e. with solid data during all the time lapse between 2014 and 2019. The final  
577 dataset consisted in 304 farms and 48 variables, referring to information on cows and artificial inseminations, as well as  
578 heifers, natural inseminations and levels of consanguinity of calves resulting from ongoing pregnancies.

579 ML techniques can provide prediction models without making any kind of a priori assumptions. On this purpose,  
580 the dataset was divided into learning and test sets, and a GP approach was proposed. The technique is a white-box  
581 method, suitable to provide performing models, while automatically selecting significant features. This characteristic  
582 was quite useful and let us develop considerations among the achieved expressions. Whenever a GP model can be  
583 expressed as a sum of terms, it is possible to perform an analysis among the standardized equation. We could deduce  
584 that the first term of the considered sum is the most important one, assuming values close to the output, whereas the  
585 other components concurred minimally in the prediction. We also recall that one of the aims of the study was to produce  
586 models that can be easily read by farmers, highlighting possible important factors, that can explain directly the measure  
587 or the performance of the farm. GP models detected the majority of information in five features, outlining their possibly  
588 crucial role in the prediction of the performance of the breeding farm. The number of calves born from natural  
589 inseminations is the most significant variables, followed by the number of cows, the total number of born calves, and  
590 the number of calves dead in the first 60 days after birth. In exactly half of the models the EBV referred to facility of  
591 parturition of the heifers was used.

592 Comparisons with other classic methods, such as k-Nearest Neighbor, Neural Network, Linear Regression, and  
593 Random Forest were developed. Compared to other techniques, GP is not the best performing method, considering the  
594 median RMSE among 30 runs. On the contrary RF produces models with the best fitness on the test. This could be mainly  
595 due to the different architecture of the algorithms and the fact that, differently from GP, the other performed methods  
596 encapsulate all the features into the prediction models. On one side, we handle with classic techniques, producing on  
597 average outperforming models, showing lower fitness but complex expressions. On the other side, evolutionary  
598 algorithms such as GP led to less accurate models, since their error is slightly greater, but easy to read and interpret. GP

---

599 can model straightforward expressions, which combine a few variables, selected during the evolution process. At the  
600 end of the procedure, the best models performed as well as those obtained with other commonly used techniques, that  
601 are however characterized by non-dynamic algorithms as evolutionary ones.

602 In conclusion, considering all the results in relation to the kind of dealt task, we could assert that GP could represent  
603 the most suitable technique. Evolutionary algorithms can be applied on zootechnical data, achieving performing models,  
604 able to learn on the available data. Further investigations are encouraged, in order to explore the role of other variables  
605 in predicting the considered output. In this sector it is common to associate cow-calf problems to genetic and  
606 pathological factors, related to pregnancy and childbirth. However, many are the factors usually considered as marginal:  
607 difficult to detect and assert as critical points, quality of water and air, illumination, available space and surface,  
608 composition of the food ration could influence the weaning period, being key information that lay into the environment  
609 of the farms. Furthermore, comparisons on other time frames are requested. The management of the farm and the  
610 choices made by the farmer drag on over time and have delayed effects. It is necessary to analyze the problem, taking  
611 into account the data on several years as learning set, to investigate whether ML techniques could detect crucial factors,  
612 that did not emerge in this study.

613

614

615 **Author Contributions:** All authors have read and agree to the published version of the manuscript.

616

617 **Funding:** This work was partially supported by FCT through funding of LASIGE Research Unit (UIDB/00408/2020) and projects BINDER  
618 (PTDC/CCI-INF/29168/2017), GADgET (DSAIPA/DS/0022/2018), AICE (DSAIPA/DS/0113/2019) and PREDICT (PTDC/CCI-  
619 CIF/29877/2017).

620

621 **Conflicts of Interest:** The authors declare no conflict of interest and no competing financial interests or personal relationships that  
622 could have appeared to influence the work reported in this paper.

623 **References**

624

625 Abbona F., Vanneschi L., Bona M., Giacobini M., "A GP approach for precision farming", *2020 IEEE Congress on Evolutionary Computation (CEC)*  
626 *Proceedings*, Glasgow, Scotland, 2020.

- 627
- 628 Abeni, F.; Petrera, F.; Galli, A. A Survey of Italian Dairy Farmers' Propensity for Precision Livestock Farming Tools. *Animals* 2019, 9, 202.
- 629 <https://doi.org/10.3390/ani9050202>
- 630
- 631 Abraham A., Nedjah N. and Mourelle L.D.M., 2006. Evolutionary Computation: from Genetic Algorithms to Genetic Programming. In: Nedjah N.,
- 632 Mourelle, L. D. M., Abraham A. (eds) *Genetic Systems Programming. Studies in Computational Intelligence*, vol 13. Springer, Berlin, Heidelberg.
- 633 [https://doi.org/10.1007/3-540-32498-4\\_1](https://doi.org/10.1007/3-540-32498-4_1)
- 634
- 635
- 636 Amrine, D. E., White, B. J., & Larson, R. L.: Comparison of classification algorithms to predict outcomes of feedlot cattle identified and treated for
- 637 bovine respiratory disease. *Computers and Electronics in Agriculture*, 105, 9-19. (2014). <https://doi.org/10.1016/j.compag.2014.04.009>
- 638
- 639 Associazione Nazionale Allevatori Bovini Razza Piemontese, <http://www.anaborapi.it>
- 640
- 641 Bhattacharya, M. (2013). Machine Learning for Bioclimatic Modelling. *International Journal of Advanced Computer Science and Applications*, 4(2), 1-
- 642 8. <http://dx.doi.org/10.14569/IJACSA.2013.040201>
- 643
- 644 Berckmans, D., General introduction to precision livestock farming, *Animal Frontiers*, Volume 7, Issue 1, January 2017, Pages 6–11.
- 645 <https://doi.org/10.2527/af.2017.0102>
- 646
- 647 Berckmans, D., Guarino, M., From the Editors: Precision livestock farming for the global livestock sector, *Animal Frontiers*, Volume 7, Issue 1, January
- 648 2017, Pages 4-5. <https://doi.org/10.2527/af.2017.0101>
- 649
- 650 Bona, M., Albera, A., Bittante, G., Moretta, A., Franco, G.: L'allevamento della manza e della vacca piemontese, *Supplemento al n. 44 dei Quaderni*
- 651 *della Regione Piemonte-Agricoltura*, pp. 65-129. (2005).
- 652
- 653 Bovine Diseases and Resources, available at: <http://www.cfsph.iastate.edu/Species/bovine.php>
- 654
- 655 Cole, J. B., Newman, S., Foertter, F., Aguilar, I., Coffey, M.,: BREEDING AND GENETICS SYMPOSIUM: Really big data: Processing and analysis of very
- 656 large data sets, *Journal of Animal Science*, Volume 90, Issue 3, March 2012, Pages 723733. <https://doi.org/10.2527/jas.2011-4584>
- 657
- 658 Cozzi G., Brscic M., Gottardo (2009) Main critical factors affecting the welfare of beef cattle and veal calves raised under intensive rearing systems in
- 659 Italy: a review, *Italian Journal of Animal Science*, 8:sup1, 67-80, <https://doi.org/10.4081/ijas.2009.s1.67>
- 660
- 661 Derner J.D., Hunt L., Filho K.E., Ritten J., Capper J., Han G. (2017) *Livestock Production Systems*. In: Briske D. (eds) *Rangeland Systems*. Springer
- 662 *Series on Environmental Management*. Springer, Cham. [https://doi.org/10.1007/978-3-319-46709-2\\_10](https://doi.org/10.1007/978-3-319-46709-2_10)

- 663
- 664 Domingos. S.P., A few useful things to know about machine learning. *Commun. ACM* 55, 10 (October 2012), 78-87.
- 665 <https://doi.org/10.1145/2347736.2347755>
- 666
- 667 González-Recio, O., Rosa, G.J.M., Gianola D., Machine learning methods and predictive ability metrics for genome-wide prediction of complex traits,
- 668 *Livestock Science* (2014). <https://doi.org/10.1016/j.livsci.2014.05.036>
- 669
- 670 Guzhva, O., Ard, H., Herlin, A., Nilsson, M., Sturm, K., Bergsten, C.: Feasibility study for the implementation of an automatic system for the detection
- 671 of social interactions in the waiting area of automatic milking stations by using a video surveillance system. *Computers and Electronics in Agriculture*,
- 672 Volume 127, Pages 506-509, ISSN 0168-1699. (2016). <https://doi.org/10.1016/j.compag.2016.07.010>.
- 673
- 674 Hastie T, Tibshirani R, Friedman J: *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. 2009, Springer, New York City, USA.
- 675 <https://doi.org/10.1007/978-0-387-84858-7>
- 676
- 677 Hessele, A., Therkildsen, M., & Arvidsson-Segerkvist, K. (2019). Beef Production Systems with Steers of Dairy and Dairy × Beef Breeds Based on Forage
- 678 and Semi-Natural Pastures. *Animals: an open access journal from MDPI*, 9(12), 1064. <https://doi.org/10.3390/ani9121064>
- 679
- 680 Koza, J.R. Genetic programming as a means for programming computers by natural selection. *Stat Comput* 4, 87-112 (1994).
- 681 <https://doi.org/10.1007/BF00175355>
- 682
- 683 Lantz, B., *Machine Learning with R*, (Second Edition), Packt Publishing. Cambridge University Press, Cambridge (2015).
- 684
- 685 Lynch E., McGee M., Earley B., Weaning management of beef calves with implications for animal health and welfare, *Journal of Applied Animal*
- 686 *Research*, 47:1, 167-175, (2019) <https://doi.org/10.1080/09712119.2019.1594825>
- 687
- 688 Lo svezzamento del vitello Piemontese [The Weaning of the Piedmontese Calf],
- 689 pp. 3-5, <http://www.anaborapi.it/images/media/pdf/rivista/2012/2012-05.pdf>
- 690 pp. 9 -11, <http://www.anaborapi.it/images/media/pdf/rivista/2012/2012-06.pdf>.
- 691
- 692 Lokhorst, C., de Mol, R.M., Kamphuis, C.: Invited review: Big Data in precision dairy farming. *Animals*. 13(7):15191528. (2019).
- 693 <https://doi.org/10.1017/S1751731118003439>
- 694
- 695 Loyola-González O., "Black-Box vs. White-Box: Understanding Their Advantages and Weaknesses From a Practical Point of View," in *IEEE Access*, vol.
- 696 7, pp. 154096-154113, 2019, <https://doi.org/10.1109/ACCESS.2019.2949286>.
- 697
- 698 Machado, G., Mendoza, M. R. & Corbellini, L. G.: What variables are important in predicting bovine viral diarrhoea virus? A random forest approach.
- 699 *Vet. Res.* 46 (2015), <https://doi.org/10.1186/s13567-015-0219-7>

- 700
- 701 Morota, G., Ventura, R. V., Silva, F. F., Koyama, M., Fernando, S. C.: BIG DATA ANALYTICS AND PRECISION ANIMAL AGRICULTURE SYMPOSIUM:  
702 Machine learning and data mining advance predictive big data analysis in precision animal agriculture. *Journal of animal science*, 96(4), 15401550.  
703 (2018). <https://doi.org/10.1093/jas/sky014>
- 704
- 705 Nasirahmadi, A., Edwards, S.A., Sturm, B., Implementation of machine vision for detecting behaviour of cattle and pigs, *Livestock Sci.*, 202 (2017), pp.  
706 25-38, <https://doi.org/10.1016/j.livsci.2017.05.014>
- 707
- 708 Ortiz-Pelaez, A., Pfeiffer, D.U.: Use of data mining techniques to investigate disease risk classification as a proxy for compromised biosecurity of cattle  
709 herds in Wales. *BMC Vet Res.*;4:24. (2008). <https://doi.org/10.1186/1746-6148-4-24>
- 710
- 711 Poli, R., Langdon, W., McPhee, N.: *A Field Guide to Genetic Programming*. Lulu Enterprises, UK Ltd. (2008).  
712 <https://doi.org/10.1007/s10710-008-9073-y>
- 713
- 714 Price E. O., Harris J. E., Borgwardt R. E., Sween M. L., Connor J. M., Fenceline contact of beef calves with their dams at weaning reduces the negative  
715 effects of separation on behavior and growth rate, *Journal of Animal Science*, Volume 81, Issue 1, January 2003, Pages 116–121.  
716 <https://doi.org/10.2527/2003.811116x>
- 717
- 718 Relazione Tecnica e Statistiche al 31.12.2018 [Technical Reports and Statistics], Anaborapi, Carrù (IT) (2018)  
719 Available at: <http://www.anaborapi.it/images/media/pdf/stat/relazionetecnica2018.pdf>  
720
- 721 Rodero, E., González, A., Luque, M., Herrera, M., Gutiérrez-Estrada, J.C., Classification of Spanish autochthonous bovine breeds. Morphometric study  
722 using classical and heuristic techniques, *Livest. Sci.*, 143 (2012), pp. 226-232 <https://doi.org/10.1016/j.livsci.2011.09.022>
- 723
- 724 Rutten CJ, Velthuis AGJ, Steeneveld W, Hogeveen H. Invited review: sensors to support health management on dairy farms. *J Dairy Sci.*  
725 2013;96(4):1928-1952. <https://doi.org/10.3168/jds.2012-6107>
- 726
- 727 Savoia S., Brugiapaglia A., Pauciullo A., Di Stasio L., Schiavon S., Bittante G., Albera A., Characterization of beef production systems and their effects  
728 on carcass and meat quality traits of Piemontese young bulls, *Meat Science*, 153 (2019), pp. 75-85. <https://doi.org/10.1016/j.meatsci.2019.03.010>.
- 729
- 730 Sistema Piemonte – UBA, available at: [http://www.sistemapiemonte.it/agricoltura/dw\\_rpu/glossario3.shtml](http://www.sistemapiemonte.it/agricoltura/dw_rpu/glossario3.shtml)
- 731
- 732 Silva, S.: GPLAB a genetic programming toolbox for Matlab, (2007). <http://gplab.sourceforge.net/index.html>
- 733
- 734 Spiess, A. N., & Neumeyer, N. (2010). An evaluation of R2 as an inadequate measure for nonlinear models in pharmacological and biochemical  
735 research: a Monte Carlo approach. *BMC pharmacology*, 10, 6. <https://doi.org/10.1186/1471-2210-10-6>

736

737 Tao, H., Guo, F., Tu, Y., Si, B. W., Xing, Y. C., Huang, D. J., & Diao, Q. Y. (2018). Effect of weaning age on growth performance, feed efficiency, nutrient  
 738 digestibility and blood-biochemical parameters in Droughtmaster crossbred beef calves. *Asian-Australasian journal of animal sciences*, 31(6), 864–  
 739 872. <https://doi.org/10.5713/ajas.17.0539>

740

741 The package 'caret', available at: <https://cran.r-project.org/web/packages/caret/caret.pdf>

742

743 Williams, M.L., Parthalin, N.M., Brewer, P., James, W.P.J., Rose, M.T.: A novel behavioral model of the pasture based dairy cow from GPS data using  
 744 data mining and machine learning techniques. *J Dairy Sci.*, 99(3):20632075. (2016). <https://doi.org/10.3168/jds.2015-10254>

745

746 Yao, C., Zhu, X., & Weigel, K. A.: Semi-supervised learning for genomic prediction of novel traits with small reference populations: an application to  
 747 residual feed intake in dairy cattle. *Genetics, selection, evolution: GSE*, 48(1), 84. (2016). <https://doi.org/10.1186/s12711-016-0262-5>

748

749

750

751

## 752 Figures Captions

753

754 **Figure 1.** Distribution of reported deaths for 304 farms during 2017, respectively at birth and after 60 days. All the breedings show extremely different  
 755 values between the dead calves at birth (in blue) and after 60 days (in red) (Kruskal-Wallis test: p-value << 0.001).

756

757 **Figure 2.** Boxplots of the distributions of the variables in Equation (5). Wilcoxon test with Bonferroni correction at  $\alpha=0.017$  reported significantly  
 758 difference between the median of the three distributions ( $p<0.001$ ). The single sample Wilcoxon test, with  $\alpha=0.05$ , finally showed for each distribution  
 759 mean values  $\mu \neq 1$  and  $\mu \neq 0$  ( $p<0.001$ ). Mean values are respectively  $\mu_1=0.9671$ ,  $\mu_2=0.0166$  and  $\mu_3=0.0163$  (red dots).

760

761 **Figure 3.** Boxplots of the distributions of the variables in Equation (10). The single sample Wilcoxon test, with  $\alpha=0.05$ , showed for each variable a  
 762 mean values  $\mu \neq 1$  and  $\mu \neq 0$  ( $p<0.001$ ).

763

764 **Figure 4.** Fitness distribution for all the applied method. Respectively for each technique, the fitness among Learning (in yellow) and Test (in blue)  
 765 sets are shown in boxplots.

766

767 **Figure 5.**

768 (a) Scatterplots for predictions among the test sets. Predicted values among test set are plotted against the corresponding observed data, for each  
 769 method on all the 30 test sets. The blue line represents the linear regression fitting line, whereas the red line is the bisector. Corresponding slopes  
 770 and intercepts are reported for each plot, as well as the corresponding intersection's abscissa on the x-axis.

---

771 (b)Q-Q plots for the fitness among the test set. Normality of the of the RMSEs obtained is visually inspected. The quantiles obtained with all the  
772 performed techniques of the fitness on the test versus the theoretical ones are plotted. The joint distribution in each case follows the diagonal and is  
773 almost entirely contained in the 95% confidence bands.

774

775