XVII^th World Congress of the International Commission of Agricultural and Biosystems Engineering (CIGR)

Hosted by the Canadian Society for Bioengineering (CSBE/SCGAB)
Québec City, Canada   June 13-17, 2010

USING DATA MINING TO EVALUATE DIFFERENT MINIMUM VENTILATION SYSTEMS IN BROILER HOUSES

D.J. MOURA[1], T.M.R. CARVALHO[1], I.A. NÄÄS[1]

[1] D.J. Moura, School of Agriculture Engineering, State University of Campinas, Brazil, daniella.moura@feagri.unicamp.br
[1] T.M.R. Carvalho, thaylamrcarvalho@hotmail.com
[1] I.A. Nääs, irenilza@feagri.unicamp.br

CSBE100657 – Section VI: Postharvest Technology and Process Engineering Conference

ABSTRACT This research analyzed a database of broiler houses climate data using data mining techniques such as attribute selection and data classification (decision tree) to model the impact of air quality and thermal comfort on broilers welfare. The objective was to evaluated four different minimum ventilation systems for broiler houses in relation to air quality and thermal comfort by studying ventilation systems. The study was carried out during the brooding phase with 1 to 14 days old broilers. The treatments were: T1: exhaust fans + curtains management on the end of the house (blue house), T2: exhaust fans + side curtains management (blue house), T3: exhaust fans + flag system (dark house) and T4: natural ventilation. The collected data were: Inside - DBT_in and outside - DBT_out dry bulb temperature, relative humidity - RH, air velocity - Vair, luminosity - L, ammonia concentration - NH_3, carbon dioxide concentration - CO_2, litter pH - pHlitter, litter surface temperature - Tlitter and litter moisture - Mlitter. Four models were fit, one for each treatment. T1 model had a precision of 73.06%, a TP rate of 83.9% for Regular class having RH as root. T2 model had an 82.64% of accuracy a TP rate of 94.8% for Regular class having pHlitter as root. T3 model had 83.13% of accuracy a TP rate of 92.40% for Regular class having NH_3 as root and T4 model had an accuracy of 84.27% a TP of 93.90% for Regular class having TempO as root. The results showed that T2 and T4 presented great percentage of predictive attributes classified as WELL in relation to BAD classification. Unexpected rules were generated by the data mining analyses.

Keywords: Poultry, welfare, data mining, winter ventilation systems, thermal comfort

INTRODUCTION The minimum ventilation may be defined as the amount of air required per hour to achieve the demand of bird's oxygen and air quality, without reducing the temperature, removing excess of dust, gases concentration and moisture, allowing ideal conditions for litter. The air quality study is an important factor to monitor the broilers welfare, health and yield, as far as the sustainability of the overall production. The hazard gases produced on broilers houses are ammonia (NH_3) and carbon dioxide (CO_2) according GLOBALGAP (2007), the upper levels of NH_3 are 20ppm and for CO_2 5000ppm.

Data mining was used to find out patterns in a pre built database, to know which variables are influencing the broiler environment. The decision tree is one of the most popular classification algorithms in current use in data mining and machine learning. Each leaf in a tree corresponds a rule, where the conditions or the rule (the IF-THEN) is formed by making a conjunction (this means that they are combined with the Boolean operator AND) of all conditions from the root to the leaf. The conclusion of the rule is, in the case of classification, to predict the most likely class.
The criteria used for selecting the model were the accuracy of the model, classes' accuracy, ability to understand the generated rules - judged by specialists and complexity of the decision tree, in function of the number of rules and tree size.
Fold cross validation was during the search process select the optimal parameters and as a verification method for the final model. Accuracy, sensitivity, specificity are generally used objective indices to estimate the performance of diagnosis results.
Vale et al. (2008) also used data mining to build a decision tree from weather data and broilers to model the impact of the incidence of heat waves on broilers mortality.
The objective of this research was use the data mining to evaluate the influence of different types of minimum ventilation such as "flag" system, exhaust fans and side curtains management on broilers brooding phase (1 to 14 days old) in relation to air quality and thermal comfort.

METHODOLOGY The present research was developed in 4 broiler housings using different minimum ventilation systems. The breeding of the birds was Cobb from 1 to 14 days old.
- The studied treatments were:
- Treatment 1: Exhaust fans+ curtain management in the end of the broiler house;
- Treatment 2: Exhaust fans + side walls curtain management;
- Treatment 3: Exhaust fans + flags
- Treatment 4: Side curtains natural ventilation.

The environment data were sampled at 80 points, in regular spacing (as a grid) at birds high at 9h00min AM, 2h00min PM, 5h00min PM and 9h00min PM.
The collected data were Dry Bulb Temperature – DBT_in (°C), relative humidity – RH (%) and luminosity intensity – L (lux) collected using the THDL 400, Instrument®, air velocity – Vair (m s^{-1}) was monitored with an anemometer VelociCalc® , TSI™ and NH_3 (ppm) and CO_2 (ppm) concentration were collected with a GasAlertMIcro 5, BW Technologies®. The outside climate data was collected using a meteorological station, HOBO®.

To determine de litter pH, a methodology proposed by Miragliotta (2005) was used with a pHmeter Anavilon® and the litter humidity was calculated by gravimetric method. The surface temperature was collected using a laser thermometer, Omegascope®.
The data mining techniques were applied according to the CRISP-DM methodology comprising the following steps: domain understanding, data acquisition, understanding, preparation, modeling and evaluation according to the knowledge from the domain experts (Chapman et al., 2000).

The software used for the analysis was Weka® 3.5.6 (Witten & Frank, 2005) which is composed of a collection of machine learning algorithms for data mining tasks (e.g., classification). In particular, the classification algorithm chosen was J48, which generates a decision tree for classifying broiler mortality as normal or high. J48 (also known as C4.5) is an algorithm introduced by Ross Quinlan (1993) for inducing Classification Models, also called decision trees. The decision tree generated by J48 can be used for classification and for this reason it is often referred to as a statistical classifier. A decision tree is the representation of recognizable patterns that describe a large number of instances of the training data in a concise and most general way to allow the best possible classification of unknown data. For the tree construction process information theoretical concepts (Shannon, 1948) are used to define the best attributes depending on the largest information gain (difference in entropy) that results from choosing an attribute for splitting the data. The attributes define the possible branches of the growing tree. Early assigned attributes are more important than attributes assigned later during the tree growth. In this way the "most important" attribute - whose values divide the data items into nearly pure subsets with respect to the classification - represents the tree root. Thus the tree construction offers a ranking in the significance of a certain attribute regarding the classification. The attribute with the highest normalized information gain is the one used to make the decision. The algorithm then recourses on the smaller sub lists. The pseudo code of the algorithm J48 can be found in (Quinlan, 1993; Quilan, 1996).

The algorithm J48 is one of best approaches for mining rules through decision trees found in the literature (Han & Kamber, 2006). Apart from that, this algorithm is available in free and commercial softwares and several experimental results using J48 for variable selection show that this algorithm maintains classification accuracy in many bench mark problems, reducing significantly running times (Martínez, & Fuentes, 2005).

Due to the large number of attributes generated in the first data pre-processing, a feature selection was used to remove the attributes with low correlation values. The tools used for the attribute selection were: (i) Principal Component Analysis (PCA) which involves a mathematical procedure to transform a number of (possibly) correlated variables into a smaller number of uncorrelated variables called principal components; (ii) Chi-squared test which evaluated the dependence between the attribute and its classifier (the class attribute); (iii) Wrapper, that evaluates the attribute cluster in a machine learning process and verifies the classifying accuracy of crossing validation; (iv) Correlation Feature Selection (CFS) that searches the cluster of correlated attributes avoiding re-use of the same information; (v) InfoGain, that evaluates the gain in information in relation to the classifier; and (vi) GainRatio that analyzes the information gain rate related to the specific class correcting impaired measurements. Alternatively, a new feature selection approach was used considering the knowledge of the domain experts who selected the main attributes based on their expertise.

The evaluation of the models was made by the fold cross validation test with ten folds. The model accuracy was calculated by a confusion matrix and it is expressed as the percentage of correctly classified test instances over all test instances, including True positives and True negatives. On the other hand, the class precision was also calculated by the confusion matrix and it is expressed as a rate raging from 0 to 1, representing the instances that were correctly classified as True positives or True negatives (Gomes, 2002). The Table 1 presents the ideals parameters of environment to the birds in different ages.

Table 1. Parameters of environment inside build to adequate the decision tree models.

Variables	1 day old	7 days old	14 days old
DBT_in [°C]	DBT_in ≥32	28 ≤ DBT_in ≤30	28 ≤ DBT_in ≤27
RH [%]	29≤RH≤50	40≤RH≤60	50≤RH≤60
ITU [adm[1]]	27,25≤ITU≤29,81	25,35≤ITU≤27,68	25,76≤ITU≤24,19
Vair [m s-1]	Vair< 0,5	Vair < 0,5	Vair < 0,5
CO_2 [ppm]	CO_2<3000	CO_2<3000	CO_2<3000
NH_3 [ppm]	NH_3<20	NH_3<20	NH_3<20
Tlitter [°C]	27 ≤ Tlitter ≤32	27 ≤ Tlitter ≤32	27 ≤ Tlitter ≤32
pHlitter [adm[1]]	pHlitter <5,5	pHlitter <5,5	pHlitter <5,5
Mlitter [%]	Mlitter < 35	Mlitter < 35	Mlitter < 35

[1] Adimensional unit.

RESULTS AND DISCUSSION The evaluation of different minimum ventilation managements was done using data mining through J48 algorithm to built the decision tree, having as target attribute values. The decision tree was made for each treatment.
The possible target-attributes were divided into four classes: very good, good, regular, and bad. From this classification, it was possible to find which factors (node leaves) are effectively influencing the broiler litter quality, the thermal environment and air quality helping the producer to make decision regarding to the minimum ventilation and heating management.

Treatment 1 This model presented an accuracy of 73.06%, and a TP Rate of 83.9% for the Regular class. Putting together the TP examples right predicted from regular, good and bad classes, it could be found that 30.05% of the examples (minimum ventilation management) were classified as "good", 47.98% of the cases were classified as "Regular" and 21.97% of the classified as "Bad". In this treatment, that was the largest tunnel ventilated broiler house in this study, the minimum ventilation was strongly associated with RH, which was considered as the root node of the tree. It shows that as larger the broiler house as difficult to renew the air. Other attributes that the model used to define the efficiency of the ventilation management were pHlitter, T_litter and U_litter, the Va , NH_3 concentration and curtains management.

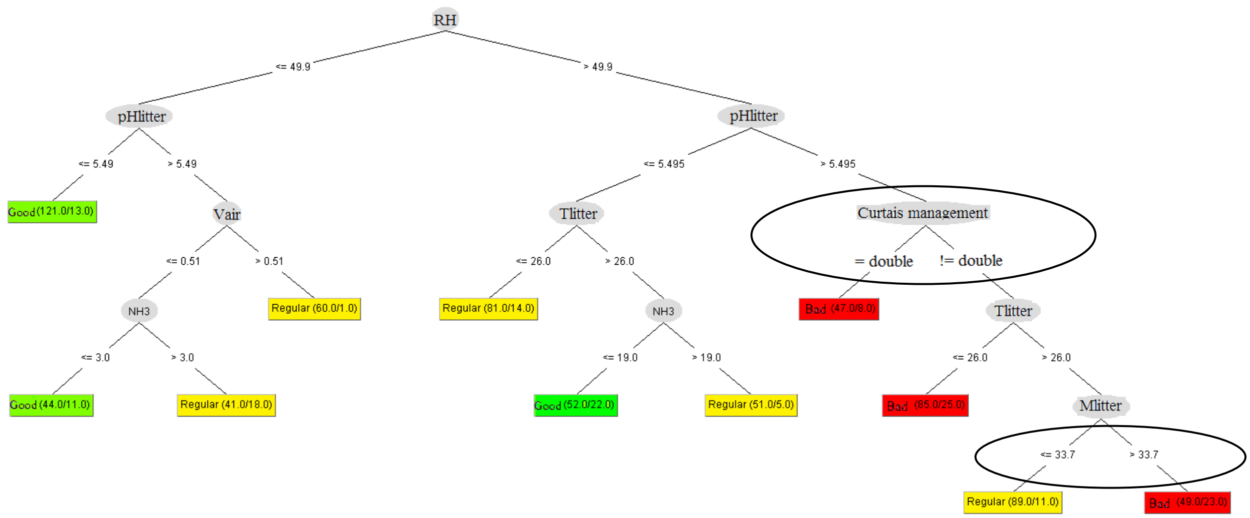

Figure 1[1]. Binary decision tree to evaluate the minimun ventilation management from treatment 1 (BinarySplits=True, minNumObj=36).

[1]RH – relative humidity, pHlitter – litter pH, Vair – air velocity, NH3 – ammonia concentration, Tlitter – superficial litter temperature, Mlitter – litter moisture

The Table 2 presents the confusion matrix of the decision tree from Figure 1 obtained by the cross validation test (10 Cross-validation Folds).

Table 2. Confusion Matrix of the decision tree from T1.

Classes	**Bad (Predicted)**	**Regular (Predicted)**	**Good (Predicted)**	**Precision Class**	**Model Precision**
Bad	88	65	0	0.58	
Regular	44	276	54	0.74	73.06 %
Good	0	31	162	0.89	

The evaluation of the rules showed that the greatest novelty was the rule: "IF RH <= 49.9 AND pHlitter <= 5.49 THEN Good", which fits the RH and pHlitter to determine the environment and minimum ventilation management. A RH less than 50% with an Ulitter lower than 35%, associated with a pHlitter <5.49 causes a decrease in NH3 volatilization of the litter, improving air quality. Gates et al. (1997) found a relationship between pH, moisture and uric acid from the broiler litter. This rule achieves the greatest sensitivity, coverage and support. Throw the fitted model it was possible to find new limits for some attributes, such as the litter moisture that was classified as a good level when lower than 33%, value different from that recommended by Almeida (1986).

Treatment 2 This model showed in Figure 2 presented an accuracy of 82.63%, and a TP Rate of 94.8% for the Regular class. Putting together the TP examples right predicted from regular, good and bad classes, it could be found that 20.06% of the cases were correctly classified as "good", 75.28% of them classified as "regular" and 4.12% of the cases classified as "bad". Through Figure 2, it can be seen that the management of minimum ventilation in this tunnel ventilated broiler house, is strongly affected by pHlitter. The other attributes that influenced the model were the Tlitter, the Vair, the

DTB_out and RH of the air. According Elliot & Collins (1983) the pHlitter strongly influences in ammonia volatilization.

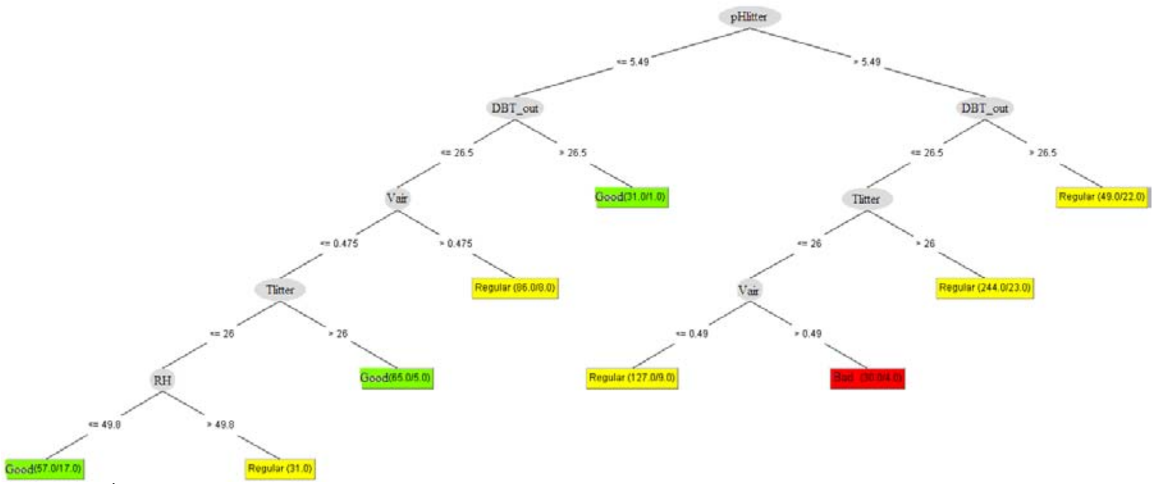

Figure 2[1]. Binary decision tree to evaluate the minimun ventiation management from treatment 2 (BinarySplits=False, minNumObj=30).

[1]RH – relative humidity, pHlitter – litter pH, Vair – air velocity, Tlitter – superficial litter temperature, DBT_out – outside dry bulb temperature

With a TP Rate value in Regular class of 94.8%, it indicates that the model presented a greater efficiency to classify the ventilation management as "regular" with indicates that only a few parameters are need to correct the classification to "good". A better curtain management, for example, could improve the classification to good.

The Table 3 shows a confusion matrix of the tree decision from Figure 2 obtained by the cross validation test (10 Cross-validation Folds).

Table 3. Confusion Matrix of the decision tree from T2.

Classes	Bad (Predicted)	Regular (Predicted)	Good (Predicted)	Precision Class	Model Precision
Regular	476	16	10	0.95	
Good	85	93	0	0.52	82.64 %
Bad	14	0	26	0.65	

The most novelty rule was "IF pHlitter> 5.49 and DBT_out <= 26.5 AND Tlitter> 26 THEN Regular," which combines the attributes of broiler litter such as pHlitter and Tlitter as well as external DBT to evaluate the thermal environment and minimum ventilation management, this rule achieves the greatest sensitivity, coverage and support.

Treatment 3 This model is illustrated in Figure 3 and presented an accuracy of 83.13%, and a TP Rate of 92.4% for the Regular class. Putting together all the TP examples right predicted from regular, good and bad classes, it could be found that 8.02% of the cases would be classified as "Good", 75.41% of the cases "Regular" and 16.57% of the cases as "Bad". Figure 3, shows that in the Dark House the NH_3 concentration had the greatest influence, which was considered the root node of the tree. The other attributes that the model showed as important were Ulitter and Tlitter, as well as the number of exhausted fans turned on and Var. Another curious attribute that defines the model was the season of the year. Therefore if the NH_3 concentration and the Ulitter were in the recommended levels it will be possible to have a good ventilation management.

Regarding the renewal of the air, the greater the number of air changes, the better the air and litter quality, however , the air velocity must be above 0.5 m s^{-1} according Barnwell and Wilson, 2005. At the right side of the decision tree is could be observed that the NH_3 concentration are above the recommended levels by GLOBALGAP (2007), and were classified as "Regular" and "Bad" management.

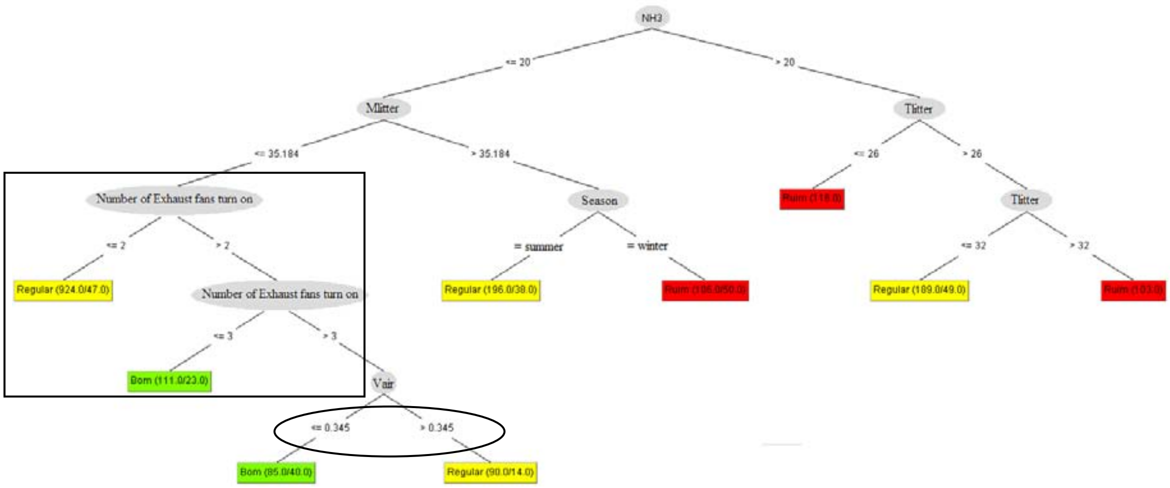

Figure 3[1]. Binary decision tree to evaluate the minimun ventiation management from treatment 3 (BinarySplits=False, minNumObj=85).

[1]Vair – air velocity, Tlitter – superficial litter temperature, NH3 – ammonia concentration, Mlitter – litter moisture

The Table 4 shows a confusion matrix of the tree decision from Figure 3 obtained by the cross validation test (10 Cross-validation Folds).

Table 4. Confusion Matrix of the decision tree from T3.

Classes	Bad (Predicted)	Regular (Predicted)	Good (Predicted)		Precision Class	Model Precision
Regular	1256	65	39	0	0.924	
Good	88	97	0	0	0.524	83.125 %
Bad	121	7	243	0	0.655	
Very Good	0	4	0	0	0	

The most novelty rule was "IF NH3 <= 20 AND Ulitter <= 35.184 ° AND Number of exhaust fans turn on <= 2 THEN Regular", that fitted variables related to air quality. This rule had the greatest sensitivity, coverage and support. In a dark house as treatment 3, the effect of exhaust fans shall be crucial to determine the quality of minimum ventilation, according Cobb (2008) decreasing the NH_3 concentration and moisture. The Dark House, being the most insulated broiler house treatment needs a better minimum ventilation system. Throw the fitted model it was possible to find new limits for some attributes, such as the air velocity for the minimum ventilation. The new values classified as good are lower than 0.345 m/s value different from that recommended by Barnwell & Wilson (2005).

Treatment 4 Figure 4, shows that in a natural ventilated broiler house the environment is strongly associated with external Tbs, which was considered the root node of the tree. Other attributes that the model took in account were related to the quality of broiler litter, (Ulitter and Tlitter), as well as parameters related to the air renewal such as the curtain management, RH and ITU. This natural ventilated broiler house didn't show any problem related to gas concentration.

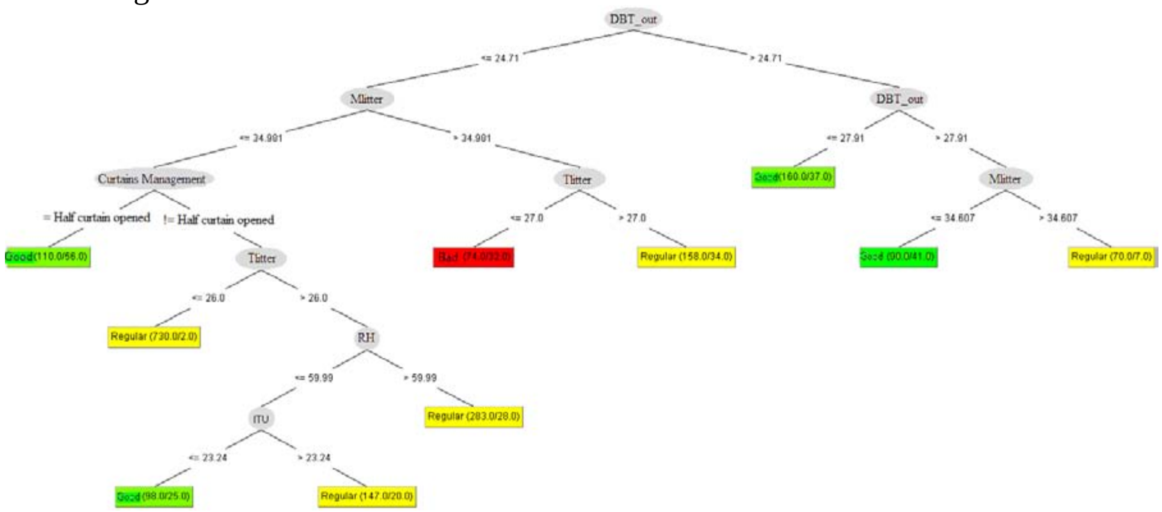

Figure 4[1]. Binary decision tree to evaluate the minimun ventiation management from treatment 4 (BinarySplits=True, minNumObj=70).

[1]Mlitter – moisture litter, DBT_out – outside dry bulb temperature, Tlitter – litter superficial temperature, RH – relative humidity, ITU – temperature and relative humidity index.

The accuracy of the model was 84.27% and TP Rate for the Regular class was 93.90%, which indicates that the model is more effective to qualify the management as "Regular. Table 5 presents the confusion matrix for the decision tree of Figure 4 obtained by the cross validation test (10 Cross-validation Folds).

Table 5. Confusion Matrix of the decision tree from T4.

Classes	Bad (Predicted)	Regular (Predicted)	Good (Predicted)		Precision Class	Model Precision
Regular	1358	88	0	0	0.939	84.27 %
Good	116	260	0	0	0.691	
Very Good	1	39	0	0	0	
Bad	58	0	0	0	0	

The rule that presented the most novelty was "IF DBT_out <= 24.71 AND Ulitter <= 34.981 and Curtain Management ! = Half curtain opened and Tlitter <= 26.0 THEN Regular", which combines attributes related to the external environment (dry bulb temperature outside) to parameters related to litter quality (Tlitter and Ulitter) and the curtains management.

CONCLUSION The results showed that T2 and T4 presented the greatest percentage of predictive attributes classified as WELL in relation to BAD classification. As expected the T1 and T3 were influenced by the moisture and NH3 because of their largest width and the greatest insulation of the dark house. Unexpected rules were generated by the data mining analyses.

Acknowledgements Acknowledgements to Fapesp and CNPq for the financial support and a master scholarship.

REFERENCES

Barnwell, B.; Wilson, M. 2005. Importance of minimum ventilation. Techinical Focul Cobb. V. 1. 2005.

Camerini, N. L.; Nascimento, J.W. B.; Fook, M. V. L.; Soares, E. A.; Silva, F. A. S. Análise de variáveis ambientais em modelos reduzidos de instalações agropecuárias com forro de resíduo de EVA. Revista eletrônica de materiais e processos, v.4.2 (2009) 40-47p.

CHAPMAN, P.; CLINTON, J.; KERBER, R.; KHABAZA, T.; REINARTZ, T.; SHEARER, C.; WIRTH, R. CRISP-DM 1.0: step-by-step data mining guide. 2000. 78p. Available at: http://www.crisp-dm.org/CRISPWP-0800.pdf. Accessed 31 Oct. 2005.

Cobb 2008 Manual de Manejo de Frangos de Corte. Cobb-Vantress

Globalg.a.p. (EUREPGAP) 2007 Pontos de Controle e Critérios de Cumprimento. Garantia Integrada da Fazenda. Aves. V.3.0-2_Sep07. 10p.

GOMES, A.K. Análise do conhecimento extraído de classificadores simbólicos utilizando medidas de avaliação e interessabilidade. São Carlos: ICMC-USP, 2002. 127 p. Dissertação (Mestrado).

Gomes, A.K. Análise do conhecimento extraído de classificadores simbólicos utilizando medidas de avaliação e de interessabilidade. 2002. 127p. Dissertação (Mestrado) – Instituto de Ciências e Matemática e de Computação, Universidade de São Paulo. São Carlos.

Han, J; Kamber, M. 2001 Data Mining: concepts and techniques. San Francisco: Morgan Kaufmann Publishers. 550p.

MARTÍNEZ, J.; FUENTES, O. Using C4.5 as variable selection criterion in classification tasks. In: ARTIFICIAL INTELLIGENCE AND SOFT COMPUTING, Benidorm, 2005. Proceedings. Benidorm: A. P. del Pobil, 2005. p.191-195.

Meira, C.A.A. Processo de descoberta de conhecimento em bases de dados para a análise e o alerta de doenças de culturas agrícolas e sua aplicação na ferrugem do cafeeiro. 2008. 17p. Tese (Doutorado) – Universidade Estadual de Campinas, Campinas –SP.

Miragliotta, M.Y. Avaliação das Condições do Ambiente Interno em dois Galpões de Produção Comercial de Frangos de Corte, com Ventilação e Densidade Populacional Diferenciados. 2005. 244p. Tese (Doutorado) - Universidade Estadual de Campinas, Campinas-SP.

QUINLAN, J.R. C4.5: programs for machine learning. San Francisco, Morgan Kaufmann, 1993. 302p.

QUINLAN, J.R. Improved use of continuous attributes in c4.5. Journal of Artificial Intelligence Research, v.4, p.77-90, 1996.

Vale, M. M. 2008 Sistema de alerta para a ocorrência de mortalidade alta em frangos de

corte devido ao calor. In: Conferência Apinco de Ciência e Tecnologia avícola. Available at: CD-ROOM.
Witten, I. H,; Frank, E. 2005. Data Mining: practical machine learning tools and techniques. 2nd ed. San Francisco: morgan Kaufmann. 525p.
Yahav, S.; Straschnow, A.; Luger, D.; Shinder, D.; Tanny, J.; Cohen, S. Ventilation, sensible heat loss, broiler energy, and water balance under harsh environment conditions. 2004 Poultry Science 83: 253:258