EXTRACTION OF RULES BY CLASSIFICATION FROM WEATHER STATION DATA TO HELP IN THE FORECAST OF TEMPERATURE AND HUMIDITY INDEX FOR DAIRY CATTLE

M. P. dos Santos*, M. M. do Vale, J. P. A. Santos, J. C. dos Santos

Departamento de Zootecnia, UFSM - Univ Federal de Santa Maria, Santa Maria, RS, Brasil

ABSTRACT

Forecasts for comfort index to dairy cattle are unavailable in Brazil and the extraction rules on weather behavior can assist in predicting the animal's comfort, especially for those who are in unprotected places. This study aims to develop a methodology for extracting predictive rules from heat stress conditions in dairy cattle. The analysis was performed using the database of the National Institute of Meteorology (INMET), referring to the hourly averages for the period between September 15th to November 13th of 2013 in Santa Maria - RS, Brazil. The input variables were time of day, air temperature, dew point temperature, relative humidity and the temperature and humidity index. The extraction of the rules was done by the technique of Data Mining and the classification task by building the J48 decision tree algorithm. The classification of the Temperature and Humidity Index (THI) was based on two classes, being NORMAL for THI values less than or equal to 74, and ALERT to values above 74, considered as a promoter of stress. Data mining has resulted in the description of 11 rules of the relationship between temperature, relative humidity and time of day with the THI. Data mining has enabled the understanding of the variables analyzed and the generated rules can help in forecasts based on meteorological forecasts and environmental temperature controllers and relative humidity schedule.

Keywords: THI, thermal comfort, heat stress, environment, data mining.

EXTRAÇÃO DE REGRAS DE CLASSIFICAÇÃO DE DADOS DE ESTAÇÃO METEOROLÓGICA PARA AUXÍLIO NA PREVISÃO DE ÍNDICE DE TEMPERATURA E UMIDADE PARA BOVINOS LEITEIROS

RESUMO

A extração de regras sobre o comportamento meteorológico pode auxiliar na previsão do conforto animal, principalmente para aqueles que estão em ambientes desprotegidos. O objetivo deste estudo foi desenvolver uma metodologia para a extração de regras de previsão das condições de estresse por calor em bovinos de leite. A análise foi realizada com o banco de dados do Instituto Nacional de Meteorologia (INMET), referente às médias horárias durante o período de 15 de Setembro a 13 de Novembro de 2013 em Santa Maria – RS, Brasil. As variáveis de entrada foram: hora do dia, temperatura do ar, temperatura de ponto de orvalho, umidade relativa do ar e o Índice de temperatura e umidade. A extração das regras foi realizada pela técnica de Mineração de Dados, sendo utilizada a tarefa de classificação

* mauricioportellamau@hotmail.com

com a construção de árvore de decisão pelo algoritmo J48. A classificação do Índice de Temperatura e Umidade (ITU) baseou-se em duas classes, sendo NORMAL, para valores de ITU menores ou iguais que 74, e ALERTA, para valores acima de 74, considerado como promotor de estresse. A mineração dos dados resultou na descrição de 11 regras da relação da temperatura, umidade relativa e hora do dia com o ITU. A mineração de dados permitiu a compreensão das variáveis analisadas e as regras geradas podem auxiliar em previsões baseados em temperatura e umidade relativa do ar.

Palavras-chave: ITU, conforto térmico, estresse térmico, ambiência, mineração de dados.

INTRODUCTION

Great economic losses in livestock production are related to the thermal environment, associated to the reduction in food consumption and weight gain (NARDONE et al., 2006). There are few estimates of losses caused by heat, in the United States are estimated about 900 million dollars a year due to effects of heat stress in livestock (ST-PIERRE et al., 2003).

When the ambient temperature is above the thermal comfort range of the animal, environmental adjustments are needed to maintain productivity standards (SOUSA et al., 2004). However, the temperature is not the only factor of the physical environment that influences the animal comfort, because it is directly related to the relative humidity. Warm, moist areas present problems for animal performance when the air temperature is above 21°C and associated with higher relative humidity to 60% (CAMPOS & PIRES, 2001). The determination of the temperature and humidity index (THI), an animal comfort index calculated from temperature and humidity can help to express adequately the degree of heat stress experienced by the animal.

For dairy production, the THI can be classified into four levels: normal, with THI less than 74; alert, between 74 and 78; moderate stress, between 78 and 84; and severe stress above 84 (BROWN-BRANDL et al., 2005). This value of THI can be obtained from sensors or weather data, however, there is no weather forecast system based on THI, even though, one can study patterns that can be applied through the meteorological data modeling summarizing the minimum information in the form of rules.

The analysis of climatic variables becomes complex as the size of databases, being necessary to resort to the use of appropriate techniques for these analyzes. Through data mining technique is possible to identify the relationship of meteorological observations with THI tracks, which can assist in the development of extreme heat mitigation strategies for obtaining classification rules that will produce decision support systems for extreme climate impacting on productivity. The objective of this study was to develop a methodology for the extraction of rules from meteorological data for predicting stress conditions by heat measured in THI for dairy cattle through data mining.

MATERIAL AND METHODS

The rules' extraction was carried out with the database of the National Institute of Meteorology (INMET). For the present study we used as reference data variables: air temperature, dew point temperature and relative humidity. The measures were the hourly average taken during the period of September 15th to November 13th of 2013 in Santa Maria - RS. The temperature and Humidity Index was calculated using Equation 1 according to JOHNSON et al. (1965).

$THI = T_{DB} + 0,36*T_{DP} + 41,2$ Eq. (1)
Where: T_{DB}= dry bulb temperature;
T_{DP}= dew point temperature

The rank attribute was the THI in two classes: NORMAL, with THI≤74; and ALERT, with THI> 74. Was adopted the value of 74 as a threshold to define the climate warning areas for dairy production, as this is the point where they can start the production losses (Silva et al, 2009).

Data were analyzed using the computer program Weka® version 3-4 (Witten & Frank, 2005), applying a classification task, with the construction of the decision tree to the rules extraction. The algorithm used was the J48, an implementation of C4.5 for classification tasks (QUINLAN, 1993; QUINLAN, 1996), which generates a classification tree, a graphic display in the form of an inverted tree in which the root node is first variable with higher power rating, presenting below it the branches formed by the other attributes to the semantic rules that allow the classification (sheet).

The models were evaluated according to VALE et al. (2008), measured on the basis of correct or incorrect ratings for the model and the accuracy of the ratings for each class. The model accuracy was measured in percentage of success and represented the class precision between 0 and 1, with 1 equaling 100% of the precision in its class. The construction of the model and feature selection was accompanied by two experts in animal production and environmental science with at least two years of experience and the rules obtained were described for variables obtained from weather stations timed with the THI.

RESULTS AND DISCUSSIONS

The database was formed by four attributes and 2,170 instances (8,680 data), and generated a classification tree for meteorological variables for THI classification as shown in Figure 1. The model accuracy of the model was 99.54% with 0.997 class precision for the classification of conditions NORMAL THI≤74 and 0.978 for the conditions of THI> 74. From the classification tree it was possible to extract 11 rules for the classification of THI from climatic variables obtained from weather stations.

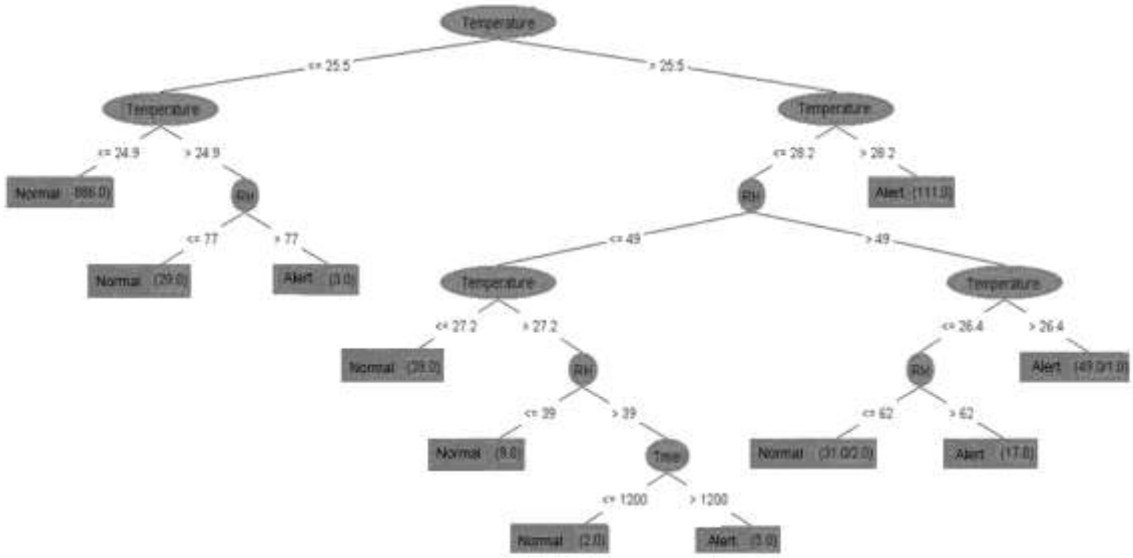

Figure 1 - Classification tree for the Temperature and Humidity Index in NORMAL and ALERT conditions.

The database presented 173h with THI values higher than 74 with an average of 76.6 in a total of 2,170h. The highest THI value reached during the period was 81.84, and the dispersal of cases is shown in Figure 2.

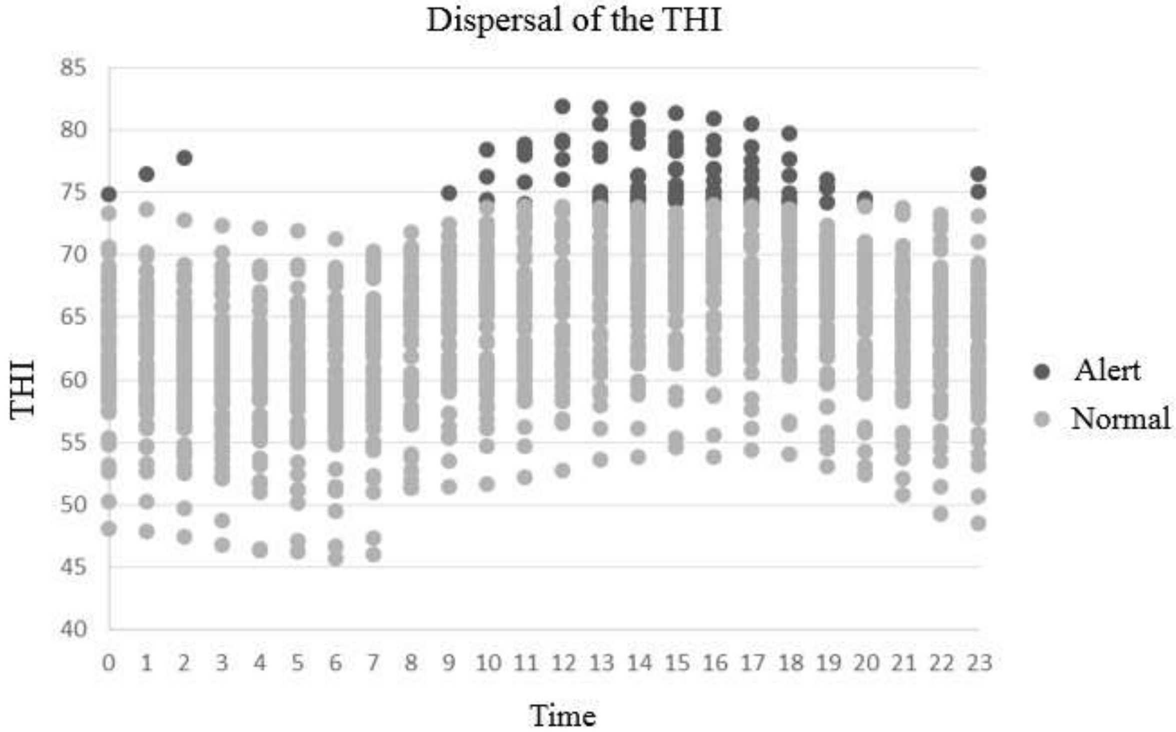

Figure 2 – Dispersal of THI in a 24h period.

According to the climatological conditions for Santa Maria (INMET, 2009), the average climatological normal temperatures during the study period (September, October, November) is 18,7ºC, considering that this average in the warmer months (December, January and February) is 23,8ºC, the number of hours in discomfort in these periods will be longer and of larger extent of occurrence.

The generated tree root node is the air temperature, the basis for classification, separating the data into two main categories, temperatures of dry bulb ≤ 25.5°C and > 25.5°C. This is an expected relationship, due to the THI be calculated by this variable, however, the model was built only in relative humidity and dry bulb temperature, this simplifies the ratings and allows to estimate conditions through simple variables and obtained from weather forecasts.

For the conditions of dry bulb temperature (T_{DB}) ≤ 25.5 ° C the following rules were extracted:

1- IF T_{DB} ≤ 24.9ºC THEN = NORMAL;
2- IF T_{DB} > 24.9ºC AND ≤ 25.5ºC AND RH ≤ 77% THEN = NORMAL;
3- IF T_{DB} > 24.9ºC AND ≤ 25.5ºC AND RH > 77% THEN = ALERT;

Rule 1 classifies the situation where T_{DB} ≤ 24.9ºC no risk exceeding their thermal comfort zone, established independent of the air RH (78% on average over the period with a minimum of 24% and up to 99% RH). In practical terms, the rule states that for weather forecasts of temperature below 25°C will not result in losses independent of RH. When the air temperature is between 24.9ºC and 25.5°C, the incidence of a THI condition ALERT (THI> 74) become dependent on the relative humidity >77%, representing a risk of heat stress.

The two climate variables used to calculate the THI, temperature and relative humidity are highly correlated to the thermal comfort, being consistent in the model. During high temperature, 80% of the heat dissipation produced by the animals occurs via evaporative mechanisms, sweating and / or breath, which depends on the relative humidity (Shearer, 1990). Thus, the higher the relative humidity, the greater the difficulty of removing heat.

The rules obtained in the model of Figure 1 highlight this relationship when associating the temperature between 24.9ºC and 25.5°C and relative humilities of the air equal or higher than 77% (Rule 3). Even at temperatures between 25.5°C and 24.9ºC, when followed by high air humidity, it can lead to productive losses.

For the rules to the conditions of temperature >25.5°C, the following rules were extracted:

4-IF T_{DB} > 28.2ºC THEN = ALERT

Rule 4 shows that temperatures > 28.2ºC present risk of exceeding their thermal comfort zone established independent of other variables.

5-IF T_{DB} ≤ 27.2ºC AND > 25.5ºC AND RH ≤ 49% THEN = NORMAL;

6-IF T_{DB} > 27.2ºC AND ≤ 28.2ºC AND RH ≤ 38% THEN = NORMAL;

Even at temperatures between 27.2 and 28.2ºC, the condition is NORMAL due to low relative humidity, for the cattle ability to withstand heat is inversely proportional to the humidity of the air.

7- IF T_{DB} > 27.2ºC AND ≤ 28.2ºC AND RH > 39% AND ≤ 49% AND time ≤ 12 THEN = NORMAL;

8- IF T_{DB} > 27.2ºC ≤ 28.2ºC AND RH > 39% AND ≤ 49% AND time > 12 THEN = ALERT;

The rules 7 and 8 relate the hours of the day as a selection variable, both in the same conditions of temperature and relative humidity; however, classes may be different by the difference in time of occurrence of the condition being NORMAL before 12 p.m. and ALERT after 12 a.m. The observation of the distribution of the data indicates that the largest THI values occur after 12 p.m., and only 15.6% of the ALERT class occur before 12 p.m., a situation covered by other rules. This relationship between the THI and the time of day is shown in Figure 2 as being in accordance to ST-PIERRE et al. (2003) work in which the dynamics of THI during the day is directly related to the time of day, increasing to a peak after 12 p.m. and returning to lower values at the end of the day.

The last branch of Figure 1 shows the rules 9, 10 and 11 as follows:

9- IF T_{DB} > 26.4ºC AND ≤ 28.2ºC AND RH > 49% THEN = ALERT;

10- IF T_{DB} ≤ 26.4ºC AND RH > 62% THEN = ALERT;

11- IF T_{DB} ≤ 26.4ºC AND RH ≤ 62% THEN = NORMAL.

These rules show the relationship between the relative humidity with temperature. In higher temperatures, relative air humidity may not be high and also trigger an ALERT situation, however, under the same conditions of air temperature, humidity can set the classification of the condition.

In the tropics and subtropics, production losses of about 10% are often observed in the production of dairy cows (BAETA & SOUZA, 2010). Evaluating the decline in milk production in Maringá/Brazil, , in Southern region, KLOSOWSKI et al. (2002) noted in December a milk decrease of approximately 0.5kg cow/day and 4.2kg cow/day for normal production levels of 10 and 35 kg cow/day, respectively. Also CAMPOS et al. (2001), in work on the Goiania/Brazil, in Center East region, obtained approximate results of reduction in milk production of 6.0 and 6,8kg/day for the months of December and March, respectively, considering a normal production level of 30 kg/day. As for the production of 10 kg/day, these same authors found reductions in animal productivity of approximately 0.8 to

1.2kg/day, for the months of December and March, respectively.

From weather forecasts, producers can anticipate the days with possibility of, ALERT situations occurrence changing managements and preparing facilities. The extraction of classification rules in THI conditions obtained in this study indicates that the technique can be used. Stations data of each location and different periods of the year should be studied for local applications.

CONCLUSION

Data mining allows establishing relationships between the main variables related to the Temperature and Humidity Index (THI) for dairy cows. Systems of rules as those presented in this study can be incorporated into computer programs or manually applied in warning systems for predicting animal stress situations based on conventional weather forecasts.

REFERENCES

BAETA, F. C.; SOUZA, C. F. Ambiência em edificações rurais - conforto animal. Viçosa: Editora da Universidade Federal de Viçosa. 2010. 269p.

CAMPOS, A. T.; PIRES, M. F. Á. Prognóstico de declínio na produção de leite em função do clima na Região de Goiás. In: REUNIÃO ANUAL DA SOCIEDADE BRASILEIRA DE ZOOTECNIA, 38., 2001, Piracicaba. **Anais.** Piracicaba: FEALQ, 2001. v.1, p. 11-13.

CURTIS, S. E. Environmental Management in Animal Agriculture. **Iowa University** Press: 1983 Ames, p.409.

BROWN-BRANDL, T. M.; EIGENBERG, R. A.; NIENABER, J. A.; HAHN, G. L. Dynamic Response Indicators of Heat Stress in Shaded and Non-shaded Feedlot Cattle, Part 1: Analyses of Indicators. **Biosystems Engineering**, v.90, n.4, p.451–462, 2005.

INMET. Organizadores; Andrea Malheiros Ramos, Luiz André Rodrigues dos Santos, Lauro Tadeu Guimarães Fortes. **Normais Climatológicas do Brasil** 1961-1990. Brasília, DF: 465p, 2009. ISBN: 978-85-62817-01-4

JOHNSON, H. D. Environmental temperature and lactation with special reference to cattle. **Internationa Jounal of Biometeorology.** 1965, v.9 p.103-116.

KLOSOWSKI, E. S.; CAMPOS, A.T.; CAMPOS.; A.T. Estimativa do declínio na produção de leite, em período de verão, para Maringá-PR, **Revista Brasileira de Agrometeorologia**, v.10,n.1, p.283-288, 2002.

NARDONE A.; RONCHI B.; LACETERA N.; BERNABUCCI U. Climatic Effects on Productive Traits in Livestock. **Veterinary Research Communications**, v. 30, p. 75-81.

NIENABER, J.A.; HAHN, G.L. Engineering and management practices to ameliorate livestock heat stress. 2004. **In: INTERNATIONAL SYMPOSIUM OF THE CIGR. NEW TRENDS IN FARM**

QUINLAN, J.R. **C4.5: Programs for machine learning**. Morgan Kaufmann Publishers, 1993, 302 p.

QUINLAN, J.R. Improved use of continuous attributes in C4.5. **Journal of Artificial Intelligence Research**, v. 4, p. 77-90, 1996.

SANTOS, P.A.; BAÊTA, F.C.; TINÔCO, I.F.F.; ALBINO, L.F.T.; CECON, P.R. Ventilação em modos túnel e lateral em galpões avícolas e seus efeitos no conforto térmico, na qualidade do ar e no desempenho das aves. **Revista CERES**, Viçosa-MG, v.56, n.2, p.172-180, 2009.

SHEARER, J. K.; BEEDE, D. K. Thermoregulation and physiological responses of dairy cattle in hot weather. **Agri-Practice**, Santa Bárbara, v. 11, n. 4, p. 5-8, 13, 16-17, 1990.

SILVA, T. G. F.; MOURA, M. S. B.; SÁ I. I. S.; ZOLNIER, E.; TURCO, S. H. N.; JUSTINO, F.; CARMO. J. F. A.; SOUZA, L. S. B. Impactos das mudanças climáticas na produção leiteira do estado de Pernambuco: análise para os cenários b2 e a2 do IPCC. **Revista Brasileira de Meteorologia**, v.24, n.4, 489 - 501, 2009

SOUZA, S. R. L.; NÄÄS, I. A.; MARCHETO, F. G.; SALGADO, D. D. Análise das condições ambientais em sistemas dede alojamento ‘‘freestall’’ para bovinos de leite. **Revista Brasileira de Engenharia Agrícola e Ambiental**, Campina Grande, PB, v.8, n.2/3, p.299-303, 2004

ST-PIERRE, N. R.; COBANOV, B.; SCHNITKEY, G. Economic losses from heat stress by livestock industries. **Journal of Dairy Science**, E. Suppl., p. 52-77, 2003.

WITTEN, I.H.; FRANK, E. Data mining: practical machine learning tools and techniques. 2nd ed. San Francisco: Morgan Kaufmann, p 525, 2005.