Mining the Impact of Climate Change on Animal Production

Laila Mohamed ElFangary
Information Systems Department
Faculty of Computers and Information, Helwan University
Cairo, Egypt

Maryam Hazman
Central Lab for Agricultural Experts Systems, Ministry of Agriculture and Land Reclamation,
Giza, Egypt

Alaa Eldin Abdallah Yassin
Central Lab for Agricultural Experts Systems, Ministry of Agriculture and Land Reclamation,
Giza, Egypt

ABSTRACT
Animal production plays an important role in the economics of any society; so this paper aims to improve animal production by studying the impact of climate change on animal production. A data mining system is constructed to discover the relation between the weather data and bovine diseases. Since, the animal database and weather database are separated; these databases are integrated using the data warehouses techniques. The proposed system represents a good base for analysis and predictions in the following time period for the purpose of quality decision-making by top management.

General Terms
Data mining.

Keywords
Regression data mining approach, data mart, climate change, animal production

1. INTRODUCTION
Animal production plays an important role in the economics of any society; therefore the governments treat the problems hampering the development of the animal production. One of the most important influences on plant and animal production is climate and weather [1]. Adaptation strategies and measures for the effect of climate on animal production include data collection; monitoring and research that help reduce vulnerability to climate risks or exploit climate opportunities. Good researching is needed for helping decision-making take strategic and tactical decisions. Analyzing how animals respond during extreme weather events, such as heat waves will help in knowing possible effects of global change as extreme events are expected to increase their frequency and severity [2]. Analyzing and interpreting past weather data can predict the future risks and its probabilities. This analysis can used to examine the dependency between the weather and disorders [1].

The general authority for veterinary services in Egypt concerned with numbering bovine and recording their information in a database. The general authority for veterinary services in collaboration with central laboratory for agriculture expert systems developed Bovine Information System (BOVIS) [3]. BOVIS records information like bovine identification number, governorate, six, species, disorder, medications given etc. Utilization such available data on BOVIS will help decision-making to make good decisions.

Mining BOVIS information with the weather database will discover useful knowledge which can be used to improve the animal production. Extracting or mining knowledge from large amounts of data is calling Data Mining (DM) [4]. Data mining provides an information technology which utilizes the data. It is very helpful for decision making since it extract regulations, patterns and models from large databases [5]. It analysis the large quantities data sets to discover relationships and to summarize the data in new ways that are understandable and useful to the data user [6]. Also, data warehouses can be used with data mining to discover knowledge from BOVIS database. Data warehouses (DW) have been an essential information technology strategy component for medium and large sized global organizations. The basis for management reports, decision support, and sophisticated on-line analytical processing and data mining are provided by the Data warehouses [7].

The goal of the research described here is to use a data mining technique to develop a system that can be used to analysis and measures the effect of climate on animal production. This system first combines the BOVIS database with the weather database using the data warehouses techniques. Then, a set of interested variables are selected and prepared. After that, the prepared set of variables will be mine to discover the relation between weather and bovine diseases. This work used the predictive data mining modeling (classification and regression). By discovering the relation between weather and bovine diseases, the system could be used to analysis and predict the when the occurrence of disease.

The structure of the paper in the following section, a brief overview of related work is presented. Section 6, presents the data mart. Section 4 describes the overall structure of the proposed system. Section 5 presents the data preparation module. Section 6, presents the data mining. Section 7 data visualization module while results and analysis presented in section 11. The final section presents the conclusion.

2. RELATED WORK
Data mining is the process of using computer algorithms to seek valuable generalizations from databases of stored information, or from continually occurring information. The aim of the data mining is to analysis existing data to discover new facts and new relationships previously unknown even to experts [5]. Data mining, as defined by [8], "is the process of non-trivial extraction of implicit, previously unknown and potentially useful information from data in large databases".

Data mining is the principal core of the knowledge discovery process. The knowledge discovery process comprises six phases: data selection, data cleansing, enrichment, data

transformation, data mining, and the reporting and display of the discovered information [9].

There are many approaches and algorithms for mining data using in agriculture. Vale and his colleges analyze databases of poultry production associated to climate data using data mining. Attribute selection, data classification and decision trees were used to model and predict the effect of heat wave incidence on broiler mortality [10]. Tripathia and his colleges use a support vector machine (SVM) approach to analyze the different possible changes of the weather scenarios [11]. They use the data pertaining to climate variables and precipitation of each meteorological sub-division to develop SVM based downscaling model for each season. The results obtained from the SVM downscaling model are then analyzed to assess the impact of climate change on precipitation over India [11].

Rajesh extracts pattern from spatial database using k-means algorithm that refers to patterns not explicitly stored in spatial databases [12]. He uses progressive refinement method. First, large data sets are mines using k-means algorithm and then improves the quality of mining in a pruned data set. He applied his work in the area of agriculture where giving the temperature and the rainfall as the initial spatial data and then by analyzing the agricultural meteorology for the enhancement of crop yields and also reduce the crop losses [12].

Ortiz-Pelaez and Pfeiffer use logistic regression, classification trees and factor analysis data mining methods to classify cattle herds according to their risk of disease presence [13]. A selected set of variables were analyzed using a classification tree algorithm in order to build classification rules. Also, they were used by the logistic regression to determine the linear relationship between a set predictors and a dichotomous outcome. The set of variables were independently analysed using factor analysis. Finally the outputs of the three methods applied were integrated in qualitative risk profiles described in the discussion [13].

Li and his colleges use pattern mining to mine animal movement data analysis for discovery of various animal movement patterns [14]. They developed two pattern mining functions: periodic behavior mining and swarm pattern mining. Mining periodic behaviors function detects the reference locations, discovers the periods in complex movements, and then finds periodic patterns by hierarchical clustering. Mining swarm patterns function uncovers flexible moving object clusters by relaxing the popularly-enforced collective movement constraints [14].

3. DATA MART

Since the goal of this work is to improve animal production and this goal can be achieved by predict the bovine diseases. The changes of temperature may cause a spread of disease and produce an increase in the incidence of disease. Analyze the effects of weather and climate on the bovine diseases. This analysis can use to predict the diseases occurrence which will lead to control a particular disease before it occurs. Analysis BOVIS database to identify interesting data for analysis from the existent BOVIS database, and the needed data that not exist. Search for the weather data that related to the animal data which can be use it in mining stage. Central Laboratory of Agriculture Climate provided us by the weather data to sixteen governorates in Egypt, this data for three years. The weather tables that related to BOVIS animal Diseases are selected to be used with the diagnosis data from BOVIS.

Data warehouse provide us by technique to integrate different data source. In order to build the data warehouse repository of integrated information, data and information are extracted from various sources. The data warehouse facilitates reporting, analysis and maintenance of organizations historical information thus it provides an adaptive source of information for decision making [15]. Data warehousing facilitates weather forecasters to accumulate, manage and use the weather data and knowledge. As weather data is extracted over time from various data sources, it is further integrated, transformed, aggregated and mined to derive associations needed by forecasters [16].

Since the interesting variables are a subset from both BOVIS and weather databases, a data mart is needed to be constructed. Data mart is a subset of the data warehouse [6]. Diagnoses data mart was built to derive associations between animal production and climate change. The weather data was modifying by adding some needed variables (like temperature average, status) to integrate it with animal disease data to obtain the used data mart.

4. THE PROPOSED SYSTEM'S ARCHITECTUR

The main objective of this work is to utilize the registered data available for the numbered bovine animals (cattle and buffalo) to help the decisions makers taking a good decision in the right time. Such as, controlling a particular disease and portend before it occurs and to prevent it. As mentioned before, the goal of this work is to measures to the effect of climate on animal production. Analyzed the data recorders for specific period can help us to predict what will happen in the future. Predicting the future events assist the top management to take a quality decision in a short time.

In order to achieve the goal of this work, the proposed system mines the database for both the animals data and the weather data. The proposed system includes 3 modules; the architecture of the proposed system is given in Figure 1.
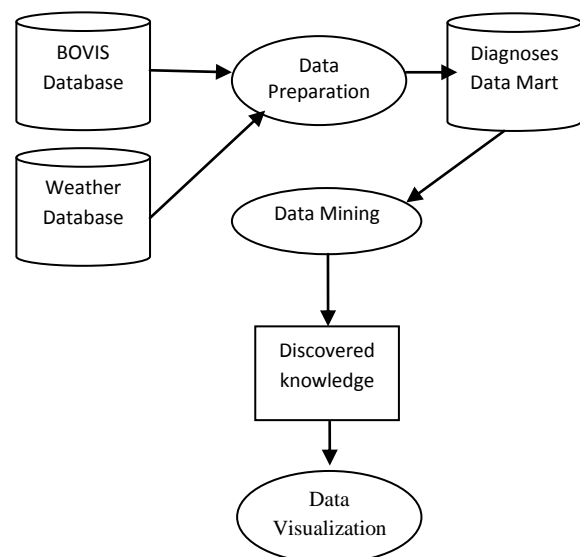

Fig 1: The architecture of the proposed system

A brief description of each of these modules is given below, while the details are described in the next sections:-

- **Data Preparation**. This module is responsible for prepare the input data in order to ensure the quality of the discovered knowledge from mining process.

- **Data Mining**. This module discovers knowledge from the data prepared in the first module. This module usually needs less time than data preparation.

- **Data Visualization**. This module is responsible for present the result data in a meaningful presentation to the user.

5. DATA PREPARATION

An important and critical step in data mining is preparing the date. It has a great importance since the discovered knowledge can only be as good as the data on which they are based [17]. So, the first module of the proposed system is data preparation module. The purpose of this module is to make sure that the used data is complete, free of noise, free of redundancy.

The data preparation involves the following steps:

1. *Data Selection.* BOVIS data base contains many fields which do not related to diagnosis. Also, there are fields do not affected by the weather changed. So, both BOVIS database and weather database were analyzed to select appropriate parts from both databases to be mined. The diagnoses data mart was built using the most significant predictive variables.

2. *Data Cleaning*. In this step, the selected data were analyzed to ensure its quality. This step is similar to what is being done in other data cleaning. It includes removal of redundancy, complete the missing values, identification and correction of data inconsistencies, identification and updating of stale data, and creating a unique record identification [17].

3. *Data discretization.* It makes learning process faster and more accurate. So, it is consider as one of the important step in data preparation for data mining. Discrete attributes will construct precise and short results compared to continuous attributes [18]. In this step, a domain expert gives us the rules that used to classify both the temperature degrees and the humidity. These rules turn the numeric value into a categorical one. Table 1 and 2 give a summarization about them.

Table 1. The Temperature Classification

Temperature	Temp. Status
T>35	Very hot
35>= T >27	Hot
27>= T >20	Good
20>= T >10	Cold
10>=T	Very Cold

Table 2. The Humidity Classification

humidity	humidity Status
H>80	High
80>= H >=50	Good
50> H	Low

4. **Data transformation.** Since different data come in unusable format (e.g. male/female, etc.), these data are converted to values that can be handled by statistical software.

6. DATA MINING

The mining goal is to forecast one variable based on others. For example, using historical data facilitates predict sales of a producer based on customer sex, income and age, geographical region, season of the year, etc. Forecast animal diagnoses based on weather attribute (e.g., temperature, humidity) and the animal attribute (e.g., status, sex, genus and species) is the goal of this work. According to this goal which is predictions, the knowledge discovered during data mining is classification tree. The data are analyzed to discovered uncover patterns and associations that may help us better understand it. Then these patterns are used to classify the data into similar groups to forecast future system behavior. In this model, the predictive data mining modeling (classification and regression) is used to analyze current and historical facts to make predictions about future events. This work uses the data mining algorithm which includes in the SQL server analysis service.

7. DATA VISUALIZATION

The data visualization is very important in any usage of the data specially in data mining. As it converts the discovered data to meaningful way can be viewed by the user. An example for chart to visualize the discovered data is given in figure 2. It presents the number of effected animals which are classified according to the disease name, temperature status, and humidity status.
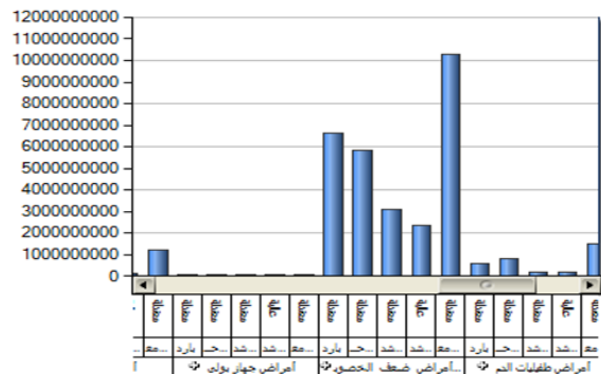

Fig 2: Chart to visualize the discovered data.

8. RESULTS ANALYSIS

The results are analyzed to verify that the discovered knowledge by the proposed data mining system is new and interesting. Through the developed data mining system, the discovered results can be analyzed in different ways. Figure 3 presents the analysis according to the temperature and the humidity on all diseases. It shows the number of affected animals for each disease when the temperature is hot and humidity is low.

Fig 3: Example of analysis the result

Analyzing can be done at the level for animal genus, the effect of the temperature on a specific genus is presented in the figures 3.
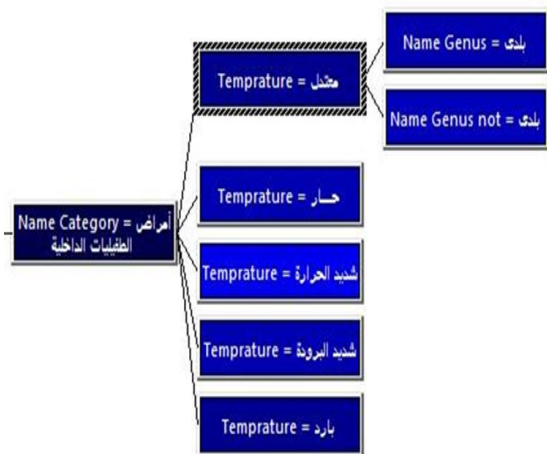
Fig 4: Example for analysis the result

This analysis led us to predict new knowledge. The following are some of the predicted knowledge:-

- The diseases that occur when the temperature is hot and humidity is low

- The genus that effect by a specific diseases categories in different humidity status

- The genus that effect by a specific diseases categories in different temperature status

9. CONCLUSION

This paper has presented an ongoing research that using data mining technique to analysis and measures the effect of climate on animal production. The developed system starts by data preparation which is an important and critical step in data mining and it has an immense impact on the success of a wide variety of complex data analysis, such as data mining [19]. After preparing the interesting data variables, the data mining module mines the built diagnosis data mart to discover the

relation between weather data and bovine diseases. This work uses the predictive data mining modeling. Present the discovered knowledge in meaningful way is done through the data visualization module. The analysis results of the discovered knowledge show that the system can be used to predict the disease occurrence.

In the future, the presented work will be extended using more attributes from the unused variables in both BOVIS and weather database like, animal feed and rain. It also can be integrated with others data mining algorithm to generate new types of discovered knowledge, like association rules.

10. ACKNOWLEDGMENTS

11. REFERENCES
[1] Silva, R. G. 2012. Weather and Climate and Animal Production. In Guide to Agricultural Meteorological Practices, World Meteorological Organization.

[2] Seo, S. N. and Mendelsohn, R. 2008. Animal husbandry in Africa: Climate change impacts and adaptations, AfJARE, Vol. 2 No 1 (Mar. 2008).

[3] Bovine Information System BOVIS, http://www.govs.gov.eg/bovis/

[4] Han, J. and Kamber, M. 2006. Data Mining: Concepts and Techniques, Morgan Kaufmann.

[5] Huang, J. Yuan, Y., Cui, W., and Zhan, Y. 2008. Development of a Data Mining Application for Agriculture Based on Bayesian Networks. In The International Federation for Information Processing, Vol. 258, 645-652.

[6] Olson, D. L. and Delen, D. 2008. Advanced data mining techniques. Springer (Mar. 2008)

[7] Tan, X. 2006. Data warehousing and its potential using in weather forecast, Institute of Urban Meteorology, CMA, Beijing, China.

[8] Piatetsky-Shapiro, G., Fayyad, U., and Smith, P. 1996. From data mining to knowledge discovery: An overview. In Advances in Knowledge Discovery and Data Mining, AAAI/MIT Press, 1-35.

[9] Elmasri, R. and Navathe, S. 2010, Fundamentals of Database Systems, Addison-Wesley Publishing Company.

[10] Vale, M. M., Moura, D. J., Naas, I. de A., Oliveira, S. R. de M., and Rodrigues, L. H. A. 2008. Data Mining to Estimate Broiler Mortality When Exposed to Heat Wave, Sci. Agric. (Piracicaba, Braz.), Vol. 65, No. 3, 223-229.

[11] Tripathia, S., Srinivasa, V.V., and Nanjundiah, R. S. 2006. Downscaling of precipitation for climate change scenarios: a support vector machine approach. J. Hydrol., 330, 621–640.

[12] Rajesh, D., 2011. Application of Spatial Data mining for Agriculture. in International Journal of Computer Applications. Vol. 15, issue 2, 7-9.

[13] Ortiz-Pelaez, A. and Pfeiffer, D. U. 2008. Use of data mining techniques to investigate disease risk classification as a proxy for compromised biosecurity of

cattle herds in Wales. In BMC Veterinary Researcer, 4-24.

[14] Li, Z, Han, J., Ji, M., Tang, L. A., Yu, Y., Ding, B., Lee, J., and Kays, R. 2011. MoveMine: Mining moving object data for discovery of animal movement patterns. ACM TIST Vol. 2, No. 4.

[15] Marakas, G. M. 2003. Modern Data Warehousing, Mining, and Visualization: Core Concepts, Kelley School of Business, Indiana University, Prentice Hall, USA, 20-60.

[16] Valtorta, S. E. 2002. Animal production in a changing climate: Impacts and mitigation, Brody Lectures at the 16th International Congress on Biometeorology, Missouri, USA.

[17] Tsui, K-L., Chen, V., Jiang, W. and slandogan, Y. A. 2006. data Mining Methods and Applications, In Springer Handbook of Engineering Statistics, Editors: Hoang Pham Prof, 651-669,

[18] Shuib, N. L. M., Bakar, A. A., and Othman, Z. A. 2009. Performance Study on Data Discretization Techniques Using Nutrition Dataset. In Proceedings of International Symposium on Computing, Communication, and Control（ISCCC 2009, 304-308.

[19] HU, X. 2003. DB-H Reduction: A Data Preprocessing Algorithm for Data Mining Applications. In applied math. Letters, Vol. 16. 889-895.