

## PERFORMANCE EVALUATION OF RULES-3 INDUCTION SYSTEM ON DATA MINING

MEHMET SABIH AKSOY<sup>1</sup>, HASSAN MATHKOUR<sup>2</sup> AND BADER ALI ALASOOS<sup>1</sup>

<sup>1</sup>Department of Information Systems

<sup>2</sup>Department of Computer Science

College of Computer and Information Sciences

King Saud University

Riyadh, Saudi Arabia

{msaksoy; mathkour}@ksu.edu.sa

Received February 2009; revised June 2009

**ABSTRACT.** *Data mining has been recognized as a key research topic in database systems and machine learning. It aims to discover a useful knowledge from large amount of data. Data mining become one of the most important tools used for solving most of today's problems that are related to different sectors of our life. Different techniques have been developed for mining data in statistics, machine learning, and other disciplines. These techniques need to be re-evaluated, and scalable algorithms should be developed for effective data mining. This paper will investigate the use of RULES-3 Inductive Learning Algorithm for data mining by comparing it with three statistical, two Lazy, and six rule-based data mining algorithms on eleven real life data sets in terms of learning rate, accuracy and robustness to noisy and incomplete data.*

**Keywords:** Inductive reasoning, Data mining, Knowledge acquisition, Rules3, Machine learning

1. **Introduction.** We live in a world where data is growing in size. Inside this data, there is some valuable information and knowledge which is hidden. This information can be extracted using a powerful technology called data mining. Data mining is the process of exploration and analysis of large volume of data in order to discover useful patterns and rules [4]. There are three major fields that have huge impact on data mining evolution. The first is the classical statistics. Most of data mining technologies are based on statistical techniques such as linear regression, standard deviation, regression analysis, standard distribution, standard variance, discriminant analysis, cluster analysis, and confidence intervals. The second field is artificial intelligence (AI). This approach was not practical until the early 1980s, since very powerful computer processing is required. AI takes advantage of reduction in cost and improvement in performance for today's computers by allowing the computer programs learn from the data. The last field is machine learning, which combines both statistics and AI techniques.

First, the data need to be prepared for data mining. That is a set of operations needs to be applied on the data. Then a data mining model has to be selected based on the task required and the type of data available. After selecting the type of the model, we have to select an algorithm to build the model that fits best to data. After the model is tested, it can be used in real world to extract significant information which is necessary to solve a business problem [11].

A lot of effort has been made for evaluating data mining and machine learning algorithms. Such work can be found in [8,9].