



UMS
UNIVERSITI MALAYSIA SABAH

**DEVELOPMENT OF A DIVERSIFIED ENSEMBLE DATA
SUMMARIZATION (DDS) TOOL FOR LEARNING MEDICAL DATA IN
A MULTI_RELATIONAL ENVIRONMENT**

PERPUSTAKAAN
UNIVERSITI MALAYSIA SABAH

FINAL REPORT

GRANT NO: SBK0023-SG-2012

RAYNER ALFRED, LEAU YU BENG AND TAN SOO FUN



UMS
UNIVERSITI MALAYSIA SABAH

SYNOPSIS

Medical or scientific data are normally stored in relational databases in which data are stored in multiple tables. A data summarization approach to knowledge discovery in structured medical datasets is often limited due to the complexity of the database schema. Since most of these data are stored in multiple tables, designing a suitable data summarization method for each individual table that is associated with the target table is required in order to get the best result in summarizing the overall data stored in a multi-relational environment. A diversified data summarization ensemble method is best applied in the task of learning data stored in multiple tables since ensemble methods improve quality and robustness of the results. This research investigates the feasibility of combining a few types of data summarization methods (e.g., DARA) in order to learn data stored in relational databases with high cardinality attributes (one-to-many relations between entities). The proposed algorithm is called a diversified data summarization ensemble method. With this new algorithm, one could facilitate the task of data modelling for data stored in a multi-relational setting by improving the predictive accuracy of the data modelling task. This can be achieved by summarizing each table that exists in the database by using a more appropriate data summarization method depending on the type of data stored in each individual table. This research helps the understanding and development of a diversified data summarization ensemble method that is able to summarize relational data. By applying a subset of data summarization methods to summarize different sets of the relational datasets, more interpretable and useful information can be extracted.

