



## Predictive Analysis Using Data Mining Techniques and SQL

Remedios de Dios Bulos, Georgio F. Go, Giselle O. Ling, Timothy C. Uy and Lawrence J. Yap

*De La Salle University*

*remedios.bulos@dlsu.edu.ph or remedisedios@yahoo.com*

**Abstract:** Most applications systems store, access and manipulate their data using relational data bases. This research aims to build a Classifier system that analyzes and mines data using Standard Query Language (SQL). Specifically, this paper discusses a Music Genre Classifier system that uses a relational database to accept tuples of audio features as input data and then uses a model that was constructed using a data mining tool but was parsed and converted to SQL statements to predict the class labels of musical compositions. In building the Music Genre Classifier system, jAudio a Digital Signal Processing tool was used to preprocess the input data through the extraction of audio features of musical compositions (songs); WEKA was used to explore several data mining algorithms and to build the prediction model; MS Access was used to accept inputs in relational format and to execute the prediction model in SQL. Classification, clustering, and association rule mining algorithms in WEKA were studied, explored, compared and then the most appropriate technique was selected to develop the system. Particularly, only algorithms that generated decision trees and rules as models were considered since these types of output can be easily parsed and then converted to SQL statements. This paper also discusses how decision trees and rules generated from WEKA are parsed and converted to SQL statements. For the comparative analysis of the several algorithms that were considered, experiments to test and measure their predictive accuracy were conducted. For the classifiers, J48 obtained the best predictive accuracy; for the Clusterers, Simple K-Means with J48 produced the highest predictive accuracy; and for Association, Predictive Apriori has the highest accuracy rate. Overall, J48 stood out to be the best algorithm for prediction of musical genre.

**Key Words:** Data Mining; Classification; Clustering; Association; SQL; Music Genre Classification, jAudio, WEKA