manos zalokostas

2011-2012

Data Mining of a Students Database

Advanced Database
Systems &
Applications

assignment one

Table of Contents

The aim of this report is to take a corrupted data set and upon applying pre-processing techniques with a data mining tool, build a model ready for classification & prediction of its attributes values.

Introduction

The following pages will describe the procedure of taking a raw set of data and apply numerous modifications upon, transforming it to a valid dataset able to provide us with classification and prediction rules.

The procedure starts, given a corrupted dataset on a spreadsheet format, that initially needs to be preprocessed through several “by hand” modifications. Then the data set becomes valid of importing it to a data mining tool. The applications that are used during this section are MS Excel and MS WordPad.

By the time the data mining tool comes into play, there will be an exploration of continuous preprocessing techniques for the attributes of the data set, in a way to construct a more effectively predictive model. The data mining application that is used during this section is WEKA. We will number 3 preprocessing models build-in to WEKA application.

Finishing the preprocessing section, the model will be considered in a ready-state for starting to actually appending on it predictive algorithms build in the WEKA application. Among those we choose to qualify on “Decision Tables”, “M5 Rules”, “J48”, “Bayes Net” and others.

Finally some of the above models will enforce us to attempt making predictions for a class attribute of a new data set.

Throughout the entire exploration, the cases will be analyzed extensively through textual and visual mediums of contents.