**Data Mining Term Project**

1. **Introduction :**

**Wine quality is widely used in machine learning and data mini-ng, therefore, this are our study subject of data mining.**

**Wine quality frequently influenced by climate, weather, temper-ature and the winemaking process that are not the indicators we u-sed in dataset, instead ,there are objectivie attributes, such as, vola-tile acidity, sulphate, sulfur dioxide.**

**After we learnt plenty of methods of data mining to find the ru-les of classification, we can now practice and compare it by Weka.**

1. **Preprocess(Related Work) :**

Because we grab the famous data set from UCI website，and we don’t find any missing data in this dataset after checking.

The following method is the Data Preprocess method we try in this dataset

* remove outliers
* discretization
* delete attribute

In Random Forest, we try all these method, the result isn’t always turn to positive way. Among these method “remove outliers” is the best method that can improve our accuracy. And other two method are always turn into lower accuracy.

In Multilayer, remove outliers and delete attribute can improve our accuracy, but discretization can’t.

In K-nearest, remove outliers and delete attribute can improve our accuracy, but discretization can’t.

1. **Data Mining :**
2. **Experiments :**
3. **Conclusion :**

**Reference :**

資料是由Paulo Cortez (Univ. Minho), Antonio Cerdeira, Fernando Almeida, Telmo Matos and Jose Reis (CVRVV) @ 2009提供

