**Data Preprocessing**

For the data preprocessing, there are three steps.

Step 1:  Delete redundant attributes.

Step 2:  Fill up the missing value.

Step 3: Adjusting the data type.

Step 4: Remove all the outliers.

Step 1

The Original data have 78 attributes and 79 tuples.

We use Wake to do the “Delete redundant attributes” processing.

#1

Graphical user interface, application

Description automatically generated

#2 Choose all attributes which we want to remove.

Graphical user interface

Description automatically generated

#3 After we remove the redundant attributes, there only have 22 attributes.

Graphical user interface, application

Description automatically generated

#4 list all attributes

Table

Description automatically generated with medium confidence

Step 2

There have an attribute have missing value, named “ aspect”

We use Excel to make the “missing value” up.

#1 if there shows “-9999”, use “null” to replace it.

Before:

Table

Description automatically generated

After replacing:

Table

Description automatically generated

#2 Then we calculate the mean value of all the numbers in this attribute, Use the mean value to fill up the missing value.

Table

Description automatically generated A picture containing table

Description automatically generated Table

Description automatically generated Table

Description automatically generated

Step 3

Adjusting the data type.

#1 Use N/Y to replace 0/1 in attribute “burned”, make it be a ‘nominal’ data.

Table

Description automatically generated Table

Description automatically generated

Table

Description automatically generated Table

Description automatically generated

Step 4

We use Weka to remove all the outliers.

#1 Open the file in Weka, and choose “InterquartileRange”

Graphical user interface, application, PowerPoint

Description automatically generated

#2 Click on “apply”

Graphical user interface, application

Description automatically generated



#3 Then we could see that “Outliers” shows up in *Attributes* list.

Graphical user interface, application

Description automatically generated

#4 Click “RemoveWithValues”

Graphical user interface, application

Description automatically generated

#5 Change the *attributesIndex* to 23, and change *nominalIndices* to last. And click OK.

Graphical user interface, application

Description automatically generated

#6 Click “Apply”.

Graphical user interface, application

Description automatically generated

#7 Click “outliers”, we could see that all the outliers have been removed.

Graphical user interface, application

Description automatically generated



#8 We need remove “Outlier” and “ ExtremeValue” from *Attributes* list. Then click “Save”.

Graphical user interface, text, application

Description automatically generated



#9 at the end, we will have an available Date for research.

Graphical user interface

Description automatically generated