

Data Mining: STATISTICA



The University of Iowa

Intelligent Systems Laboratory

Outline

- Prepare the data
- Classification and regression

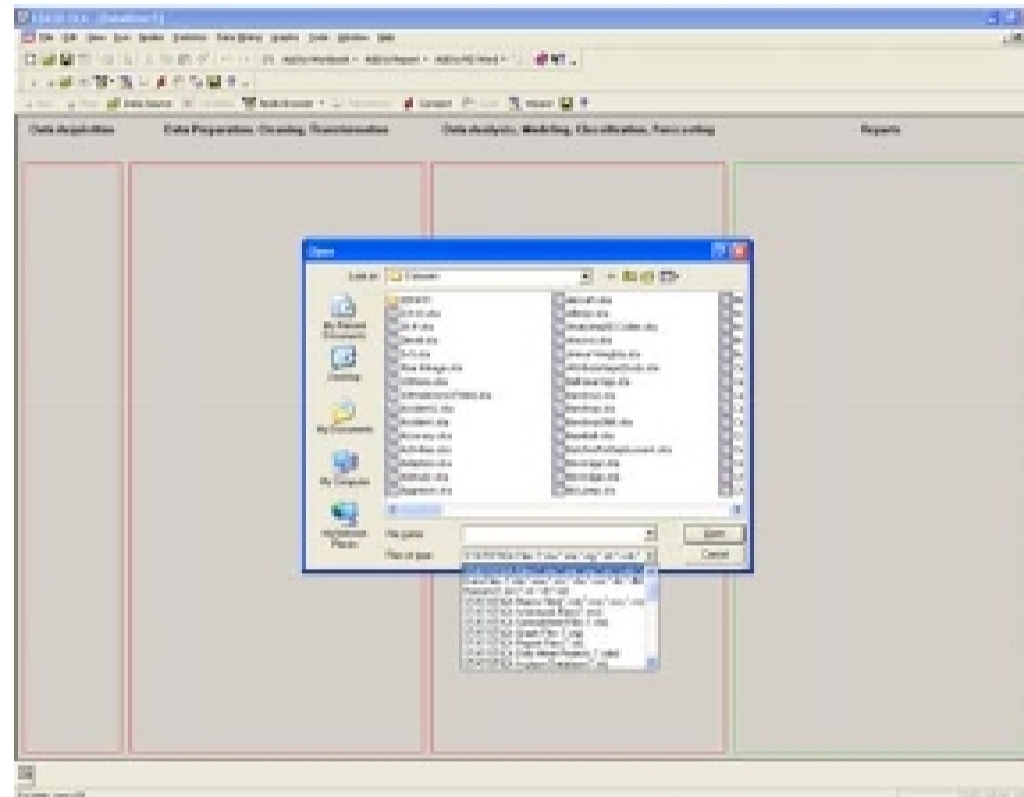


The University of Iowa

Intelligent Systems Laboratory

Prepare the Data

- Statistica can read from Excel, .txt and many other types of files
- Compared with WEKA, Statistica is much easier in terms of data preparing

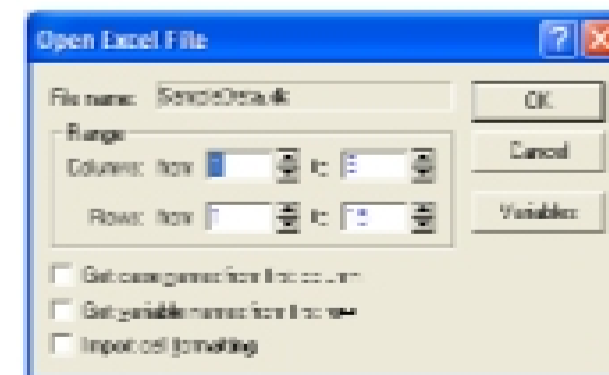
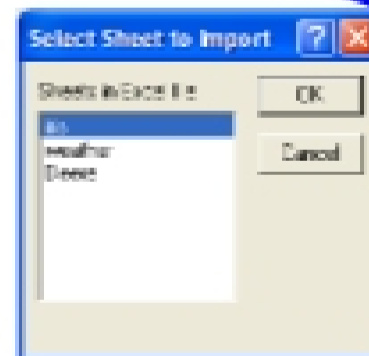
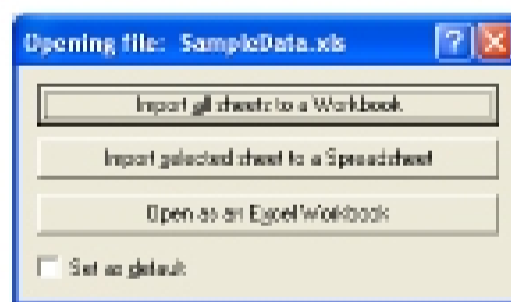


The University of Iowa

Intelligent Systems Laboratory

Open an Excel File

- Click the “Import selected sheet to Spreadsheet”
- Select the desired Excel sheet where your data is stored
- Get variable names from the first row



The University of Iowa

Intelligent Systems Laboratory

Open an Excel File

- Change variable type

	1	2	3	4	5
	outlook	temperature	humidity	windy	play
1	sunny	hot	high	0	no
2	sunny	hot	high	1	no
3	overcast	hot	high	0	yes
4	windy	mild	high	1	yes
5	rainy	cool	normal	0	yes
6	windy	cool	normal	1	no
7	overcast	cool	normal	1	yes
8	windy	mild	high	1	no
9	sunny	cool	normal	0	yes
10	windy	mild	normal	1	yes
11	sunny	mild	normal	1	yes
12	overcast	mild	high	1	yes
13	overcast	hot	normal	0	yes
14	windy	mild	high	1	no



The University of Iowa

Intelligent Systems Laboratory

Open an Excel File

- Change variable type

Variable 1

Name: outlook Type: Text OK

Measurement Type: auto Length: 8 Cancel

Excluded Label Overwrite HD code: 000000000 < >

Display format

- General
- Number
- Date
- Time
- Scientific
- Currency
- Percentage
- Fraction
- Custom

Alt Space... Text Label... Values/Stats... Expand... [Default]...

Long name (label or formula with EuroSign): Function guide

Labels: use any text. Formula: use +variable name or +1, +2, ..., +0 to case B.
Example: (a) = mean(v1+v3, opt)+71, AGE) (b) = v1+v2, comment (info)



The University of Iowa

Intelligent Systems Laboratory