

# Introduction to WEKA

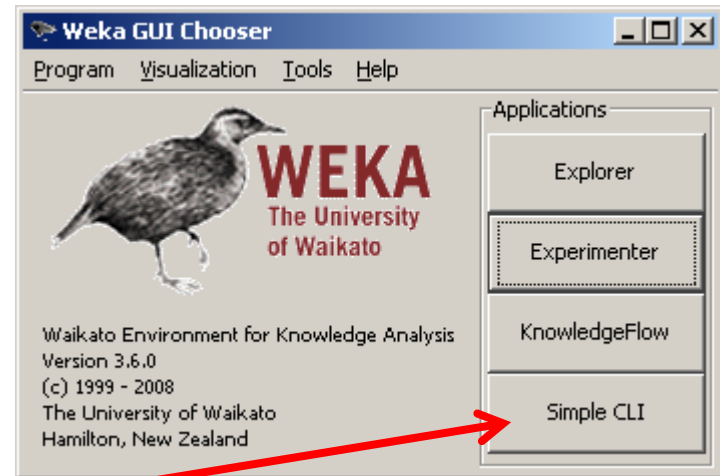
## Lab 1

# WEKA – Waikato Environment for Knowledge Analysis

- Build model
- Predict class
- Evaluate model
- Compare different models
  
- Java Code:
  - Insert into your java application
  - Use GUI for your own classifier

# User Interface

- GUI
  - Explorer
  - Knowledge Flow
  - Experimenter
- Command line



# Input:

## ARFF—Attribute Relation File Format

Structure:

- @relation
- @attribute
- @data

% - comments line

? – missing value

# @relation

@relation <relation-name>

where <relation-name> is a string. The string must be quoted if the name includes spaces.

## Example:

@relation golfWeatherMichigan\_1988/02/10\_14days

@relation 'golfWeatherMichigan 1988/02/10 14 days'

# @attribute

@attribute <attribute-name> <datatype>

The <datatype> can be:

1. numeric
2. nominal
3. string
4. date

# Numeric and nominal attributes

## 1. Numeric:

**@attribute** <attribute-name> <numeric || integer || real>

**@attribute** salary *numeric*

**@attribute** salary *real*

**@attribute** age *integer*

## 2. Nominal:

**@attribute** <attribute-name> <nominal-specification>

**@attribute** outlook {*sunny,rainy,overcast*}

# String and date attributes

3. String – text of unlimited length:

**@attribute** <attribute-name> *string*

**@attribute** comments *string*

4. Date:

**@attribute** <attribute-name> *date* [<date-format>]

**@ attribute** timestamp *date*

Default format:

ISO-8601 combined date and time format: yyyy-MM-d'T'HH:mm:ss

**@ attribute** timestamp *date* "yyyy-MM-dd"



# ARFF file example

```
% This is a toy example, the UCI weather dataset.  
% Any relation to real weather is purely coincidental.  
@relation golfWeatherMichigan_1988/02/10_14days  
  
@attribute outlook {sunny, overcast, rainy}  
@attribute temperature real  
@attribute humidity real  
@attribute windy {TRUE, FALSE}  
@attribute play {yes, no}  
  
@data  
sunny,85,85,FALSE,no  
sunny,80,90,TRUE,no  
overcast,83,86,FALSE,yes  
rainy,70,96,FALSE,yes  
rainy,68,80,FALSE,yes  
rainy,65,70,TRUE,no  
overcast,64,65,TRUE,yes
```

# Simple CLI

(commands are in 'CLIcommands.txt')

- Show some statistics:

```
Java weka.core.Instances filename.arff
```

Example:

```
java weka.core.Instances data/weather.arff
```

- Convert from CSV to ARFF:

**CSV file: the first line is assumed to contain attribute names**

```
java weka.core.converters.CSVLoader filename.csv > filename.arff
```

- Example:

```
java weka.core.converters.CSVLoader data/bankruptcy.csv >  
data/bankruptcy.arff
```