**GUDLAVALLERU ENGINEERING COLLEGE** 

(An Autonomous Institute with Permanent Affiliation to JNTUK, Kakinada) Seshadri Rao Knowledge Village, Gudlavalleru – 521 356.

**Department of Computer Science and Engineering** 

# **III Year II Semester**

2019 - 2020



# DATA MINING-LAB FACULTY MANUAL

# Prepared by

Dr. G.V.S.N.R.V.Prasad Professor & Vice Prinicpal(Academics)

Mr. K. Bala Brahmeswara Assistant Professor Dr.A.Jagadeswara Rao Associate Professor

Mrs. Y. Aditya Assistant Professor

# **GUDLAVALLERU ENGINEERING COLLEGE**

# (An Autonomous Institution with Permanent Affiliation to JNTUK, Kakinada) Seshadri Rao Knowledge Village, Gudlavalleru – 521356 DEPARTMENT OF COMPUTER SCIENCE AND ENGIEERING

# **INSTITUTE VISION & MISSION**

#### **Institute Vision:**

To be a leading institution of engineering education and research, preparing students for

leadership in their fields in a caring and challenging learning environment.

#### **Institute Mission:**

- To produce quality engineers by providing state-of-the-art engineering education.
- To attract and retain knowledgeable, creative, motivated and highly skilled individuals whose leadership and contributions uphold the college tenets of education, creativity, research and responsible public service.
- To develop faculty and resources to impart and disseminate knowledge and information to students and also to society that will enhance educational level, which in turn, will contribute to social and economic betterment of society.
- To provide an environment that values and encourages knowledge acquisition and academic freedom, making this a preferred institution for knowledge seekers.
- To provide quality assurance.
- To partner and collaborate with industry, government, and R&D institutes to develop new knowledge and sustainable technologies and serve as an engine for facilitating the nation's economic development.
- To impart personality development skills to students that will help them to succeed and lead.
- To instil in students the attitude, values and vision that will prepare them to lead lives of personal integrity and civic responsibility.
- To promote a campus environment that welcomes and makes students of all races, cultures and civilizations feel at home.
- Putting students face to face with industrial, governmental and societal challenges.

## DEPARTMENT VISION & MISSION

### **Vision**

To be a Centre of Excellence in Computer Science and Engineering education and training to meet the challenging needs of the industry and society.

### **Mission**

- To impart quality education through well-designed curriculum in tune with the growing software needs of the industry.
- To serve our students by inculcating in them problem solving, leadership, teamwork skills and the value of commitment to quality, ethical behavior & respect for others.
- To foster industry-academia relationship for mutual benefit and growth.

# **PROGRAMME EDUCATIONAL OBJECTIVES (PEOs)**

- **PEO 1:** Identify, analyze, formulate and solve computing problems both independently and in a team environment using appropriate modern tools.
- **PEO 2:** Develop software systems with significant technical, legal, ethical, social, environmental and economic considerations.
- **PEO 3:** Exhibit commitment in lifelong learning, professional development and leadership and communicate effectively with professional clients and the public.

# **PROGRAM SPECIFIC OUTCOMES (PSOs)**

- 1. Design, develop, test and maintain reliable software systems and intelligent systems.
- 2. Design and develop websites, web apps and mobile apps.

# **Program Outcomes (Pos)**

(B.Tech in Computer Science and Engineering)

Engineering students will be able to:

- **1. Engineering knowledge**: Apply the knowledge of mathematics, science, engineering fundamentals, and an engineering specialization to the solution of complex engineering problems.
- 2. **Problem analysis**: Identify, formulate, review research literature, and analyze complex engineering problems reaching substantiated conclusions using first principles of mathematics, natural sciences, and engineering sciences.
- **3. Design/development of solutions**: Design solutions for complex engineering problems and design system components or processes that meet the specified needs with appropriate consideration for the public health and safety, and the cultural, societal, and environmental considerations.
- 4. Conduct investigations of complex problems: Use research-based knowledge and research methods including design of experiments, analysis and interpretation of data, and synthesis of the information to provide valid conclusions.
- **5. Modern tool usage**: Create, select, and apply appropriate techniques, resources, and modern engineering and IT tools including prediction and modelling to complex engineering activities with an understanding of the limitations.
- 6. The engineer and society: Apply reasoning informed by the contextual knowledge to assess societal, health, safety, legal and cultural issues and the consequent responsibilities relevant to the professional engineering practice.
- **7. Environment and sustainability**: Understand the impact of the professional engineering solutions in societal and environmental

contexts, and demonstrate the knowledge of, and need for sustainable development.

- **8. Ethics:** Apply ethical principles and commit to professional ethics and responsibilities and norms of the engineering practice.
- **9. Individual and team work:** Function effectively as an individual, and as a member or leader in diverse teams, and in multidisciplinary settings.
- **10. Communication:** Communicate effectively on complex engineering activities with the engineering community and with society at large, such as, being able to comprehend and write effective reports and design documentation, make effective presentations, and give and receive clear instructions.
- **11. Project management and finance**: Demonstrate knowledge and understanding of the engineering and management principles and apply these to one's own work, as a member and leader in a team, to manage projects and in multidisciplinary environments.
- **12. Life-long learning**: Recognize the need for, and have the preparation and ability to engage in independent and life-long learning in the broadest context of technological change.

# **Department of Computer Science and Engineering** GENERAL LABORATORY INSTRUCTIONS

- Students are advised to come to the laboratory at least 5 minutes before (to the starting time), those who come after 5 minutes will not be allowed into the lab.
- 2. Plan your task properly much before to the commencement, come prepared to the lab with the synopsis / program / experiment details.
- 3. Student should enter into the laboratory with:
- a. Laboratory observation notes with all the details (Problem statement, Aim, Algorithm, Procedure, Program, Expected Output, etc.,) filled in for the lab session.
- b. Laboratory Record updated up to the last session experiments and other utensils (if any) needed in the lab. c. Proper Dress code and Identity card.
- 4. Sign in the laboratory login register, write the TIME-IN, and occupy the computer system allotted to you by the faculty.
- 5. Execute your task in the laboratory, and record the results / output in the lab observation note book, and get certified by the concerned faculty.
- 6. All the students should be polite and cooperative with the laboratory staff, must maintain the discipline and decency in the laboratory.
- 7. Computer labs are established with sophisticated and high end branded systems, which should be utilized properly.
- 8. Students / Faculty must keep their mobile phones in SWITCHED OFF mode during the lab sessions. Misuse of the equipment, misbehaviors with the staff and systems etc., will attract severe punishment.
- 9. Students must take the permission of the faculty in case of any urgency to go out ; if anybody found loitering outside the lab / class without permission during working hours will be treated seriously and punished appropriately.

10. Students should LOG OFF/ SHUT DOWN the computer system before he/she leaves the lab after completing the task (experiment) in all aspects. He/she must ensure the system / seat is kept properly.

#### COURSE NAME: DATA MINING LAB

#### COURSE CODE: CS2514

#### **COURSE OBJECTIVES:**

To exercise the data mining techniques such as classification, clustering, pattern mining etc with different datasets and dynamic parameters using WEKA tool.

#### **COURSE OUTCOMES:**

Upon successful completion of the course, the students will be able to

- Learn to execute data mining tasks using a data mining toolkit (such as WEKA) and visualize the results.
- Demonstrate the working of algorithms for data mining tasks such association rule mining, classification, clustering and regression.

#### **MAPPING OF COURSE OUTCOMES WITH PROGRAM OUTCOMES:**

COURSE OUTCOMES	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9	P10	P11	P12
C01	$\checkmark$	$\checkmark$		$\checkmark$	$\checkmark$							$\checkmark$
CO2	$\checkmark$	$\checkmark$			$\checkmark$							$\checkmark$

# **DATAMINING LAB INDEX**

SNO	Experiment Name	Page No
1	Explore WEKA Data Mining/Machine Learning Toolkit	11
2	Perform data preprocessing tasks oni. Add attributeii. Add expressioniii. Copy attributeiv. Remove attribute	31
3	Demonstrate performing classification on data sets	49
4	Demonstrate performing association rule mining on data sets	62
5	Demonstrate performing regression on data sets	78
6	Demonstrate performing SVM classification on data sets	88
7	Demonstrate performing clustering on data sets	96
8	Demonstrate performing knowledge flow on WEKA	113

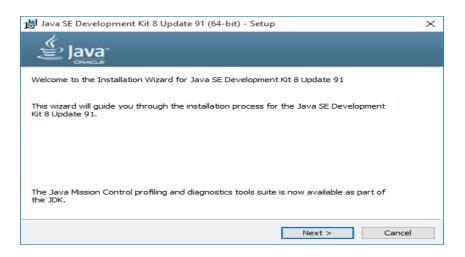
# **ADDITIONAL – EXPERIMENT LIST**

SNO	Experiment Name	Page No
1	FILE FORMATS FOR WEKA	154
2	Perform Data Preprocessing tasks on Data Cleaning & Noisy Data.	156
3	Demonstrate performing Hierarchal clustering on data sets	161

### AIM: i) Downloading and/or installation of WEKA data mining toolkit

**Note:** if weka installation package don't have java included in it then compatible version of the java has to be installed initially as given below.

Step 1: Select appropriate JAVA version (Here we are using:jdk1.8). Operate the executable file of welcome window to the installation wizard of Development kit ,click next button



Step 2: Custom setup window to select path and click next.

讃 Java SE Development Kit 8 Update 91	(64-bit) - Custo	m Setup	$\times$
Java	_	_	
Select optional features to install from the installation by using the Add/Remove Progr			of features after
Development Tools Source Code Public JRE		Feature Descripti Java SE Developr Update 91 (64-bi JavaFX SDK, a pr the Java Mission suite. This will re your hard drive.	nent Kit 8 t), including the ivate JRE, and Control tools
Install to: C:\Program Files\Java\jdk1.8.0_91\			Change
	< Back	Next >	Cancel

Step 3: In the destination folder window we check java file stored location and click next.



Java SE Development Kit 8 Update 91 (64-bit) - Complete

 Second
 X

 Java SE Development Kit 8 Update 91 (64-bit) Successfully Installed

 Java SE Development Kit 8 Update 91 (64-bit) Successfully Installed

 Click Next Steps to access tutorials, API documentation, developer guides, release notes and more to help you get started with the JDK.

 Next Steps

Step 4: Installation process completion window is displayed and click close.

#### Weka –Installation:

Go to the Weka website, http://www.cs.waikato.ac.nz/ml/weka/, and download the software.On the left hand side, click on the link that says download. Select the appropriate link corresponding to the version of the software based on your operating system and whether or not you already have Java VM running on your machine.Save the self extracting executable to the disk and double-click on it to install weka.

Step 1: Welcome setup wizard window is viewed where next button has to be clicked



Step2: In the license agreement window accept the agreement with "I Agree" button.

Weka 3.8.3 Setup			_		
<b>Weka</b>	License Agree Please review t	<b>ment</b> he license terms b	efore installing V	Veka 3.8.3.	
Press Page Down to se	the rest of the agree	ement.			
	RAL PUBLIC LICENSE 3, 29 June 2007				^
Copyright (C) 2007 Fr Everyone is permitted of this license docume	to copy and distribute nt, but changing it is n	e verbatim copies			
Prear The GNU General Pub software and other kin	lic License is a free, co	opyleft license for			~
If you accept the terms agreement to install We		ck I Agree to cont	tinue. You must a	accept the	
ullsoft Install System v08	-Mar-2013.cvs ———				
		< Back	I Agree	Cance	

Step 3: The component window is displayed where the type of installation must be "Full" and click on next.

🗃 Weka 3.8.3 Setup		
	Choose Components Choose which features of	Weka 3.8.3 you want to install.
install. Click Next to continue.		the components you don't want to
Select the type of install:	Full	$\sim$
Or, select the optional components you wish to install:	Associate Files	Description Position your mouse over a component to see its description.
Space required: 107.3MB		
Nullsoft Install System v08-Mar-2	2013.cvs	
	< Ba	ack Next > Cancel

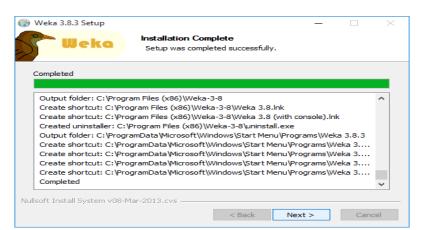
Step 4: In the installation window to select the path where the software loads can be selected through Browse button to the destination folder and click next button.

🌍 Weka 3.8.3 Setup			_		$\times$
Weka	Choose Insta Choose the fo	Il Location	nstall Weka 3.8.3	i.	
Setup will install Weka 3. and select another folder	8.3 in the following r. Click Next to con	) folder. To install tinue.	in a different fol	der, click Bro	owse
Destination Folder	6)\Weka-3-8		В	rowse	]
Space required: 107.3ME Space available: 354.6GE					
Nullsoft Install System v08-1	Mar-2013.cvs ——	< Back	Next >	Can	cel

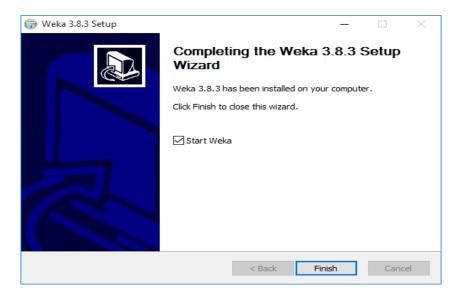
Step 5: To add to the list of programs from the start menu check the create shortcut and press the Install button

🗊 Weka 3.8.3 Setup			-		$\times$
Weka	Choose Start N Choose a Start I	<b>lenu Folder</b> Menu folder for the	Weka 3.8.3 sho	ortcuts.	
Select the Start Menu fo			program's shor	tcuts. You	
can also enter a name to	create a new folder.				_
Weka 3.8.3 Accessibility					<b>.</b>
Accessories					
Administrative Tools Adobe					
AOMEI Backupper					
doPDF 7					
Java Development Kit					
Macromedia					
Maintenance Microsoft Office					v
Do not create shortc	uts				
Vullsoft Install System v08-	Mar-2013.cvs ———				
		< Back	Install	Cance	-
		< DOCK	Install	Cance	

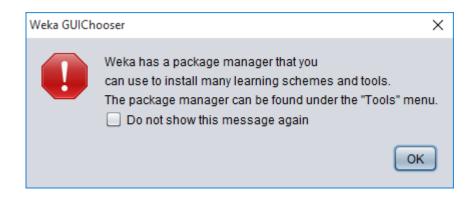
Step 6: completion of installation is indicated in the window where the next button is selected.



Step 7: The setup wizard with the start up of Weka is chosen and finish the process.



Step 8: An warning window tells about the feature of package manager, click ok.



Step 9: Weka home page is viewed.



Step 10: The installation process is completed fully.

#### 2. Understand the feature of Weka toolkit

The Weka GUI Chooser (class weka.gui.GUIChooser) provides a starting point for launching Weka's main GUI applications and supporting tools. If one prefers a MDI ("multiple document interface") appearance, then this is provided by an alternative launcher called "Main" (class weka.gui.Main). The GUI Chooser consists of four buttons—one for each of the four major Weka applications—and four menus.



#### The buttons can be used to start the following applications:

• **Explorer** An environment for exploring data with WEKA (the rest of this documentation deals with this application in more detail).

• **Experimenter** An environment for performing experiments and conduct ing statistical tests between learning schemes.

• KnowledgeFlow This environment supports essentially the same functions as the Explorer but with a drag-and-drop interface. One advantage is that it supports incremental learning.

• **SimpleCLI** Provides a simple command-line interface that allows direct execution of WEKA commands for operating systems that do not provide their own command line interface.

#### The menu consists of four sections:

1. Program

Program	
LogWindow	Ctrl+L
Memory usage	Ctrl+M
E×it	Ctrl+E

- LogWindow Opens a log window that captures all that is printed to stdout or stderr. Useful for environments like MS Windows, where WEKA is normally not started from a terminal.
- Exit Closes WEKA.

#### 2. Visualization

Visualization	
Plot	Ctrl+P
ROC	Ctrl+R
TreeVisualizer	Ctrl+T
GraphVisualizer	Ctrl+G
BoundaryVisualizer	Ctrl+B

- **Plot** For plotting a 2D plot of a dataset.
- **ROC** Displays a previously saved ROC curve.
- **TreeVisualizer** For displaying directed graphs, e.g., a decision tree.
- **GraphVisualizer** Visualizes XML BIF or DOT format graphs, e.g., for Bayesian networks.
- **BoundaryVisualizer** Allows the visualization of classifier decision boundaries in two dimensions.

### **3.** Tools

Tools '	
ArffViewer	Ctrl+A
SqlViewer	Ctrl+S
Bayes net editor	Ctrl+N

- ArffViewer An MDI application for viewing ARFF files in spread- sheet format.
- SqlViewer Represents an SQL worksheet, for querying databases via JDBC.
- Bayes net editor An application for editing, visualizing and learn- ing Bayes nets.

### 4. Help

Help		
We	eka homepage	Ctrl+H
HC	WTOs, code snippets, etc.	Ctrl+W
We	eka on Sourceforge	Ctrl+F
Sys	stemInfo	Ctrl+I

- Weka homepage Opens a browser window with WEKA's home-page.
- HOWTOs, code snippets, etc. The general WekaWiki [2], con-taining lots of examples and HOWTOs around the development and use of WEKA.
- Weka on SourceforgeWEKA's project homepage on Sourceforge.net.
- SystemInfo Lists some internals about the Java/WEKA environment, e.g., the CLASSPATH.
- **3.** Navigate the options available in the WEKA (ex. Select attributes panel, preprocess panel, classify panel, cluster panel, associate panel and visualize panel).

# Explorer

### 1. The user interface

#### 1.1 Section Tabs

At the very top of the window, just below the title bar, is a row of tabs. When the Explorer is first started only the first tab is active; the others are greyed out. This is because it is necessary to open (and potentially pre-process) a data set before starting to explore the data. The tabs are as follows:

- 1. **Preprocess**. Choose and modify the data being acted on.
- 2. Classify. Train and test learning schemes that classify or perform regression.
- 3. Cluster. Learn clusters for the data.
- 4. Associate. Learn association rules for the data.
- 5. **Select attributes**. Select the most relevant attributes in the data.
- 6. Visualize. View an interactive 2D plot of the data.

#### 1.2 Status Box

The status box appears at the very bottom of the window. It displays messages that keep you informed about what's going on. For example, if the Explorer is busy loading a file, the status box will say that.

- **Memory information**. Display in the log box the amount of memory available to WEKA.
- **Run garbage collector**. Force the Java garbage collector to search for memory that is no longer needed and free it up, allowing more memory for new tasks. Note that the garbage collector is constantly running as a background task anyway.

#### 1.3 Log Button

Clicking on this button brings up a separate window containing a scrollable text field. Each line of text is stamped with the time it was entered into the log. As you perform actions in WEKA, the log keeps a record of what has happened.

#### 1.4 WEKA Status Icon

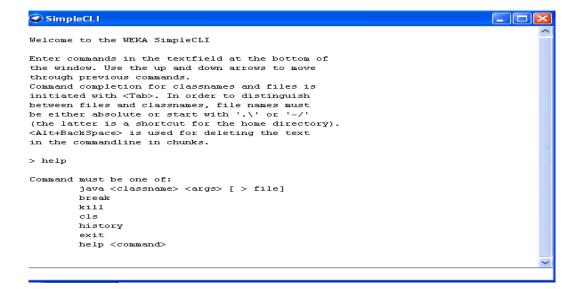
To the right of the status box is the WEKA status icon. When no processes are running, the bird sits down and takes a nap. The number beside the × symbol gives the number of concurrent processes running.

#### 1.5 Graphical output

Most graphical displays in WEKA, e.g., the Graph Visualizer or the Tree Visualizer, support saving the output to a file. A dialog for saving the output can be brought up with Alt+Shift+left-click. Supported formats are currently Windows Bitmap, JPEG, PNG and EPS (encapsulated Postscript). The dialog also allows you to specify the dimensions of the generated image.

## **Simple CLI**

The Simple CLI provides full access to all Weka classes, i.e., classifiers, filters, clusterers, etc., but without the hassle of the CLASSPATH (it facilitates the one, with which Weka was started). It offers a simple Weka shell with separated command line and output.



### Commands

The following commands are available in the Simple CLI:

 java <classname> [<args>] invokes a java class with the given arguments (if any)

- break stops the current thread, e.g., a running classifier, in a friendly manner.
- Kill stops the current thread in an unfriendly fashion
- **cls** clears the output area
- exit exits the Simple CLI
- help [<command>] provides an overview of the available commands if without a command name as argument, otherwise more help on the specified command.

#### Invocation

In order to invoke a Weka class, one has only to prefix the class with "java". This command tells the Simple CLI to load a class and execute it with any given parameters. E.g., the J48 classifier can be invoked on the iris dataset with the following command:

java weka.classifiers.trees.J48 -t c:/temp/iris.arff

```
📥 SimpleCLI
                                                                          _ 🗆 ×
 50
    0 0 | a = Iris-setosa
  0 49 l | b = Iris-versicolor
  0 2 48 | c = Iris-virginica
=== Stratified cross-validation ==
                                    144
                                                        96
Correctly Classified Instances
                                                                ÷
                                     6
Incorrectly Classified Instances
                                                         4
Kappa statistic
                                       0.94
Mean absolute error
                                       0.035
Root mean squared error
                                       0.1586
7.8705 %
Relative absolute error
Root relative squared error
Total Number of Instances
                                     33.6353 %
Total Number of Instances
                                    150
=== Confusion Matrix ===
          <-- classified as
  a b c
 49 1 0 | a = Iris-setosa
 0 47 3 | b = Iris-versicolor
  0 2 48 | c = Iris-virginica
java weka.classifiers.trees.J48 -t ./data/iris.arff
```

#### **Command redirection**

Starting with this version of Weka one can perform a basic redirection:

java weka.classifiers.trees.J48 -t test.arff > j48.txt

Note: the > must be preceded and followed by a space, otherwise it is not recognized as redirection, but part of another parameter.

#### <u>WEEK 2:</u>

- iv) Study the arff file format
- v) Explore the available data sets in WEKA.
- vi) Load a data set (ex. Weather dataset, Iris dataset, etc.)
- vii) Load each dataset and observe the following:

- a) List the attribute names and they types
- b) Number of records in each dataset
- c) Identify the class attribute (if any)
- d) Plot histogram
- e) Determine the number of records for each class.
- f) Visualize the data in various dimensions
- iv) Study the arff file format

#### **Attribute-Relation File Format (ARFF)**

An ARFF (Attribute-Relation File Format) file is an ASCII text file that describes a list of instances sharing a set of attributes. ARFF files were developed by the Machine Learning Project at the Department of Computer Science of The University of Waikato for use with the Weka machine learning software. This document describes the version of ARFF used with Weka versions 3.2 to 3.3; this is an extension of the ARFF format as described in the data mining book written by Ian H. Witten and Eibe Frank (the new additions are string attributes, date attributes, and sparse instances).

This explanation was cobbled together by Gordon Paynter (gordon.paynter at ucr.edu) from the Weka 2.1 ARFF description, email from Len Trigg (lenbok at myrealbox.com) and Eibe Frank (eibe at cs.waikato.ac.nz), and some datasets. It has been edited by Richard Kirkby (rkirkby at cs.waikato.ac.nz). Contact Len if you're interested in seeing the ARFF 3 proposal.

#### Overview

ARFF files have two distinct sections. The first section is the **Header** information, which is followed the **Data** information.

The **Header** of the ARFF file contains the name of the relation, a list of the attributes (the columns in the data), and their types. An example header on the standard IRIS dataset looks like this:

% 1. Title: Iris Plants Database
%
% 2. Sources:
% (a) Creator: R.A. Fisher
% (b) Donor: Michael Marshall (MARSHALL%PLU@io.arc.nasa.gov)
% (c) Date: July, 1988
%

```
@RELATION iris
```

@ATTRIBUTE sepallength NUMERIC
@ATTRIBUTE sepalwidth NUMERIC
@ATTRIBUTE petallength NUMERIC
@ATTRIBUTE petalwidth NUMERIC
@ATTRIBUTE class {Iris-setosa,Iris-versicolor,Iris-virginica}

The **Data** of the ARFF file looks like the following:

@DATA
5.1,3.5,1.4,0.2,Iris-setosa
4.9,3.0,1.4,0.2,Iris-setosa
4.7,3.2,1.3,0.2,Iris-setosa
4.6,3.1,1.5,0.2,Iris-setosa
5.0,3.6,1.4,0.2,Iris-setosa
5.4,3.9,1.7,0.4,Iris-setosa
4.6,3.4,1.4,0.3,Iris-setosa
4.6,3.4,1.5,0.2,Iris-setosa
4.4,2.9,1.4,0.2,Iris-setosa
4.9,3.1,1.5,0.1,Iris-setosa
Lines that begin with a % are comments. The @RELATION, @ATTRIBUTE and @DATA
declarations are case insensitive.

#### Examples

Several well-known machine learning datasets are distributed with Weka in the \$WEKAHOME/data directory as ARFF files.

#### The ARFF Header Section

The ARFF Header section of the file contains the relation declaration and attribute declarations.

#### The @relation Declaration

The relation name is defined as the first line in the ARFF file. The format is:

@relation <relation-name>

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

#### The @attribute Declarations

Attribute declarations take the form of an orderd sequence of **@attribute** statements. Each attribute in the data set has its own **@attribute** statement which uniquely defines the name of that attribute and it's data type. The order the attributes are declared indicates the column position in the data section of the file. For example, if an attribute is the third one declared then Weka expects that all that attributes values will be found in the third comma delimited column.

The format for the @attribute statement is:

@attribute <attribute-name> <datatype>

where the *<attribute-name>* must start with an alphabetic character. If spaces are to be included in the name then the entire name must be quoted.

The *<datatype>* can be any of the four types currently (version 3.2.1) supported by Weka:

- numeric
- <nominal-specification>
- string
- date [<date-format>]

where <nominal-specification> and <date-format> are defined below. The keywords **numeric**, **string** and **date** are case insensitive.

#### Numeric attributes

Numeric attributes can be real or integer numbers.

#### Nominal attributes

Nominal values are defined by providing an <nominal-specification> listing the possible values: {<nominal-name1>, <nominal-name2>, <nominal-name3>, ...}

For example, the class value of the Iris dataset can be defined as follows:

@ATTRIBUTE class {Iris-setosa,Iris-versicolor,Iris-virginica}

Values that contain spaces must be quoted.

#### String attributes

String attributes allow us to create attributes containing arbitrary textual values. This is very useful in text-mining applications, as we can create datasets with string attributes, then write

Weka Filters to manipulate strings (like StringToWordVectorFilter). String attributes are declared as follows:

@ATTRIBUTE LCC string

#### **Date attributes**

Date attribute declarations take the form:

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

where <name> is the name for the attribute and <date-format> is an optional string specifying how date values should be parsed and printed (this is the same format used by SimpleDateFormat). The default format string accepts the ISO-8601 combined date and time format: "yyyy-MM-dd'T'HH:mm:ss".

Dates must be specified in the data section as the corresponding string representations of the date/time (see example below).

#### **ARFF Data Section**

The ARFF Data section of the file contains the data declaration line and the actual instance lines.

#### The @data Declaration

The **@data** declaration is a single line denoting the start of the data segment in the file. The format is:

@data

#### The instance data

Each instance is represented on a single line, with carriage returns denoting the end of the instance.

Attribute values for each instance are delimited by commas. They must appear in the order that they were declared in the header section (i.e. the data corresponding to the nth @attribute declaration is always the nth field of the attribute).

Missing values are represented by a single question mark, as in:

GEC

@data 4.4,?,1.5,?,Iris-setosa

Values of string and nominal attributes are case sensitive, and any that contain space must be quoted, as follows:

@relation LCCvsLCSH

@attribute LCC string@attribute LCSH string

@data
AG5, 'Encyclopedias and dictionaries.;Twentieth century.'
AS262, 'Science -- Soviet Union -- History.'
AE5, 'Encyclopedias and dictionaries.'
AS281, 'Astronomy, Assyro-Babylonian.;Moon -- Phases.'
AS281, 'Astronomy, Assyro-Babylonian.;Moon -- Tables.'

Dates must be specified in the data section using the string representation specified in the attribute declaration. For example:

@RELATION Timestamps

@ATTRIBUTE timestamp DATE "yyyy-MM-dd HH:mm:ss"

@DATA "2001-04-03 12:12:12" "2001-05-03 12:59:55"

#### **Sparse ARFF files**

Sparse ARFF files are very similar to ARFF files, but data with value 0 are not be explicitly represented.

Sparse ARFF files have the same header (i.e @**relation** and @**attribute** tags) but the data section is different. Instead of representing each value in order, like this:

@data 0, X, 0, Y, "class A"

GEC

0, 0, W, 0, "class B"

the non-zero attributes are explicitly identified by attribute number and their value stated, like this:

@data {1 X, 3 Y, 4 "class A"} {2 W, 4 "class B"}

Each instance is surrounded by curly braces, and the format for each entry is: <index> <space> <value> where index is the attribute index (starting from 0).

Note that the omitted values in a sparse instance are **0**, they are not "missing" values! If a value is unknown, you must explicitly represent it with a question mark (?).

**Warning**: There is a known problem saving SparseInstance objects from datasets that have string attributes. In Weka, string and nominal data values are stored as numbers; these numbers act as indexes into an array of possible attribute values (this is very efficient). However, the first string value is assigned index 0: this means that, internally, this value is stored as a 0. When a SparseInstance is written, string instances with internal value 0 are not output, so their string value is lost (and when the arff file is read again, the default value 0 is the index of a different string value, so the attribute value appears to change). To get around this problem, add a dummy string value at index 0 that is never used whenever you declare string attributes that are likely to be used in SparseInstance objects and saved as Sparse ARFF files.

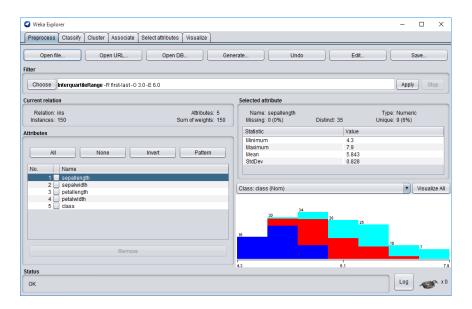
v) Explore the available data sets in WEKA.

🕝 Open			×
Look <u>I</u> n: ( 📄 d	lata	•	
<ul> <li>airline.arff</li> <li>breast-can</li> <li>contact-len</li> <li>cpu.arff</li> <li>cpu.with.ve</li> <li>credit-g.arf</li> <li>diabetes.a</li> <li>glass.arff</li> <li>hypothyroid</li> <li>ionospher</li> <li>iris.2D.arff</li> <li>labor.arff</li> </ul>	dicer.arff hses.arff endor.arff ff d.arff e.arff	<ul> <li>ReutersCorn-test.arff</li> <li>ReutersCorn-train.arff</li> <li>ReutersGrain-test.arff</li> <li>ReutersGrain-train.arff</li> <li>segment-challenge.arff</li> <li>segment-test.arff</li> <li>soybean.arff</li> <li>supermarket.arff</li> <li>unbalanced.arff</li> <li>vote.arff</li> <li>weather.nominal.arff</li> <li>weather.numeric.arff</li> </ul>	Invoke options dialog Note: Some file formats offer additional options which can be customized when invoking the options dialog.
File <u>N</u> ame:	C:\Program F	Files\Weka-3-8\data	
Files of <u>T</u> ype:	All Files		•
			Open Cancel

vi) Load a data set (ex. Weather dataset, Iris dataset, etc.)

Preprocess Classify Cluster Associate Select attributes Visualize			
Open file Open L Discover association rules 3 Gen	erate Undo	Ed	it Save
ilter			
Choose InterquartileRange -R first-last -O 3.0 -E 6.0			Apply Stop
urrent relation	Selected attribute		
Relation: weather.symbolic         Attributes: 5           Instances: 14         Sum of weights: 14	Name: outlook Missing: 0 (0%)	Distinct: 3	Type: Nominal Unique: 0 (0%)
ttributes	No. Label	Count	Weight
All None Invert Pattern	1 sunny 2 overcast 3 rainy	5 4 5	5.0 4.0 5.0
No. Name			
2 temperature 3 humidity 4 windy	Class: play (Nom)		Visualize A
5 🔲 play	5		5
		4	
Remove			

0	Viewer					$\times$
elat	tion: weath	er.symbolic				
No.	1: outlook Nominal	2: temperature Nominal	3: humidity Nominal	4: windy Nominal		
1	sunny	hot	high	FALSE	no	
2	sunny	hot	high	TRUE	no	
3	overcast	hot	high	FALSE	yes	
4	rainy	mild	high	FALSE	yes	
5	rainy	cool	normal	FALSE	yes	
6	rainy	cool	normal	TRUE	no	
7	overcast	cool	normal	TRUE	yes	
8	sunny	mild	high	FALSE	no	
9	sunny	cool	normal	FALSE	yes	
10	rainy	mild	normal	FALSE	yes	
11	sunny	mild	normal	TRUE	yes	
12	overcast	mild	high	TRUE	yes	
13	overcast	hot	normal	FALSE	yes	
14	rainy	mild	high	TRUE	no	
		dd instance	Undo	ок	Cance	×1



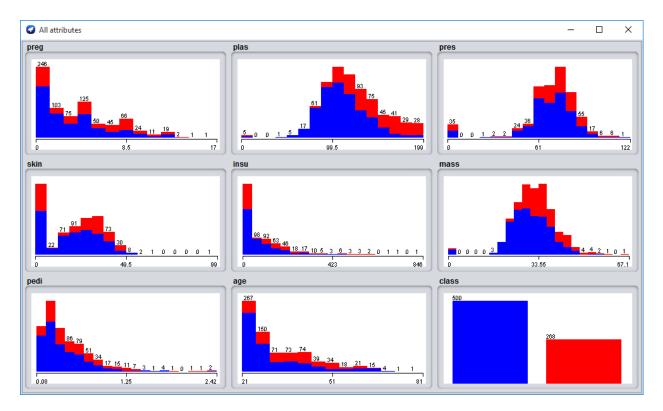
	on: iris					
JO. 1		2: sepalwidth 3:				
	Numeric	Numeric	Numeric	Numeric	Nominal	
1	5.1	3.5	1.4	0.2	Iris-s	L .
2	4.9	3.0	1.4	0.2	Iris-s	
3	4.7	3.2	1.3	0.2	Iris-s	
1	4.6	3.1	1.5	0.2	Iris-s	
5	5.0	3.6	1.4	0.2	Iris-s	
5	5.4	3.9	1.7	0.4	Iris-s	
~	4.6	3.4	1.4	0.3	Iris-s	
3	5.0	3.4	1.5	0.2	Iris-s	
•	4.4	2.9	1.4	0.2	Iris-s	
10	4.9	3.1	1.5	0.1	Iris-s	
11	5.4	3.7	1.5	0.2	Iris-s	
12	4.8	3.4	1.6	0.2	Iris-s	
13	4.8	3.0	1.4	0.1	Iris-s	
14	4.3	3.0	1.1	0.1	Iris-s	
15	5.8	4.0	1.2	0.2	Iris-s	
16	5.7	4.4	1.5	0.4	Iris-s	
17	5.4	3.9	1.3	0.4	Iris-s	
18	5.1	3.5	1.4	0.3	Iris-s	
19	5.7	3.8	1.7	0.3	Iris-s	
20	5.1	3.8	1.5	0.3	Iris-s	
21	5.4	3.4	1.7	0.2	Iris-s	
22	5.1	3.7	1.5	0.4	Iris-s	
23	4.6	3.6	1.0	0.2	Iris-s	
24	5.1	3.3	1.7	0.5	Iris-s	
25	4.8	3.4	1.9	0.2	Iris-s	
26	5.0	3.0	1.6	0.2	Iris-s	
27	5.0	3.4	1.6	0.4	Iris-s	
28	5.2	3.5	1.5	0.2	Iris-s	
29	5.2	3.4	1.4	0.2	Iris-s	
30	4.7	3.2	1.6	0.2	Iris-s	
3 1	4.8	3.1	1.6	0.2	Iris-s	
32	5.4	3.4	1.5	0.4	Iris-s	
33	5.2	4.1	1.5	0.1	Iris-s	
34	5.5	4.2	1.4	0.2	Iris-s	
35	4.9	3.1	1.5	0.1	Iris-s	
36	5.0	3.2	1.2	0.2	Iris-s	
37	5.5	3.5	1.3	0.2	Iris-s	

27

- vii) Load each dataset and observe the following:
  - a) List the attribute names and they types
  - b) Number of records in each dataset
  - c) Identify the class attribute (if any)
  - d) Plot histogram
  - e) Determine the number of records for each class.
  - f) Visualize the data in various dimensions

🖉 Weka Explorer		– 🗆 X
Preprocess Classify Cluster Associate Select attributes Visualize		
Open file Open URL Open DB Gene	erate Undo	Edit Save
Choose None		Apply Stop
Current relation	Selected attribute	
Relation: weather.symbolic Attributes: 5 Instances: 14 Sum of weights: 14	Name: outlook Missing: 0 (0%) Dist	Type: Nominal inct: 3 Unique: 0 (0%)
Attributes	No. Label	Count Weight
All     None     Invert     Pattern       No.     Name       1     outlook       2     temperature       3     humidity       4     windy       5     play	1 sunny 2 overcast 3 rainy Class: play (Nom)	5 5.0 4 4.0 5 5.0 Visualize All
Remove	5	
OK		
🗄 Search the web and Windows 🔲 🤤	📄 🌞 🖨 🍅 📄	- 💿 🖳 🥫 🕢

# d) Plot histogram:



e) Determine the number of records for each class.

🥥 Weka Explorer					- 🗆 X
Preprocess Classify Cluster Associate Select attrit	outes Visualize				
Open file Open URL Open URL	Gene	erate	Und	o Edit.	. Save
Choose None					Apply Stop
Current relation		Selected at	tribute		
Relation: weather.symbolic Instances: 14	Attributes: 5 Sum of weights: 14	Name: Missing:		Distinct: 2	Type: Nominal Unique: 0 (0%)
Attributes		No.	Label	Count	Weight
All None Invert	Pattern		yes no	9 5	9.0 5.0
1outlook 2temperature 3humidity 4windy		Class: play	(Nom)		Visualize All
5 ♥ play		9		5	
Remove					
OK					Log 📣 x0

#### f) Visualize the data in various dimensions

🚱 Weka Explorer	- D X
Preprocess Classify Cluster Associate Select attributes Visualize	
Open IIIe Open URL Open DB Gene	erate Undo Edit Save
Choose None	Apply Stop
Current relation	Selected attribute
Relation: pima_diabetes     Attributes: 9       Instances: 768     Sum of weights: 768	Name: insu Type: Numeric Missing: 0 (0%) Distinct: 186 Unique: 93 (12%)
Attributes	Statistic Value
All     None     Invert     Pattern       No.     Name       1     preg       2     plas       3     pres       4     Skin       5     Insu       6     mass       7     pedi       8     age       9     class	Minimum         0           Maximum         846           Mean         79,799           StdDev         115.244             Class: insu (Num)         Visualize All
Remove Status	$\begin{array}{ c c c c c c c c c c c c c c c c c c c$
ок	Log 🛷 x0

#### **VIVA-QUESTIONS**

- 1. WEKA Stands for
- 2. ARFF Stands for
- 3. CSV Stands for
- 4. Present Version of WEKA Software?
- 5. Which of the following is not a DATA Mining Software ?

A. WEKAB. SPSSC. OrangeD. Cogno's6. Which of the following is not a Data Type in Weka ?

- A. Numeric B. String C. Date D. Real
- 7. Which of the following is not a support files in Weka ? A. ARFF B. CSV C..cpp
- 8. The full form of KDD is
- 9. Data Mining helps in

10. Extreme values that occur infrequently are called as

### EXP2: Perform data preprocessing tasks on

i.Add attribute ii. Add expression iii. Copy attribute iv. Remove attribute

#### **Filters**

Step-1: - Go to start button then select All Programs and then select weka 3.8.3



#### Click on Explorer

🚖 Weka Explorer			
Preprocess Classify Cluster Associate Select attributes Visualize			
Open file Open URL Open DB	Undo	Edit	Save
Filter			
Choose Add -N unnamed -C last			Apply
Current relation	Selected attribute		
Relation: STUDENT	Name: SNO		Type: Numeric
Instances: 5 Attributes: 3	Missing: 0 (0%)	Distinct: 5	Unique: 5 (100%)
Attributes	Statistic	Value	
	Minimum	1	
All None Invert	Maximum	5	
	Mean	3	
No. Name	StdDev	1.581	
1 SNO			
2 SNAME			
3 SAL			
	Class: SAL (Num)		Visualize All
	5		
Remove			
	1	3	5
Status			
ок			Log 💉 × O

31

🗁 Open							$\mathbf{X}$
Look in:	C WEKA			*	၈ ⊳[		
My Recent Documents Desktop My Documents	<ul> <li>emp.arff</li> <li>student1.a</li> <li>student2.a</li> <li>student3.a</li> <li>student4.a</li> </ul>	arff					
My Computer My Network Places	File <u>n</u> ame: Files of <u>typ</u> e:	stu.arff Arff data files	 		~	_	Open

#### ADD ATTRIBUTE

### Step-2: - Go to preprocess menu and choose then select attribute of Add option.

👻 Weka Explorer			- D ×
Preprocess Classify Cluster Associate Select attributes Visualize			
Open file Open URL Open DB	Undo	Edit	Save
Filter Choose Add -N unnamed -C last			Apply
Current relation Relation: STUDENT Instances: 5 Attributes: 3	Selected attribute Name: SNO Missing: 0 (0%)	Distinct: 5	Type: Numeric Unique: 5 (100%)
Attributes	Statistic Minimum	Value	
All None Invert	Maximum Mean	5	
No.     Name       1     SNO       2     SNAME       3     SAL	Class: SAL (Num)	1.581	Visualize All
Remove Status OK	1	3	6 Log x0

<
$\sim$
111

Step-3: -To enter the required fields is added to table.

🚔 weka.gui.GenericObjectEditor	
weka.filters.unsupervised.attribute.Add	
About	
An instance filter that adds a new attribute to the dataset.	More
attributeIndex first	
attributeName regno	
nominalLabels	
Open Save OK	Cancel

#### Click on ok button

👙 Weka Explorer	
Preprocess Classify Cluster Associate Select attributes Visualize	
Open file Open URL Open DB	Undo Edit Save
Filter Choose Add -N regno -C first	Apply
Current relation Relation: STUDENT-weka.filters.unsupervised.attribute.Add-Nregno Instances: 5 Attributes: 4	Selected attribute Name: regno Type: Numeric Missing: 5 (100%) Distinct: 0 Unique: 0 (0%)
Attributes	Statistic Value
	Minimum NaN
All None Invert	Maximum NaN
	Mean NaN
No. Name	StdDev NaN
1 regno 2 5NO 3 5NAME	
4 SAL	Class: SAL (Num) Visualize All
Remove	<u>0</u>
	NaN NaN Na
Status OK	Log ×1

Click on Apply button.

	iewer				X			
Relation: STUDENT-weka.filters.unsupervised.attribute.Add-Nregno-Cfirst								
No.	regno	SNO	SNAME	SAL				
	Numeric	Numeric	String	Numeric				
1			DEEPTHI	1000.0				
2			SIRI	2000.0				
з			RAJI	4000.0				
4			LAKS	3000.0				
5		5.0	PRAS	5000.0				
				D in	ght click (or left+alt) for context menu			
					gric click (or lerc+alc) for concext mena			
					Undo OK Cancel			

Step-4: - if the index position is last to enter the fields

🛓 weka.gui.G	enericObjectEditor
weka.filters.unsup	pervised.attribute.Add
About	
An instance f	Iter that adds a new attribute to the dataset. More
attributeIndex	last
attributeName	avg
nominalLabels	Set the new attribute's nam
Open	Save OK Cancel

## Click on Ok button.

😩 Weka Explorer			- D X						
Preprocess Classify Cluster Associate Select attributes Visualize									
Open file Open URL Open DB	Undo	Edit	Save						
Filter									
Choose Add -N avg -C last			Apply						
Current relation	Selected attribute								
Relation: STUDENT-weka.filters.unsupervised.attribute.Add-Nregno Instances: 5 Attributes: 5	Name: regno Missing: 5(100%)	Distinct: 0	Type: Numeric Unique: 0 (0%)						
Attributes	Statistic	Value							
	Minimum	NaN							
All None Invert	Maximum	NaN							
	Mean	NaN							
No. Name	StdDev	NaN							
1 regno									
2 5NO 3 SNAME									
3 SNAME 4 SAL									
5 avg	Class: avg (Num)		Visualize All						
Remove	0								
	NaN	NaN	NaN						
Status									
ок			Log 💉 ×0						

Click on Apply button.

Viewer Relation: STUDENT-weka.filters.unsupervised.attribute.Add-Nregno-Cfirst-weka.filters.uns								
No.	regno Numeric	SNO Numeric	SNAME String	SAL Numeric	<b>avg</b> Numeric			
1		1.0	DEEPTHI	1000.0				
2		2.0	SIRI	2000.0				
3		3.0	RAJI	4000.0				
4		4.0	LAKS	3000.0				
5		5.0	PRAS	5000.0				
						Undo OK Cancel		

Click on ok button.

Step-5: -if the index position is middle

🛎 weka.gui.G	enericObjectEditor	
weka.filters.unsu	pervised.attribute.Add	
About		
An instance f	ilter that adds a new attribute to the dataset.	More
attributeIndex	4	
attributeName	max sal	
nominalLabels		
Open	Save OK	Cancel

Click on ok button.

👙 Weka Explorer	
Preprocess Classify Cluster Associate Select attributes Visualize	
Open file Open URL Open DB	Undo Edit Save
Filter	
Choose Add -N "max sal" -C 4	Apply
Current relation	Selected attribute
Relation: STUDENT-weka.filters.unsupervised.attribute.Add-Navg-Cla Instances: 5 Attributes: 6	Name: regno Type: Numeric Missing: 5 (100%) Distinct: 0 Unique: 0 (0%)
Attributes	Statistic Value
	Minimum NaN
All None Invert	Maximum NaN
	Mean NaN
No. Name Invert	rts the current attribute selection NaN
1 regno	
2SNO	
3 SNAME	
4 max sal'	Class: avg (Num) Visualize All
5 SAL	
6 avg	
Remove	
	0 
	NaN NaN NaN
OK OK	Log ×0

Click on Apply button.

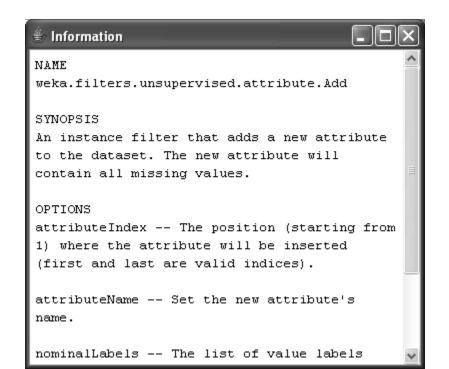
🐇 Viewer 🔀								
Relation: STUDENT-weka.filters.unsupervised.attribute.Add-Navg-Clast-weka.filters.unsup								
No.	regno Numeric	SNO Numeric	SNAME String	'max sal' Numeric	SAL Numeric	<b>avg</b> Numeric		
1		1.0	DEEPTHI		1000.0			
2		2.0	SIRI		2000.0			
3		3.0	RAJI		4000.0			
4		4.0	LAKS		3000.0			
5		5.0	PRAS		5000.0			
						Undo	OK Cancel	

Click on ok button.

### **Add Expression**

Step-6: - click on choose button and then select AddExpression option.

👻 Weka Explorer									
Preprocess Classify Cluster Associate Select attributes Visualize									
Open file Open URL Open DB	Undo	Edit	Save						
Filter									
Choose AddExpression -E a1^2 -N expression			Apply						
Current relation	Selected attribute								
Relation: STUDENT-weka.filters.unsupervised.attribute.Add-Navg-Cla Instances: 5 Attributes: 6	Name: regno Missing: 0(0%)	Distinct: 5 L	Type: Numeric Inique: 5 (100%)						
Attributes	Statistic	Value							
	Minimum	12							
All None Invert	Maximum	16							
	Mean 14								
No. Name	StdDev	1.581							
1 regno 2 SNO 3 SNAME									
4 [] 'max sal' 5 [] SAL	Class: avg (Num)		Visualize All						
6 avg	5								
	<u>°</u>								
Remove									
	12	14	16						
Status OK			Log ×0						



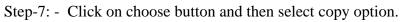
👻 Weka Explorer	
Preprocess Classify Cluster Associate Select attributes Visualize	
Open file Open URL Open DB	Undo Edit Save
Filter Choose AddExpression -E a4+a5 -N addtion	Apply
Current relation Relation: STUDENT-weka.filters.unsupervised.attribute.Add-Navg-Cla Instances: S Attributes: 7	Selected attribute Name: regno Type: Numeric Missing: 0 (0%) Distinct: 5 Unique: 5 (100%)
Attributes	Statistic Value
All None Invert	Maximum 16 Mean 14
No.         Name           1         regno           2         SNO           3         SNAME	StdDev 1.581
4 [_'max sal' 5SAL 6avg 7addtion	Class: addtion (Num) Visualize All
	<b>o</b>
Remove	12 14 16
Status OK	

Click on Apply option.

14.0         3.0         RAJI         6000.0         4000.0         13.3         10000.0           15.0         4.0         LAK5         6000.0         3000.0         14.2         9000.0
14.0         3.0         RAJI         6000.0         4000.0         13.3         10000.0           15.0         4.0         LAK5         6000.0         3000.0         14.2         9000.0
15.0 4.0 LAK5 6000.0 3000.0 14.2 9000.0
5 16.0 5.0 PRAS 10000.0 5000.0 15000.0

🚔 Weka Explorer	
Preprocess Classify Cluster Associate Select attributes Visualize	
Open file Open URL Open DB	Undo Edit Save
Filter Choose AddExpression -E a4-a5 -N subtraction	Apply
Current relation	Selected attribute
Relation: STUDENT-weka.filters.unsupervised.attribute.Add-Navg-Cla Instances: 5 Attributes: 8	Name: regno Type: Numeric Missing: 0 (0%) Distinct: 5 Unique: 5 (100%)
Attributes	Statistic Value
	Minimum 12
All None Invert	Maximum 16
	Mean 14 StdDev 1.581
No. Name	Studev 11.561
1 regno 2 SNO	
3 SNAME	
4 []'max sal'	Class: subtraction (Num) Visualize All
5 SAL	
6 avg	
7 addtion 8 subtraction	5
8 subtraction	
Remove	
<u></u>	12 14 16
OK	Log ×0

## <u>COPY</u>



eprocess Classify Cluster Associate Select attributes Visualize			الكرك
Open file Open URL Open DB		Edit	Save
		Laterri	
er			
Choose Copy			Apply
rrent relation	Selected attribute		
Relation: STUDENT-weka.filters.unsupervised.attribute.Add-Navg-Cl. instances: 5 Attributes: 8			Type: Numeric
	Missing: 0 (0%)	Distinct: 5	Unique: 5 (100%)
tributes	Statistic Minimum	Value	
All None Invert	Maximum	16	
	Mean	14	
lo. Name	StdDev	1.581	
1 🔤 regno 2 🔽 SNO			
3 SNAME			
4 max sal	Class: subtraction (Num)		Visualize All
5 5AL 6 avg			
7 addtion	5		
8 subtraction			
Remove			
Kellüve		1	
	12	14	
itus		14	
		14	Log
atus <		17	Log 🛷
<			
👻 Information			
Information conjunction with other	r filters that o	verwrite	
Information conjunction with othe attribute values duri	r filters that o ng the course of	verwrite their	
Information conjunction with othe attribute values duri operation this fil	r filters that o ng the course of ter allows the o	verwrite their riginal	
Information conjunction with othe attribute values duri	r filters that o ng the course of ter allows the o	verwrite their riginal	
Information conjunction with othe attribute values duri operation this fil	r filters that o ng the course of ter allows the o	verwrite their riginal	
Information conjunction with othe attribute values duri operation this fil attributes to be kept	r filters that o ng the course of ter allows the o	verwrite their riginal	
Information conjunction with othe attribute values duri operation this fil attributes to be kept	r filters that o ng the course of ter allows the o	verwrite their riginal	
Information conjunction with othe attribute values duri operation this fil attributes to be kept attributes. OPTIONS	r filters that o ng the course of ter allows the o as well as the a	verwrite their riginal	
Information conjunction with othe attribute values duri operation this fil attributes to be kept attributes. OPTIONS attributeIndices S	r filters that o ng the course of ter allows the o as well as the s pecify range of	verwrite their riginal new	
Information conjunction with othe attribute values duri operation this fil attributes to be kept attributes. OPTIONS attributeIndices S attributes to act on.	r filters that o ng the course of ter allows the o as well as the pecify range of This is a comma	verwrite their riginal new	
Information conjunction with othe attribute values duri operation this fil attributes to be kept attributes. OPTIONS attributeIndices S attributes to act on. separated list of att	r filters that o ng the course of ter allows the o as well as the pecify range of This is a comma ribute indices,	verwrite their riginal new	
Information conjunction with othe attribute values duri operation this fil attributes to be kept attributes. OPTIONS attributeIndices S attributes to act on.	r filters that o ng the course of ter allows the o as well as the pecify range of This is a comma ribute indices,	verwrite their riginal new	
Information conjunction with othe attribute values duri operation this fil attributes to be kept attributes. OPTIONS attributeIndices S attributes to act on. separated list of att	r filters that o ng the course of ter allows the o as well as the pecify range of This is a comma ribute indices, lid values. Spec	verwrite their riginal new	
Information conjunction with othe attribute values duri operation this fil attributes to be kept attributes. OPTIONS attributeIndices S attributes to act on. separated list of att "first" and "last" va inclusive range with	r filters that o ng the course of ter allows the o as well as the pecify range of This is a comma ribute indices, lid values. Spec "-". E.g:	verwrite their riginal new	
Information conjunction with othe attribute values duri operation this fil attributes to be kept attributes. OPTIONS attributeIndices S attributes to act on. separated list of att "first" and "last" va	r filters that o ng the course of ter allows the o as well as the pecify range of This is a comma ribute indices, lid values. Spec "-". E.g:	verwrite their riginal new	
Information conjunction with othe attribute values duri operation this fil attributes to be kept attributes. OPTIONS attributeIndices S attributes to act on. separated list of att "first" and "last" va inclusive range with "first-3,5,6-10,last"	r filters that o ng the course of ter allows the o as well as the pecify range of This is a comma ribute indices, lid values. Spec "-". E.g:	verwrite their riginal new with	
<pre>Information Conjunction with othe attribute values duri operation this fil attributes to be kept attributes. OPTIONS attributeIndices S attributes to act on. separated list of att "first" and "last" va inclusive range with "first-3,5,6-10,last" invertSelection Se</pre>	r filters that o ng the course of ter allows the o as well as the pecify range of This is a comma ribute indices, lid values. Spec "-". E.g: ts copy selected	verwrite their riginal new with ify an	
Information conjunction with othe attribute values duri operation this fil attributes to be kept attributes. OPTIONS attributeIndices S attributes to act on. separated list of att "first" and "last" va inclusive range with "first-3,5,6-10,last"	r filters that o ng the course of ter allows the o as well as the pecify range of This is a comma ribute indices, lid values. Spec "-". E.g: ts copy selected	verwrite their riginal new with ify an	

# If the index place is any

🖆 weka.gui.GenericObjectEditor	- OX
weka.filters.unsupervised.attribute.Copy	
About	
An instance filter that copies a range of attributes in the dataset.	More
attributeIndices first	
invertSelection False	*
Open Save OK	Cancel

## Click on ok button.

👙 Weka Explorer	
Preprocess Classify Cluster Associate Select attributes Visualize	
Open file Open URL Open DB	Undo Edit Save
Choose Copy -R first	Apply
Current relation Relation: STUDENT-weka.filters.unsupervised.attribute.Add-Navg-Cl Instances: 5	Selected attribute Name: regno Type: Numeric Missing: 0 (0%) Distinct: 5 Unique: 5 (100%)
Attributes	Statistic Value
	Minimum 12
All None Invert	Maximum 16
No. Name	Mean 14 StdDev 1.581
1 = regno 2 = SNO 3 = SNAME 4 = "max sal"	Class: Copy of regno (Num)
S         SAL           6         avg           7         addtion           8         subtraction	
9 Copy of regno	
Remove	
	12 14 10
Status OK	Log ×0

## Click On ok button.

	🗄 Viewe	r						$\mathbf{\times}$
Re	elation: ST	UDENT-W	eka.filters	unsupervi	ised.attrib	ute.Add-N	lavg-Clast-we	ka.filters.unsup
	SNO	SNAME	'max sal'	SAL	avg	addtion	subtraction	Copy of regno
	Numeric	String	Numeric	Numeric	Numeric	Numeric	Numeric	Numeric
ο		DEEPTHI	2000.0	1000.0	10.2	3000.0	1000.0	12.0
0		SIRI	4000.0	2000.0	12.2	6000.0	2000.0	13.0
Ο		RAJI	6000.0	4000.0	13.3	10000.0	2000.0	14.0
Ο		LAKS	6000.0	3000.0	14.2	9000.0	3000.0	15.0
Ο	5.0	PRAS	10000.0	5000.0		15000.0	5000.0	16.0
	•							>

Click on ok button.

If the last position.

👻 weka.gui.GenericObjectEditor	
weka.filters.unsupervised.attribute.Copy	
About	
An instance filter that copies a range of attributes in dataset.	the More
attributeIndices last	
invertSelection False	~
Open Save OK	Cancel

🔹 Weka Explorer	
Preprocess Classify Cluster Associate Select attributes Visualize	
Open file Open URL Open DB	Undo Edit Save
Filter Choose Copy -R last	Apply
Current relation Relation: STUDENT-weka.filters.unsupervised.attribute.Add-Navg-Cla Instances: 5 Attributes: 10	Selected attribute Name: regno Type: Numeric Missing: 0 (0%) Distinct: 5 Unique: 5 (100%)
Attributes	Statistic Value Minimum 12
All None Invert	Maximum         16           Mean         14
No. Name I regno SNO SNO SNAME	StdDev 1.581
4 ['max sal' 5 [5AL 6 [6] 7 [addtion	Class: Copy of Copy of regno (Num) Visualize All
8 subtraction 9 Copy of regno 10 Copy of Copy of regno	
Remove	12 14 16
Status OK	

Click on Apply option.

🛎 Vi	ewer							$\mathbf{X}$
Relation	h: STUDE	NT-weka.fil	ters.unsu	pervised.a	attribute.A	dd-Navg-Clas	t-weka.filters.un	sup
SNO Jmeric	SNAME	'max sal' Numeric	SAL Numeric	avg Numeric	addtion Numeric	subtraction Numeric	Copy of regno Numeric	Сору
1.0	DEEPTHI	2000.0	1000.0	10.2	3000.0	1000.0	12.0	
2.0	SIRI	4000.0	2000.0	12.2	6000.0	2000.0	13.0	
3.0	RAJI	6000.0	4000.0	13.3	10000.0	2000.0	14.0	
4.0	LAKS	6000.0	3000.0	14.2	9000.0	3000.0	15.0	
5.0	PRAS	10000.0	5000.0		15000.0	5000.0	16.0	
<						Undo		>

If the index position is specified.

👻 weka.gui.GenericObjectEditor	- D ×
weka.filters.unsupervised.attribute.Copy	
About	
An instance filter that copies a range of attributes in the dataset.	More
attributeIndices 5	
invertSelection False	~
Open Save OK	Cancel

Click on ok button.

🔮 Weka Explorer	
Preprocess Classify Cluster Associate Select attributes Visualize	
Open file Open URL Open DB	Undo Edit Save
Filter Choose Copy -R 5	Apply
Current relation Relation: STUDENT-weka.filters.unsupervised.attribute.Add-Navg-Cla Instances: 5 Attributes: 11	Selected attribute Name: regno Type: Numeric Missing: 0 (0%) Distinct: 5 Unique: 5 (100%)
Attributes	Statistic Value Minimum 12
All None Invert	Maximum 16 Mean 14
No.     Name       1     regno       2     SNO       3     SNAME       4     'max sal'       5     SAL       6     avg       7     addition       8     subtraction       9     Copy of regno       10     Copy of regno	StdDev 1.581 Class: Copy of SAL (Num) Visualize All 6
11     Copy of SAL       Remove       Status	
ок	

Click on Apply option.

the V	iewer										$\overline{\mathbf{x}}$
Relatio	on: STUDE	NT-weka.	filters.uns	supervised.	attribute.	Add-Navg	-Clast-wei	a.filters.unsu	pervised.attribu	te.Add-Nregno-0	first-weka.fil
No.	regno Numeric	SNO Numeric	SNAME	'max sal' Numeric	SAL Numeric	avg Numeric	addtion Numeric	subtraction Numeric	Copy of regno Numeric	Copy of Copy of regno	Copy of SAL Numeric
1	12.0	1.0	DEEPTHI	2000.0	1000.0	10.2	3000.0	1000.0	12.0	12.0	1000.0
2	13.0	2.0	SIRI	4000.0	2000.0	12.2	6000.0	2000.0	13.0	13.0	2000.0
3	14.0	3.0	RAJI	6000.0	4000.0	13.3	10000.0	2000.0	14.0	14.0	4000.0
4	15.0	4.0	LAKS	6000.0	3000.0	14.2	9000.0	3000.0	15.0	15.0	3000.0
5	16.0	5.0	PRAS	10000.0	5000.0		15000.0	5000.0	16.0	16.0	5000.0
									Un	do OK	] [ Cancel ]

If the index position range.

🚔 weka.gui.G	enericObjectEditor									
weka.filters.unsup	ervised.attribute.Copy									
About										
An instance fil dataset.	An instance filter that copies a range of attributes in the dataset.									
attributeIndices	5-8									
invertSelection	False Specify ra	ange of attributes								
Open	Save OK	Cancel								

Click on ok button.

👙 Weka Explorer						
Preprocess Classify Cluster Associate Select attributes Visualize						
Open file Open URL Open DB		Undo	Edit	Save		
Choose Copy -R 5-8				Apply		
Current relation Relation: STUDENT-weka.filters.unsupervised.attribute.Add-Navg-Cla Instances: 5 Attributes: 15 Missing: 0 (0%) Distinct: 5 Unique: 5 (100%)						
Attributes		Missing: 0 (0%) Disting	Value	Unique: 5 (100%)		
Attributes		Minimum	12			
All None Invert	ר	Maximum	16			
		Mean	14			
No. Name		StdDev	1.581			
Interview         Interview           6         avg           7         addition           8         subtraction           9         Copy of regno           10         Copy of copy of regno           11         Copy of SAL           12         Copy of SAL		Class: Copy of subtraction (Num)		Visualize All		
	~					
Remove						
		12	14	16		
OK				Log ×0		

	_						
*	Viewer						$\left[\times\right]$
Rela	ation: STU	DENT-web	a.filters.ur	nsupervise	d.attribute.A	dd-Navg-Clast-w	eka.filters.unsup
1E	'max sal'	SAL	avg	addtion	subtraction	Copy of regno	
2	Numeric	Numeric	Numeric	Numeric	Numeric	Numeric	Numeric
ΉI	2000.0	1000.0	10.2	3000.0	1000.0	12.0	
	4000.0	2000.0	12.2	6000.0	2000.0	13.0	
	6000.0	4000.0	13.3	10000.0	2000.0	14.0	
· ·	6000.0	3000.0	14.2	9000.0	3000.0	15.0	
	10000.0	5000.0		15000.0	5000.0	16.0	
<		_		1111			>
						Undo Of	Cancel

## **REMOVE**

Step-8:- Click on choose button and then select Remove option

👙 Weka Explorer							
Preprocess Classify Cluster Associate Select attributes Visualize							
Open IRL Open URL Open DB		Undo	Edit	Save			
Filter							
Choose Remove				Apply			
Current relation		Selected attribute					
Relation: STUDENT-weka.filters.unsupervised.attribute.Add-Navg-Cl Instances: 5 Attributes: 15	a	Name: regno Missing: 0 (0%)	Distinct: 5	Type: Numeric Unique: 5(100%)			
Attributes		Statistic	Valu	Je III			
		Minimum	12				
All None Invert		Maximum	16				
	_	Mean	14				
No. Name		StdDev	1.581	L			
4 max sal	~						
5 SAL							
6 avg							
7 addtion		Class: Copy of subtraction (Num) Visualize All					
8 subtraction							
9 Copy of regno	_						
10 Copy of Copy of regno	=	5					
11 Copy of SAL	-111						
12 Copy of SAL 13 Copy of avg	-111						
14 Copy of addition	-						
15 Copy of subtraction	-						
Remove							
		12	14	16			
Status							
ок				Log 💉 × 0			

🔹 Information	$\times$
NAME	^
weka.filters.unsupervised.attribute.Remove	
SYNOPSIS	
An instance filter that removes a range of	
attributes from the dataset.	
OPTIONS	
attributeIndices Specify range of attributes to act on. This is a comma	
separated list of attribute indices, with	
"first" and "last" valid values. Specify an	
inclusive range with "-". E.g:	
"first-3,5,6-10, last".	
1113C-3,3,0-10,1ast .	
invertSelection Determines whether action	
is to select or delete. If set to true, only	~

If the index position is first.

👻 weka.gui.GenericObjectEditor	- <b>-</b> ×
weka.filters.unsupervised.attribute.Remove	
About	
An instance filter that removes a range of attributes from the dataset.	More
attributeIndices first	
invertSelection False	*
Open Save OK	Cancel

Click on ok button.

🏶 Weka Explorer		
Preprocess Classify Cluster Associate Select attributes Visualize		
Open file Open URL Open DB	Undo Edit	Save
Choose Remove -R first		Apply
Current relation	<ul> <li>Selected attribute</li> </ul>	
Relation: STUDENT-weka.filters.unsupervised.attribute.Add-Navg-Cl Instances: 5 Attributes: 14	Name: SNO Missing: 0 (0%) Distinct: 5	Type: Numeric 5 Unique: 5(100%)
Attributes	Statistic	Value
	Minimum	1
All None Invert		3
No. Name		1.581
1     SNO       2     SNAME       3     'max sal'       4     SAL       5     avg       6     addition       7     subtraction       8     Copy of regno       9     Copy of SAL       112     Copy of sAL       122     Copy of sAL	Class: Copy of subtraction (Num)	Visualize All
Remove		
	1 3	3 5
Status OK		Log ×0

# Click on Apply option.

If the index position is last.

👻 weka.gui.GenericObjectEditor	- D ×							
weka.filters.unsupervised.attribute.Remove								
About								
An instance filter that removes a range of attributes from the dataset.	More							
attributeIndices last								
invertSelection False	~							
Open Save OK	Cancel							

## Click on ok button.

🐇 Weka Explorer				
Preprocess Classify Cluster Associate Select attributes Visualize				
Open file Open URL Open DB		Undo	Edit	Save
Filter				
Choose Remove -R last				Apply
Current relation Relation: STUDENT-weka.filters.unsupervised.attribute.Add-Navg-Cla Instances: 5 Attributes: 13		Selected attribute Name: SNO Missing: 0 (0%)	Distinct: 5	Type: Numeric Unique: 5(100%)
Attributes		Statistic	Value	
		Minimum	1	
All None Invert		Maximum	5	
		Mean	3	
No. Name		StdDev	1.581	
2 SNAME 3 max sal' 4 5AL 5 avg		Class: Copy of addtion (N	um)	Visualize All
6 addtion 7 subtraction		_		
8 Copy of regno 9 Copy of Copy of regno		Ê		
10 Copy of SAL	1			
11 Copy of SAL				
12 Copy of avg				
13 Copy of addtion	~			
Remove				
Status		201	3	6
OK				Log ×0

If the range .

👙 weka.gui.Ge	enericObjectEditor	- DX							
weka.filters.unsupervised.attribute.Remove									
About									
An instance filt dataset.	An instance filter that removes a range of attributes from the More dataset.								
attributeIndices	8-13								
invertSelection	False	*							
Open	Save OK C	ancel							

Click on ok button.

A			
🖆 Weka Explorer			
Preprocess Classify Cluster Associate Select attributes Visualize			
Open file Open URL Open DB	Undo	Edit	Save
Filter			
Choose Remove -R 8-13			Apply
Current relation	Selected attribute		
Relation: STUDENT-weka.filters.unsupervised.attribute.Add-Navg-Cla Instances: 5 Attributes: 7	Name: SNO Missing: 0(0%)	Distinct: 5	Type: Numeric Unique: 5 (100%)
Attributes	Statistic	Value	
	Minimum	1	
All None Invert	Maximum	5	
	Mean	3	
No. Name	StdDev	1.581	
1 SNO			
2 SNAME			
3 🔄 'max sal'			
4 SAL	Class: subtraction (Num)		Visualize All
5 avg			
6 addtion			
7 subtraction	5 [1,5]		
Remove			
	1	3	5
Status			
ок			Log 💉 🕬

Click on Apply option.

No.         SNO         SNAME         'max sal'         SAL         avg         addtion         subtraction           1         1.0         DEPTHI         2000.0         1000.0         1000.0         1000.0           2         2.0         SIRI         4000.0         2000.0         12.2         6000.0         2000.0           3         3.0         RAJI         6000.0         4000.0         14.2         9000.0         2000.0           4         4.0         LAKS         6000.0         3000.0         14.2         9000.0         3000.0           5         5.0         PRAS         10000.0         5000.0         15000.0         5000.0	🔹 Viewer 🗙									
Numeric         String         Numeric         Numeric         Numeric         Numeric         Numeric         Numeric           1         1.0         DEEPTHI         2000.0         1000.0         10.2         3000.0         1000.0           2         2.0         SIRI         4000.0         2000.0         12.2         6000.0         2000.0           3         3.0         RAJI         6000.0         4000.0         13.3         10000.0         2000.0           4         4.0         LAKS         6000.0         3000.0         14.2         9000.0         3000.0	Relation: STUDENT-weka.filters.unsupervised.attribute.Add-Navg-Clast-weka.filters.unsup									
1         1.0         DEEPTHI         2000.0         1000.0         10.2         3000.0         1000.0           2         2.0         SIRI         4000.0         2000.0         12.2         6000.0         2000.0           3         3.0         RAJI         6000.0         4000.0         13.3         10000.0         2000.0           4         4.0         LAK5         6000.0         3000.0         14.2         9000.0         3000.0	No.									
2         2.0         SIRI         4000.0         2000.0         12.2         6000.0         2000.0           3         3.0         RAJI         6000.0         4000.0         13.3         10000.0         2000.0           4         4.0         LAK5         6000.0         3000.0         14.2         9000.0         3000.0		Numeric	String	Numeric	Numeric	Numeric	Numeric	Numeric		
2         2.0 SIRI         4000.0         2000.0         12.2         6000.0         2000.0           3         3.0 RAJI         6000.0         4000.0         13.3         10000.0         2000.0           4         4.0 LAK5         6000.0         3000.0         14.2         9000.0         3000.0	1	1.0	DEEPTHI	2000.0	1000.0	10.2	3000.0	1000.0		
4 4.0 LAK5 6000.0 3000.0 14.2 9000.0 3000.0	2	2.0	SIRI	4000.0	2000.0	12.2		2000.0		
		3.0	RAJI		4000.0	13.3		2000.0		
5 5.0 PRAS 10000.0 5000.0 15000.0 5000.0						14.2				
	5	5.0	PRAS	10000.0	5000.0		15000.0	5000.0		

Click on ok button.

47

## **VIVA – QUESTIONS**

- 1. When to apply the data preprocessing techniques for mining the data
- 2. Which of the following is not a data preprocessing methods:
  - A. Data Visualization B. Data Discretization
  - C. Data Cleaning D. Data Reduction
- 3. Use the attribute mean to fill the missing value of data 1,2,3,4,5,6,\_\_,7,8,9,10.
- 4. Data for Attendance : 50,55,60,65,70,75,80,85,90,95 Partition the above attendance data into equidepth bins of depth 5.
- 5. Data for Attendance : 4,8,15 Smoot by bin boundaries
- 6. Incorrect or invalid data is known as\_\_\_\_\_
- 7. PCA stands for
- The Minimum and maximum values for the attribute income are \$12,000and \$98,000, respectively. We would like to map income to the range[0.0,1.0].By min-max normalization, a value of \$73,600 for income is transformed to
- The mean and standard deviation of the values for the attribute income are \$54,000 and \$16,000 respectively. With Z-Score normalization a value of \$73,600 for income is transformed to
- 10. \_\_\_\_\_ reduces the data set size by removing irrelevant or redundant attributes(dimensions)

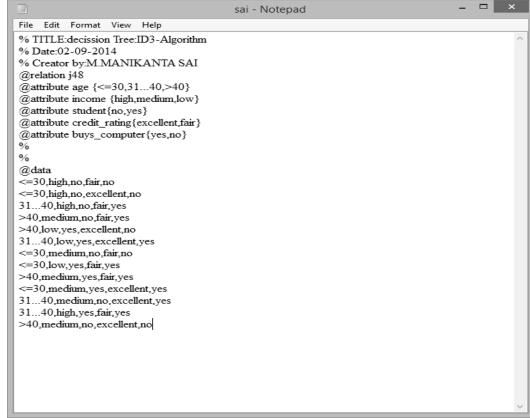
## **Experiment No. 3: Demonstrate performing classification on data sets**

A.

- i. Load weather dataset into WEKA and run Id3, J48 classification algorithm. Study the classifier output. Compute entropy values, kappa statistic.
- ii. Extract if-then rules from the decision tree generated by the classifier, observe the confusion matrix and derive Accuracy, F-measure, TPrate, FPrate, precision and recall values. Apply cross-validation strategy with various fold levels and compare the accuracy results

#### ID3 Algorithm

#### Step1:open the Notepad and create the ARFF File.



**Step 2:**To open All Programms → weka3.8.3 → weka3.8(with console) **STEP 3:** Click on **EXPLORER**.



Relation: J48									
No.	age Nominal	income Nominal	student Nominal	credit-rating Nominal	<b>buys-computer</b> Nominal				
1	<=30	high	no	fair	no				
2	<=30	high	no	excellent	no				
3	3140	high	no	fair	yes				
4	>40	medium	no	fair	yes				
5	>40	low	yes	fair	yes				
6	>40	low	yes	excellent	no				
7	3140	low	yes	excellent	yes				
8	<=30	medium	no	fair	no				
9	<=30	low	yes	fair	yes				
10	>40	medium	yes	fair	yes				
11	<=30	medium	yes	excellent	yes				
12	3140	medium	no	excellent	yes				
13	3140	high	yes	fair	yes				
14	>40	medium	no	excellent	no				

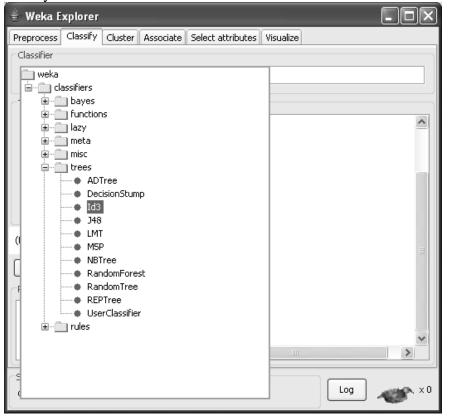
Undo

ОK

Cancel

Step 4:open the ARFF file from the folder and click open.Step 5: In the Preprocessor menu select open file and select Edit

Step 6: Go to classify menu click choose on that select trees Click ID3



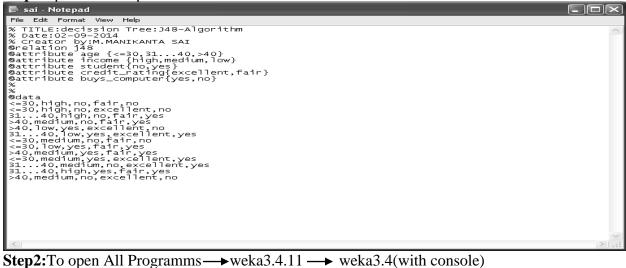
👙 Weka Explorer	
Preprocess Classify Cluster Associate	Select attributes Visualize
Classifier	
Choose Id3	
Test options	Classifier output
O Use training set	
O Supplied test set Set	
⊙ Cross-validation Folds 10	
O Percentage split % 66	
More options	
(Nom) buys-computer	
Start Stop	
Result list (right-click for options)	
Status	
ОК	Log 🛷 ×0

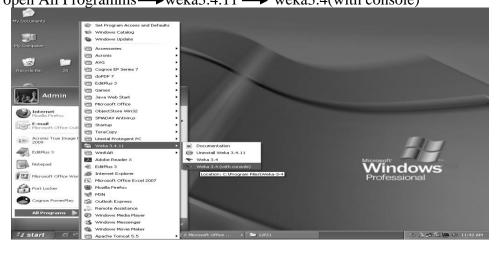
# Step 7: And then click start button.

👙 Weka Explorer		- D X
Preprocess Classify Cluster Associate S	ielect attributes Visualize	
Classifier		
Choose Id3		
Test options	Classifier output	
<ul> <li>Use training set</li> </ul>	Kappa statistic	0.6889
O Supplied test set Set	Mean absolute error	0.1429 0.378
Cross-validation Folds 10	Root mean squared error Relative absolute error	0.378 30 %
Ŭ T	Root relative squared error	°° 76.6097 %
O Percentage split % 66	Total Number of Instances	14
More options		
	=== Detailed Accuracy By Class ===	
(Nom) buys-computer 🗸 🗸 🗸	TP Rate FP Rate Precision Recall	F-Measure Class
	0.889 0.2 0.889 0.889	0.889 ves
Start Stop	0.8 0.111 0.8 0.8	0.8 no
Result list (right-click for options)		
11:43:44 - trees.Id3	=== Confusion Matrix ===	
	a b < classified as	=
	8 l   a = yes l 4   b = no	
11:43:44 - trees.Id3		
		~
	<	>
<u></u> )(		
Status		
ок		

#### J48

Step1:open the Notepad and create the ARFF File.



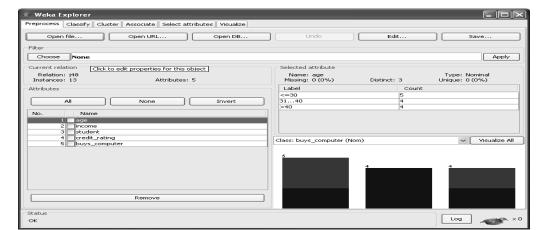


## STEP 3: Click on EXPLORER .

🜍 Weka GUI Chooser	– 🗆 X
Program Visualization Tools Help	
	Applications Explorer
WEKA The University of Waikato	Experimenter
	KnowledgeFlow
Waikato Environment for Knowledge Analysis Version 3.8.3	Workbench
(c) 1999 - 2018 The University of Waikato Hamilton, New Zealand	Simple CLI

🚔 Open					$\left  \times \right $
Look in:	🗀 12f21		~	1	
My Recent Documents	<u>∿</u> ∽ sai				
Desktop					
My Documents					
My Computer					
My Network Places	File <u>n</u> ame: Files of <u>typ</u> e:	sai.arff Arff data files		~	Open Cancel

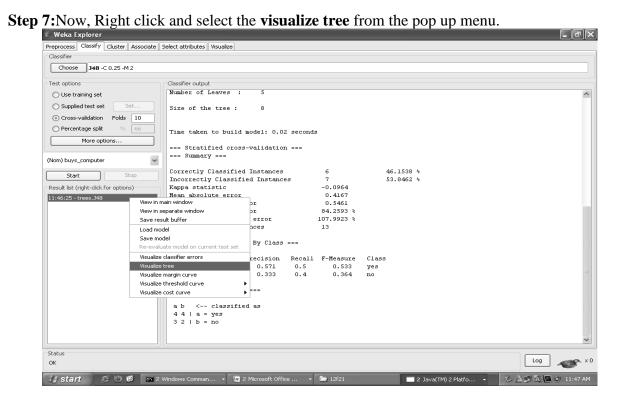
## **Step 4:**open the ARFF file from the folder and click open.



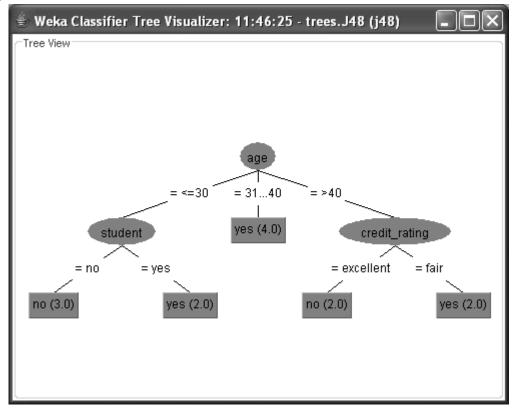
Step 5:click on the classify from the menu and click on the choose.

## Step 6:select the j48 from the tree and click on the start button.

Preprocess Classify Cluster Associate	Select attribute	s Visualize						
Classifier								
Choose 348 -C 0.25 -M 2								
Test options	Classifier out	out						
	Classifier ouq	_						
<ul> <li>Use training set</li> </ul>								
O Supplied test set Set			ed Instances		6	46.153		
Cross-validation Folds 10			fied Instance	3	7	53.846	52 %	
<u> </u>	Kappa sta				-0.0964			
O Percentage split % 66		olute erro:			0.4167			
More options		n squared			0.5461			
- More optionsm		absolute			84.2593 %			
		ative squa: wher of In:	red error		107.9923 % 13			
(Nom) buys_computer	Total Nu	mer or in:	stances		13			
Start Stop	=== Detai	iled Accura	acy By Class					
Result list (right-click for options)	TP Bate	FP Rate	Precision	Recall	F-Measure	Class		
11:46:25 - trees.J48	0.5	0.6	0.571	0.5	0.533	ves		
	0.4	0.5	0.333	0.4		no		
	Confu	usion Matr	ix ===					
	ab <-	classif:	ied as					
	441a							
	3215							
	J[[							
Status								



#### **Output:**



#### B.

i. Load weather dataset into WEKA and perform Naive-Bayes classification

ii. K-nearest neighbour classification. Interpret the results obtained.

#### **Naïve Bayesian Classification**

The naïve Bayesian classifier, or simple Bayesian classifier, works as follows:

**1.** Let *D* be a training set of tuples and their associated class labels. As usual, each tuple is represented by an *n*-dimensional attribute vector, X = (x1, x2, :::, xn), depicting *n* measurements made on the tuple from *n* attributes, respectively, *A*1, *A*2, :::, *An*.

2. Suppose that there are *m* classes, C1, C2, :::, Cm. Given a tuple, X, the classifier will predict that X belongs to the class having the highest posterior probability, conditioned on X. That is, the naïve Bayesian classifier predicts that tuple X belongs to the class Ci if and only if

$$P(C_i|X) > P(C_j|X) \quad \text{for } 1 \le j \le m, j \ne i.$$
$$P(C_i|X) = \frac{P(X|C_i)P(C_i)}{P(X)}.$$

**3.** As P(X) is constant for all classes, only P(XjCi)P(Ci) need be maximized. If the class prior probabilities are not known, then it is commonly assumed that the classes are

$$P(X|C_i) = \prod_{k=1}^n P(x_k|C_i)$$
  
=  $P(x_1|C_i) \times P(x_2|C_i) \times \cdots \times P(x_n|C_i).$ 

**4.** Given data sets with many attributes, it would be extremely computationally expensive to compute  $P(X_jC_i)$ . In order to reduce computation in evaluating  $P(X_jC_i)$ , the naive assumption of class conditional independence is made. This presumes that the values of the attributes are conditionally independent of one another, given the class label of the tuple (i.e., that there are no dependence relationships among the attributes). Thus,

#### We wish to predict the class

label of a tuple using naïve Bayesian classification, given the same training data as in Example weather data set for decision tree induction. The training data are in weather data set. The data tuples are described by the attributes *age*, *income*, *student*, and *credit rating*. The class label attribute, *buys computer*, has two distinct values (namely, *yes*, *no*). Let *C*1 correspond to the

class *buys computer* = yes and C2 correspond to *buys computer* = no. The tuple we wish to classify is

X = (age = youth, income = medium, student = yes, credit rating = fair)

 $P(buys \ computer = yes) = 9=14 = 0:643$ 

 $P(buys \ computer = no) = 5=14 = 0:357$ 

To compute  $PX_jCi$ ), for i = 1, 2, we compute the following conditional probabilities:

 $P(age = youth/buys \ computer = yes) = 2=9 = 0:222$ 

P(age = youth | buys computer = no) = 3=5 = 0.600

P(income = medium | buys computer = yes) = 4=9 = 0:444

P(income = medium | buys computer = no) = 2=5 = 0:400

P(student = yes | buys computer = yes) = 6=9 = 0:667

 $P(student = yes \mid buys \ computer = no) = 1=5 = 0:200$ 

P(credit rating = fair | buys computer = yes) = 6=9 = 0:667

P(credit rating = fair | buys computer = no) = 2=5 = 0:400

We need to maximize  $P(X_jC_i)P(C_i)$ , for i = 1, 2.  $P(C_i)$ , the prior probability of each class, can be computed based on the training tuples:

Using the above probabilities, we obtain

 $P(X|buys \ computer = yes) = P(age = youth j \ buys \ computer = yes)$ 

P(income = mediu | buys computer = yes)

*P*(*student* = *yes* | *buys computer* = *yes*) \_

*P*(*credit rating = fair | buys computer = yes*)

 $= 0:222_0:444_0:667_0:667 = 0:044.$ 

Similarly,

 $P(X|buys\ computer = no) = 0.600_{0.400}_{0.200}_{0.200}_{0.400} = 0.019.$ 

To find the class, Ci, that maximizes P(XjCi)P(Ci), we compute

 $P(X|buys\ computer = yes)P(buys\ computer = yes) = 0:044\_0:643 = 0:028$ 

 $P(X|buys\ computer = no)P(buys\ computer = no) = 0:019_{0:357} = 0:007$ 

Therefore, the naïve Bayesian classifier predicts *buys computer* = yes for tuple X.

Step1: Open Notepad and write the program and save with .arm		_
Ď bayes - Notepad		<
File Edit Format View Help		
<pre>% TITLE:decission Tree:Bayes % Date:09-09-2014 % Creator by:M.MANIKANTA SAI @relation Bayes @attribute income {high,medium,low} @attribute income {high,medium,low} @attribute credit_rating{excellent,fair} @attribute credit_rating{excellent,fair} @attribute buys_computer{yes,no} % % @data &lt;=30,high,no,fair,no &lt;=30,high,no,fair,yes &gt;40,medium,no,fair,yes &gt;40,now;yes,excellent,no 3140,low;yes,excellent,yes &lt;=30,medium,no,fair,no &lt;=30,how;yes,fair,yes &gt;40,medium,no,fair,yes &gt;40,medium,no,excellent,yes 3140,medium,no,excellent,yes 3140,medium,no,excellent,yes 3140,medium,no,excellent,no</pre>	2	
	≥	::

Step1: Open Notepad and write the program and save with .arff

Step2: Go to Start menu and select Weka 3.8.3, on that select Weka 3.8(with console).



Step3:Go to Preprocessor and select the open file.

👙 Open		_	_			×
Look in:	🛅 12f21			*	1 🖻 🖽	
My Recent Documents	🐦 bayes 🜪 ID3 🜪 J48 🜪 regression					
Desktop						
My Documents						
My Computer						
My Network Places	File <u>n</u> ame: Files of <u>t</u> ype:	bayes.arff Arff data files			*	Open Cancel

Step4: In the Preprocessor menu selectEdit.

塗 Weka Explorer	
Preprocess Classify Cluster Associate Select attributes Visualize	
Open file Open URL Open DB	Undo Edit Save Open the current dataset in a Viewer for editi
Choose None	Apply
Current relation Relation: Bayes Instances: 13 Attributes: 5	Selected attribute Name: age Type: Nominal Missing: 0 (0%) Distinct: 3 Unique: 0 (0%)
Attributes	Label Count
All         None         Invert           No.         Name         Invert         Invert	3140 4 >40 4
1 age 2 income	
3	Class: buys_computer (Nom) Visualize All
Remove	5 
Status OK	Log ×0

No.	age Nominal	income Nominal	student Nominal	credit-rating Nominal	<b>buys-computer</b> Nominal	
1	<=30	high	no	fair	no	
2	<=30	high	no	excellent	no	
3	3140	high	no	fair	yes	
4	>40	medium	no	fair	yes	
5	>40	low	yes	fair	yes	
6	>40	low	yes	excellent	no	
7	3140	low	yes	excellent	yes	
8	<=30	medium	no	fair	no	
9	<=30	low	yes	fair	yes	
10	>40	medium	yes	fair	yes	
11	<=30	medium	yes	excellent	yes	
12	3140	medium	no	excellent	yes	
13	3140	high	yes	fair	yes	
14	>40	medium	no	excellent	no	

Step5: Go to **classify** tab menu ,click no **choose** button and then select **Trees** from the list Click on the **NaiveBayes** option.

🖆 Weka Explorer					
Preprocess Classify Cluster Associate Select attributes	Visualize				
Classifier					
i weka					
classifiers					
AODE	ier output				
<ul> <li>BayesNet</li> </ul>					
ComplementNaiveBayes NaiveBayes					
<ul> <li>NaiveBayesMultinomial</li> </ul>					
NaiveBayesSimple     NaiveBayesUpdateable					
lazy					
() meta misc					
C trees					
L					
Status					
ок					Log 💉 × 0
🔐 start 🖉 😕 🔞 📁 12f21	2-9-14 - Micros	regression - Wo	C:\WINDOWS\s	🗖 2 Java(TM) 2 👻	🛆 🕰 📰 🕲 😵 11:22 AM

Step 6:click on the **start** button to view the result.

Weka Explore	;		
Preprocess Classify	Cluster Associate	Select attributes Visualize	
Classifier			
Choose Naiv	reBayes		
Test options		Classifier output	
🔘 Use training se	c .		·
<ul> <li>Supplied test s</li> </ul>	et Set		
Cross-validatio	n Folds 10		
O Percentage spl	it % 66	1	
More op	tions	1	
		1	
(Nom) buys-compute	er 🗸		
Start	Stop	1	
Result list (right-clic	(for options)		
			-
Status			
ок			Log 🔬

# Output:

o mp an		
🛎 Weka Explorer		
Preprocess Classify Cluster Associate S	Select attributes   Visualize	
Classifier		
Choose NaiveBayes		
Test options	Classifier output	
O Use training set		1.00
	=== Run information ===	<u> </u>
O Supplied test set Set	Scheme: weka.classifiers.bayes.NaiveBayes	
<ul> <li>Cross-validation Folds 10</li> </ul>	Relation: Bayes	
O Percentage split % 66	Instances: 13 Attributes: 5	
More options	Attributes: 5	
	income	
(Nom) buys_computer 🗸	student	
Start Stop	credit_rating buys_computer	
	Test mode: 10-fold cross-validation	
Result list (right-click for options)		
11:22:56 - bayes.NaiveBayes	=== Classifier model (full training set) ===	
	Naive Bayes Classifier	
	Class yes: Prior probability = 0.6	
	age: Discrete Estimator. Counts = 3 5 3 (Total = 11)	
	income: Discrete Estimator. Counts = 3 5 3 (Total = 11)	
	student: Discrete Estimator. Counts = $4.6$ (Total = $10$ ) credit rating: Discrete Estimator. Counts = $4.6$ (Total = $10$ )	
	credit_rating: Discrete Estimator. Counts = 4 6 (Total = 10)	
	Class no: Prior probability = 0.4	
	age: Discrete Estimator. Counts = 4 1 3 (Total = 8) income: Discrete Estimator. Counts = 3 3 2 (Total = 8)	
	student: Discrete Estimator. Counts = 5.2 (Total = 7)	
	credit_rating: Discrete Estimator. Counts = 4 3 (Total = 7)	
Status OK		Log 💉 🕬
12 start 8 8 6 12	121 🕎 2-9-14 - Micros 💾 regression - Wo 🗛 C:\/WINDOWS\s 🗖 2 Java(TM) 2 🔹	A A B A A A

🏝 Weka Explorer		_ IJI×
Preprocess Classify Cluster Associate :	Select attributes Visualize	
Classifier		
Choose NaiveBayes		
Test options	Classifier output	
<ul> <li>Use training set</li> </ul>	income: Discrete Estimator. Counts = 3 3 2 (Total = 8) student: Discrete Estimator. Counts = 5 2 (Total = 7)	^
O Supplied test set Set	credit_rating: Discrete Estimator. Counts = 43 (Total = 7)	
Cross-validation Folds 10		
O Percentage split % 66	Time taken to build model: 0 seconds	
More options	=== Stratified cross-validation ===	
(Nom) buys_computer ~	Summary	
Start Stop	Correctly Classified Instances 8 61.5385 %	
Result list (right-click for options)	Incorrectly Classified Instances 5 38.4615 % Kappa statistic 0.0845	
11:22:56 - bayes.NaiveBayes	Mean absolute error 0.4907	
	Root mean squared error 0.5251	
	Relative absolute error 99.2301 % Root relative squared error 103.8335 %	
	Total Number of Instances 13	
	Detailed Accuracy By Class	
	TP Rate FP Rate Precision Recall F-Measure Class 0.875 0.8 0.636 0.875 0.737 ves	
	0.2 0.125 0.5 0.2 0.286 no	
	Confusion Matrix	
	a b < classified as 7 l   a = yes	
	41 + b = no	
	1	~
Status	Log	× 0
5		

## **VIVA-QUESTIONS**

- 1. Formulae for Information gain=\_\_\_\_\_
- 2. Formulae for GainRatio=\_\_\_\_\_
- 3. Compute entropy value of(9,5) = \_\_\_\_\_
- 4. Classification is supervised learning.
- 5. \_\_\_\_\_ measure is used to select the test attribute at each node in the decision tree.
- 6. Posterior probability can be calculated by \_\_\_\_\_\_ theorem.
- 7. Decision tree is a type of \_\_\_\_\_\_ algorithm.
- 8. KNN stand for
- 9. How does KNN calculate distance?
- 10. ID3 stands for

#### **Experiment No. 4: Demonstrate performing association rule mining on data sets**

- i. Load transaction dataset into WEKA and run Aprori algorithm with different support and confidence values.
- ii. The Apriori algorithm uses a generate and count strategy for deriving frequent items sets and generate association rules.

#### The Apriori Algorithm

Apriori is a seminal algorithm proposed by R. Agrawal and R. Srikant in 1994 for mining frequent itemsets for Boolean association rules. The name of the algorithm is based on the fact that the algorithm uses *prior knowledge* of frequent itemset properties, as we shall see following. Apriori employs an iterative approach known as a *level-wise* search, where *k*-itemsets are usedtoexplore (k+1)-itemsets. First, the setof frequent 1-itemsets is found by scanning the database to accumulate the count for each item, and collecting those items that satisfy minimum support. The resulting set is denoted *L*1.Next, *L*1 is used to find *L*2, the set of frequent 2-itemsets, which is used to find *L*3, and so on, until no more frequent *k*-itemsets can be found. The finding of each *Lk* requires one full scan of the database.

**Apriori property**: All nonempty subsets of a frequent itemset must also be frequent. TheApriori property is based on the following observation. By definition, if an itemset I does not satisfy the minimum support threshold, min sup, then I is not frequent; that is,  $P(I) < \min$  sup. If an item A is added to the itemset I, then the resulting itemset (i.e., I [A) cannot occur more frequently than I. Therefore, I [A is not frequent either; that is,  $P(I) < \min$  sup

**The join step:** To find *Lk*, a set of candidate *k*-itemsets is generated by joining *Lk* $\Box$ 1 with itself. This set of candidates is denoted *Ck*. Let *l*1 and *l*2 be itemsets in *Lk* $\Box$ 1. The notation *li*[*j*] refers to the *j*th item in *li* (e.g., *l*1[*k* $\Box$ 2] refers to the second to the last item in *l*1). For the (*k* $\Box$ 1)-itemset, *li*, this means that the items are sorted such that *li*[1] < *li*[2] < : : : < *li*[*k* $\Box$ 1]. The join, *Lk* $\Box$ 1 on *Lk* $\Box$ 1, is performed, where members of *Lk* $\Box$ 1 are joinable if their first (*k* $\Box$ 2) items are in common. That is, members *l*1 and *l*2 of *Lk* $\Box$ 1 are joined if (*l*1[1] = *l*2[1]) ^ (*l*1[2] = *l*2[2]) ^: : :^(*l*1[*k* $\Box$ 2] = *l*2[*k* $\Box$ 2]) ^(*l*1[*k* $\Box$ 1] < *l*2[*k* $\Box$ 1]). The condition *l*1[*k* $\Box$ 1] < *l*2[*k* $\Box$ 1] simply ensures that no duplicates are generated. The resulting itemset formed by joining *l*1 and *l*2 is *l*1[1], *l*1[2], :::, *l*1[*k* $\Box$ 2], *l*1[*k* $\Box$ 1], *l*2[*k* $\Box$ 1].

The prune step:Ck is a superset of Lk, that is, its members may or may not be frequent, but all of the frequent *k*-itemsets are included inCk.Ascan of the database to determine the count of each

candidate in Ck would result in the determination of Lk (i.e., all candidates having a count no less than the minimum support count are frequent by definition, and therefore belong to Lk).

<i>ics</i> branch.	
TID	List of item_IDs
T100	I1, I2, I5
T200	I2, I4
T300	I2, I3
T400	I1, I2, I4
T500	I1, I3
T600	I2, I3
T700	I1, I3
T800	I1, I2, I3, I5
Т900	I1, I2, I3

Transactional	data for an AllElectron-
ics branch.	

In the first iteration of the algorithm, each item is a member of the set of candidate 1-itemsets, C1. The algorithm simply scans all of the transactions in order to count the number of occurrences of each item.

2. Suppose that the minimum support count required is 2, that is, min sup = 2. (Here, we are referring to *absolute* support because we are using a support count. The corresponding relative support is 2/9 = 22%). The set of frequent 1-itemsets, *L*1, can then be determined. It consists of the candidate 1-itemsets satisfying minimum support. In our example, all of the candidates in *C*1 satisfy minimum support.

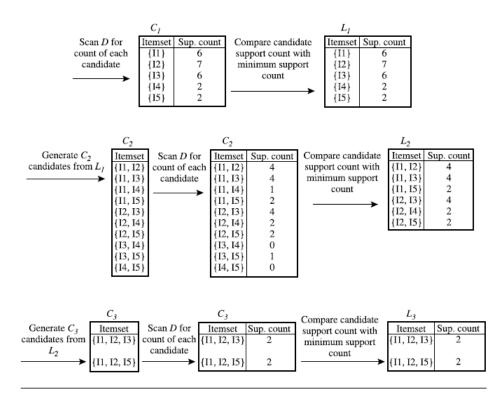
**3.** To discover the set of frequent 2-itemsets, *L*2, the algorithm uses the join *L*1 on *L*1 to generate a candidate set of 2-itemsets, *C*2.8 *C*2 consists of  $\Box jL1j$  2 2-itemsets. Note that no candidates are removed from *C*2 during the prune step because each subset of the candidates is also frequent. 4.Next, the transactions in *D* are scanned and the support count of each candidate itemset in *C*2 is accumulated.

5. The set of frequent 2-itemsets, *L*2, is then determined, consisting of those candidate 2-itemsets in *C*2 having minimum support.

6. The generation of the set of candidate 3-itemsets, C3,. From the join step, we first get C3 =L2 on L2 = ffI1, I2, I3g, fI1, I2, I5g, fI1, I3, I5g, fI2, I3, I4g, fI2, I3, I5g, fI2, I4, I5gg. Based on the Apriori property that all subsets of a frequent itemsetmust also be frequent, we can determine that the four latter candidates cannot possibly be frequent. We therefore remove them from C3, thereby saving the effort of unnecessarily obtaining their counts during the subsequent scan of *D* to determine *L*3. Note that when given a candidate *k*-itemset, we only need to check if its  $(k\Box 1)$ -subsets are frequent since the Apriori algorithm uses a level-wise search strategy.

7. The transactions in D are scanned in order to determine L3, consisting of those candidate 3itemsets in C3.

8. The algorithm uses L3 on L3 to generate a candidate set of 4-itemsets, C4. Although the join results in I1, I2, I3, I5gg, this itemset is pruned because its subset I2, I3, I5 is not frequent. Thus, C4 = f, and the algorithm terminates, having found all of the frequent itemsets.



Algorithm: Apriori. Find frequent itemsets using an iterative level-wise approach based on candidate generation.

Input:

```
    D, a database of transactions;

   min_sup, the minimum support count threshold.
Output: L, frequent itemsets in D.
Method:
(1)
         L_1 = find\_frequent\_1-itemsets(D);
        for (k = 2; L_{k-1} \neq \phi; k++) {

C_k = \operatorname{apriori}_{gen}(L_{k-1});
(2)
(3)
            for each transaction t \in D { // scan D for counts
(4)
(5)
                 C_t = \text{subset}(C_k, t); // \text{ get the subsets of } t \text{ that are candidates}
                 for each candidate c \in C_t
(6)
(7)
                      c.count++;
(8)
(9)
            L_k = \{c \in C_k | c.count \ge min\_sup\}
(10)
        -}
        return L = \bigcup_k L_k;
(11)
procedure apriori_gen(L_{k-1}:frequent (k-1)-itemsets)
(1)
        for each itemset l_1 \in L_{k-1}
(2)
            for each itemset l_2 \in L_{k-1}
                 if (l_1[1] = l_2[1]) \land (l_1[2] = l_2[2]) \land ... \land (l_1[k-2] = l_2[k-2]) \land (l_1[k-1] < l_2[k-1]) then {
(3)
                         = l_1 \bowtie l_2; // join step: generate candidates
(4)
(5)
                      if has_infrequent_subset(c, L_{k-1}) then
(6)
                           delete c; // prune step: remove unfruitful candidate
(7)
                      else add c to C_k;
(8)
        return C_k;
(9)
procedure has_infrequent_subset(c: candidate k-itemset;
            L_{k-1}: frequent (k-1)-itemsets); // use prior knowledge
(1)
         for each (k-1)-subset s of c
(2)
           if s \notin L_{k-1} then
(3)
                 return TRUE;
         return FALSE;
(4)
```

Association rule mining using the Apriori-Algorithm

**Step-1:** To Create an ARFF File.

```
🛤 ITEM1 - Notepad
File Edit Format View Help
%TITLE : TRANSACTION -ITEM
%
%DATE :05-08-2014
%
@RELATION ITEM1
@ATTRIBUTE A1{I1,I2,I3,I4,I5}
@ATTRIBUTE A2{I1,I2,I3,I4,I5}
@ATTRIBUTE A3{I1,I2,I3,I4,I5}
@ATTRIBUTE A4{I1,I2,I3,I4,I5}
@ATTRIBUTE A5{I1,I2,I3,I4,I5}
%
%
@DATA
I1,I2,?,?,I5
?,I2,?,I4,?
?,12,13,?,?
I1,I2,?,I4,?
I1,?,I3,?,?
?,I2,I3,?,?
I1,?,I3,?,?
```

Step-2: -Save as the file .ARFF file format

\_ **-** | ×

**Step-3:** - Go to All Programs in the start menu bar and then choose weeka 3.8.3 in the weeka 3.8(with console).

**Step-4:** In the weeka window to choose the Explorer and then select the Preprocess in the weeka explorer window.

Open file Open URL Open DB General	ate	Jndo Edit.	Save
Current relation	Selected attribute		
Relation: ITEM1 Attributes: 5 Instances: 9 Sum of weights: 9	Name: A1 Missing: 3 (33%)	Distinct: 1	Type: Nominal Unique: 0 (0%)
Attributes	No. Labe		Weight
	1 I1	6	6.0
All None Invert Pattern	2 12	0	0.0
	3 I3 4 I4	0	0.0
No.         Name           1         A1           2         A2           3         A3	5 15	0	0.0
4_A4 5_A5	Class: A5 (Nom)		Visualize All
Remove	6		
	0	0	0 0
Status OK			Log 💉 V O

**Step-5:** Go to Association option in the menu and then select choose button to select the Apriori option.

🐨 Weka Explorer	
Preprocess Classify Cluster Associate Select attributes	Visualize
Associator	
a weka	-S -1.0 -c -1
Apriori     FPGrowth	ator output
Close	
	a 
Status OK	Log 💉 O

Step-6: - Click on Apriori field to configure the various options.

🕝 weka.gui.Generic		$\mathbf{X}$				
weka.associations.Apriori						
About	About					
Class implementing	g an Apriori-type algorithm. More Capabilities					
car	False	~				
classIndex	-1					
delta	0.05					
lowerBoundMinSupport	0.2					
metricType	Confidence	~				
minMetric	0.7					
numRules	10					
outputItemSets	False	~				
removeAllMissingCols	False	*				
significanceLevel	-1.0					
treatZeroAsMissing	False	~				
upperBoundMinSupport	1.0					
verbose	False	*				
Open (	Save OK Cancel					

## Click on OK button.

**Step-7:** - Click on Start button in the Association menu then displays the output.

	lek on Start Sation in the Association mend then displays the Suppl.	
🐨 Weka Explorer		
Preprocess Classify Clu Associator	Isser Associate Select attributes Visualize	
Choose Apriori -	N 10 -T 0 -C 0.7 -D 0.05 -U 1.0 -M 0.2 -S -1.0 -c -1	
Start Stop	Associator output	
Result list (right-click for	=== Run information ===	^
11:36:54 - Apriori	Scheme:       weka.associations.Apriori -N 10 -T 0 -C 0.7 -D 0.05 -U 1.0 -M 0.2 -S -1.0 -c -1         Relation:       ITEM1         Instances:       9         Attributes:       5         Al         A2         A3         A4	
	A5 === Associator model (full training set) ===	=
	Apriori	
	Minimum support: 0.2 (2 instances) Minimum metric <confidence>: 0.7 Number of cycles performed: 16</confidence>	
	Generated sets of large itemsets:	
	Size of set of large itemsets L(1): 5	
	Size of set of large itemsets L(2): 6	
	Size of set of large itemsets L(3): 2	
	Best rules found:	
	1. A5=I5 2 ==> A1=I1 2 <conf;(1)> lift:(1.5) lev:(0.07) [0] conv:(0.67)</conf;(1)>	~
Status OK	Log	• • • • •

lep-o v	-	s options are ca eka.gui.Generic	ObjectEditor	
		associations.Apriori		
	Abo	ut		
	Cla	ass implementin <u>c</u>	g an Apriori-type algorithm. More Capabilities	
		car	False	
		classIndex	-1	
		delta	0.05	
	lower	rBoundMinSupport	0.3	
		metricType	Confidence	
		minMetric	0.7	
		numRules	10	
		outputItemSets	False	
	rem	noveAllMissingCols	False	
		significanceLevel	-1.0	
	tr	reatZeroAsMissing	False	
	upper	rBoundMinSupport	1.0	
		verbose	False	
		Open	Save OK Cancel	
AW-L- FL				
		ster Associate Select .	attributes Visualize	- DX
		ster Associate Select	attributes Visualize	
Preprocess Clas	ssify Clu		attributes Visualize U 1.0 -M 0.3 -S -1.0 -c -1	
Preprocess Clas Associator Choose	ssify Clu			
Preprocess Clas Associator Choose	ssify Clu: Apriori -N Stop	Associator output	U 1.0 -M 0.3 -5 -1.0 -c -1	
Associator Choose	Ssify Clus Apriori - N Stop - click for pri	A 10 -T 0 -C 0.7 -D 0.05 -I Associator output A4 A5	U 1.0 -M 0.3 -5 -1.0 -c -1	
Preprocess Clas Associator Choose Start Result list (right 11:36:54 - Apric	Ssify Clus Apriori - N Stop - click for pri	A 10 -T 0 -C 0.7 -D 0.05 -I Associator output A4 A5	U 1.0 -M 0.3 -5 -1.0 -c -1	
Preprocess Clas Associator Choose Start Result list (right 11:36:54 - Apric	Ssify Clus Apriori - N Stop - click for pri	A 10 -T 0 -C 0.7 -D 0.05 - Associator output A4 A5 === Associator m Apriori ====== Minimum support:	U1.0-M0.3-5-1.0-c-1	
Preprocess Clas Associator Choose Start Result list (right 11:36:54 - Apric	Ssify Clus Apriori - N Stop - click for pri	A10 -T 0 -C 0.7 -D 0.05 -1 Associator output A4 A5 === Associator m Apriori ====== Minimum support: Minimum metric < Number of cycles	U1.0-M0.3-5-1.0-c-1	
Preprocess Clas Associator Choose Start Result list (right 11:36:54 - Apric	Ssify Clus Apriori - N Stop - click for pri	A10-T0-C0.7-D0.05-1 Associator output A4 A5 === Associator m Apriori ====== Minimum support: Minimum metric < Number of cycles Generated sets of	U1.0-M0.3-5-1.0-c-1	
Preprocess Clas Associator Choose Start Result list (right 11:36:54 - Apric	Ssify Clus Apriori - N Stop - click for pri	A10-T0-C0.7-D0.05- Associator output A4 A5 === Associator m Apriori ====== Minimum support: Minimum metric < Number of cycles Generated sets of Size of set of 1	U1.0-M0.3-5-1.0-c-1 u1.5 model (full training set) === c 0.3 (3 instances) cconfidence>: 0.7 s performed: 14 of large itemsets: large itemsets L(1): 3	
Preprocess Clas Associator Choose Start Result list (right 11:36:54 - Apric	Ssify Clus Apriori - N Stop - click for pri	A10-T0-C0.7-D0.05-1 Associator output A4 A5 === Associator m Apriori ======= Minimum support: Minimum metric < Number of cycles Generated sets of Size of set of 1 Size of set of 1	U1.0-M0.3-5-1.0-c-1	
Preprocess Clas Associator Choose Start Result list (right 11:36:54 - Apric	Ssify Clus Apriori - N Stop - click for pri	A10-T0-C0.7-D0.05- Associator output A4 A5 === Associator m Apriori ====== Minimum support: Minimum metric < Number of cycles Generated sets of Size of set of 1	U1.0-M0.3-5-1.0-c-1	
Preprocess Clas Associator Choose Start Result list (right 11:36:54 - Apric	Ssify Clus Apriori - N Stop - click for pri	A10-T0-C0.7-D0.05-1 Associator output A4 A5 === Associator m Apriori ======= Minimum support: Minimum metric < Number of cycles Generated sets of Size of set of 1 Size of set of 1	U1.0-M0.3-5-1.0-c-1	
Preprocess Clas Associator Choose Start Result list (right 11:36:54 - Apric	Ssify Clus Apriori - N Stop - click for pri	A10-T0-C0.7-D0.05-1 Associator output A4 A5 === Associator m Apriori ======= Minimum support: Minimum metric < Number of cycles Generated sets of Size of set of 1 Size of set of 1	U1.0-M0.3-5-1.0-c-1	
Preprocess Clas Associator Choose Start Result list (right 11:36:54 - Apric	Ssify Clus Apriori - N Stop - click for pri	A10-T0-C0.7-D0.05-1 Associator output A4 A5 === Associator m Apriori ====== Minimum support: Minimum metric < Number of cycles Generated sets of Size of set of 1 Size of set of 1 Best rules found	U1.0-M0.3-5-1.0-c-1	

**Step-8:** - Various options are calculated for minimum support value like this.

🕝 weka.gui.Generic	ObjectEditor	X						
weka.associations.Apriori								
About	About							
Class implementing an Apriori-type algorithm. Capabi								
car	False	~						
classInde×	-1							
delta	0.05							
lowerBoundMinSupport	0.4							
metricType	Confidence	~						
minMetric	0.7							
numRules	10							
outputItemSets	False	~						
removeAllMissingCols	False	~						
significanceLevel	-1.0							
treatZeroAsMissing	False	~						
upperBoundMinSupport	1.0							
verbose	False	~						
Open	Save OK Cancel							

🕑 Weka Explorer		- <b>D</b> ×
Preprocess Classify Clu	ster Associate Select attributes Visualize	
Associator		
Choose Apriori -N	N 10 -T 0 -C 0.7 -D 0.05 -U 1.0 -M 0.4 -S -1.0 -c -1	
	Associator output	
Start Stop		
Result list (right-click for	A4 A5	<u> </u>
11:36:54 - Apriori	AS === Associator model (full training set) ===	
11:38:19 - Apriori	( (	
11:38:40 - Apriori 11:38:54 - Apriori		
	Apriori	
	Minimum support: 0.4 (4 instances)	
	Minimum metric <confidence>: 0.7</confidence>	
	Number of cycles performed: 12	
	Generated sets of large itemsets:	
		=
	Size of set of large itemsets L(l): 3	
	Size of set of large itemsets L(2): 3	
	Best rules found:	
	<	>
Status	Log	
ок		

💌 weka.g	eui.Generio	:ObjectEditor				
	iations.Apriori					
About						
	mplementing	g an Apriori-type algo	rithm.	More Capabilities		
	car	False		~		
	classInde×	-1				
	delta	0.05				
lowerBour	ndMinSupport	0.5				
	metricType	Confidence				
	minMetric		0.7			
	numRules		10 False			
	outputItemSets		False			
	removeAllMissingCols		False V			
_	ificanceLevel	-1.0				
	eroAsMissing	False		~		
upperBoun	adMinSupport	1.0	1.0			
	verbose False					
Оре		Save	ОК	Cancel		
Yeka Explorer rocess Classify Cluster A pociator	Associate Select	attributes Visualize				
Choose Apriori -N 10 -T	0 -⊂ 0.7 -D 0.05 -	U 1.0 -M 0.5 -S -1.0 -c -1				
	ciator output					
ult list (right-click for	Run informat	cion ===				
38:19 - Apriori       Rel:         38:40 - Apriori       Ins:         38:54 - Apriori       Ins:         39:29 - Apriori       Att:	ation: IT tances: 9 ributes: 5 Al A2 A3 A4 A5 Associator m	TEM1 2 3 4 5 model (full training	ori -N 10 -T 0 -C 0.7 set) ===	-D 0.05 -U 1.0 -M 0.5	-S -1.0 -	
No .	large itemset	ts and rules found!				

Status

ОК

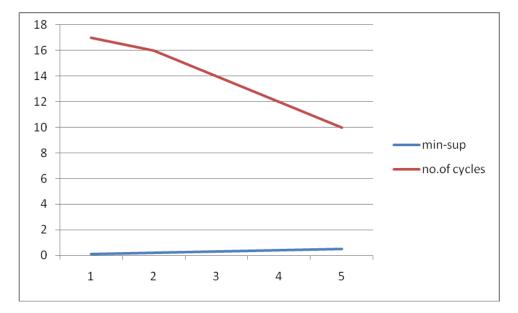
<

>

Log

#### Graph for Apriori Algorithm: -

min-sup	no.of cycles
0.1	17
0.2	16
0.3	14
0.4	12
0.5	10

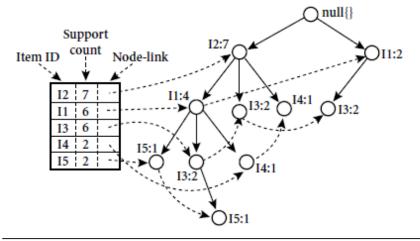


#### FP-growth.

The first scan of the database is the same as Apriori, which derives the set of frequent items (1itemsets) and their support counts (frequencies). Let the minimum support count be 2. The set of frequent items is sorted in the order of descending support count. This resulting set or *list* is denoted *L*. Thus, we have L = I2:7, I1: 6, I3: 6, I4: 2, I5: 2.

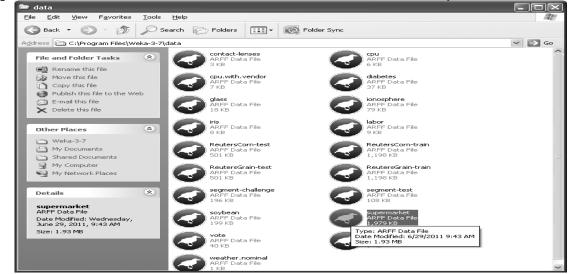
An FP-tree is then constructed as follows. First, create the root of the tree, labeled with "null." Scan database D a second time. The items in each transaction are processed in L order (i.e., sorted according to descending support count), and a branch is created for each transaction. For example, the scan of the first transaction, "T100: I1, I2, I5," which contains three items (I2, I1, I5 in L order), leads to the construction of the first branch of the tree with three nodes, I2: 1, I1:1, and I5: 1, where I2 is linked as a child of the root, I1 is linked to I2, and I5 is linked to I1. The second transaction, T200, contains the items I2 and I4 in L order, which would result in a branch where I2 is linked to the root and I4 is linked to I2. However, this branch would share a common prefix, I2, with the existing path for T100. Therefore, we instead increment the count of the I2

node by 1, and create a newnode, I4: 1, which is linked as a child of I2: 2. In general, when considering the branch to be added for a transaction, the count of each node along a common prefix is incremented by 1, and nodes for the items following the prefix are created and linked accordingly.



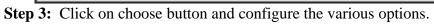
ltem	Conditional Pattern Base	Conditional FP-tree	Frequent Patterns Generated
15	$\{\{I2, I1; 1\}, \{I2, I1, I3; 1\}\}$	(I2: 2, I1: 2)	$\{12, 15; 2\}, \{11, 15; 2\}, \{12, 11, 15; 2\}$
I4	$\{\{I2, I1; 1\}, \{I2; 1\}\}$	(I2: 2)	{12, 14: 2}
13	$\{\{\mathrm{I2},\mathrm{I1};2\},\{\mathrm{I2};2\},\{\mathrm{I1};2\}\}$	$\langle \mathrm{I2:}4,\mathrm{I1:}2\rangle,\langle \mathrm{I1:}2\rangle$	$\{12, 13; 4\}, \{11, 13; 4\}, \{12, 11, 13; 2\}$
11	{{I2: 4}}	$\langle I2:4\rangle$	{12, 11: 4}

Step 1: Select the ARFF file in the weka3-7 folder and in data folder supermarket.



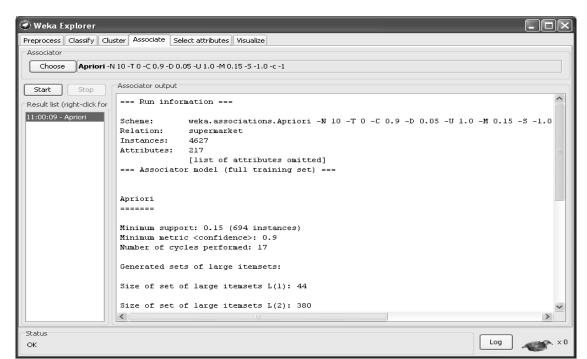
Double Click on the supermarket file and then open.

Preprocess classifier chaster Accessible Colect attributes (Revelation		_	
Preprocess Classify Cluster Associate Select attributes Visualize			
Open file Open URL Open DB Gener	ate Undo	Edit	Save
Filter			
Choose None			A
	Selected attribute		
Relation: supermarket Attributes: 217	Name: department1		Type: Nominal
Instances: 4627 Sum of weights: 4627	Missing: 3580 (77%)	Distinct: 1	Unique: 0 (0%)
Attributes	No. Label	Count	Weight
All None Invert Pattern	1 t	1047	1047.0
No. Name			
1 department1			
3 department3			
4 department4 5 department5	Class: total (Nom)		Visual
6 department6			
7 department7	1047		
8 department8 9 department9			
10 grocery misc			
11 department11 12 baby needs			
Remove			
Status			
OK			Log
	u bar		
<b>2:</b> Go to the Association option in the men	u bar.		
2: Go to the Association option in the men	u bar.	_	
2: Go to the Association option in the men Weka Explorer Preprocess Classify Cluster Associate Select attributes Visualize	u bar.		
2: Go to the Association option in the men	u bar.		
2: Go to the Association option in the men Weka Explorer Preprocess Classify Cluster Associate Select attributes Visualize Associator Choose Apriori -N 10 -T 0 -C 0.9 -D 0.05 -U 1.0 -M 0.1 -S -1.0 -c -1	u bar.		
2: Go to the Association option in the men Weka Explorer Preprocess Classify Cluster Associate Select attributes Visualize Associator Choose Apriori -N 10 -T 0 -C 0.9 -D 0.05 -U 1.0 -M 0.1 -S -1.0 -c -1 Start Stop Associator output	u bar.		
2: Go to the Association option in the men Weka Explorer Preprocess Classify Cluster Associate Select attributes Visualize Associator Choose Apriori -N 10 -T 0 -C 0.9 -D 0.05 -U 1.0 -M 0.1 -S -1.0 -c -1	u bar.		
2: Go to the Association option in the men Weka Explorer Preprocess Classify Cluster Associate Select attributes Visualize Associator Choose Apriori -N 10 -T 0 -C 0.9 -D 0.05 -U 1.0 -M 0.1 -S -1.0 -c -1 Start Stop Associator output	u bar.		
2: Go to the Association option in the men Weka Explorer Preprocess Classify Cluster Associate Select attributes Visualize Associator Choose Apriori -N 10 -T 0 -C 0.9 -D 0.05 -U 1.0 -M 0.1 -S -1.0 -c -1 Start Stop Associator output	u bar.		
2: Go to the Association option in the men Weka Explorer Preprocess Classify Cluster Associate Select attributes Visualize Associator Choose Apriori -N 10 -T 0 -C 0.9 -D 0.05 -U 1.0 -M 0.1 -S -1.0 -c -1 Start Stop Associator output	u bar.		
2: Go to the Association option in the men Weka Explorer Preprocess Classify Cluster Associate Select attributes Visualize Associator Choose Apriori -N 10 -T 0 -C 0.9 -D 0.05 -U 1.0 -M 0.1 -S -1.0 -c -1 Start Stop Associator output	u bar.		
2: Go to the Association option in the men Weka Explorer Preprocess Classify Cluster Associate Select attributes Visualize Associator Choose Apriori -N 10 -T 0 -C 0.9 -D 0.05 -U 1.0 -M 0.1 -S -1.0 -c -1 Start Stop Associator output	u bar.		
2: Go to the Association option in the men Weka Explorer Preprocess Classify Cluster Associate Select attributes Visualize Associator Choose Apriori -N 10 -T 0 -C 0.9 -D 0.05 -U 1.0 -M 0.1 -S -1.0 -c -1 Start Stop Associator output	u bar.		
2: Go to the Association option in the men Weka Explorer Preprocess Classify Cluster Associate Select attributes Visualize Associator Choose Apriori -N 10 -T 0 -C 0.9 -D 0.05 -U 1.0 -M 0.1 -S -1.0 -c -1 Start Stop Associator output	u bar.		
2: Go to the Association option in the men Weka Explorer Preprocess Classify Cluster Associate Select attributes Visualize Associator Choose Apriori -N 10 -T 0 -C 0.9 -D 0.05 -U 1.0 -M 0.1 -S -1.0 -c -1 Start Stop Associator output	u bar.		
2: Go to the Association option in the men Weka Explorer Preprocess Classify Cluster Associate Select attributes Visualize Associator Choose Apriori -N 10 -T 0 -C 0.9 -D 0.05 -U 1.0 -M 0.1 -S -1.0 -c -1 Start Stop Associator output	u bar.		
2: Go to the Association option in the men Weka Explorer Preprocess Classify Cluster Associate Select attributes Visualize Associator Choose Apriori -N 10 -T 0 -C 0.9 -D 0.05 -U 1.0 -M 0.1 -S -1.0 -c -1 Start Stop Associator output	u bar.		
2: Go to the Association option in the men Weka Explorer Preprocess Classify Cluster Associate Select attributes Visualize Associator Choose Apriori -N 10 -T 0 -C 0.9 -D 0.05 -U 1.0 -M 0.1 -S -1.0 -c -1 Start Stop Associator output	u bar.		
2: Go to the Association option in the men Weka Explorer Preprocess Classify Cluster Associate Select attributes Visualize Associator Choose Apriori -N 10 -T 0 -C 0.9 -D 0.05 -U 1.0 -M 0.1 -S -1.0 -c -1 Start Stop Associator output	u bar.		
2: Go to the Association option in the men Weka Explorer Preprocess Classify Cluster Associate Select attributes Visualize Associator Choose Apriori -N 10 -T 0 -C 0.9 -D 0.05 -U 1.0 -M 0.1 -S -1.0 -c -1 Start Stop Associator output	u bar.		



73

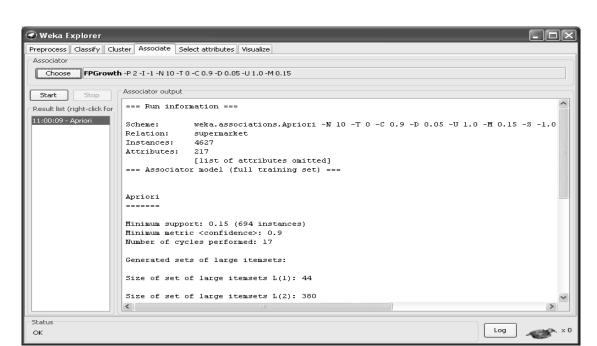
About Class implementing an Apriori-type algorithm. More Capabilities car False classindes -1 delte 0.05 lowerBoundMinSupport 0.15 netrictype Confidence minMetrie 0.9 numRules 10 outputItemSets False significanceLevel -1.0 treatZeroAsMissing False open OK Cancel Verbose False open OK Cancel Woka Leptorer Chees Apriori N10-T0-C.0.9-D.0.65-U.1.0.40.15-5-1.0-c-1 Set 200 Associator output Associator output	weka.associations.Apriori				
Capabilities          car       False         classIndex       -1         delta       0.05         lowerBoundMinSupport       0.15         metricType       Confidence         minMetric       0.9         numRules       10         outputItemSets       False         removeAllMissingCols       False         significanceLevel       -1.0         treatZeroAsMissing       False         Open       Save         Open       Save         Open       Save         Okts tplorer	About				
classIndex -1   delta 0.05   lowerBoundMinSupport 0.15   metricType Confidence   minMetric 0.9   numRules 10   outputItemSets False   removeAllMissingCols False   significanceLevel -1.0   treatZeroAsMissing False   open Save   OK Cancel   Weka Explorer reprocess Classfy Cluster Associator Choose Apriori -N 10 -T 0 -C 0.9 -D 0.05 -U 1.0 -M 0.15 -5 -1.0 -c -1 Start Stop Associator output	Class implementing	g an Apriori-type algo	prithm.		
delta       0.05         lowerBoundMinSupport       0.15         metricType       Confidence         minMetric       0.9         numRules       10         outputItemSets       False         removeAllMissingCols       False         significanceLevel       -1.0         treatZeroAsMissing       False         upperBoundMinSupport       1.0         verbose       False         Open       Save         Open       Save         Veka Explorer	car	False		~	
lowerBoundMinSupport 0.15 metricType Confidence minMetric 0.9 numRules 10 outputItemSets False removeAllMissingCols False significanceLevel -1.0 treatZeroAsMissing False upperBoundMinSupport 1.0 verbose False Open Save OK Cancel Open Save OK Cancel Weka Explore reprocess Classify Cluster Associate Select attributes Visualize Associator Choose Apriori +N 10 -T 0 -C 0.9 -D 0.05 -U 1.0 -M 0.15 -5 -1.0 -c -1	classIndex	-1			
metricType       Confidence         minMetric       0.9         numRules       10         outputItemSets       False         removeAllMissingCols       False         significanceLevel       -1.0         treatZeroAsMissing       False         upperBoundMinSupport       1.0         verbose       False         Open       Save         Open       Save         Ok       Cancel	delta	0.05			
minMetric 0.9 numRules 10 outputItemSets False removeAllMissingCols False significanceLevel -1.0 treatZeroAsMissing False upperBoundMinSupport 1.0 verbose False Open Save OK Cancel Open Save OK Cancel Meka Explore reprocess Classify Cluster Associate Select attributes Visualize Associator Choose Apriori -N 10 -T 0 - C 0.9 -D 0.05 -U 1.0 - M 0.15 - 5 - 1.0 - c -1 Start Stop Associator output	lowerBoundMinSupport	0.15			
numRules 10   outputItemSets False   removeAllMissingCols False   significanceLevel -1.0   treatZeroAsMissing False   upperBoundMinSupport 1.0   verbose False   Open Save   Ok Cancel   Open OK Cancel     Oke Explorer reprocess   Classify Cluster   Associator   Choose Apriori -N 10 -T 0 - C 0.9 -D 0.05 -U 1.0 -M 0.15 - 5 - 1.0 - c -1     Start Stop	metricType	Confidence		~	
outputItemSets False   removeAllMissingCols False   significanceLevel -1.0   treatZeroAsMissing False   upperBoundMinSupport 1.0   verbose False   Open Save   Ok Cancel   Ok Cancel Ok Save Ok Cancel Associator Choose Apriori -N 10 -T 0 - C 0.9 - D 0.05 -U 1.0 -M 0.15 - S - 1.0 - c -1 Start Stop Associator output	minMetric	0.9			
removeAllMissingCols False significanceLevel -1.0 treatZeroAsMissing False upperBoundMinSupport 1.0 verbose False Open Save OK Cancel Open Save OK Cancel Weka Explorer reprocess Classify Cluster Associate Select attributes Visualize Associator Choose Apriori -N 10 -T 0 -C 0.9 -D 0.05 -U 1.0 -M 0.15 -5 -1.0 -c -1 Start Stop Associator output	numRules	10			
significanceLevel -1.0 treatZeroAsMissing False upperBoundMinSupport 1.0 verbose False Open Save OK Cancel Open Save OK Cancel Meka Explore reprocess Classify Cluster Associate Select attributes Visualize Associator Choose Apriori -N 10 -T 0 -C 0.9 -D 0.05 -U 1.0 -M 0.15 -S -1.0 -c -1 Start Stop Associator output	outputItemSets	False		~	
treatZeroAsMissing       False         upperBoundMinSupport       1.0         verbose       False         Open       Save         Ok       Cancel         Ok       Select attributes         Visualize       Associator         Choose       Apriori -N 10 -T 0 -C 0.9 -D 0.05 -U 1.0 -M 0.15 -S -1.0 -c -1         Start       Stop	removeAllMissingCols	False		~	-
upperBoundMinSupport 1.0 verbose False Open Save OK Cancel Open OK Cancel OK Cancel Meka Explorer reprocess Classify Cluster Associate Select attributes Visualize Associator Choose Apriori -N 10 -T 0 -C 0.9 -D 0.05 -U 1.0 -M 0.15 -S -1.0 -c -1 Start Stop Associator output	significanceLevel	-1.0			
verbose False          Open       Save       OK       Cancel         on OK button.         Weka Explorer       Performed and the second and the s	treatZeroAsMissing	False		~	
Open       Save       OK       Cancel         on OK button.         Weka Explorer	upperBoundMinSupport	1.0			
on OK button. Weka Explorer reprocess Classify Cluster Associate Select attributes Visualize Associator Choose Apriori -N 10 -T 0 -C 0.9 -D 0.05 -U 1.0 -M 0.15 -S -1.0 -c -1 Start Stop Associator output	verbose	False		~	-
on OK button. Weka Explorer reprocess Classify Cluster Associate Select attributes Visualize Associator Choose Apriori -N 10 -T 0 -C 0.9 -D 0.05 -U 1.0 -M 0.15 -S -1.0 -c -1 Start Stop Associator output					- 1
Start Stop Associator output	on OK button. Weka Explorer		ОК	Cancel	
Start Stop	on OK button. Weka Explorer eprocess Classify Cluster Associate Sele Associator	ect attributes Visualize	OK	Cancel	
Result list (right-click for options)	on OK button. Weka Explorer eprocess Classify Cluster Associate Sele Associator	ect attributes Visualize	<u>ОК</u>	Cancel	
	on OK button. Weka Explorer eprocess Classify Cluster Associate Sele associator Choose Apriori -N 10 -T 0 -C 0.9 -D 0.0 Start Stop	ct attributes Visualize	OK	Cancel	



## Give the lowerBoundMinSupport value.

🕢 weka.gui.GenericObjectEditor					
weka.associations.FPGrowth					
About					
Class implementing the sets without candidate g	FP-growth algorithm for finding large item More eneration. Capabilities				
delta	0.05				
findAllRulesForSupportLevel	False	~			
lowerBoundMinSupport	0.15				
maxNumberOfItems	-1				
metricType	Confidence	~			
minMetric	0.9				
numRulesToFind	10				
positiveIndex	2				
rulesMustContain					
transactionsMustContain					
upperBoundMinSupport	1.0				
useORForMustContainList	False	~			
Open	Save OK Cancel				

**Step 4:** Click on choose button and then select the FP-growth Algorithm and configure to various minimum support values.



## Click on start

🐨 Weka Explorer		
Preprocess Classify Clu	Ister Associate Select attributes Visualize	
Associator		
Choose FPGrowt	h -P 2 -I -1 -N 10 -T 0 -C 0.9 -D 0.05 -U 1.0 -M 0.15	
Start Stop	Associator output	
Result list (right-click for	=== Run information ===	<u>^</u>
11:00:09 - Apriori 11:02:09 - FPGrowth	Scheme: weka.associations.FPGrowth -P 2 -I -1 -N 10 -T 0 -C 0.9 -D 0.05 -U 1.	) -M (
11:02:09 - PPGrowth	Relation: supermarket	
	Instances: 4627	
	Attributes: 217 [list of attributes omitted]	
	=== Associator model (full training set) ===	
	ASSOCIACOL MODEL (IULI CLAINING SEC)	
	FPGrowth found 16 rules (displaying top 10)	
	1. [fruit=t, frozen foods=t, biscuits=t, total=high]: 788 ==> [bread and cake=t]: '	723
	2. [fruit=t, baking needs=t, biscuits=t, total=high]: 760 ==> [bread and cake=t]:	
	3. [fruit=t, baking needs=t, frozen foods=t, total=high]: 770 ==> [bread and cake=	t]: 71
	<ol> <li>Ifruit=t, vegetables=t, biscuits=t, total=high]: 815 ==&gt; [bread and cake=t]: 74</li> </ol>	
	5. [fruit=t, party snack foods=t, total=high]: 854 ==> [bread and cake=t]: 779 <	
	6. [vegetables=t, frozen foods=t, biscuits=t, total=high]: 797 ==> [bread and cake	
	<ol> <li>[vegetables=t, baking needs=t, biscuits=t, total=high]: 772 ==&gt; [bread and cake= 8. [fruit=t, biscuits=t, total=high]: 954 ==&gt; [bread and cake=t]: 866 <conf:(0.9)< li=""> </conf:(0.9)<></li></ol>	
	<ol> <li>[Iffuit=t, vegetables=t, focal=nigh]: 93s=t, total=high]: 83s =&gt;&gt; [Dread and caze=t]: obs </li> <li>[fruit=t, vegetables=t, frozen foods=t, total=high]: 83s =&gt;&gt; [Dread and caze=t]</li> </ol>	
	10. [fruit=t, frozen foods=t, total=high]: 969 ==> [bread and cake=t]: 877 < conf:	
		~
		>
Status		
ок	Log	×0

min-sup	apriori	fp-growth
0.15	0.44	0.2
0.2	0.3	0.1
0.3	0.2	0.1
0.4	0.15	0.1
0.5	0.11	0.1
0.6	0.8	0.1
0.7	0.6	0.1
0.8	0.5	0.1
0.9	0.3	0.1

#### **VIVA-QUESTIONS**

- 1. Market-basket problem was formulated by\_\_\_\_\_.
- 2. The Market Basket Analysis is a typical example of\_\_\_\_\_
- 3. An itemset satisfying the support criterion is known as:
- 4. In Apriori algorithm, for generating k-item sets, we use:
- 5. Mathematically, Support( $\{X\} \rightarrow \{Y\}$ ) is
- 6. All nonempty subset of a frequent itemset must also be frequent is \_\_\_\_\_ property
- 7. FP-Growth algorithm is based on \_\_\_\_\_\_.
- 8. The formula for Support(A  $\Rightarrow$  B)=
- 9. The formula for Confidence  $(A \Rightarrow B)$ =
- 10. Apriori method mines the frequent itemsets without candidate generation

## **Experiment No. 5: Demonstrate performing regression on data sets**

i. Load numeric dataset into WEKA and build linear regression model.

ii. Explore simple linear regression technique that only looks at one variable

#### **Linear Regression**

Straight-line regression analysis involves a response variable, *y*, and a single predictor variable, *x*. It is the simplest form of regression, and models *y* as a linear function of *x*. That is,

# y = b + wx;

where the variance of y is assumed to be constant, and b and w are regression coefficients specifying the Y-intercept and slope of the line, respectively. The regression coefficients, w and b, can also be thought of as weights, so that we can equivalently write,

$$y = w_0 + w_1$$

These coefficients can be solved for by the method of least squares, which estimates the bestfitting straight line as the one that minimizes the error between the actual data and the estimate of the line.

$$w_1 = \frac{\sum_{i=1}^{|D|} (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{|D|} (x_i - \bar{x})^2}$$

$$w_0 = \overline{y} - w_1 \overline{x}$$

where x is the mean value of  $x_1, x_2, \ldots, x_jD_j$ , and y is the mean value of  $y_1, y_2, \ldots, y_jD_j$ . The coefficients w0 and w1 often provide good approximations to otherwise complicated regression equations.

x years experien	ice y salary	(in \$1000s	)	
3		30	_	
8		57		
9		64		
13		72		
3		36		
6		43		
11		59		
21		90		
1		20		
16		83		
80 - (s000'f ui) (1000'f ui) 40 - 40 - 20 - ♦	<ul><li>◇</li><li>◇</li></ul>	* *	\$	
o   o	5 10 Years	15 experience	20	

25

Given the above data, we compute x = 9:1 and y = 55:4. Substituting these values into Equations (6.50) and (6.51), we get w1 = 3:5w0 = 23:6Thus, the equation of the least squares line is estimated by y = 23:6+3:5x. Using this equation, we can predict that the salary of a college graduate with, say, 10 years of experience

Step1: Open Notepad and write the program and save with .arff

							regres	sion -	Note	pad			-	×
File	Edit	Format	View	Help										
			ation r	egress	ion									
		-09-20												
			MANI	KAN?	fa sa	Ι								
		regress												
		X num												
	ibute	Y num	eric											
%														
%														
@dat	a													
3,30														
8,57														
9,64														
13,72	2													
3,36														
6,43														
11,59														
21,90														
16,83	\$													
											 	 		_

Step2: Go to Start menu and select Weka 3.8.3, on that select Weka 3.8(with console).select Knowledge Flow.



Step 3: In Weka Select Explorer window select **preprocessor** tab and select the **open file** for selection of desired file.

is \$58,600.

Preprocess Classify Cl.					
	ister Associate Sele	ct attributes Visualize			
Open file	Open URL	Open D8	Undo	Edt	Seve
Filter Open a set	t of instances from a file	•			
Choose None					Apply
Current relation			Selected attribute		
Relation: None Instances: None	Attribu	tes: None	Name: None Missing: None	Distinct: None	Type: None Unique: None
Attributes					
Al	None	Invert			
			,		
					Visualize Al
			L		Visualize All
-			1		
	Remove				
Status			2		
Welcome to the Welka Ex	plorer				Log 🛷 ×0
👙 Open		_			×
Look in:	🛅 12f21			¥ 🗊 🖻	
Look in:	📄 12f21			• 0 8	
Look in:				• 0 0	
My Recent	🜪 ID3			• 🕸 🗆	
Ì	🌪 ID3 🜪 J48			• 1	
My Recent	🌪 ID3 🜪 J48			• Ø B	
My Recent	🌪 ID3 🜪 J48			✓ Ø B	
My Recent	🌪 ID3 🜪 J48				
My Recent Documents	🌪 ID3 🜪 J48				
My Recent Documents	🌪 ID3 🜪 J48				
My Recent Documents Desktop	🌪 ID3 🜪 J48				
My Recent Documents	🌪 ID3 🜪 J48				
My Recent Documents Desktop	🌪 ID3 🜪 J48				
My Recent Documents Desktop	🌪 ID3 🜪 J48				
My Recent Documents Desktop My Documents	🌪 ID3 🜪 J48				
My Recent Documents Desktop	🌪 ID3 🜪 J48				
My Recent Documents Desktop My Documents	TD3	regression arff			
My Recent Documents Desktop My Documents	🌪 ID3 🜪 J48	regression.arff			

Step 4: In the Preprocessor menu select open file and select Edit

No.         X         Y           Numeric         Numeric         Numeric           1         3.0         30.0           2         8.0         57.0           3         9.0         64.0           4         13.0         72.0           5         3.0         36.0           6         6.0         43.0           7         11.0         59.0           8         21.0         90.0           9         16.0         83.0	<b>≝</b> V	'iewe r		
Numeric         Numeric           1         3.0         30.0           2         8.0         57.0           3         9.0         64.0           4         13.0         72.0           5         3.0         36.0           6         6.0         43.0           7         11.0         55.0           8         21.0         90.0	Relatio	on: regres	sion	
3.0         30.0           2         8.0         57.0           3         9.0         64.0           4         13.0         72.0           5         3.0         36.0           6         6.0         43.0           7         11.0         59.0           8         21.0         90.0		X	Y	
4         13.0         72.0           5         3.0         36.0           6         6.0         43.0           7         11.0         59.0           8         21.0         90.0				
4         13.0         72.0           5         3.0         36.0           6         6.0         43.0           7         11.0         59.0           8         21.0         90.0	1	3.0		
4     13.0     72.0       5     3.0     36.0       6     6.0     43.0       7     11.0     59.0       8     21.0     90.0	2			
4         13.0         72.0           5         3.0         36.0           6         6.0         43.0           7         11.0         59.0           8         21.0         90.0	3	9.0		
6 6.0 43.0 7 11.0 59.0 8 21.0 90.0	4			
7 11.0 59.0 8 21.0 90.0		3.0	36.0	
8 21.0 90.0		6.0	43.0	
	7	11.0	59.0	1
9 16.0 83.0		21.0	90.0	
	9	16.0	83.0	

Step 5: Go to classify menu click choose classifiers on that functions and that Linear Regression.

🗄 Weka Explorer	
Preprocess Classify Cluster Associate Select attributes	Visualize
Classifier	
Wela         Gosfiers         Hototons         Hototons         Hototons         Hototons         Hototons         Hototons         Hototons         Hototons         Hototons         Hotospission         Hotospission         SingleCogatic         Storeg         WitchProception         WitchProcepti	
ок	Log 🛷 ×0

81

🚔 Weka Explorer	
Preprocess Classify Cluster Associate Select attributes Visualize	
Classifier Choose LinearRegression -S 0 -R 1.0E-8	
Test options Classifier output	
O Use training set	~
Use training set         Supplied test set         Set         Cross-validation         Folds         Percentage split         More options	
Cross-validation Folds 10	
O Percentage split % 66	
More options	
(Num) y	
Start Stop	
Result list (right-click for options)	
	~
Status	
OK	

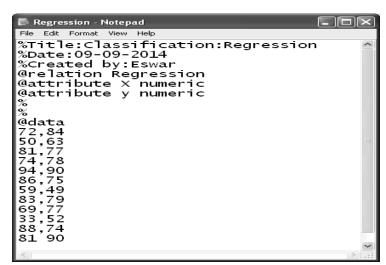
👻 weka.gui.GenericObjectEditor							
weka.classifiers.functions.Line	earRegression						
About							
Class for using linear re	egression for prediction.	More					
attributeSelectionMethod	M5 method	*					
debug	False	~					
eliminateColinearAttributes	True	*					
ridge	1.0E-8						
Open S	ave OK	Cancel					

When we click on the Linear Regression option on the above the dialog box appears then click OK, and click on start button.

# **Output:**

😒 Weka Explorer			
Preprocess Classify Cluster Associate	Select attributes Visualize		
			]
Choose LinearRegression -5.0	-R 1.0E-8		
Test options	Classifier output		
<ul> <li>Use training set</li> </ul>	Linear Regression Model		~
O Supplied test set Set			
	¥ =		
O Percentage split % 66	3.5375 * X +		
More options	23.209		
	Time taken to build model: 0 seconds		
(Num) y	The sale of party model, o peoondp		
(······, /	=== Cross-validation ===		
Start Stop	=== Summary ===		
Result list (right-click for options)	Correlation coefficient	0.9474	
22:32:43 - functions.LinearRegression	Mean absolute error	6.0423	
	Root mean squared error	7.2002	
	Relative absolute error	29.3632 %	
	Root relative squared error Total Number of Instances	29.7254 %	
		10	~
			>
Status			
ок		Log .	
5			

Step1: Open Notepad and write the program and save with .arff



Step2: Go to Start menu and select Weka 3.4.11, on that select Weka 3.4(with console) In Weka Select Explorer and on that select preprocessor and select the desired file

😒 Weka Explorer			- O ×
Preprocess Classify Cluster Associate Select attribut	es Visualize		
Open file Open UR Open DB	Undo	Edit	Save
Filter			]
Choose None			Apply
Current relation	Selected attribute		
Relation: Regression Instances: 10 Attributes: 2	Name: X Missing: 0 (0%)	Distinct: 9	Type: Numeric Unique: 8 (80%)
Attributes	Statistic	Value	
	Minimum	1	~
All None Invert	Maximum	21 9.1	
No. Name	Mean StdDev	6.315	~
	Class: y (Num)		Visualize All
	7		
		3	
Remove			
	1	11	21
Status			
ок			Log 💉 🔨 × O

In the Preprocess open file is present on that select the desired file

<b>4</b>								- DX
About								
Reads a source	e that is in arff	(attribute relati	on file forma	t) format.				More
Look in:	🛅 12f21				~	Ð	2	1
My Recent Documents	🜪 ID3 🏷 J48							
Desktop								
My Documents								
My Computer								
My Network Places	File <u>n</u> ame: Files of <u>t</u> ype:	J48.arff Arff data files					<ul> <li>I</li> <li>I</li> </ul>	<u>O</u> pen ancel

Step3: In the Preprocessor menu select open file and select Edit.

¥	iewer		×
Relatio	on: Regres	ssion	
No.	×	Y	
	Numeric	Numeric	
1	72.0	84.0	
2 3	50.0	63.0	
3	81.0	77.0	
4	74.0	78.0	
5	94.0	90.0	
6	86.0	75.0	
7	59.0	49.0	
8	83.0	79.0	
9	69.0	77.0	
10	33.0	52.0	
11	88.0	74.0	
12	81.0	90.0	
	Undo	ОК	Cancel

Step4: Go to classify menu click choose classifiers on that functions and that Linear Regression

84

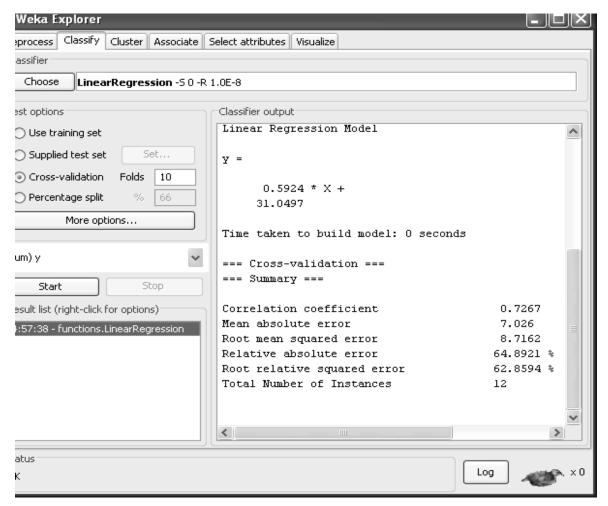
😌 Weka Explorer			
Preprocess Classify Cluster Associate	Select attributes	Visualize	
Preprocess Classify Cluster Associate Classifier  weka classifiers			
د		Log	×۵ 🔪
Weka Explorer Preprocess Classify Cluster Associate	Select attributes	Visualize	
Classifier	Soloce dealbaces		
Choose LinearRegression -5 0	-R 1.0E-8		
Test options	Classifier output	t properties for this object	
<ul> <li>○ Use training set</li> <li>○ Supplied test set</li> <li>○ Cross-validation</li> <li>○ Percentage split</li> <li>?%</li> <li>66</li> <li>More options</li> <li>(Num) y</li> <li>Start</li> <li>Stop</li> <li>Result list (right-click for options)</li> </ul>			
Status OK		Log	×0

Step 5:When we click on the Linear Regression option on the above the dialog box appears then click O.K

👙 weka.gui.GenericOb	jectEditor	- DX
weka.classifiers.functions.Line	earRegression	
About		
Class for using linear re	gression for prediction.	More
attributeSelectionMethod	M5 method	~
debug	False	*
eliminateColinearAttributes	True	~
ridge	1.0E-8	
Open S	ave OK	Cancel

Step 6: And then click OK. button, and then click start in the classify menu present in Weka Explorer window.

Output:



# **VIVA-QUESTIONS**

- 1. The process of constructing a mathematical model or function that can be used to predict or determine one variable by another variable is called.
- 2. In the regression equation Y = 12 21X, the slope is
- 3. In the regression equation Y = 55.65 + 0.50X, the intercept is
- 4. Which of the following methods do we use to find the best fit line for data in Linear Regression
- 5. In the mathematical Equation of Linear Regression  $Y = \beta 1 + \beta 2X + \epsilon$ , ( $\beta 1$ ,  $\beta 2$ ) refers to
- 6. A regression between foot length (response variable in cm) and height (explanatory variable in inches) for 33 students resulted in the following regression equation.  $y^{2} = 10.9 + 0.23 x$

One student in the sample was 100 inches tall with a foot length of 29 cm. What is the predicted foot length for this student?

 Considerx1, x2 to be the independent variables and y the dependent variable, which of the following represents a linear regression model?

represents a linear regression model?

- 8. The sum of the squared residuals is called as:
- 9. The value of the coefficient of determination ranges between
- 10. How many coefficients do you need to estimate in a simple linear regression model (One independent variable)?

#### **Experiment No. 6: Demonstrate performing SVM classification on data sets**

i. Load each dataset into WEKA and run proximal SVM to find a classifiers.

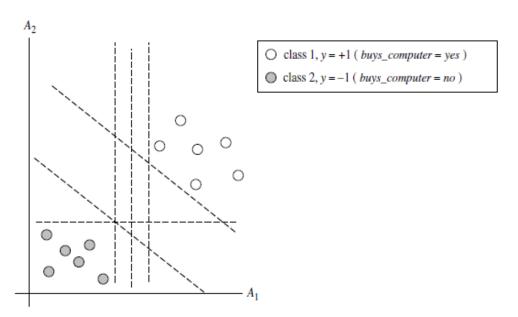
ii. Load each dataset into WEKA and run liner classifier using linear Programming.

## SVM

Support Vector Machines, a promising new method for the classification of both linear and nonlinear data. In a nutshell, a support vector machine (or SVM) is an algorithm that works as follows. It uses a nonlinear mapping to transform the original training data into a higher dimension. Within this new dimension, it searches for the linear optimal separating hyperplane (that is, a "decision boundary" separating the tuples of one class from another). With an appropriate nonlinear mapping to a sufficiently high dimension, data from two classes can always be separated by a hyperplane.

## The Case When the Data Are Linearly Separable

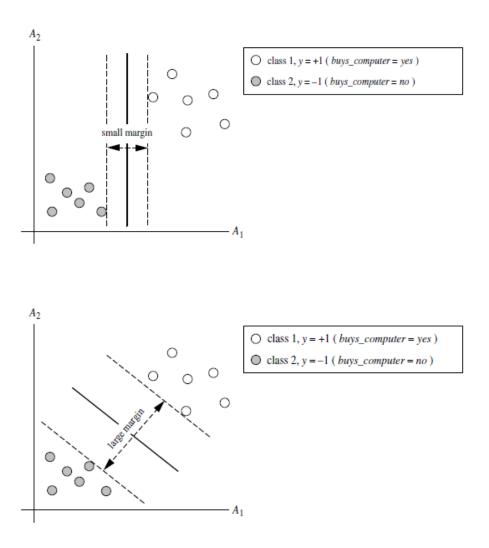
Let the data set *D* be given as (*X*1, *y*1), (*X*2, *y*2), : : : , (*X*j*D*j, *y*j*D*j),where *Xi* is the set of training tuples with associated class labels, *yi*. Each *yi* can take one of two values, either+1 or  $\Box$ 1 (i.e., *yi* 2 f+1,  $\Box$ 1g), corresponding to the classes *buys computer* = *yes* and *buys computer* = *no*, respectively.



Before we get into the definition of margins, let's take an intuitive look at this figure. Both hyperplanes can correctly classify all of the given data tuples. Intuitively, however, we expect the hyperplane with the larger margin to be more accurate at classifying future data tuples than the hyperplane with the smaller margin. This is why (during the learning or training phase), the SVM searches for the hyperplane with the largest margin, that is, the *maximum marginal hyperplane* (MMH). The associated margin gives the largest separation between classes. Getting to an informal definition of margin, we can say that the shortest distance from a hyperplane to one side of its margin is equal to the shortest distance from the hyperplane to the other side of its margin, where the "sides" of the margin are parallel to the hyperplane. When dealing with the MMH, this distance is, in fact, the shortest distance from the MMH to the closest training tuple of either class.

# $W \cdot X + b = 0,$

$$w_0 + w_1 x_1 + w_2 x_2 = 0.$$



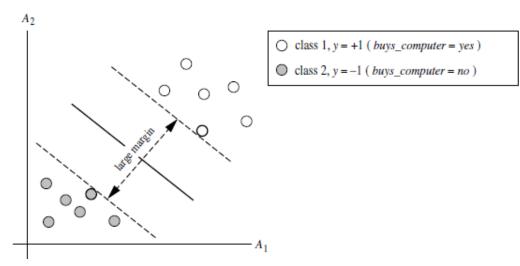
Data Mining-Lab

Thus, any point that lies above the separating hyperplane satisfies

$$w_0 + w_1 x_1 + w_2 x_2 > 0.$$

Similarly, any point that lies below the separating hyperplane satisfies

$$w_0 + w_1 x_1 + w_2 x_2 < 0.$$



The weights can be adjusted so that the hyperplanes defining the "sides" of the margin can be written as

$$H_1: w_0 + w_1x_1 + w_2x_2 \ge 1$$
 for  $y_i = +1$ , and  
 $H_2: w_0 + w_1x_1 + w_2x_2 \le -1$  for  $y_i = -1$ .

<sup>9</sup>If 
$$W = \{w_1, w_2, \dots, w_n\}$$
 then  $\sqrt{W \cdot W} = \sqrt{w_1^2 + w_2^2 + \dots + w_n^2}$ .

#### Algorithm

□ Define an optimal hyper plane: maximize margin

 $\hfill\square$  Extend the above definition for non-linearly separable problems: have a penalty term for misclassifications.

 $\Box$  Map data to high dimensional space where it is easier to classify with linear decision surfaces: reformulate problem so that data is mapped implicitly to this space.

## **PROCEDURE:**

1. We begin the experiment by loading the data (vote.arff) into weka.

2. Next we select the classify tab and click choose function button to select the Support vector machine (SMO).

3. Now we specify the various parameters. These can be specified by clicking in the text box to the right of the chose button.

4. Under the "text "options in the main panel. We select the 10-fold cross validation as our evaluation approach. Since we don't have separate evaluation data set, this is necessary to get a reasonable idea of accuracy of generated model.

5. We now click "start" to generate the model .the ASCII version of the tree as well as evaluation statistic will appear in the right panel when the model construction is complete.

6. Note that the classification accuracy of model is about 69%.this indicates that we may find more work. (Either in preprocessing or in selecting current parameters for the classification)

7. The run information of the support vector classifier will be displayed with the correctly and

incorrectly classified instances.

## **STEPS:**

Open Weka tool – click Explorer in Weka GUI Chooser.

	Weka GUI Chooser			
	Program Visualization Tools Help			
	Waikato Environment for Knowledge Analysis	A	Applications Explorer Experimenter KnowledgeFlow Workbench	
	Version 3.8.3 (c) 1999 - 2018 The University of Waikato Hamilton, New Zealand		Simple CLI	
Weka Explorer reprocess Classify Cluste Open file O	r Associate Select attributes Visualize pen URL Open DB Gener	rate	ndo Edit	Save
Choose None				Apply
Current relation Relation: None Instances: None	Attributes: None Sum of weights: None	Selected attribute Name: None Missing: None	Distinct: None	Type: None Unique: None
Attributes	one Invert Pattern			
				▼ Visualize All

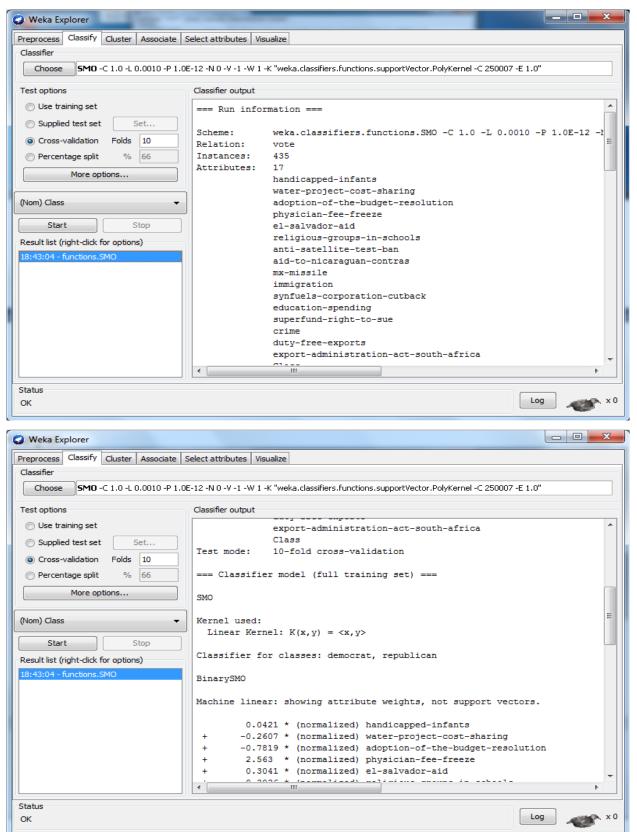
In **preprocess** tab click **Open file..** – choose **file name** (vote.arff).

No     No     No     No       <		<u>ن</u> -						ex8 - Micr	osoft Word								0 2
Image: International and the second secon	X Cut	Insert Page Layo	ut References								АаВІ	AaBbCcl	AaBbCcD	AaBbCcD	AaBbCcD	A	과 Find
Outer       Atthute:       2         No.       None       Invested faits         1       None       Pattern         2       No.       None         1       None       Pattern         2       No.       None         3       doption-roting       Sector         10       Interreter bandon       10         10       None       Pattern         10       Interreter bandon       10         10       None       Interreter bandon         10       Interreter bandon       10         10       Interreter bandon       Interreter bandon         10       Interreter bandon       Interreter bandon         11       None       Interreter bandon         10       Interreter bandon       Interreter bandon         11       Interreter bandon       Interreter bandon         11       Interreter bandon       Intereteret bandon <t< td=""><td>🗧 🍼 Forma</td><td></td><td>Cluster, Associate</td><td>Select attributes</td><td>Visualize</td><td>-</td><td>-</td><td>_</td><td></td><td>g 2</td><td></td><td></td><td></td><td></td><td></td><td>Change Styles *</td><td>Select</td></t<>	🗧 🍼 Forma		Cluster, Associate	Select attributes	Visualize	-	-	_		g 2						Change Styles *	Select
Wind Explorer         Vecka Explorer         Prevoces         Copen Rie         Open Rie <tr< td=""><td>Clipboard</td><td></td><td>-</td><td></td><td>-</td><td>-</td><td></td><td></td><td>A. 11 A</td><td></td><td>× Style</td><td></td><td></td><td></td><td></td><td>G</td><td>Editing</td></tr<>	Clipboard		-		-	-			A. 11 A		× Style					G	Editing
Version of the state of th			Look in:	: 🚺 data					•	2 🕫 🛄 🔤							
With a second			C.		ses 😧 vote	e ather			[[]] Inv	voke options dialog							
With a bit of the second decision of	1	Relation: None		Cpu.with.ver			al		Note:		¥						
Image: Select attribute:       Select attribute:         Vexia Explorer       Image: Select attribute:         Preprocess       Cleasify         Current relation       Remove         All       None         Inverte       Pattribute:         Vexia Explorer       All         None       Inverte         Pattributes       Selected attribute:         None       Inverte         Pattributes       Count         Vexia Explorer       All         None       Inverte         Pattributes       Selected attribute:         None       Inverte         Pattributes       Selected attribute:         None       Inverte         Pattributes       Count         Vexia Service on ontrabasion       Inverte         Pattributes       Count         Vexia Service on ontrabasion       Inverte         Pattributes       Count       Vexia Service on ontrabasion         No       None       Inverte         Pattributes       Class (Monn)       Vaculate All         Open pattributes       Class (Monn)       Vaculate All         Open pattributes       Class (Class (Nonn)       Vaculate All				G glass					option	is which can be custon	ized						
Webs Explorer         Webs Explorer         Preproces         Cassfy Cluster Associate Select attributes Vasalize         Open IRI:       Open URL:         Open IRI:       Open URL:         Open IRI:       Open URL:         Open IRI:       Open URL:         Selected attributes       Selected attribute         None       Invert         Pattor       Selected attribute         Selection of the classion of the classion       Selected attribute         None       Invert       Patter         Open Selection       Selected attribute         Selected attrinute <th></th> <th></th> <th>Desktop</th> <th>iris</th> <th></th> <th></th> <th></th> <th></th> <th>when</th> <th>invoking the options d</th> <th>alog.</th> <th></th> <th></th> <th></th> <th></th> <th></th> <th></th>			Desktop	iris					when	invoking the options d	alog.						
New Explorer         Veka Explorer         Proprocess         Corrent relation         Open Ric			-	ReutersCorn	n-test												
Image: State of the state											-						
Vector				C ReutersGrain	n-train hallenge												
Image: State Stat				Segment-ter													
Version Versio Version Version Versio Version Version Version Version Version V				supermarket	t												
Image: Status       Image: Status																	
Image: International internatinternatintentity international international internat				Files of type:	Arff data files (*.a	arff)				✓ Canc	el xo						
Weeke Explorer     Image: 1       Preprocess     Gassify       Current relation     Apply       Relation: vote     Apply       Current relation     Apply       None     Invert       Pattern     Invert       All     None       Invert     Pattern       Image: 1203/00     Undo       Edit     Selected attribute       None     Invert       Pattern     Invert       Image: 1203/00     Distinct:       Image:			Remove					-									
Image: 1 km/mixture       Image: 1 km/mixture         Image: 1 km/mixture       Image: 1 km/mixture <td></td> <td>Status</td> <td></td>		Status															
A tributes   Open file   Open URL			Explorer						Log	×**							
Weka Explorer       Image: Constraint of the second s	L. L.									)							
Weka Explorer       Image: Constraint of the second s																	
Weka Explorer       Image: Constraint of the second s																	
Weka Explorer   Preprocess   Classify   Open URL																	
Weka Explorer   Preprocess   Classify   Open IRL   Open URL   Open IRL   All   No.   No.   Index open Information   In						_		_			_						
Weka Explorer       Image: Classify Cluster Associate Select attributes Visualize         Open file       Open URL       Open DB       Generate       Undo       Edit       Save         Filter       Choose None       Apply         Current relation       Attributes: 17       Selected attribute       Man: Type: Nominal         Instances: 435       Sum of weights: 435       Attributes: 12       None: Type: Nominal         All       None       Invert       Pattern       No:       Iabel       Count       Weight         1       hand.capped-infants       Type: Nominal       Invert       Pattern       Invert       Invert </td <td>e: 2 of 2 Wor</td> <td>rds: 422 🥳</td> <td></td> <td></td> <td></td> <td></td> <td>_</td> <td></td> <td></td> <td></td> <td></td> <td></td> <td></td> <td>804</td> <td>  ∋   =   140</td> <td>% 😑</td> <td>Ū</td>	e: 2 of 2 Wor	rds: 422 🥳					_							804	∋   =   140	% 😑	Ū
Weka Explorer       Image: Classify Cluster Associate Select attributes Visualize         Open file       Open URL       Open DB       Generate       Undo       Edit       Save         Filter       Choose None       Apply         Current relation       Attributes: 17       Selected attribute       Man: Type: Nominal         Instances: 435       Sum of weights: 435       Attributes: 12       None: Type: Nominal         All       None       Invert       Pattern       No:       Iabel       Count       Weight         1       hand.capped-infants       Type: Nominal       Invert       Pattern       Invert       Invert </th <th></th> <th>😼 🦽 🚸</th> <th>6:35 P</th>																😼 🦽 🚸	6:35 P
Preprocess Classify Cluster Associate Select attributes Visualize Preprocess Classify Cluster Associate Select attributes Visualize Open file Open URL Open DB Generate Undo Edit Save Filter Choose None Current relation Relation: vote Attributes: 17 Instances: 435 Sum of weights: 435 Attributes Attributes Invert Pattern Selected attribute None Invert Pattern I handcapped-infants 2 water project-cost-sharing 3 adoption-of-the-budget-resolution 4 physican-fee-freeze 5 el-salvador-aid 8 ad-to-nicaraguan-contras 9 mx-missie Remove Status																	
Preprocess Classify Cluster Associate Select attributes Visualize  Open file Open URL Open DB Generate Undo Edit Save  Filter  Choose None Current relation Relation: vote Attributes: 17 Instances: 435 Sum of weights: 435 Attributes Attributes Attributes Attributes Attributes Attributes Attributes Invert Pattern  No. Name  I handcapped-infants 2 water project-cost-sharing 3 adoption-of-the-budget-resolution 4 physican-fee affects 5 el-salvador-aid 6 religious groups-in-schools 7 anti-satellite-test-ban 8 add-to-ricaraguan-contras 9 mx-missile  Remove Status	Weka	Explorer															X
Open file       Open DB       Generate       Undo       Edit       Save         Filter       Choose       None       Apply         Current relation       Relation: vote       Attributes: 17       Selected attribute       Name: handicapped-infants       Type: Nominal         Instance: 435       Sum of weights: 435       Missing: 12 (3%)       Distinct: 2       Unique: 0 (0%)         Attributes       In       236       236.0       1 <td< th=""><th>- HCKG</th><th>Explorer</th><th>-</th><th>-</th><th></th><th></th><th></th><th></th><th></th><th></th><th></th><th></th><th></th><th></th><th>_</th><th></th><th></th></td<>	- HCKG	Explorer	-	-											_		
Filter  Choose None  Current relation Relation: voite Attributes: 17 Instances: 435 Sum of weights: 435 Attributes  All None Invert Pattern  No. Name  I handicapped-infants I h	Deere							100									
Filter  Choose None  Current relation Relation: voite Attributes: 17 Instances: 435 Sum of weights: 435 Attributes  All None Invert Pattern  No. Name  I handicapped-infants I h	Preproces	ss Classify	Cluster As	sociate S	elect attri	ibutes	Visualize	-									
Choose     None     Apply       Current relation     Relation: vote     Attributes: 17       Instances: 435     Sum of weights: 435       Attributes     Name: handicapped-infants     Type: Nominal       All     None     Invert       All     None     Invert       Pattern     210     Label       Count     Weight       1     handcapped-infants       2     water-project-cost-sharing       3     adoption of-the-budget-resolution       4     physician-fee-freeze       5     el-salvador-aid       6     religious-groups-in-schools       7     anti-satellite-test-ban       8     aid-to-nicaraguan-contras       9     mx-missle       11     synfuels-corporation-cutback       12     eduction-spending       13     superfund-kinb-to-sue	reproces	ss Classify	Cluster As	sociate S	elect attri	ibutes	Visualize		_								
Choose     None     Apply       Current relation     Relation: vote     Attributes: 17       Instances: 435     Sum of weights: 435       All     None     Invert       All     None     Invert       Pattern     236     236.0       1     handicapped-infants     Type: Nominal       0.     Name     1       1     handicapped-infants     236.0       2     water-project-cost-sharing     187       3     adoption of the-budget-resolution     187       4     physician-fee-freeze     187       5     el-agivador-aid     187       6     religious-groups-in-schools     187       11     synfuels-corporation-cutback     187       12     eduction-spending     187       13     sumerfund-kinktorsus     187	-						Visualize	Genera	ate	Und	0		Edit		S	ave	
Current relation Relation: vote Attributes: 17 Instances: 435 Sum of weights: 435 Attributes Attrib	Oper						Visualize	Genera	ate	Und	0		Edit		S	Gave	
Relation: vote Attributes: 17   Instances: 435 Sum of weights: 435     Attributes     All   No.     Status     Count     Weight:     No.     No.     No.     No.     No.     No. <th></th> <th></th> <th></th> <th></th> <th></th> <th></th> <th>Visualize</th> <th>Genera</th> <th>ate</th> <th>Und</th> <th>0</th> <th></th> <th>Edit</th> <th></th> <th>S</th> <th>ave</th> <th></th>							Visualize	Genera	ate	Und	0		Edit		S	ave	
Relation: vote Attributes: 17   Instances: 435 Sum of weights: 435     Attributes     All   No.     Status     Count     Weight:     No.     No.     No.     No.     No.     No. <td>Oper</td> <td>n file</td> <td></td> <td></td> <td></td> <td></td> <td>Visualize</td> <td>Genera</td> <td>ate</td> <td>Und</td> <td>0</td> <td></td> <td>Edit</td> <td></td> <td>S</td> <td></td> <td>ply</td>	Oper	n file					Visualize	Genera	ate	Und	0		Edit		S		ply
Instances: 435 Sum of weights: 435 Missing: 12 (3%) Distinct: 2 Unique: 0 (0%) Attributes Attributes Attribut	Oper Filter Choos	n file					Visualize	Genera			0		Edit		S		ply
Attributes          All       None       Invert       Pattern         No.       Name       1       n       236       236.0         1       handicapped-infants       2       y       187       187.0         2       water-project-cost-sharing       adoption-of-the-budget-resolution       4       physician-fee-freeze       Class: Class (Nom)       Visualize All         5       el-salvador-aid       e       236       236       236         6       religious-groups-in-schools       e       236       236       236         7       anti-satellite-test-ban       aid-to-inicaraguan-contras       9       mx-missile       187         10       mmigration       1       187       187       187         12       education-spending       e       187       187         13       cuperfund-induct-to-sue       e       187       187         Status	Oper Filter Choos Current r	n file se None				n DB			Selected a	attribute			Edit			App	ply
All     None     Invert     Pattern       1     handicapped-infants       2     water-project-cost-sharing       3     adoption-of-the-budget-resolution       4     physician-fee-freeze       5     el-salvador-aid       6     religious-groups-in-schools       7     anti-satellite-test-ban       8     aid-to-nicaraguan-contras       9     mx-missile       11     synfuels-corporation-cutback       12     education-spending       13     eucerfund-cicht-to-eue	Oper Filter Choos Current r Relati	n file se None relation ion: vote			Oper	n DB	ttributes:	17	Selected a	attribute handicappe	d-infants				/pe: No	App minal	ply
All None Invert Pattern     No. Name     1 handicapped-infants   2 water-project-cost-sharing   3 adoption-of-the-budget-resolution   4 physician-fee-freeze   5 el-salvador-aid   6 religious-groups-in-schools   7 anti-satellite-test-ban   8 aid-to-nicaraguan-contras   9 mx-missile   10 mmigration   11 synfuels-corporation-cutback   12 education-spending   13 superfund-cidbt-to-sue   Status	Oper Filter Choos Current r Relati Instanc	n file None relation ion: vote ces: 435			Oper	n DB	ttributes:	17	Selected a	attribute handicappe	d-infants				/pe: No	App minal	ply
No. Name   1 handicapped-infants   2 water-project-cost-sharing   3 adoption-of-the-budget-resolution   4 physician-fee-freeze   5 el-salvador-aid   6 religious-groups-in-schools   7 anti-satellite-test-ban   8 aid-to-nicaraguan-contras   9 mx-missile   10 immigration   11 synfuels-corporation-cutback   12 education-spending   13 eunerfund-cicht-to-sue	Oper Filter Choos Current r Relati Instanc	n file None relation ion: vote ces: 435			Oper	n DB	ttributes:	17	Selected a Name: Missing:	attribute handicappe 12 (3%)	d-infants	Distinct:	2		/pe: No que: 0 (	App minal (0%)	ply
1 handicapped-infants   2 water-project-cost-sharing   3 adoption-of-the-budget-resolution   4 physician-fee-freeze   5 el-salvador-aid   6 religious-groups-in-schools   7 anti-satellite-test-ban   8 aid-to-nicaraguan-contras   9 mx-missile   10 immigration   11 synfuels-corporation-cutback   12 education-spending   13 superfund-ciobt-to-sue	Oper Filter Choos Current r Relati Instanc Attribute	n file None relation ion: vote res: 435 es			Oper	n DB	ttributes: fweights:	17	Selected a Name: Missing:	attribute handicappe 12 (3%) Label	d-infants	Distinct: Cou	2		/pe: No que: 0 ( Weight	App minal (0%)	ply
1 handicapped-infants   2 water-project-cost-sharing   3 adoption-of-the-budget-resolution   4 physician-fee-freeze   5 el-salvador-aid   6 religious-groups-in-schools   7 anti-satellite-test-ban   8 aid-to-nicaraguan-contras   9 mx-missile   10 immigration   11 synfuels-corporation-cutback   12 education-spending   13 sumerfund-cicht-to-sue	Oper Filter Choos Current r Relati Instanc Attribute	n file None relation ion: vote res: 435 es	Open U	RL	Oper	n DB	ttributes: fweights:	17	Selected a Name: Missing:	attribute handicappe 12 (3%) Label 1 n	d-infants	Distinct: Cou 236	2		ype: No que: 0 ( Weight 236.0	App minal (0%)	ply
2 water-project-cost-sharing   3 adoption-of-the-budget-resolution   4 physician-fee-freeze   5 el-salvador-aid   6 religious-groups-in-schools   7 anti-satellite-test-ban   aid-to-nicaraguan-contras   9 mx-missile   10 immigration   11 synfuels-corporation-cutback   12 education-spending   13 sumerfund-right-to-sue	Oper Filter Choos Current r Relati Instanc Attribute	n file None relation ion: vote res: 435 es	Open U	RL	Oper	n DB	ttributes: fweights:	17	Selected a Name: Missing:	attribute handicappe 12 (3%) Label 1 n	d-infants	Distinct: Cou 236	2		ype: No que: 0 ( Weight 236.0	App minal (0%)	ply
3 adoption-of-the-budget-resolution   4 physician-fee-freeze   5 el-salvador-aid   6 religious-groups-in-schools   7 anti-satellite-test-ban   aid-to-inicaraguan-contras   9 mx-missile   10 immigration   11 synfuels-corporation-cutback   12 education-spending   13 sumerfund-tidbt-to-sue   Status	Oper Filter Choos Current r Relati Instanc Attribute	n file se None relation ion: vote ces: 435 es All	Open U	RL	Oper	n DB	ttributes: fweights:	17	Selected a Name: Missing:	attribute handicappe 12 (3%) Label 1 n	d-infants	Distinct: Cou 236	2		ype: No que: 0 ( Weight 236.0	App minal (0%)	ply
3 adoption-of-the-budget-resolution   4 physician-fee-freeze   5 el-salvador-aid   6 religious-groups-in-schools   7 anti-satellite-test-ban   aid-to-inicaraguan-contras   9 mx-missile   10 immigration   11 synfuels-corporation-cutback   12 education-spending   13 sumerfund-tidbt-to-sue   Status	Oper Filter Chooss Current r Relati Instanc Attribute	n file se None relation ion: vote ces: 435 es All Name	Open U	RL	Oper	n DB	ttributes: fweights:	17 435	Selected a Name: Missing:	attribute handicappe 12 (3%) Label 1 n	d-infants	Distinct: Cou 236	2		ype: No que: 0 ( Weight 236.0	App minal (0%)	ply
5     el-salvador-aid       6     religious-groups-in-schools       7     anti-satellite-test-ban       8     aid-to-nicaraguan-contras       9     mx-missile       10     immigration       11     synfuels-corporation-cutback       12     education-spending       13     eunerfund-right-to-sue	Oper Filter Choos Current r Relati Instanc Attribute	n file se None relation ion: vote res: 435 es All Name 1 handica	Open Uf None	RL	Oper	n DB	ttributes: fweights:	17 435	Selected a Name: Missing:	attribute handicappe 12 (3%) Label 1 n	d-infants	Distinct: Cou 236	2		ype: No que: 0 ( Weight 236.0	App minal (0%)	ply
3     et-salvador-adu       6     religious-groups-in-schools       7     anti-satellite-test-ban       8     aid-to-nicaraguan-contras       9     mx-missile       10     immigration       11     synfuels-corporation-cutback       12     education-spending       13     superfund-sight-to-supe	Oper Filter Choos Current r Relati Instanc Attribute	n file se None relation ion: vote ces: 435 es All Name 1 handica 2 water-1	None	RL	Oper	n DB	ttributes: fweights:	17 435	Selected a Name: Missing:	attribute handicappe 12 (3%) Label 1 n	d-infants	Distinct: Cou 236	2		ype: No que: 0 ( Weight 236.0	App minal (0%)	ply
6     religious-groups-in-schools       7     anti-satellite-test-ban       8     aid-to-nicaraguan-contras       9     mx-missile       10     immigration       11     synfuels-corporation-cutback       12     education-spending       13     sumerfund-cidbt-to-sue	Oper Filter Chooss Current r Relati Instanc Attribute	n file se None relation ion: vote ces: 435 es All Name 1 handicz 2 water- 3 adoptic 4 physicia	None None	RL	Oper	n DB	ttributes: fweights:	17 435	Selected a Name: Missing: No.	attribute handicappe 12 (3%) Label 1 n 2 y	d-infants	Distinct: Cou 236	2		ype: No que: 0 ( Weight 236.0	App minal 0%) t	
8     aid-to-nicaraguan-contras       9     mx-missile       10     immigration       11     synfuels-corporation-cutback       12     education-spending       13     superfund-cicht-to-sue       Remove	Oper Filter Chooss Current r Relati Instanc Attribute	n file se None relation ion: vote ces: 435 es All Name Name Name Name All Adoptic Gel-salva	None None None noject-cost- no-f-the-bu an-fee-freez ador-aid	RL	Oper	n DB	ttributes: fweights:	17 435	Selected a Name: Missing: No.	attribute handicappe 12 (3%) Label 1 n 2 y	d-infants	Distinct: Cou 236	2		ype: No que: 0 ( Weight 236.0	App minal 0%) t	
9 mx-missile 10 immigration 11 synfuels-corporation-cutback 12 education-spending 13 superfund-right-to-sue Remove	Oper Filter Choos Current r Relati Instanc Attribute	n file None relation ion: vote ces: 435 as All Name 1 handice 2 water-1 3 adoptio 4 physici 5 el-salvz 6 religiou	None None None apped-infant project-cost- project-cost- project-cost- ador-aid s-groups-in-	RL	Oper	n DB	ttributes: fweights:	17 435	Selected a Name: Missing: No.	attribute handicappe 12 (3%) Label 1 n 2 y	d-infants	Distinct: Cou 236	2		ype: No que: 0 ( Weight 236.0	App minal 0%) t	
10     immigration       11     synfuels-corporation-cutback       12     education-spending       13     superfund-sight-to-sue       Remove	Oper Filter Chooss Current r Relati Instanc Attribute	n file None relation ion: vote ces: 435 es All Name 1 handice 2 water - 3 adoptic 4 physici 5 el-salvz 6 el-salvz 6 el-salvz 7 anti-sa	None None None None None None None None	RL	Oper	n DB	ttributes: fweights:	17 435	Selected a Name: Missing: No.	attribute handicappe 12 (3%) Label 1 n 2 y	d-infants	Distinct: Cou 236	2		ype: No que: 0 ( Weight 236.0	App minal 0%) t	
11     synfuels-corporation-cutback       12     education-spending       13     euperfund-right-to-sue       Remove	Oper Filter Chooss Current r Relati Instanc Attribute	n file se None relation ion: vote res: 435 ses All Name 1 handica 2 water- 3 adoptic 4 physicia 5 el-salva 6 religiou 7 anti-sa 8 aid-to-1	None None None None- nof-the-bu an-fee-freez ador-aid s-groups-in- tellite-test-b nicaraguan-co	RL	Oper	n DB	ttributes: fweights:	17 435	Selected a Name: Missing: No.	attribute handicappe 12 (3%) Label 1 n 2 y	d-infants	Distinct: Cou 236	2 nt		ype: No que: 0 ( Weight 236.0	App minal 0%) t	
12     education-spending       13     euperfund-right-to-sue       Remove	Oper Filter Choos Current r Relati Instanc Attribute	n file se None relation ion: vote ces: 435 ses All Name 1 handica 2 water-1 3 adoptic 4 physici 5 el-salva 6 religiou 7 anti-sa 8 aid-to-1 9 mx-mis	None None None None None None None None	RL	Oper	n DB	ttributes: fweights:	17 435	Selected a Name: Missing: No.	attribute handicappe 12 (3%) Label 1 n 2 y	d-infants	Distinct: Cou 236	2 nt		ype: No que: 0 ( Weight 236.0	App minal 0%) t	
13 Remove	Oper Filter Choos Current r Relati Instanc Attribute	n file se None relation ion: vote ces: 435 es All Name 1 handica 2 water-1 3 adoptic 4 physici 5 el-salvz 6 religiou 7 anti-sa 8 aid-to-1 9 mx-mis 10 immigra	None None None None None- the-bu an-fee-freez ador-aid s-groups-in- tellite-test-b nicaraguan-c sile ation	RL	[ Oper	n DB	ttributes: fweights:	17 435	Selected a Name: Missing: No.	attribute handicappe 12 (3%) Label 1 n 2 y	d-infants	Distinct: Cou 236	2 nt		ype: No que: 0 ( Weight 236.0	App minal 0%) t	
Remove Status	Oper Filter Choos Current r Relati Instance Attribute	n file se None relation ion: vote ces: 435 25 All Name 1 handica 2 water-1 3 adoptic 4 physici 5 el-salvz 6 religiou 7 anti-sa 8 aid-to-1 9 mx-mis 10 immigra 11 synfuel	None None None None None- the-bu an-fee-freez ador-aid s-groups-in- tellite-test-b nicaraguan-c sile stion s-corporatio	RL	[ Oper	n DB	ttributes: fweights:	17 435	Selected a Name: Missing: No.	attribute handicappe 12 (3%) Label 1 n 2 y	d-infants	Distinct: Cou 236	2 nt		ype: No que: 0 ( Weight 236.0	App minal 0%) t	
Status	Oper Filter Choos Current r Relati Instanc Attribute	n file se None relation ion: vote ces: 435 es All Name Name Name Name Name Name Name Name	None None None Non-of-the-bu an-fee-freez ador-aid s-groups-in- tellite-test-b nicaraguan-c sile sile sicorporatio ion-spending	RL	[ Oper	n DB	ttributes: fweights:	17 435	Selected a Name: Missing: No.	attribute handicappe 12 (3%) Label 1 n 2 y	d-infants	Distinct: Cou 236	2 nt		ype: No que: 0 ( Weight 236.0	App minal 0%) t	
	Oper Filter Choos Current r Relati Instanc Attribute	n file se None relation ion: vote ces: 435 es All Name Name Name Name Name Name Name Name	None None None Non-of-the-bu an-fee-freez ador-aid s-groups-in- tellite-test-b nicaraguan-c sile sile sicorporatio ion-spending	RL	[ Oper	n DB	ttributes: fweights:	17 435	Selected a Name: Missing: No.	attribute handicappe 12 (3%) Label 1 n 2 y	d-infants	Distinct: Cou 236	2 nt		ype: No que: 0 ( Weight 236.0	App minal 0%) t	
	Oper Filter Choos Current r Relati Instanc Attribute	n file se None relation ion: vote ces: 435 es All Name Name Name Name Name Name Name Name	None None None None None None None None	ks -sharing idget-resolu re -schools van contras	[ Oper	n DB	ttributes: fweights:	17 435	Selected a Name: Missing: No.	attribute handicappe 12 (3%) Label 1 n 2 y	d-infants	Distinct: Cou 236	2 nt		ype: No que: 0 ( Weight 236.0	App minal 0%) t	
	Oper Filter Choos Current r Relati Instanc Attribute	n file se None relation ion: vote ces: 435 es All Name Name Name Name Name Name Name Name	None None None None None None None None	ks -sharing idget-resolu re -schools van contras	[ Oper	n DB	ttributes: fweights:	17 435	Selected a Name: Missing: No.	attribute handicappe 12 (3%) Label 1 n 2 y	d-infants	Distinct: Cou 236	2 nt		ype: No que: 0 ( Weight 236.0	App minal 0%) t	
	Oper Filter Choos Current r Relati Instanc Attribute	n file se None relation ion: vote ces: 435 es All Name Name Name Name Name Name Name Name	None None None None None None None None	ks -sharing idget-resolu re -schools van contras	[ Oper	n DB	ttributes: fweights:	17 435	Selected a Name: Missing: No.	attribute handicappe 12 (3%) Label 1 n 2 y	d-infants	Distinct: Cou 236	2 nt		ype: No que: 0 ( Weight 236.0	App minal 0%) t	
	Oper Filter Choos Current r Relati Instanc Attribute	n file se None relation ion: vote ces: 435 es All Name Name Name Name Name Name Name Name	None None None None None None None None	ks -sharing idget-resolu re -schools van contras	[ Oper	n DB	ttributes: fweights:	17 435	Selected a Name: Missing: No.	attribute handicappe 12 (3%) Label 1 n 2 y	d-infants	Distinct: Cou 236	2 nt	Unic	ype: No que: 0 ( Weight 236.0	App minal 0%) t	≥ All

Goto  $Classify \ \text{tab}$  - click  $Choose \ \text{button} - \text{select} \ SMO \ \text{option}.$ 

🕘 Weka Explorer	
Preprocess Classify Cluster Associate Select attribut Classifier	visualize
<pre>veka classifiers classifi</pre>	-K "weka.classifiers.functions.supportVector.PolyKernel -C 250007 -E 1.0" r output
Status OK	Log ×0
Weka Explorer	
	tes Visualize 1 -W 1 -K "weka.classifiers.functions.supportVector.PolyKernel -C 250007 -E 1.0" lassifier output
<ul> <li>Use training set</li> <li>Supplied test set Set</li> <li>Cross-validation Folds 10</li> <li>Percentage split % 66</li> <li>More options</li> </ul>	
(Nom) Class	
Status OK	Log 💉 x 0

#### Click Start button.



reprocess Classify Cluster Associate	Select attributes Visua	alize						
Classifier								
Choose SMO -C 1.0 -L 0.0010 -P 1.0	DE-12 -N 0 -V -1 -W 1 -K	"weka.classifi	ers.functions.s	upportVector.Pol	yKernel -⊂ 2	50007 -E 1.0"		
Fest options	Classifier output							
O Use training set	Time taken to	build mode	el: 0.05 se	conds				-
<ul> <li>Supplied test set</li> <li>Set</li> <li>Cross-validation</li> <li>Folds 10</li> </ul>	=== Stratified === Summary ==		lidation ==					
Percentage split % 66	Correctly Clas	sified In	stances	418		96.092	\$	
More options	Incorrectly Cl			17		3.908	-	
	Kappa statisti			0.91	78		-	
Nom) Class	Mean absolute			0.03	91			
	Root mean squa	red error		0.19	77			
Start Stop	Relative absol	ute error		8.24	05 %			
Result list (right-click for options)	Root relative	squared en	rror	40.60	18 %			
8:43:04 - functions.SMO	Coverage of ca	ses (0.95	level)	96.09	2 %			
8:43:04 - functions.SMO	Mean rel. regi	on size (	0.95 level)	50	8			
	Total Number o	f Instance	28	435				ſ
	=== Detailed A	ccuracy By	y Class ===					:
		TP Rate	FP Rate	Precision	Recall	F-Measure	ROC Are	28
		0.963	0.042	0.973	0.963	0.968	0.96	
			0.037		0.958			
	Weighted Avg.	0.961	0.04	0.961	0.961	0.961	0.96	
	•	111					Þ	1
tatus								_

# **VIVA-QUESTIONS**

- 1. Support vector machine may be termed as:
- 2. What do you mean by generalization error in terms of the SVM?
- 3. Margin of a hyperplane is defined as:
- 4. In a hard margin support vector machine
- 5. The effectiveness of an SVM depends upon:
- 6. If the hyper plane WTX+b=0 coorectly classifies all the training points (Xi,Yi), where  $Yi=\{+1,-1\}$ , then
- 7. Support vectors are the data points that lie closest to the decision surface
- 8. Consider a two class problem, whose training points are distributed in the figure below. One possible separating hyperplane.
- 9. For a two-class classification problem, we use an SVM classifier and obtain the following separating hyperplane. We have marked 4 instances of the training data. Identify the point which will have the most impact on the shape of the boundary on it's removal.

III B.Tech II Sem

## **Experiment No. 7: Demonstrate performing clustering on data sets**

- i. Load each dataset into WEKA and run simple k-means clustering algorithm with different values of k (number of desired clusters).
- ii. Explore visualization features of WEKA to visualize the clusters. Derive interesting insights and explain.

# CLUSTERING

Cluster analysis or clustering is the assignment of a set of observations into subsets (called clusters) so that observations in the same cluster are similar in some sense. Clustering is a method of unsupervised learning, and a common technique for statistical data analysis used in many fields, including machine learning, data mining, pattern recognition, image analysis and bioinformatics.

# **Types of clustering**

Data clustering algorithms can be hierarchical. Hierarchical algorithms find successive clusters using previously established clusters. These algorithms can be either agglomerative ("bottom-up") or divisive ("top-down"). Agglomerative algorithms begin with each element as a separate cluster and merge them into successively larger clusters. Divisive algorithms begin with the whole set and proceed to divide it into successively smaller clusters.

Partitional algorithms typically determine all clusters at once, but can also be used as divisive algorithms in the hierarchical clustering.

Density-based clustering algorithms are devised to discover arbitrary-shaped clusters. In this approach, a cluster is regarded as a region in which the density of data objects exceeds a threshold. DBSCAN and OPTICS are two typical algorithms of this kind.

Two-way clustering, co-clustering or bi-clustering are clustering methods where not only the objects are clustered but also the features of the objects, i.e., if the data is represented in a data matrix, the rows and columns are clustered simultaneously.

Many clustering algorithms require specification of the number of clusters to produce in the input data set, prior to execution of the algorithm. Barring knowledge of the proper value beforehand, the appropriate value must be determined, a problem for which a number of techniques have been developed.

# *k*-means clustering:

In statistics and machine learning, k-means clustering is a method of cluster analysis which aims to partition n observations into k clusters in which each observation belongs to the cluster with the nearest mean.

# ALGORITHM:

Regarding computational complexity, the *k*-means clustering problem is:

- NP-hard in general Euclidean space *d* even for 2 clusters
- NP-hard for a general number of clusters *k* even in the plane

• If k and d are fixed, the problem can be exactly solved in time  $O(ndk+1 \log n)$ , where n is the number of entities to be cluster

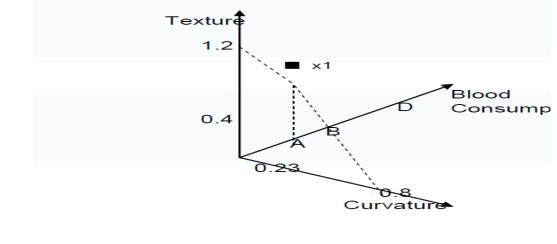
## Example

for example, if our class (decision) attribute is tumor Type and its values are: malignant, benign, etc. - these will be the classes. They will be represented by cluster1, cluster2, etc. However, the class information is never provided to the algorithm. The class information can be used later on, to evaluate how accurately the algorithm classified the objects.

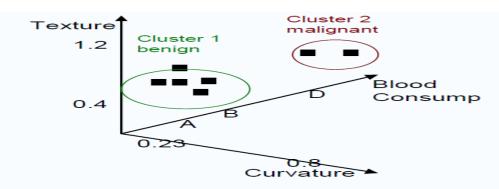
Curvature	Texture	Blood	Tumor	
		Consump	Туре	
0.8	1.2	А	Benign	
0.75	1.4	В	Benign	
0.23	0.4	D	Malignant	
0.23	0.5	D	Malignant	

	Curvature	Texture	Blood	Tumor
			Consump	Туре
x1	0.8	1.2	А	Benign
x2	0.75	1.4	В	Benign
x3	0.23	0.4	D	Malignant
x4	0.23	0.5	D	Malignant

The way we do that, is by plotting the objects from the database into space. Each attribute is one dimension



After all the objects are plotted, we will calculate the distance between them, and the ones that are close to each other – we will group them together, i.e. place them in the same cluster.



**Problem:** Cluster the following eight points (with (x, y) representing locations) into three clusters A1(2, 10) A2(2, 5) A3(8, 4) A4(5, 8) A5(7, 5) A6(6, 4) A7(1, 2) A8(4, 9). Initial cluster centers are: A1(2, 10), A4(5, 8) and A7(1, 2).

The distance function between two points a=(x1, y1) and b=(x2, y2) is defined as:  $\rho(a, b) = |x2 - x1| + |y2 - y1|$ .

Use k-means algorithm to find the three cluster centers after the second iteration. Iteration 1

		(2, 10)	(5, 8)	(1, 2)	
	Point	Dist Mean 1	Dist Mean 2	Dist Mean 3	Cluster
A1	(2, 10)				
A2	(2, 5)				
A3	(8, 4)				
A4	(5, 8)				
A5	(7, 5)				
A6	(6, 4)				
A7	(1, 2)				
A8	(4, 9)				

First we list all points in the first column of the table above. The initial cluster centers – means, are (2, 10), (5, 8) and (1, 2) - chosen randomly. Next, we will calculate the distance from the first point (2, 10) to each of the three means, by using the distance function:

point	mean2
x1, y1	x2, y2
(2, 10)	(5, 8)

 $\rho(a, b) = |x2 - x1| + |y2 - y1|$   $\rho(point, mean2) = |x2 - x1| + |y2 - y1|$  = |5 - 2| + |8 - 10| = 3 + 2 = 5

point	mean3
x1, y1	x2, y2
(2, 10)	(1, 2)

 $\rho(a, b) = |x2 - x1| + |y2 - y1|$   $\rho(point, mean2) = |x2 - x1| + |y2 - y1|$  = |1 - 2| + |2 - 10| = 1 + 8 = 9

So, we fill in these values in the table:

Data Mining-Lab

		(2, 10)	(5, 8)	(1, 2)	
	Point	Dist Mean	Dist Mean	Dist Mean	Cluster
		1	2	3	
A1	(2, 10)	0	5	9	1
A2	(2, 5)				
A3	(8, 4)				
A4	(5, 8)				
A5	(7, 5)				
A6	(6, 4)				
A7	(1, 2)				
A8	(4, 9)				

So, which cluster should the point (2, 10) be placed in? The one, where the point has the shortest distance to the mean – that is mean 1 (cluster 1), since the distance is 0.

Cluster 1 Cluster 2 Cluster 3

(2, 10)

So, we go to the second point (2, 5) and we will calculate the distance to each of the three means,

by using the distance function:

point	mean1
x1, y1	<i>x</i> 2, <i>y</i> 2
(2, 5)	(2, 10)
	x2 - x1  +  y2 - y1  ean1) =  x2 - x1  +  y2 - y1  0 - 5
point	mean2
x1, y1	x2, y2
(2, 5)	(5, 8)
	x2 - x1  +  y2 - y1  ean2) =  x2 - x1  +  y2 - y1  - 5

point mean3 x1, y1 x2, y2 (2, 5) (1, 2)  $\rho$  (a, b) = |x2 - x1| + |y2 - y1|  $\rho$  (point, mean2) = |x2 - x1| + |y2 - y1| = |1 - 2| + |2 - 5| = 1 + 3= 4

So, we fill in these values in the table: Iteration 1

		(2, 10)	(5, 8)	(1, 2)	
	Point	Dist Mean	Dist Mean	Dist Mean	Cluster
		1	2	3	
A1	(2, 10)	0	5	9	1
A2	(2, 5)	5	6	4	3
A3	(8, 4)				
A4	(5, 8)				
A5	(7, 5)				
A6	(6, 4)				
A7	(1, 2)				
A8	(4, 9)				

So, which cluster should the point (2, 5) be placed in? The one, where the point has the shortest distance to the mean – that is mean 3 (cluster 3), since the distance is 0.

Cluster 1 Cluster 2 Cluster 3

(2, 10) (2, 5)

Analogically, we fill in the rest of the table, and place each point in one of the clusters: Iteration 1

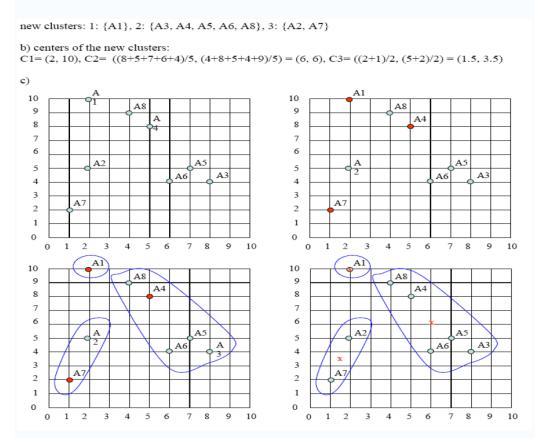
		(2, 10)	(5, 8)	(1, 2)	
	Point	Dist Mean	Dist Mean	Dist Mean	Cluster
		1	2	3	
A1	(2, 10)	0	5	9	1
A2	(2, 5)	5	6	4	3
A3	(8, 4)	12	7	9	2
A4	(5, 8)	5	0	10	2
A5	(7, 5)	10	5	9	2
A6	(6, 4)	10	5	7	2
A7	(1, 2)	9	10	0	3
A8	(4, 9)	3	2	10	2

Cluster 1	Cluster 2	Cluster 3
(2, 10)	(8, 4)	(2, 5)
	(5, 8)	(1, 2)
	(7, 5)	
	(6, 4)	
	(4, 9)	

Next, we need to re-compute the new cluster centers (means). We do so, by taking the mean of all points in each cluster.

For Cluster 1, we only have one point A1(2, 10), which was the old mean, so the cluster center remains the same.

For Cluster 2, we have ((8+5+7+6+4)/5, (4+8+5+4+9)/5) = (6, 6)For Cluster 3, we have ((2+1)/2, (5+2)/2) = (1.5, 3.5)

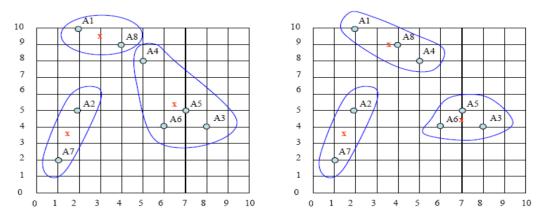


The initial cluster centers are shown in red dot. The new cluster centers are shown in red x. That was Iteration1 (epoch1). Next, we go to Iteration2 (epoch2), Iteration3, and so on until the means do not change anymore.

In Iteration2, we basically repeat the process from Iteration1 this time using the new means we computed.

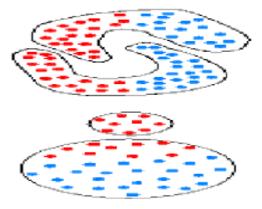
## d)

We would need two more epochs. After the  $2^{nd}$  epoch the results would be: 1: {A1, A8}, 2: {A3, A4, A5, A6}, 3: {A2, A7} with centers C1=(3, 9.5), C2=(6.5, 5.25) and C3=(1.5, 3.5). After the  $3^{rd}$  epoch, the results would be: 1: {A1, A4, A8}, 2: {A3, A5, A6}, 3: {A2, A7} with centers C1=(3.66, 9), C2=(7, 4.33) and C3=(1.5, 3.5).



# **K-means: Problems and Limitations**

Based on minimizing within cluster error -a criterion that is not appropriate for many situations.– Unsuitable when clusters have widely different sizes or have convex shapes.



# K-mediod method:

The *k*-medoids algorithm is a clustering algorithm related to the *k*-means algorithm and the medoidshift algorithm. Both the *k*-means and *k*-medoids algorithms are partitional (breaking the dataset up into groups) and both attempt to minimize squared error, the distance between points labeled to be in a cluster and a point designated as the center of that cluster. In contrast to the *k*-means algorithm *k*medoids chooses datapoints as centers (medoids or exemplars).

# ALGORITHM:

The most common realization of *k*-medoid clustering is the Partitioning Around Medoids (PAM) algorithm and is as follows:

1. Initialize: randomly select k of the n data points as the mediods

2. Associate each data point to the closest medoid. ("closest" here is defined using any valid distance metric, most commonly Euclidean distance, Manhattan distance or Minkowski distance) 3. For each mediod m

1. For each non-mediod data point *o* 

1. Swap *m* and *o* and compute the total cost of the configuration

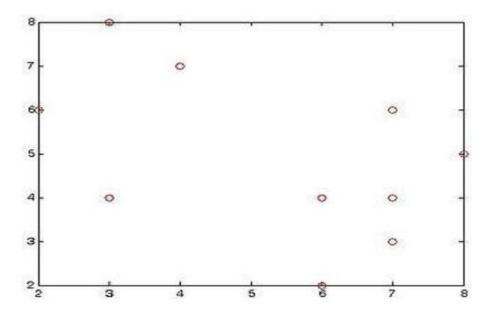
4. Select the configuration with the lowest cost.

5. Repeat steps 2 to 5 until there is no change in the medoid.

# EXAMPLE:

Cluster the following data set of ten objects into two clusters i.e k = 2. Consider a data set of ten objects as follows:

0110 11 5.		
X1	2	6
X2	2 3	4
X3	3	8
X4	4	7
X5	6	2
X6	6	4
X7	7	3
X8	7	4
X9	8	5
X10	7	6



# Step 1

Initialise *k* centre

Let us assume c1 = (3,4) and c2 = (7,4)

So here c1 and c2 are selected as medoid.

Calculating distance so as to associate each data object to its nearest medoid. Cost is calculated using Minkowski distance metric with r = 1.

C1		Data Objects (X <sub>i</sub> )		Cost (distance)
		(.	$\Delta_i$	(uistance)
3	4	2	6	3
3	4	3	8	4
3	4	4	7	4
3	4	6	2	5
3	4	6	4	3
3	4	7	3	5
3	4	8	5	6
3	4	7	6	6

C2		Data Objects		Cost
		$(X_i)$		(distance)
7	4	2	6	7
7	4	3	8	8
7	4	4	7	6
7	4	6	2	3
7	4	6	4	1
7	4	7	3	1
7	4	8	5	2
7	4	7	6	2

Then so the clusters become:

Cluster1 = {(3,4)(2,6)(3,8)(4,7)}

Cluster2 = {(7,4)(6,2)(6,4)(7,3)(8,5)(7,6)}

Since the points (2,6) (3,8) and (4,7) are close to c1 hence they form one cluster whilst remaining points form another cluster.

So the total cost involved is 20.

Where cost between any two points is found using formula

 $cost(x,c)=\Sigma|x-c|$  from i=1 to d

where x is any data object, c is the medoid, and d is the dimension of the object which in this case is 2.

Total cost is the summation of the cost of data object from its medoid in its cluster so here:

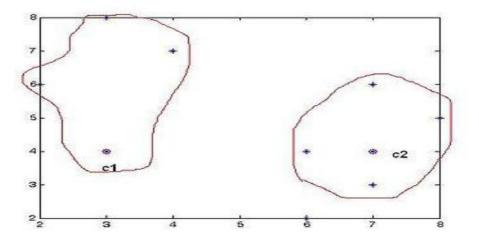
 $Totaol cost = \{ cost((3,4),(2,6)) + cost((3,4),(3,8)) + cost((3,4),(4,7)) \}$ 

+cost((7,4),(6,2))+cost((7,4),(6,4))+cost((7,4)(7,3)) +cost((7,4),(8,5))+cost((7,4),(7,6))}

=(3+4+4)+(3+1+1+2+2)

-(3+4+4)+(3+1+1+2+2)

=20



# Step 2

Selection of nonmedoid O' randomly

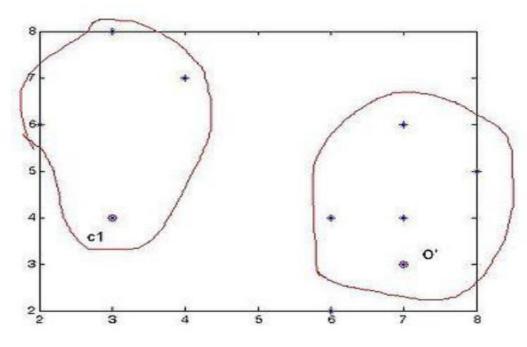
Let us assume O' = (7,3)

So now the medoids are c1(3,4) and O'(7,3)

If c1 and O' are new medoids, calculate the total cost involved By using the formula in the step 1

C1		Data Objects		Cost
		$(X_i)$		(distance)
3	4	2	6	3
3	4	3	8	4
3	4	4	7	4
3	4	6	2	5
3	4	6	4	3
3	4	7	4	4
3	4	8	5	6
3	4	7	6	6

01		Data Objects		Cost
		$(X_i)$		(distance)
7	3	2	6	8
7	3	3	8	9
7	3	4	7	7
7	3	6	2	2
7	3	6	4	2
7	3	7	4	1
7	3	8	5	3
7	3	7	6	3



total cost=3+4+4+2+2+1+3+3 =22 So cost of swapping medoid from c2 to O' is S=current total cost-past total cost =22-20 =2

So moving to O' would be bad idea, so the previous choice was good and algorithm terminates here (i.e there is no change in the medoids).

It may happen some data points may shift from one cluster to another cluster depending upon their closeness to medoid.

1.Open weka tool and click Explorer.

🚱 Weka GUI Chooser	– 🗆 X
Program Visualization Tools Help	
	Applications
	Explorer
WEKA The University of Waikato	Experimenter
And a	KnowledgeFlow
Waikato Environment for Knowledge Analysis	Workbench
Version 3.8.3 (c) 1999 - 2018 The University of Waikato Hamilton, New Zealand	Simple CLI

😅 Weka Explorer	<b>ne</b> in Preprocess tab choose vo		
	Classify Cluster Associate Select attributes		
Open file Filter O	Open URL Open DB Gen	Und	do Edit Save
Choose No			Apply
Current relation		Selected attribute	
Relation: No Instances: No		Name: None Missing: None	Type: None Distinct: None Unique: None
Attributes			
All	None Invert Pattern		
			Visualize All
	Remove		
Status		1	
Welcome to the	Weka Explorer		Log 💉 X
🕘 Open			×
6			
Look In:	data		] [ 🏠 ] [ 🕍 ] [ 🛄 ] 🚇 ]
Ê	Recent	*	
-		rff	Invoke options dialog
	Desktop	L	
	Computer	<b>M</b> ff	
	Local Disk (C:)		Note:
Corn toot		ol off	
Corn-test.:	📄 Program Files	hal.arff	Some file formats offer additional
Corn-train	🚔 Weka-3-8	ric.arff	options which can be customized
			when invoking the options dialog.
	fi data		
	CD Drive (D:)	V	
File Name:	vote.arff		
r no <u>re</u> anno.	roto.am		
Files of Type:	Arff data files (*.arff)		<b>v</b>
- noo or <u>1</u> )po.			
			Open Cancel
🗿 Open			<b>X</b>
- open			
Look In: 📋	data	<b></b>	) 🟠 📋 🔳
			🗌 Inveka antiana dialag
rff	ReutersGrain-test.arff	market.arff	Invoke options dialog
	📄 ReutersGrain-train.arff 🛛 📄 unbala	anced.arff	
-	segment-challenge.arff 🕒 vote.ar	ff	Note:
Corn-test.arff		er.nominal.arff	
			Some file formats offer additional
Corn-train.arff	🕒 soybean.arff 🛛 📄 weathe	er.numeric.arff	options which can be customized
4			when invoking the options dialog.
File <u>N</u> ame:	vote.arff		
. no <u>H</u> ame.			
Files of Type:	Arff data files (*.arff)		<b>v</b>
			<u>Open</u> Cancel

2. click**Open file**...in Preprocess tab.- choose **vote.arff** 

Weka Explorer				
Preprocess Classify Cluster Associate Select attributes	/isualize			
Open file Open URL Open DB Gene	rate	Undo	Edit	Save
filter				
Choose None				Apply
Current relation	Selected a	ttribute		
Relation: vote         Attributes: 17           Instances: 435         Sum of weights: 435		: handicapped : 12 (3%)	-infants Distinct: 2	Type: Nominal Unique: 0 (0%)
Attributes	No.	Label	Count	Weight
		1 n	236	236.0
All None Invert Pattern		2 у	187	187.0
1 handicapped-infants	Class; Cla	ss (Nom)		Visualize Al
2 water-project-cost-sharing 3 adoption-of-the-budget-resolution				
4 physician-fee-freeze	236			
5 el-salvador-aid	230			
6 religious-groups-in-schools			187	
7 anti-satemite-test-ban				
Remove				
Remove				

# Choose cluster tab – click choose button – choose SimpleKmeans

-	Weka Explore	- '					
F	Preprocess	Classify	Cluster	Associate	Select attributes	Visualize	
Clu	isterer						
	🔻 🚞 weka				10 -perioc	ic-pruning 10	0000 -min-density 2.0 -t1 -1.25 -t2 -1.0 -N 2 -A "weka.core.Eu
-		isterers					
CI		Canopy			Cluster	er output	
		Cobweb EM					A
		FarthestFi	ret.				
		FilteredClu					
		Hierarchic		r			
		MakeDens	ityBasedC	lusterer			
		SimpleKM	eans				
-							
Re							
				C	lose		<b>T</b>
St	nus						
							Log x0
•	ж						
_							
0	Weka Explore	er					
_	Weka Explore	Classify	Cluster	Associate	Select attributes	Visualize	
F			Cluster	Associate	Select attributes	Visualize	
F	Preprocess Isterer	Classify			L	1	]]
F	Preprocess	Classify			L	1	0000 -min-density 2.0 -t1 -1.25 -t2 -1.0 -N 2 -A "weka.core.Eu
Ch	Preprocess Isterer Choose	Classify			dates 100 -perioc	ic-pruning 10	]]
Ch	Preprocess Isterer Choose	Classify SimpleKM			dates 100 -perioc	1	]]
Ch	Preprocess Isterer Choose Ister mode Suse train	Classify SimpleKM		0 -max-candi	dates 100 -perioc	ic-pruning 10	]]
Ch	Preprocess Isterer Choose	Classify SimpleKM			dates 100 -perioc	ic-pruning 10	]]
Ch	Preprocess Isterer Choose Ister mode Use train Supplied	Classify SimpleKM ning set I test set		0 -max-candi Set	dates 100 -perioc	ic-pruning 10	]]
Chu	Preprocess Isterer Choose Ister mode Suse train	Classify SimpleKM ing set I test set ge split	leans -init i	0 -max-candi Set %	dates 100 -perioc	ic-pruning 10	]]
Chu	Preprocess Isterer Choose Ister mode Use train Supplied Percenta Classes	Classify SimpleKM ling set l test set ge split to clusters	leans -init i	0 -max-candi Set %	dates 100 -perioc	ic-pruning 10	]]
Ch	Preprocess Insterer Choose Inster mode Use train Use train Use train Classes (Nom)C	Classify SimpleKM hing set I test set to clusters class	leans -init evaluation	0 -max-candi Set %	dates 100 -perioc	ic-pruning 10	]]
Ch	Preprocess Isterer Choose Ister mode Use train Supplied Percenta Classes	Classify SimpleKM hing set I test set to clusters class	leans -init evaluation	0 -max-candi Set %	dates 100 -perioc	ic-pruning 10	]]
Ch	Preprocess Insterer Choose Inster mode Use train Use train Use train Classes (Nom)C	Classify SimpleKM ing set I test set ige split to clusters class sters for vis	evaluation	0 -max-candi Set %	dates 100 -perioc	ic-pruning 10	]]
Ch	Preprocess Insterer Choose Inster mode Use train Use train Use train Classes (Nom)C	Classify SimpleKM ing set I test set ige split to clusters class sters for vis	leans -init evaluation	0 -max-candi Set %	dates 100 -perioc	ic-pruning 10	]]
Ch	Preprocess Insterer Choose Inster mode Use train Use train Use train Classes (Nom)C	Classify SimpleKM ing set i test set to clusters class sters for vis	evaluation	0 -max-candi Set %	dates 100 -perioc	ic-pruning 10	]]
	Preprocess isterer Choose ister mode Use train Supplied Percenta Classes (Nom) C Store du Sta	Classify SimpleKM I test set to clusters Class sters for vis Ignore	evaluation	0 -max-candi Set %	dates 100 -perioc	ic-pruning 10	]]
	Preprocess interer Choose inter mode Use train Use train Use train Chomy Classes (Nom) V Store clu	Classify SimpleKM I test set to clusters Class sters for vis Ignore	evaluation	0 -max-candi Set %	dates 100 -perioc	ic-pruning 10	]]
	Preprocess isterer Choose ister mode Use train Supplied Percenta Classes (Nom) C Store du Sta	Classify SimpleKM I test set to clusters Class sters for vis Ignore	evaluation	0 -max-candi Set %	dates 100 -perioc	ic-pruning 10	]]
	Preprocess isterer Choose ister mode Use train Supplied Percenta Classes (Nom) C Store du Sta	Classify SimpleKM I test set to clusters Class sters for vis Ignore	evaluation	0 -max-candi Set %	dates 100 -perioc	ic-pruning 10	]]
	Preprocess isterer Choose ister mode Use train Supplied Percenta Classes (Nom) C Store du Sta	Classify SimpleKM I test set to clusters Class sters for vis Ignore	evaluation	0 -max-candi Set %	dates 100 -perioc	ic-pruning 10	]]
	Preprocess isterer Choose ister mode Use train Supplied Percenta Classes (Nom) C Store du Sta	Classify SimpleKM I test set to clusters Class sters for vis Ignore	evaluation	0 -max-candi Set %	dates 100 -perioc	ic-pruning 10	]]
	Preprocess isterer Choose ister mode ● Use train ● Supplied ● Percenta ● Classes (Nom) © ■ Store clu Star sult list (right)	Classify SimpleKM I test set to clusters Class sters for vis Ignore	evaluation	0 -max-candi Set %	dates 100 -perioc	ic-pruning 10	]]
Ch	Preprocess isterer Choose ister mode ● Use train ○ Supplied ○ Percenta ○ Classes (Nom) ○ Store clu Star sult list (right)	Classify SimpleKM I test set to clusters Class sters for vis Ignore	evaluation	0 -max-candi Set %	dates 100 -perioc	ic-pruning 10	0000 -min-density 2.0 -t1 -1.25 -t2 -1.0 -N 2 -A "weka.core.E.
Ch	Preprocess isterer Choose ister mode ● Use train ● Supplied ● Percenta ● Classes (Nom) © ■ Store clu Star sult list (right)	Classify SimpleKM I test set to clusters Class sters for vis Ignore	evaluation	0 -max-candi Set %	dates 100 -perioc	ic-pruning 10	]]

109

# Click Start button

	attributes Visualize	
lusterer		
Choose SimpleKMeans -init 0 -max-candidates 10	0 -periodic-pruning 10	0000 -min-density 2.0 -t1 -1.25 -t2 -1.0 -N 2 -A "weka.core.EuclideanDistance -R first-last" -I 500 -num-si
luster mode	Clusterer output	
<ul> <li>Use training set</li> </ul>	Run info	Fmation ===
O Supplied test set Set		
	Scheme:	weka.clusterers.SimpleKMeans -init 0 -max-candidates 100 -periodic-pruning 10
O Percentage split % 66	Relation:	vote
<ul> <li>Classes to clusters evaluation</li> </ul>	Instances:	435
(Nom) Class	Attributes:	17
✓ Store clusters for visualization		handicapped-infants
Store clusters for visualization		water-project-cost-sharing adoption-of-the-budget-resolution
		adoption-of-the-budget-resolution
Ignore attributes		el-salvador-ald
		religious-groups-in-schools
Start Stop		anti-satelite-test-ban
		ald-to-nicaraguan-contras
sult list (right-click for options)		mx-missile
13:40:24 - SimpleKMeans		immigration
13:40:24 - SimpleKMeans		synfuels-corporation-cutback
		education-spending
		superfund-right-to-sue
		crime
		duty-free-exports
		export-administration-act-south-africa
		Class
	Test mode:	evaluate on training data
	Clusteri	ng model (full training set)
atus		
ок		Log

Weka Explor	Classify	Cluster	Associate	Select attribu	es í Visu	alize						-    -
lusterer	,	) <u>C</u>										
Choose	SimpleKM	leans -init 0	-max-cand	idates 100 - per	odic-prupi	ing 10000 -min-de	uneity 2 0 -t1 -1 25 -t	2 -1.0 -N 2 -A "weka	core Euclidea	Distance -R	firet-lact" -L6	500 - num-s
		eena -mit e		dates 100-per	oure-pram	ing rooto -min-de	1101ty 2:0 -11 -1:20 -1	2-110-142-75 Works	.core.corrore ar	instance -re	morroot -re	/00 -mann-a
uster mode				CI	sterer ou	tput						
Ose train	ning set				Means							
Supplied	testset		Set									
					Jumber o	f iterations:	3					
O Percenta			%	66	fithin c	luster sum of	squared errors:	1510.0				
O Classes	to clusters	evaluation										
(Nom) C	Class		<b>T</b>		Initial	starting point	s (random):					
Store clu	sters for vis	sualization					. n. n. y. n. n. n. y.					
							, n, n, y, n, n, n, n, y,					
	lane	re attributes										
	igito	re aurioutes			dissing '	values globall	y replaced with	mean/mode				
Sta	rt		Stop									
					final cl	uster centroid	a:		Cluster#			
esult list (righ	IT-CIICK TOP	options)			ttribut	•		Full Data	Cluster#	1		
13:40:24 - S	impleKMe:	ins			10022040	-		(435.0)	(214.0)	(221.0)		
						ped-infants		n	n	Y		
						oject-cost-sha		У	Y	n		
						-of-the-budget n-fee-freeze	-resolution	y D	n	Y		
					nysicia 1-salva			n	y y	n		
						s-groups-in-sc	hools	, s	×.	n		
						ellite-test-ba		Ŷ	'n	Y		
					aid-to-n	icaraguan-cont	ras	Y	n	Y		
					nx-missi			Y	n	Y		
					mmigrat	ion		У	У	У		
					-							7 -
atus												
OK											Log	ALC: NO.

	Classify Cluster	Associate	Select attributes	Visualize				
usterer		· · · · · · · · · · · · · · · · · · ·		· · · ·				
Choose	impleKMeans -ini	0 -max-candid	ates 100 -period	c-pruning 10000 -min-density 2.0 -t1 -1.25 -t2 -1.0	0 -N 2 -A "weka.core	EuclideanDi	stance -R first-lasť"	-I 500 -num
uster mode			Cluste	erer output				
Use training	qset		phy	sician-fee-freeze	n	Y	n	
	-		el·	-salvador-aid	У	У	n	
<ul> <li>Supplied te</li> </ul>	estset	Set	re	igious-groups-in-schools	У	У	n	
Percentage	solit	% 6		i-satellite-test-ban	У	n	y	
			aid	l-to-nicaraguan-contras	У	n	Y	
O Classes to	clusters evaluatio	n		missile	У	n	У	
(Nom) Cla	ISS	~		nigration	У	У	Y	
Store cluste	ers for visualizatio			fuels-corporation-cutback	n	n	n	
Store cluste	ers for visualization	•		cation-spending	n	Y	n	
				erfund-right-to-sue	У	У	n	
	Ignore attribut	es		me	У	У	n	
	-			y-free-exports	n	n	Y	
Start		Stop		oort-administration-act-south-africa	y democrat repu	У	У	
13:40:24 - Sim	click for options)							
			Ti	ae taken to build model (full training	data) : 0.02 se	conds		
				Model and evaluation on training set				
				stered Instances				
			C11	214 (49%)				
			Cla					
			C10 0 1	214 (49%)				
			C11	214 (49%)				
			C10 0 1	214 (49%)				
atus			C10 0 1	214 (49%)				

Goto - **Visualize** tab- click one box any visualize.

Preprocess	Classify	Cluster	Associate	Select attribu	tes Visualize	1							
ot Matrix	handic	apped-infar	ntswater-pro	oject-cost-aluluj	itipn-of-the-buc	pbysicialutium	freeze el-s	alvador-aid	religious-g	oups-in-e	nticaaltellit	e-test-banid-l	to-nicaraguar
ass												- <b>1</b>	
		X: ha	andicapped	infants Y: Clas	s (click to enlarg	ie)		1			1		1997 - 1997 1997 - 1997
port-adminis	tra												
												- A. (1)	· ·
ty-free-expo	rts												
										1.0			
A 7	-		1	I		1			1	11		1	
tSize: [100]	0		0					Update	ng (uses moi	e memory	)		
	0							Select Attribu	$\equiv$ $-$				
colour: Class	(Nom)							SubSample	%: 100				
lass Colour													
						democrat re	publican						
atus													
ок													. og
													1:47

handicapped-infants (Nom)	Y: Class (Nom)	
olour: Class (Nom)	Select Instance	
Reset Clear Open	Save Jitter	
*	× × × × × × × × × × × × × × × × × × ×	
	× × × × × × × × × × × × × × × × × × ×	5 5 6 7
ss colour		
	democrat republican	

### **VIVA-QUESTIONS**

- 1. A good clustering is one having
- 2. Which of the following is an exploratory data mining technique?
  - A. Classification B. Clustering C. Regression
- 3. Which of the following tasks can be best solved using Clustering.
  - A. Predicting the amount of rainfall based on various cues
  - B. Detecting fraudulent credit card transactions
  - C. Training a robot to solve a maze
- 4. Given two objects represented by the tuples (22, 1, 42, 10) and (20, 0, 36, 8);
- 5.

Compute the Manhattan distance between the two objects

6. Consider a set of five 2-dimensional points p1=(0, 0), p2=(0, 1), p3=(5, 8), p4=(5, 7), and p5=(0, 0.5).
Euclidean distance is the distance function. The k-means algorithm is used to cluster the

points into two clusters. The initial cluster centers are p1 and p4. The clusters after two iterations of k-means are:

- 7. Clustering is a
- 8. Which of the following algorithm is most sensitive to outliers?

K-means clustering algorithm

K-medians clustering algorithm

K-modes clustering algorithm

K-medoids clustering algorithm

9. K means and K-medioids are example of which type of clustering method?

P1 P2 P3 P4				
P1	0	1.000	1.414	1.000
P2	1.000	0	1.000	1.414
P3	1.414	1.000	0	1.000
P4	1.000	1.414	1.000	0

The Euclidean distance matrix between four 2-dimensional points, p1, p2, p3, and p4, is shown below. A possible set of co-ordinate values of these points are:

10. Consider a set of five 2-dimensional points D1=(2,0), D2=(1,3), D3=(3,5), D4=(2,2) and D5=(4,6).

Euclidean distance is the distance function. The k-means algorithm is used to cluster the points into two clusters. The initial cluster centers are D2 and D4. The clusters after two iterations of k-means are:

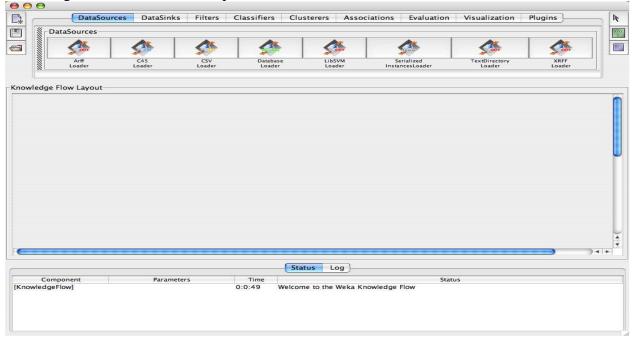
# Experiment No. 8: Demonstrate performing knowledge flow on WEKA

i. Perform pre-processing tasks. ii. Perform decision-tree iii. Perform clustering

# KnowledgeFlow

# Introduction

The KnowledgeFlow provides an alternative to the Explorer as a graphical front end to WEKA's core algorithms. The KnowledgeFlow is a work in progress so some of the functionality from the Explorer is not yet available. On the other hand, there are things that can be done in the KnowledgeFlow but not in the Explorer.



The Knowledge Flow presents a data-flow inspired interface to WEKA. The user can select WEKA components from a tool bar, place them on a layout canvas and connect them together in order to form a knowledge flow for processing and analyzing data. At present, all of WEKA's classifiers, filters, clusters, loaders and savers are available in the Knowledge Flow along with some extra tools.

WEKA there are ten classifiers that can handle data incrementally:

- AODE
- IB1
- IBk
- KStar
- NaiveBayesMultinomialUpdateable
- NaiveBayesUpdateable
- NNge
- Winnow

#### Features

The Knowledge Flow offers the following features:

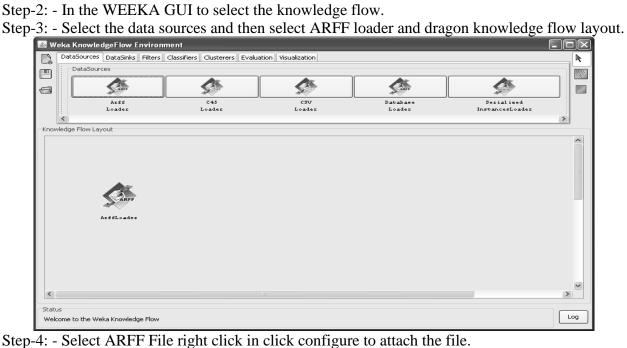
• intuitive data flow style layout

- process data in batches or incrementally
- process multiple batches or streams in parallel (each separate flow executes in its own thread)
- chain filters together
- view models produced by classifiers for each fold in a cross validation
- visualize performance of incremental classifiers during processing (scrolling plots of classification accuracy, RMS error, predictions etc.)
- plugin facility for allowing easy addition of new components to the Knowledge Flow

### **KNOWLEDGE FLOW FOR PRE-PROCESSING** ADD ATTRIBUTE

Step-1: - Click on start button and then select All Programs and choose WEKA 3.8.3 in the WEKA 3.8(with console).

🥥 Weka GUI Chooser	– 🗆 ×
Program Visualization Tools Help	
	Applications
	Explorer
WEKA The University of Waikato	Experimenter
he -	KnowledgeFlow
Waikato Environment for Knowledge Analysis	Workbench
Version 3.8.3 (c) 1999 - 2018 The University of Waikato Hamilton, New Zealand	Simple CLI



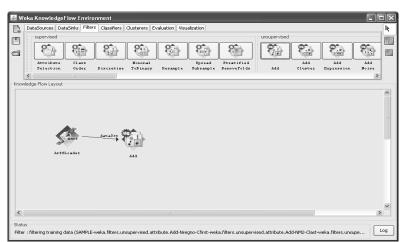
About Reads a sourc	e that is in anff	(attribute relatio	on file format)	format.			More
Look jn:	24-6-14				~	¢ 🕫	
My Recent Documents Desktop My Documents My Computer	👻 isema SEM2 👻 sem21 student						
My Network Places	File <u>n</u> ame: Files of <u>t</u> ype:	student.arff Arff data files				~	 Open Open selecte

Click on open button.

Step-5: - Go to Filters and select unsupervised in the Add Field and dragon knowledge flow layout.



Step-6: - Select ARFF loader mouse right click in the connection to select data set and connection between ARFF loader to Add.

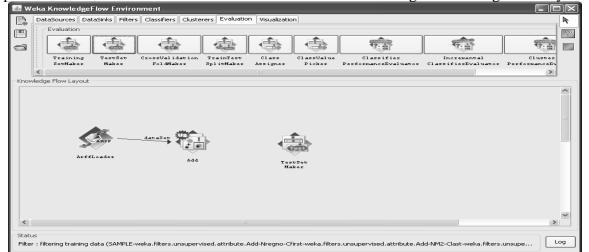


Step-7: - Select Add and mouse right click on in the select configure.

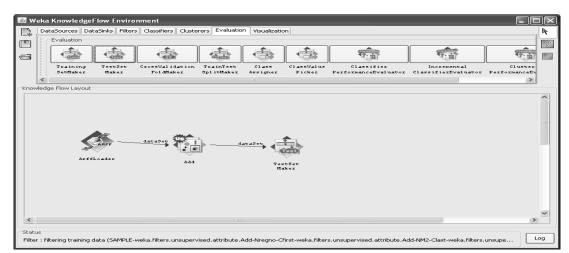
<b>\$</b>	
Choose weka.filters.unsupervised.attribute.Add	
About	
An instance filter that adds a new attribute to the dataset.	More
attributeIndex last	
attributeName total	
nominalLabels	
Open Save OK	Cancel

Click on ok button.

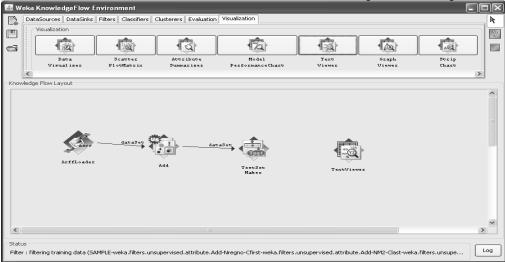
Step-8: - Go to Evaluation menu and select Test set Maker and dragon knowledge flow layout.



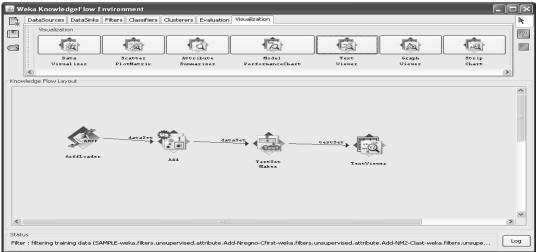
Step-9: - Select Add in the knowledge flow layout and mouse right click in the connection data set and connection between Add to Test Set Maker.



Step-10: - Go to Visualization menu to select Test Viewer and dragon knowledge flow layout.

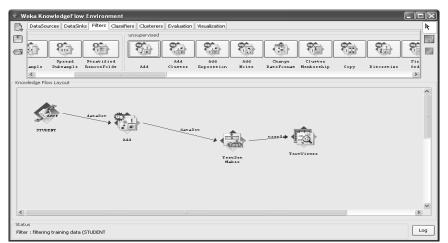


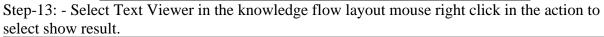
Step-11: - Select Test Set Maker in knowledge flow layout and mouse right click in connection to select test set and connection between test set to Text viewer.



Step-12: - Select ARFF Loader in the knowledge flow layout in the action select start loading.

117





Text Viewer		∎₽
Result list	T Text	
04:38:27 - STUDENT-weka.filter	Grelation STUDENT-weka.filters.unsupervised.attribute.Add-Ntotal-Clast	
	Battribute SNO string	
	Battribute SNAME string	
	Gattribute ENGLISH numeric	
	Ratribute CPDS numeric	
	Ratribute CO numeric	
	Rattribute DSGT numeric	
	Rattribute PsS numeric	
	Gattribute A&FM numeric	
	Attribute CPDS-LAB numeric	
	Rattribute CO-LAB numeric	
	Battribute total numeric	
	Ødata	
	12481F0001,RLAK3HMI,74,75,79,80,85,78,74,80,?	
	12481F0002,BLAKSHMI,74,75,79,80,85,78,74,80,?	
	12481F0003, JEEVAN, 84, 85, 89, 80, 86, 77, 74, 80, ?	
	12481F0004,PHANI,74,88,69,80,65,88,84,80,?	
	12481F0005,PAVANI,74,75,79,80,85,78,74,72,?	
	12481F0006,TRINADH,74,75,79,80,85,78,74,75,2	
	12481F0007,GOUTHAMI,74,75,79,80,85,78,74,80,2	
	12481F0008,LAKSHMI,74,75,79,80,85,78,74,60,2	
	12481F0009,GIRISH,74,75,79,80,85,78,74,70,?	
	12481F0010,GRACE,74,75,79,80,85,78,74,76,?	
	12481F0011,DLAKSHMI,74,75,79,80,85,78,74,75,2	
	12481F0012,SYAM,84,85,79,80,62,88,74,85,?	
	12481F0013,SPARJAN,74,75,79,80,85,78,74,70,?	
	12481F0014,ESWAR,74,75,79,80,85,78,74,70,?	
	12481F0015,SUBHADRA,74,75,69,80,85,78,74,70,?	
	12481F0016,PHANI,74,75,79,60,85,78,74,70,?	
	12481F0017,BRAMAHYA,74,75,89,80,85,78,74,70,?	
	12481F0018,DEEPTHI,74,75,70,80,85,78,74,70,?	
	12481F0019,SRAVYALATHA,74,75,79,80,85,78,74,70,?	
	12481F0020,NAGABABU,74,75,79,80,85,78,74,70,?	
	12481F0021,'MANIKANTA SAI',74,75,79,80,85,78,74,70,?	
	12481F0022,RAMAKRISHNA,74,75,89,80,85,78,74,70,?	
	12481F0023,NAGARJUNA,74,75,79,80,85,78,74,70,?	
	12481F0024,BHULAKSHMI,74,75,89,80,85,78,74,70,?	
	12481F0025,PLAKSHMI,74,75,79,70,85,78,74,70,?	

### Add Expression.

Step-1: - Click on start button and then select All Programs and choose WEKA 3.8.3 in the WEKA 3.8(with console).



Step-2: - In the WEEKA GUI to select the knowledge flow.

Step-3: - Select the data sources and then select ARFF loader and dragon knowledge flow layout.

🛎 w	eka	1 KnowledgeFlow Environ	ment					X
	Da		Classifiers Clusterers Evalua	tion Visualization				R
		DataSources						2
6		<b>A</b>	<b>S</b>	\$	<u>s</u>	SERIEL		
		Arff Loader	C45 Loader	CSV Loader	Database Loader	Serial ised InstancesLoader		
	Ś						>	
Know	led	ge Flow Layout						
								^
								=
		ARFF						
		ArffLoader						
		AffiLoader						
<							>	~
	_							
- Statu Weld		e to the Weka Knowledge Flow						.og

Step-4: - Select ARFF File right click in click configure to attach the file.

÷			
About			
Reads a sourc	e that is in arff (	attribute relation file format) format.	More
Look in:	24-6-14	· Ø P	1
Ì	💎 iisema		
My Recent Documents	🜪 sem21 🜪 student		
G			
Desktop			
My Documents			
My Documents			
My Computer			
My Computer			
My Network	File <u>n</u> ame:	SEM2.ARFF	Open
Places	Files of <u>type</u> :	Arff data files	Cancel

Click on open button. Step-5: Go to Filters and then select Add Expression and dragon knowledge flow layout.

📓 Weka KnowledgeFlow Environment											
DataSources DataSinks Filters Classifiers Clusterers Evaluation Visualization	R										
	2										
Attribute Class Nominal Spread Stratified Add Add Add Selection Order Discretise ToBinary Resample Subsample RemoveFolds Add Cluster Expression Noise											
C Knowledge Flow Layout	<u> </u>										
ArffLoader Add Expression											
Status Welcome to the Weka Knowledge Flow	Log										

Step-6: - Select Add Expression and right click on select configure.

<b>4</b>									
Choose	weka.filters.unsupervised.attribute.AddExpression								
About									
	An instance filter that creates a new attribute by applying a More mathematical expression to existing attributes.								
debug	False	*							
expression	a3+a4+a5+a6+a7								
name	total								
Open	Save OK	Cancel							

Click on ok button.

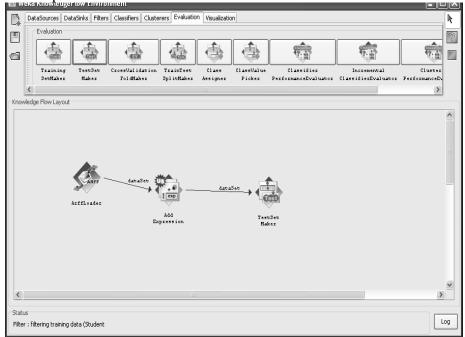
Step-7: - Select ARFF file to connection between the Add Expression to right click on ARFF file to select Data set.

													لكاك
	Da	taSources D	ataSinks Filter	rs Classifiers	Clusterers Ev	aluation Visu	ualization						R
	1	supervised							unsupervi	unsupervised			
													2
		Attribute Selection		Discretize	Nominal ToBinary	Resample	Spread Subsample	Stratified RemoveFolds	Add	Add Cluster	Add Expression	Add Noise	
	: <											>	
~Kno	wled	ge Flow Layou	t										
			ArtéEloader	dara3	Add Expressio	n							< )
<												>	
-Stat Filte		ltering training	) data (Student										.og

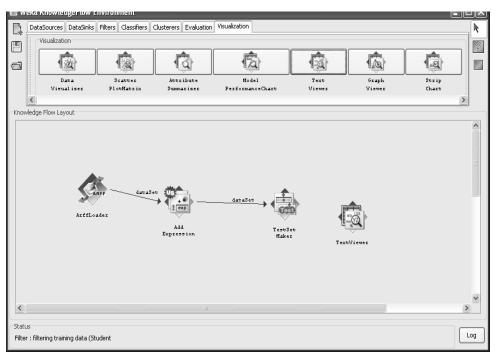
Step-8: - Go to Evaluation and select Test Set Maker drag on knowledge flow layout.

	_		DataSinks Filters	Classifiers Cluste	rers Evaluation	י Visualizati	on				
	ir	Evaluation									2
		Training SetMaker		CrossValidation FoldMaker	TrainTest SplitMaker	Class Assigner	ClassValue Picker	Classifier PerformanceEvaluator	Incremental ClassifierEvaluator	Cluster PerformanceEv	
СКлем	/ledo	ge Flow Layo	ut							7	1
			ARTELoader	data3et Ex	Add pression			r3et ker			III
<										>	~
Statu Filter		ltering trainin	ıg data (Student								Log

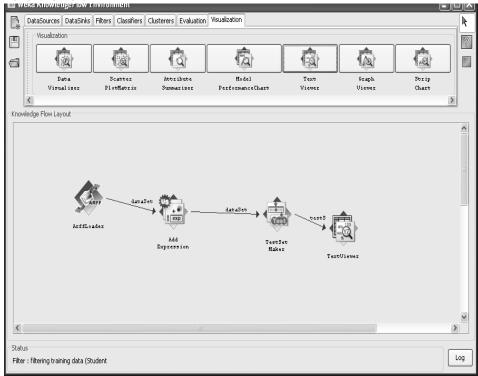
Step-9: - Select Add Expression and right click on to select Data set option to connection between the Add Expression and Test Set Maker.



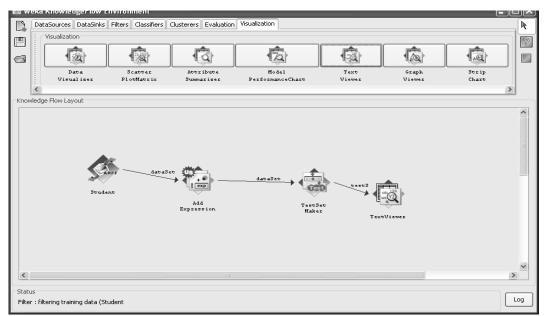
Step-10: - Go to visualization and select Text Viewer to drag on knowledge flow layout.



Step-11: - select Test Set maker to right click on to select test set to connetion between Test Set Maker and Text Viewer.



Step-12: - select ARFF Loader to right click on start loading.



Step-13: - select Text Viewer to right Click on show result.

Text Viewer	
Result list	Text
04:43:55 - STUDENT-weka.filter	$\ensuremath{\emptyset}$ relation STUDENT-weka.filters.unsupervised.attribute.AddExpression-Ea3+a4+a5+a6+a7-Nexpression
	Battribute SNO string Battribute SNAME string
	Rattribute JAVA numeric
	Gattribute COBOL numeric
	@attribute 05 numeric
	@attribute OSHRM numeric
	@attribute OR numeric
	@attribute a3+a4+a5+a6+a7 numeric
	Ødata
	12481F0001, FLAKSHMI, 85, 86, 87, 79, 89, 426
	12481F0002,BLAKSHMI,80,81,82,82,70,395 12481F0003,JEEVAN,79,75,85,86,90,415
	12481F0004, PHAN, 65, 79, 78, 75, 92, 410
	12481F0005, PAVANT, 85, 84, 85, 84, 83, 421
	12481F0006,TRINADH,79,75,76,74,80,384
	12481F0008, LAKSHMI, 78,86,88,75,90,417
	12481F0009,GIRISH,80,80,80,85,85,410
	12481F0010,GRACE,85,86,85,80,85,421
	12481F0011,LAKSHMI,85,81,90,79,80,415
	12481F0012, SYAM, 79,80,75,80,79,393
	12481F0013,S74R74N,95,76,80,75,79,405
	12481F0014,ESWAR,90,79,79,76,80,404 12481F0015,SUBADHRA,85,80,76,80,70,391
	12401F0013,308#DHR4,03,00,70,00,70,391

# Copy Attribute.

Step-1: - Click on start button and then select All Programs and choose WEKA 3.8.3 in the WEEKA 3.8(with console).

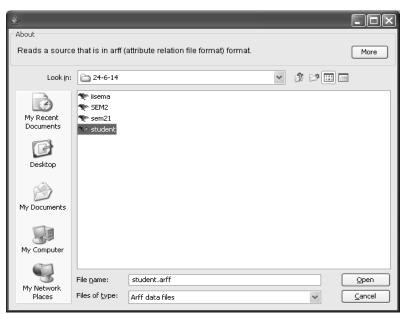


Step-2: - In the WEEKA GUI to select the knowledge flow.

Step-3: - Select the data sources and then select ARFF loader and dragon knowledge flow layout.

🛓 W	/eka	a KnowledgeFlow Environ	ment				-DX
	Da	taSources DataSinks Filters	Classifiers Clusterers Evalua	tion Visualization			₹
		DataSources					
		<b>A</b>	Ś	Ś	<u>Sar</u>	and a second	
		Arff Loader	C45 Loader	CSV Loader	Database Loader	Serialised InstancesLoader	
	: <						>
Knov	wled	ge Flow Layout					
							^
							≡
		A.C.					
		ARFF					
		ArffLoader					
							~
<				1111			>
Stat							Log
Wel	com	e to the Weka Knowledge Flow					

Step-4: - Select ARFF File right click in click configure to attach the file.



# Click on open button.

Step-5: - Go to Filters and then select copy option and dragon knowledge flow layout.

<u>ل</u> ا ھ	eka Knov	wledgeFlow	Environmen	it				0		0		
	DataSour	ces DataSink	s Filters Clar	sifiers Clusterer	rs Evaluation Vi	isualization						k
					unsupervised							
6												8
	¶ominal oBinary	Resample	Spread Subsample	Stratified RemoveFolds	Add	Add Cluster	Add Expression	Add Noise	Change DateFormat	Cluster Membership		Disc
	< ob inary	Resampie	Subsample	Removeroras	aaa	Cluster	Expression	Noise	Daterormat	Membership	Copy	Disc >
Know	wledge Flow	/ Layout										
												^
												≡
			ARFF									
			ArffLoader									
					Сору							
												~
<								_			_	>
State												
Filte	r : filtering t	training data (S	5tudent									Log

Step-6: - select the ARFF Loader and right click to choose Data Set to create connection between ARFF Loader and copy.

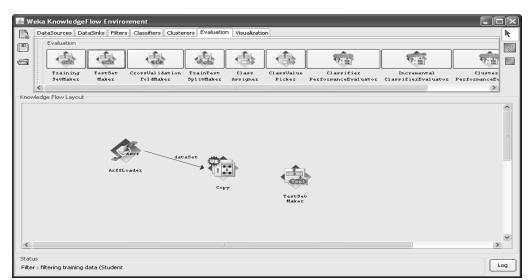
l≝ w	/eka Know	ledgeFlow	Environmen	nt									IX
				ssifiers Clusterer	s Evaluation Vi	isualization							
		_			unsupervised								Ľ
e												R	2
	Nominal oBinary	Resample	Spread Subsample	Stratified RemoveFolds	Add	Add Cluster	Add Expression	Add Noise	Change DateFormat	Cluster Membership	Сору	Disc	
	<	And D Houge a C	JUDSANDIE	Kunorerere	Daa	CA GOVER	DAPIGE TO.		DROGLOUND	theme a service	~~F7	>	
Kno	wledge Flow I	Layout											
			Arff Arff Arff Arff Arff Arff Arff Arff	dat 23	er Lupy Copy								
<												>	*
Stat Filte		raining data (S	itudent									Lc	og

Step-7: - select copy and right click on to choose configure option .

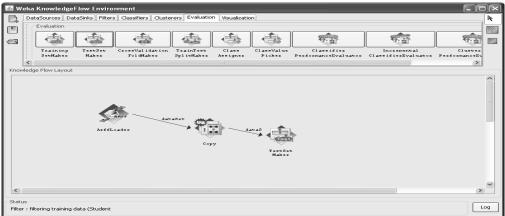
<u>é</u>	- DX
Choose weka.filters.unsupervised.attribute.Copy	
About	
An instance filter that copies a range of attributes in the dataset.	More
attributeIndices 1-2	
invertSelection False	*
Open Save OK	Cancel

Click on ok button.

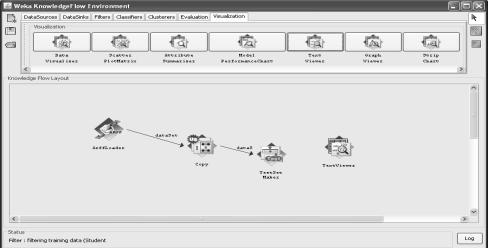
Step-8: - Go to Evaluation and select Test Set Maker drag on to knowledge flow layout.



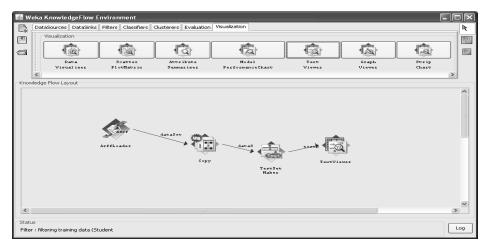
Step-9: - select copy and right click on to choose Data Set to connection between the copy and Test Set Maker.



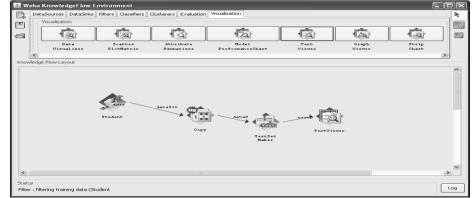
Step-10: - Go to Visualization and select Text Viewer drag on to knowledge flow layout.



Step-11: - select Test Set Maker and right click on to choose Test Set to connection between Test Set Maker to Text Viewer.

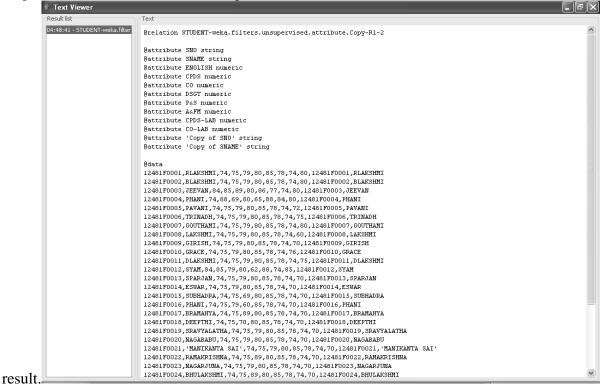








- PX



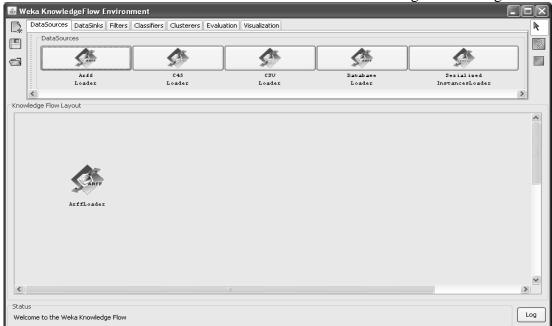
#### Remove Attribute

Step-1: - Click on start button and then select All Programs and choose WEKA 3.8.3 in the WEKA 3.8(with console).



Step-2: - In the WEEKA GUI to select the knowledge flow.

Step-3: - Select the data sources and then select ARFF loader and dragon knowledge flow layout.



Step-4: - Select ARFF File right click in click configure to attach the file.

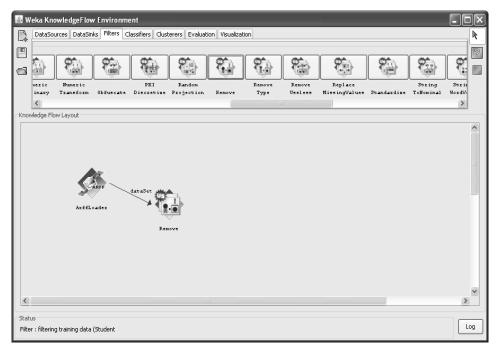
<b>4</b> 0				_		- DX
About						
Reads a sourc	e that is in arπ i	attribute relation	n file format) foi	mat.		More
Look in:	24-6-14			~	ø Þ.	
My Recent Documents	💎 iisema SEM2 💎 sem21					
Desktop	💎 student					
My Documents						
My Computer						
My Network Places	File <u>n</u> ame: Files of <u>t</u> ype:	SEM2.ARFF Arff data files			~	Open Cancel

Click on open button.

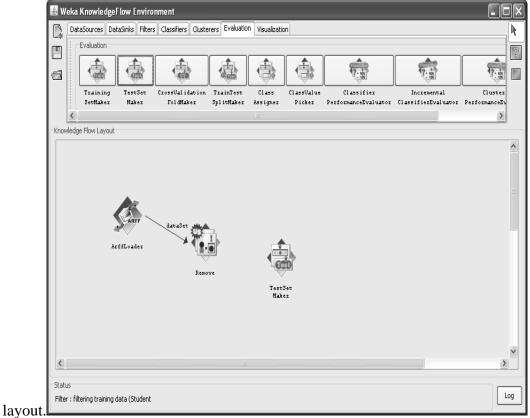
Step-5: - Go to Filters and then select Remove and dragon knowledge flow layout.

🛃 w	/eka Kn	owledgeFlov	w Environme	ent									
	DataSo	urces DataSir	nks Filters C	lassifiers Clust	erers Evaluati	on Visualizat	ion					[	<b>₽</b>
													2
												- CC	
	meric inary	Numeric Transform	Obfuscate	PKI Discretize	Random Projection	Remove	Remove Type	Remove Useless	Replace MissingValues	Standardise	String ToNominal	Stri: WordVe	
	<							1111				>	
Know	vledge Flo	ow Layout											
		Arfel	ARFF oadez	Rem	eve								
<												>	*
Stati Filte		ig training data	(Student									Log	•

Step-6: - select ARFF Loader and right click on to choose Data Set to connection between the ARFF Loader and Remove.





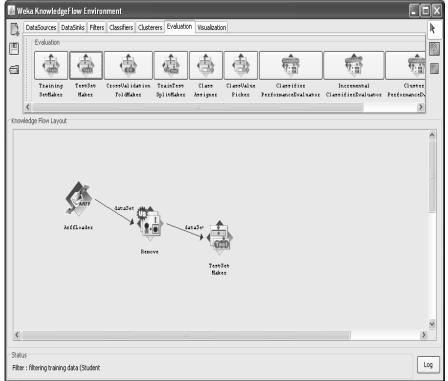


Step-8: -select Remove and right click on to choose configure .

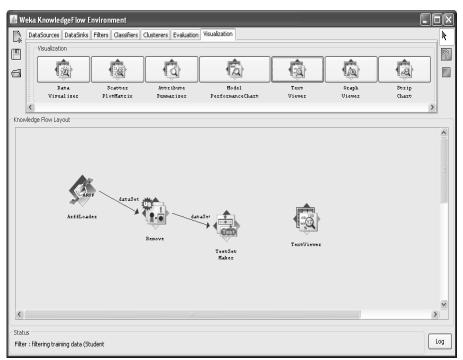
÷ >	<
Choose weka.filters.unsupervised.attribute.Remove	
About	1
An instance filter that removes a range of attributes from the More dataset.	
attributeIndices 1-2	
invertSelection False	
Open Save OK Cancel	)

Click on ok button.

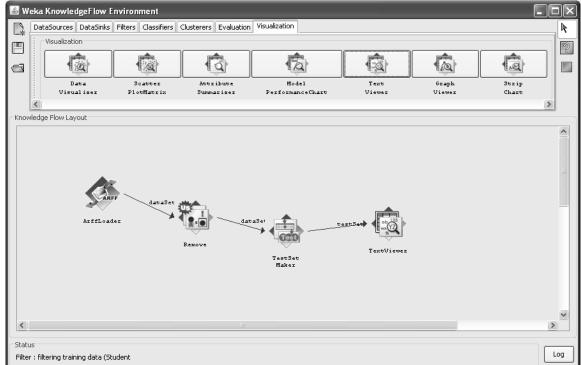
Step-9: - select Remove and right click on to choose Data Set to connection between to Remove and Test Set Maker.



Step-10: -Go to Visualization and select Text Viewer drag on to knowledge flow layout.

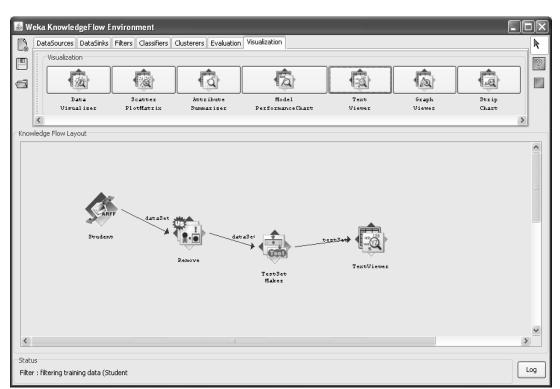


Step-11: - select Test Set Maker and right click on choose Test Set to connection between Test Set Maker to Text Viewer.



Step-12: - Select ARFF Loader and right click on start loading.

134



# Step-13: - select Text Viewer and right click on the show result.

🚔 Text Viewer	
Result list	Text
04:52:32 - STUDENT-weka.filter 04:53:11 - STUDENT-weka.filter	Grelation STUDENT-weka.filters.unsupervised.attribute.Remove-R1-2
	Gattribute JAVA numeric
	Gattribute COBOL numeric
	Gattribute 05 numeric
	Gattribute OSHRM numeric
	Øattribute OR numeric
	Ødata
	85,86,87,79,89
	80,81,82,82,70
	79,75,85,86,90
	86,79,78,75,92
	85,84,85,84,83
	79,75,76,74,80 78,86,88,75,90
	80,80,80,85,85
	65,86,80,85
	85,81,90,79,80
	79,80,75,80,79
	95,76,80,75,79
	90,79,79,76,80
	85,80,76,80,70
	T

### **KNOWLEDGE FLOW FOR DECISSION TREE**

#### Knowledge flow for ID3.

Step 1: open the Notepad and create the ARFF File.

🗖 ID3 - Notepad	
File Edit Format View Help	
<pre>% TITLE:decission Tree:ID3-Algorithm % Date:02-09-2014 % Creator by:M.MANIKANTA SAI @relation ID3 @attribute age {&lt;=30,3140,&gt;40} @attribute student{no,yes} @attribute student{no,yes} @attribute credit_rating{excellent,fair} @attribute buys_computer{yes,no} % % @data &lt;=30,high,no,fair,no &lt;=30,high,no,fair,yes &gt;40,medium,no,fair,yes &gt;40,medium,no,fair,yes &gt;40,low,yes,excellent,no 3140,low,yes,excellent,yes &lt;=30,medium,no,fair,no &lt;=30,medium,yes,fair,yes &gt;40,medium,yes,fair,yes &gt;40,medium,yes,fair,yes &gt;40,medium,yes,fair,yes &gt;40,medium,no,fair,no &lt;=30,medium,yes,fair,yes &gt;40,medium,yes,fair,yes &gt;40,medium,yes,fair,yes &gt;40,medium,no,excellent,yes 3140,high,yes,fair,yes &gt;40,medium,no,excellent,no</pre>	

Step2: To open All Programms  $\rightarrow$  weka3.8.3  $\rightarrow$  weka3.8(with console). Step 3:click on the **Knowledge Flow.** 



Step 4: click on the **Datasources** tab and drag the **Arff Loader**, Right click over the Arff loader and select "configure" from the pop menu.

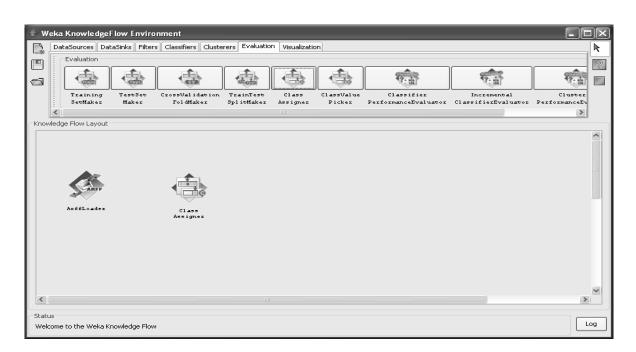
DataSe	ources DataSinks Filters	Classifiers Clusterers Evalu	ation Visualization			
Da	taSources					
	Carer		\$	<b>S</b>		
	Arff Loader	C45 Loader	C3V Loader	Database Loader	Serialized InstancesLoader	_
< edge F	low Layout					>
	10					
	ARFF					
Ar	ffLoader					

Click on open.

40						
About						
Reads a source	e that is in arff	attribute relation	file format) forn	nat.		More
Look in:	🛅 12f21			~	1 🕫 🛄	
My Recent Documents	❤ ID3 ❤ J48					
My Documents						
My Computer						
My Network Places	File <u>n</u> ame: Files of <u>t</u> ype:	ID3.ARFF Arff data files			~	Open Cancel

Step 5:click on the Evaluation tab at top of window, choose class Assigner,

137



### Step 6: Right click on Arff loader and select the "Data Set".

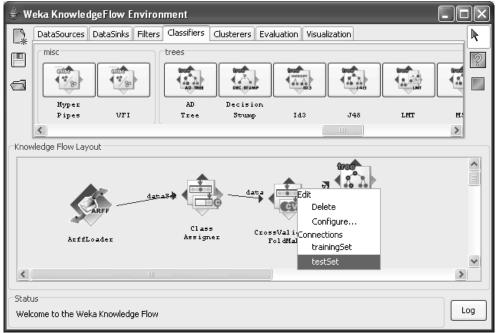
	dgeFlowEnvi DataSinks Fil	ironment ters Classifiers Cluste	rers Evaluation	י Visualizatio		_		- 0
Evaluation Traini SetMak	ng TestSet er Maker	CrossVal idation FoldMaker	TrainTest SplitMaker	Class Assigner	ClassValue Picker	Classifier PerformanceEvaluator	Incremental ClassifierEvaluator	Cluster PerformanceEv
Arfflos		Class Assigner		sAssignerCus hoose class al Nom) buys_co	tribute			
tus			111					

Step 7 : Choose cross validation fold maker, component from the evaluation toolbar Step 8: connect class assigner to cross validation fold maker and right click over class assigner, select dataset under the connection menu.

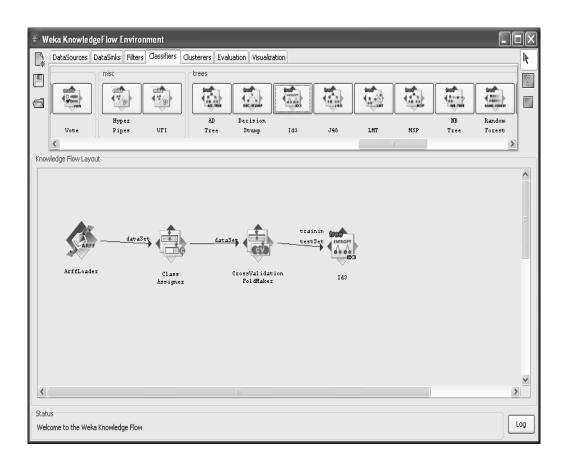
<		aSinks Filter:	s Classifiers Cluste	rers Evaluation	n Visualizatior	1			,	
	Evaluation							~		2
1								17:11	97: m	
	Training SetMaker	TestSet Maker	CrossValidation FoldMaker	TrainTest SplitMaker	Class Assigner	ClassValue Picker	Classifier PerformanceFueluato	Incremental c ClassifierEvaluator	Cluster	
-	Securates	UAKEI	rordnaker	Spi ionaker	ASSIGNED	FICKEL	Performancervaruaco	Classifierryal acor	PerformanceL	
	e Flow Layout									
owledge										
owledge										^
owledge										^
owledge										^
owledge		dataS		dataSet						
owledge	Case -	<u>da</u> taS		dataSey .						
	ArffLoader	dat at								III
	Garer -	data5	Class Assigner	CIO	SSUal idation FoldMakez					
	Garer -	dat at	Cl ass	CIO						
	Garer -	dat = 2	Cl ass	CIO						
	Garer -	dat at	Cl ass	CIO						
	Garer -	data3	Cl ass	CIO						
	Garer -	data	Cl ass	CIO						•

Step 9: click on classifiers tab on the window scroll the toolbar, specify ID3.

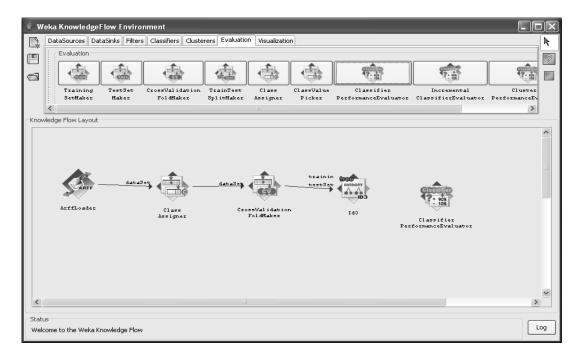
Step 10: connect the cross validation fold maker to ID3 by first choosing training set and then text set for the cross validation fold maker.



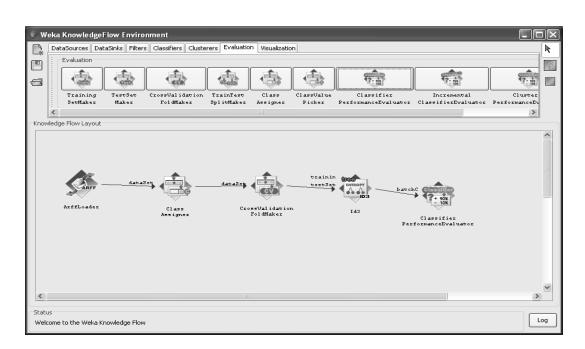
139



Step 11: Go to evaluation tab and place a classifier performance evaluator component on the layout Connect ID3 to this component by selecting the batch classifier entry from the popup menu for ID3.

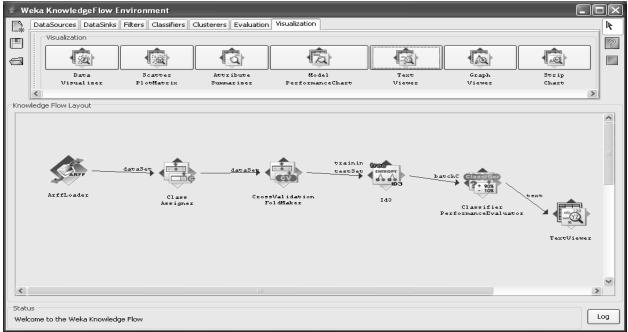


140



Step13: Go to visualization toolbar and place text viewer component on the layout. connect the classifier performance evaluator to the text viewer by selecting the text entry form the popup for classifier performance evaluator.

step14: Now start the flow executing by selecting start loading from the popup menu for artfloader we also see some progress information in the status bar and log at bottom of the window.



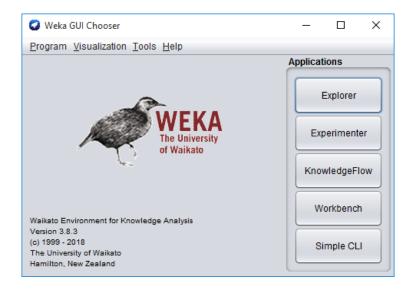
step15: when finished, you can view the results by choosing show results from the popup menu for the test viewer component. Output:

\_ PX

esult list	Text							
):40:23 - Id3	=== Evalu	ation res	ult ===					
	Scheme: 1	:d3						
	Relation: ID3							
	Correctly	· Classifi	ed Instances		10	76.9231 %		
			ied Instanc		3	23.0769 %		
	Kappa sta	tistic			0.4935			
		lute error	<u>-</u>		0.2308			
	Root mear	squared (	error		0.4804			
	Relative	absolute (	error		46.1538 %			
	Root rela	tive squar	ed error		96.0769 %			
	Total Number of Instances 13							
	=== Detailed Accuracy By Class ===							
	TP Rate	FP Rate	Precision	Recall	F-Measure	Class		
	0.875					yes		
	0.6	0.125	0.75	0.6	0.667	no		
	=== Confu	sion Matr:	ix ===					
	ab <-	- classif:	ied as					
	71 a							
	23   b	= no						

# Knowledge flow for decision tree J48.

Step 1: Select knowledge flow in Weka GUI chooser and click on it.

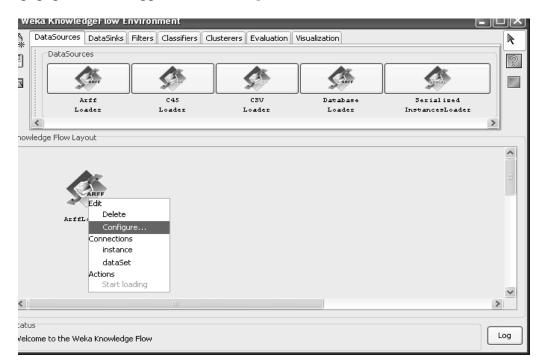


Step 2: click on data source tab and choose arff loader from the tool bar

DataSources DataSinks	Filters Classifiers Clu	usterers Evaluation	Visualization		
DataSources					
				L'ERTAL	
Arff Loader	C45 Loader	CSV Loader	Database Loader	Serialized InstancesLoader	>
ledge Flow Layout					2
ArffLoader					
ArffLoader					~

# Step 3: The arff loader component on arff area clicking somewhere on layout

Step 4: specify j48. arff file to load by right clicking mouse over the arffloader icon on the layout. A popup menu will appear select **configure**.



\$							
About							
Reads a sourc	e that is in arff (	attribute relation file	e format) format.				More
Look in:	🛅 12f21			~	Ð	P 🔛 📟	
My Recent Documents	💎 ID3						
Desktop							
My Documents							
My Computer							
My Network	File <u>n</u> ame:	J48.arff					Open
Places	Files of <u>type</u> :	Arff data files				~	⊆ancel

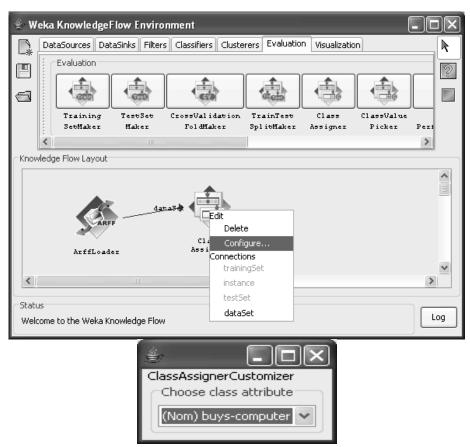
Click on open button.

Step 5: Click on evaluation tab, at the top of window, choose Class Assigner.

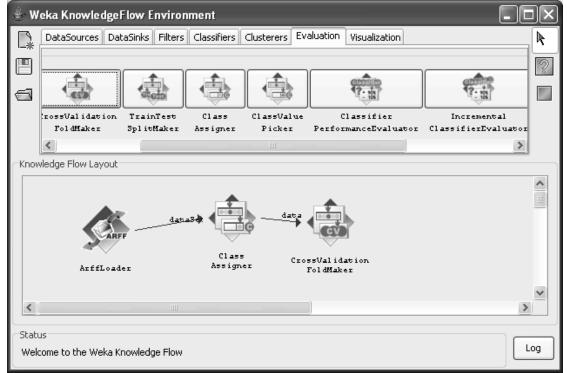
Step6: Now connect arff loader to Class Assigner, first right click over Arffloader and select **dataset** under connection in the menu.

👙 Weka Kno	wledgeFlo	w Enviror	nment			[		<
DataSou	rces DataSi	nks Filters	Classifiers Cluste	rers Evaluation	n Visualization	1	₽	
	uation						2	1
								1
	aining 1	[estSet	CrossValidation	TrainTest	Class	ClassValue		1
	Maker	Maker	FoldMaker	SplitMaker	Assigner	Picker	Peri	
<				]			>	
Knowledge Flov	w Layout							
							^	
	Carry .	<u>dat</u>	-34					
	AREF		~					
	ArffLoader		Class Assigner					
							~	
<							>	
Status								
Welcome to th	e Weka Know	vledge Flow					Log	

Step7:click on class assigner configure menu, choose specify columns is the class in our data given as(buys-computer).



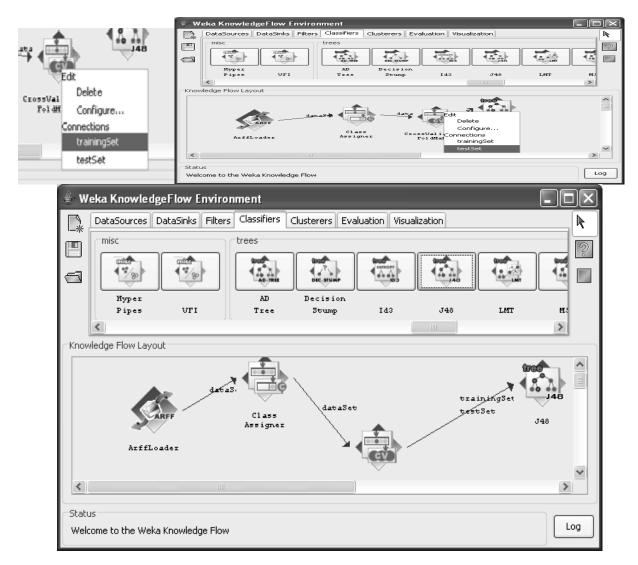
Step 8: Choose **cross validation fold maker**, component from the evaluation toolbar Step 9: connect class assigner to cross validation fold maker and right click over class assigner, select dataset under the connection menu.



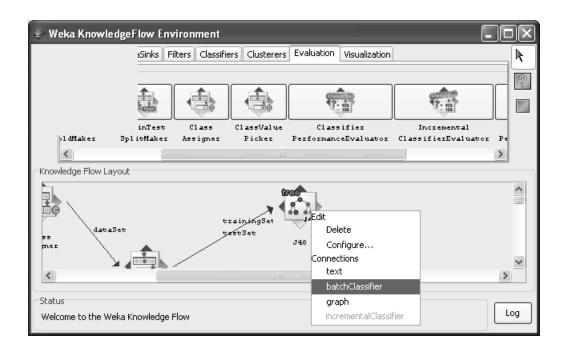
145

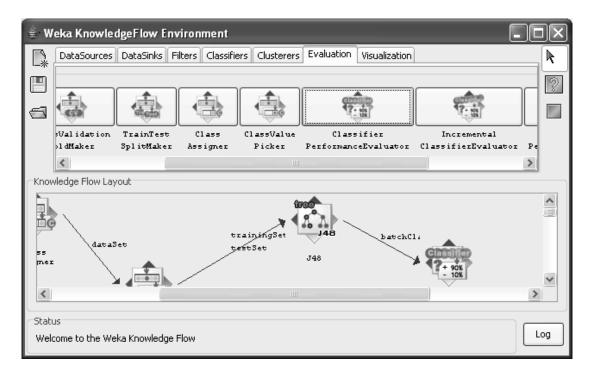
Step 10: click on classifiers tab on the window scroll to the toolbar, specify j48

Step 11: connect the cross validation fold maker to j48 by first choosing **training set** and then **text set** for the cross validation fold maker



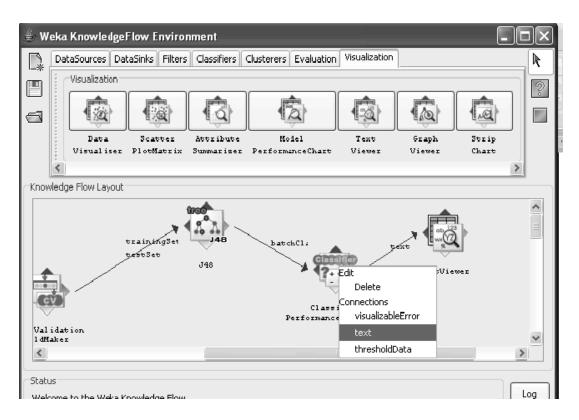
Step 12: Go to evaluation tab and place a classifier performance evaluator component on the layout Connect j48 to this component by selecting the **batch classifier** entry from the popup menu for j48



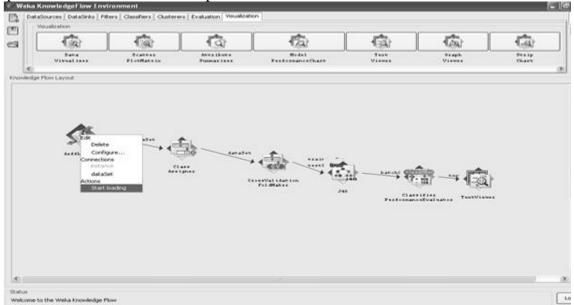


Step 13: Go to visualization toolbar and place **text viewer** component on the layout.connect the classifier performance evaluator to the text viewer by selecting he text entry form the popup for classifier performance evaluator

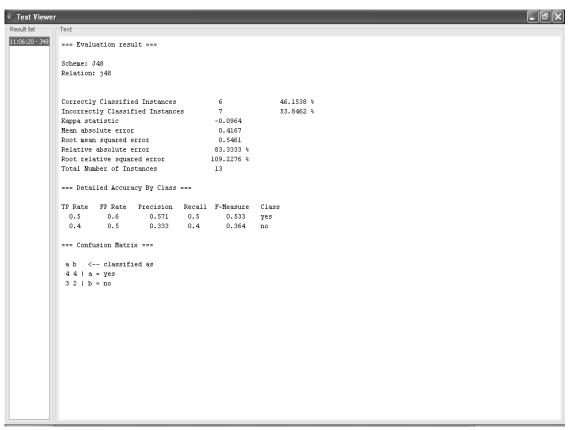
147



step14: Now start the flow executing by selecting **start loading** from the popup menu of arffloader. step15: when finished, you can view the results by choosing show results from the popup menu for the test viewer component



### **OUTPUT:**



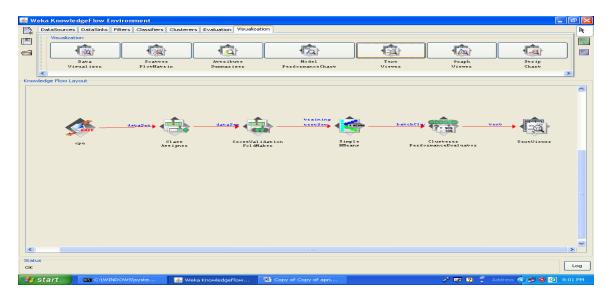
## **KNOWLEDGE FLOW FOR CLUSTERING**

- First start the Knowledge Flow.
- Next click on the Data Sources tab and choose "ArffLoader" from the toolbar (the mouse pointer will change to a "cross hairs").
- Next place the ArffLoader component on the layout area by clicking somewhere on the layout (A copy of the ArffLoader icon will appear on the layout area).
- Next specify an arff file to load by first right clicking the mouse over the ArffLoader icon on the layout. A pop-up menu will appear. Select "Configure" under "Edit" in the list from this menu and browse to the location of your arff file.
- Next click the "Evaluation" tab at the top of the window and choose the "Class Assigner" (allows you to choose which column to be the class) component from the toolbar. Place this on the layout.
- Now connect the ArffLoader to the Class Assigner: first right click over the ArffLoader and select the "data Set" under "Connections" in the menu. A "rubber band" line will appear. Move the mouse over the Class Assigner component and left click - a red line labeled "data Set" will connect the two components.

150

- Next right click over the Class Assigner and choose "Configure" from the menu. This will pop up a window from which you can specify which column is the class in your data (last is the default).
- Next grab a "Cross Validation Fold Maker" component from the Evaluation toolbar and place it on the layout. Connect the Class Assigner to the Cross Validation Fold Maker by right clicking over "Class Assigner" and selecting "dataset" from under "Connections" in the menu.
- Next click the "Clustrer" tab the top of the window and choose the "SimpleKMeans" component from the toolbar. Place this on the layout.
- Connect the CrossValidationFoldMaker to SimpleKMeans by first choosing "training Set" and then "test Set" from the pop-up menu for the CrossValidationFoldMaker
- Next go back to the "Evaluation" tab and place a "ClustererPerformanceEvaluator" component on the layout. Connect SimpleKMeans to this component by selecting the "batch Clusterer" entry from the pop-up menu for simpleKMeans.
- Next go to the "Visualization" toolbar and place a "Text Viewer" component on the layout. Connect the ClustererPerformanceEvaluator to the Text Viewer by selecting the "text" entry from the pop-up menu for ClustererPerformanceEvaluator.
- Now start the flow executing by selecting "Start loading" from the pop-up menu for ArffLoader.
- When finished you can view the results by choosing show results from the pop-up menu for the TextViewer component.

# Output:



ext Viewer		
ult list	Text	
0:00:04 - SimpleKMeans	=== Evaluation result for training instances ===	
	Scheme: SimpleKMeans	
	Relation: cpu	
	kMeans	
	Number of iterations: 17 Within cluster sum of squared errors: 17.681682670930194	
	Cluster centroids:	
	Cluster 0	
	Mean/Mode: 256.7075 1549.2789 7129.6327 10.1088 2.3537 11.9184	
	Std Devs: 285.9676 1349.555 4969.8984 16.307 2.2446 17.1098	
	Cluster 1 Mean/Mode: 49.1707 7862.3415 28345.8537 80.5854 11.9024 40.2927	
	Std Devs: 34.008 6109.2962 14976.0028 57.47 9.1373 38.7132	
	Clustered Instances	
	Clustered instantes	
	Unclustered instances : 188	
	kNeans	
	Number of iterations: 4	
	Within cluster sum of squared errors: 19.50617730173854	
	Cluster centroids:	
	Cluster 0	
	Mean/Mode: 121.8471 3159.9294 12690.6118 27.4824 5 19.4765 Std Devs: 95.3727 4131.1216 11539.21 41.1982 6.3925 24.085	
	Cluster 1	
	Mean/Mode: 855.5556 653.3333 4584 2.8889 0.9444 2.1667	
	Std Devs: 202.491 448.0924 3048.8873 7.7375 0.6391 1.7573	
	C:\WINDOWS\syste 📓 Weka KnowledgeFilo 📓 Text Viewer 🖾 Copy of Copy of apri 🦯 👳 😰 😤	Address < 🗾 🖁 🚺 8:01 PM

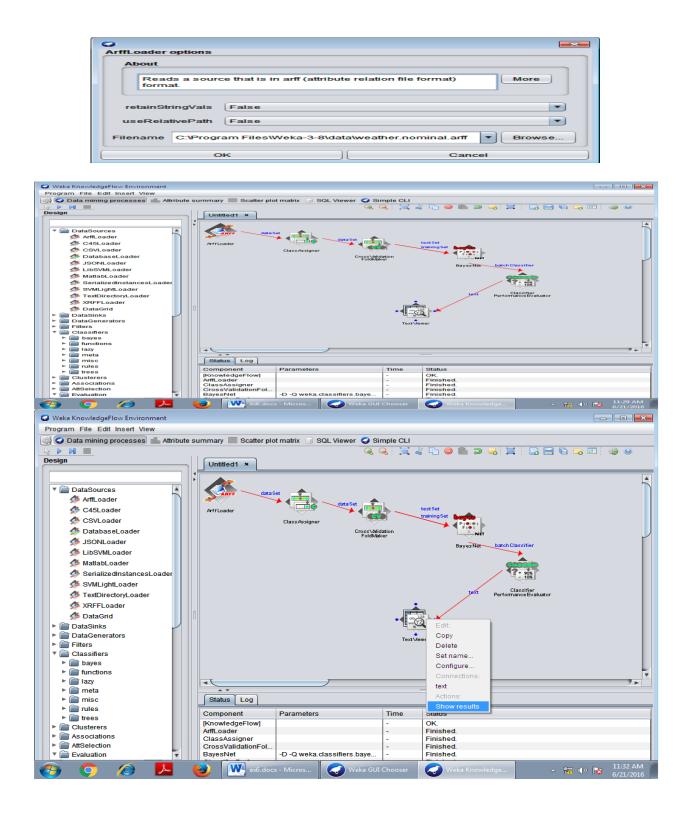
Naïve Bayes

Click Knowledge Flow in Weka GUI Chooser



To create the following and click **Run** button and **Right click** the **TextViwer** and select **Show Result.** 

Draw ArrfLoader and select the filename.



esult list	. 1	Text							
11:22:39.615 - BayesNet		=== Evaluation	result ===						ľ
		Scheme: BayesNe	t						
		Options: -D -Q	weka.class	ifiers.ba	yes.net.sea	rch.local	1.K2P 1	-S BAYE	S-Ewek
		Relation: weat)	er.symboli	.c					
		Correctly Class	ified Inst	ances	8		57.1429	ł	
		Incorrectly Classified Instances		6		42.8571	8		
	n	Kappa statistic		-0.02	44				
	U	Mean absolute error		0.41	5				
		Root mean squared error		0.49	09				
		Relative absolu	te error		87.15	01 %			
		Root relative :	-		99.51	04 %			i i
		Total Number of	Instances	r	14				
		=== Detailed Ad	curacy By	Class ===					
			TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Ar
			0.778	0.800	0.636	0.778	0.700	-0.026	0.622

### **VIVA-QUESTIONS**

1\_\_\_\_\_ is the graphical representation of information and data.

2Which types of chart shows the relationship between a numerical variable and categorical variable?

C. Pie A. Line B. Bar D. x-y plot 3Which of the following are Data sources of Knowledge Flow? A. CSV B. Notepad C. Excel D. Acess 4Data Visualization in mining cannot be done using 5A bubble chart is a variation of 6Which method shows hierarchical data in a nested format? 7Data visualization is also an element of the broader \_ 8Which of the following is a Trees in data visualization A. B-Tree C. Heap-Tree D. AD-Tree B. B+-Tree 9which of the following not support data source

A. ARFF Loader B. CSV LoaderC. Access LoaderD. XML Loader10what is use of class assigner in weka

# ADDITIONAL EXPERIMENTS

### FILE FORMATES FOR WEKA

1. Create CSV(Comma Separated Values) file.

**Step1:** Create an excel file and save with specified format as CSV(Comma Delimited). **Step2:** Now open with notepad and check the values.Here,the fields of data in each row are delimited with a comma and individual rows are separated by new line.

**2.** Create arff(Attribute Relation File Format) file.

**Step1:** Open a notepad and type the data as instructed below:

ARFF files have two distinct sections. The first section is the **Header** information, which is followed the **Data** information.

The **Header** of the ARFF file contains the name of the relation, a list of the attributes (the columns in the data), and their types. An example header on the standard IRIS dataset looks like this:

% 1. Title: Iris Plants Database % % 2. Sources: (a) Creator: R.A. Fisher % % (b) Donor: Michael Marshall (MARSHALL%PLU@io.arc.nasa.gov) % (c) Date: July, 1988 % @RELATION iris @ATTRIBUTE sepallength NUMERIC @ATTRIBUTE sepalwidth NUMERIC @ATTRIBUTE petallength NUMERIC @ATTRIBUTE petalwidth NUMERIC @ATTRIBUTE class {Iris-setosa,Iris-versicolor,Iris-virginica} The **Data** of the ARFF file looks like the following: @DATA 5.1,3.5,1.4,0.2, Iris-setosa 4.9,3.0,1.4,0.2, Iris-setosa 4.7,3.2,1.3,0.2, Iris-setosa 4.6,3.1,1.5,0.2, Iris-setosa

155

5.0,3.6,1.4,0.2,Iris-setosa

5.4,3.9,1.7,0.4, Iris-setosa

4.6,3.4,1.4,0.3, Iris-setosa

5.0,3.4,1.5,0.2, Iris-setosa

4.4,2.9,1.4,0.2, Iris-setosa

4.9,3.1,1.5,0.1, Iris-setosa

Lines that begin with a % are comments. The **@RELATION**, **@ATTRIBUTE** and **@DATA** declarations are case insensitive.

Step 2: Save the file as .arff.Step 3: Open with Weka Explorer and check the file values.

# **3.** Convert CSV to ARFF file format.

Step 1: Open CSV file with notepadStep 2: To fill header and data section in CSV file.Step 3: save the file type as arff.Step 4: Open with Weka Explorer.

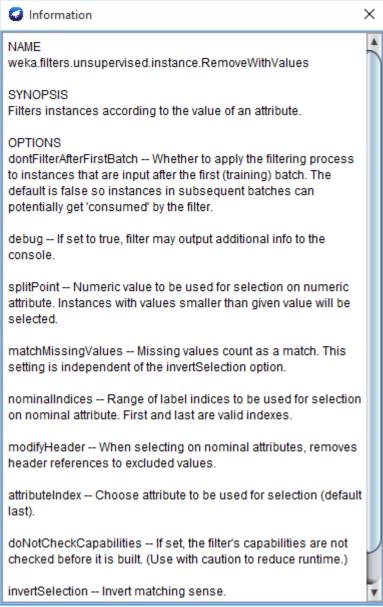
Missing Values

#### Ignore the tuple (Remove Missing Data)

1.Open the Weka Explorer.

2.Load the Student Data set

3.Click the "Choose" button for the filter and select "Remove with values" as under Unsupervised instance.Remove with Values.



4. Set Match Missing value to "True"

😋 weka.gui.GenericObjectEditor X							
weka.filters.unsupervised.instance.RemoveWithValues							
About							
About							
Filters instances according to the value of an attribute.							
	Capabilitie	es					
attributeIndex	2						
debug	False						
doNotCheckCapabilities	False	•					
dontFilterAfterFirstBatch	False						
invertSelection	True						
matchMissingValues	True						
modifyHeader	False						
moullyneader							
nominalIndices	first-last						
splitPoint	0.0						
Open	Save OK Cancel						
	n to use the configuration for the filter. tton to apply the filter click.						
Viewer	stance RemoveWithValues-S0.0-C2-Lfirst-last-V-weka filters.unsupervised instance.Remov	veWithValues-S0 0-C2-L first-last-V-M					
No.         1: SNO         2: SNAME         3: M1         4: M2         5: M           Numeric         Nominal         Numeric         Numeric         Numeric         Numeric           1         1.0         balu         10.0         20.0         30	3 ric .0						
2 2.0 siva 100.0 40.0 50 3 3.0 raja 40.0 40.0 60							

 $\times$ 

Add instance Undo OK Cancel

### Impute Missing Values

How to Impute mean values for missing values

1.Open the Weka Explorer.

2.Load the Student Data set

3.Click the "Choose" button for the filter and select "ReplaceMissingValues" as under

Unsupervised attribute.ReplaceMissingValues.

Information X								
NAME weka.filters.unsupervised.attribute.ReplaceMissingValues								
SYNOPSIS Replaces all missing values for nominal and numeric attributes in a dataset with the modes and means from the training data. The class attribute is skipped by default.								
OPTIONS debug If set to true, filter may output additional info to the console.								
doNotCheckCapabilities If set, the filter's capabilities are not checked before it is built. (Use with caution to reduce runtime.)								
ignoreClass The class index will be unset temporarily before the filter is applied.								
weka.gui.GenericObjectEditor	×							
weka.filters.unsupervised.attribute.ReplaceMissingValues								
About								
in a dataset with the modes and means from the training data.	lore							
debug False								
debug (False	•							
doNotCheckCapabilities False	•							
ignoreClass True	•							
Open         Save         OK         C	ancel							
5. Click the "OK" button to use the configuration for the filter.								

6. Click the "Apply" button to apply the filter click .

1.Open the Weka Explorer.

2.Load the Student Data set

3.Click the "Choose" button for the filter and select "ReplaceMissingWithUserConstant" as under

Unsupervised attribute.ReplaceMissingWithUserConstant.

Information	×
SYNOPSIS Replaces all missing values for nominal, string, numeric and date attributes in the dataset with user-supplied constant values.	
OPTIONS numericReplacementValue The constant to replace missing values in numeric attributes with	
debug If set to true, filter may output additional info to the console.	
nominalStringReplacementValue The constant to replace missing values in nominal/string attributes with	
doNotCheckCapabilities If set, the filter's capabilities are not checked before it is built. (Use with caution to reduce runtime.)	
dateFormat The formatting string to use for parsing the date replacement value	
ignoreClass The class index will be unset temporarily before the filter is applied.	
dateReplacementValue The constant to replace missing values in date attributes with	
attributes Specify range of attributes to act on. This is a comma separated list of attribute indices, with "first" and "last" valid values. Specify an inclusive range with "-". E.g: "first-3,5,6-10,last". Can alternatively specify a comma separated list of attribute names. Note that you can't mix indices and attribute names in the same list	
	•

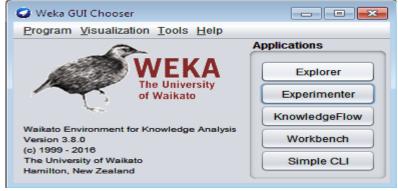
weka.gui.GenericObjectEditor	>	×					
weka.filters.unsupervised.attribute.F	ReplaceMissingWithUserConstant						
About							
Deploase of missing volues for	r pominal atting numeric and data						
	Replaces all missing values for nominal, string, numeric and date attributes in the dataset with user-supplied constant values. Capabilities						
attributes	first-last						
dateFormat	yyyy-MM-dd'T'HH:mm:ss						
dateReplacementValue							
debug	False	J					
doNotCheckCapabilities	False	)					
ignoreClass	True	h					
nominalStringReplacementValue	Nan						
numericReplacementValue	0	٦					
Open Sav	e OK Cancel						

- 5. Click the "OK" button to use the configuration for the filter.6. Click the "Apply" button to apply the filter click .

## HIERARCHICAL CLUSTERING

Hierarchical clustering is a method of cluster analysis which seeks to build a hierarchy of clusters. Strategies for hierarchical clustering generally fall into two types. Agglomerative is a "bottom up" approach: each observation starts in its own cluster, and pairs of clusters are merged as one moves up the hierarchy. Divisive is a "top down" approach: all observations start in one cluster, and splits are performed recursively as one moves down the hierarchy.

1.OpenWeka tool and choose Explorer.



2.Click - Open file... in preprocess tab -choose vote.arff.

G Weka Explorer				
Preprocess Classify Cluster Associate Select attributes	Visualize			
Open file Open URL Open DB Gene	erate	Undo	Edit	Save
Choose None				Apply
Current relation	Selected at	tribute		
Relation: vote         Attributes: 17           Instances: 435         Sum of weights: 435		handicapped 12 (3%)	l-infants Distinct: 2	Type: Nominal Unique: 0 (0%)
Attributes	No.	Label	Count	Weight
	1		236	236.0
All None Invert Pattern	2	У	187	187.0
No. Name				
1 handicapped-infants	Class: Clas	s (Nom)		Visualize All
2 water-project-cost-sharing 3 adoption-of-the-budget-resolution				
4 physician-fee-freeze				
5 el-salvador-aid	236			
6 religious-groups-in-schools			187	
7 anti-satellite-test-ban				
Remove				
Status				
ок				Log 💉 x 0

3.Goto Cluster tab – click choose button - select HierarchicalClusterer.

Preprocess	Classify	Cluster	Associate	Select attri	ibutes	Visualize	]		
usterer									
Choose	Hierarchic	alCluster	er -N 2 -L SIN	IGLE -P -A "	weka.co	re.Euclidea	anDistance	-R first-last"	
uster mode				C	lusterer	output			 
Use train	ing set								
Supplied	testset		Set						
O Percenta	ge split		%	66					
	to clusters	evaluation							
(Nom) (	Class		-						
Store clu	sters for vis	ualization							
	Ignore	attributes							
Sta	4		Stop						
sult list (rigi			Stop						
suit list (rigi	IL-CIICK IOF	opuons)							
									-
atus									

# 5.Click Start button

🖸 Weka Explorer							
Preprocess Classify Cluster Associate Select a	ttributes Visualize						
Clusterer							
Choose HierarchicalClusterer -N 2 -L SINGLE -P -A "weka.core.EuclideanDistance -R first-last"							
Cluster mode	Clusterer output						
<ul> <li>Use training set</li> <li>Supplied test set</li> <li>Set</li> </ul>	Test mode: evaluate on training data						
Percentage split % 66     Classes to clusters evaluation	Clustering model (full training set)						
(Nom) Class ▼ ✓ Store clusters for visualization	Cluster 0 ((((((((((((((((((((((((((((((((((((						
Ignore attributes	Time taken to build model (full training data) : 0.85 second						
Start Stop	=== Model and evaluation on training set ===						
Result list (right-click for options)	Clustered Instances						
14:51:46 - HierarchicalClusterer	0 435 (100%)						
Status OK	Log x0						

6. Visualize the tree by **right click**ing and choose **Visualize Tree** option.

💭 Weka Explorer		
Preprocess Classify Cluster Assoc	ate Select attributes Visualize	
Clusterer Choose HierarchicalClusterer -N 2 -	SINGLE -P -A "weka.core.EuclideanDistance -R first-last"	AaBbCc Heading 2 v Change Styles v Editing
Cluster mode	Clusterer output	La catalig
Use training set     Supplied test set     Set     Percentage split     Classes to clusters evaluation     (Nom) Class     Store clusters for visualization	Test mode: evaluate on training data Clustering model (full training set) Cluster D (((((((((((((((((((((((((((((((((((	
Ignore attributes Start Stor Result list (right-click for options)	Time taken to build model (full training data) : 0.8 Model and evaluation on training set Clustered Instances View in main window Yiew in main window Save result buffer Delete result buffer	5 seco
Status OK	Load model Save model Re-evaluate model on current test set Re-apply this model's configuration Log	×0
Visuali	Visualize cluster assignments Visualize tree	
Page: 4 of 5 Words: 246 3 Dinamalar 6	≽ 🥑 Weka GUI 🕢 Weka Expl 🕅 🚧 es5.docx	E:\dm lab\

🍰 Weka Clusterer Visualize: 14:51:46 - Hierard	chicalClusterer (vote)			
X: Instance_number (Num)	Y: religious-groups-in-schools (Nom)			
Colour: Cluster (Nom)	Select Instance			
Reset Clear Open Save	Jitter O			
Plot: vote_clustered				
Υ ····································	× 434			
Class colour				
cluster0 cluster1				