JEPPIAAR INSTITUTE OF TECHNOLOGY

SELF BELIEF | SELF DISCIPLINE | SELF RESPECT

KUNNAM, SUNGUVARCHATRAM, SRIPERUMPUDHUR, CHENNAI - 631 604

ISO 9001:2015

DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING

CCS341- DATA WAREHOUSING LABORATORY

REGULATION 2021
LAB MANUAL

NAME :

REG NO :

YEAR : III

SEMESTER : 06

INSTITUTE VISION

Jeppiaar Institute of Technology aspires to provide technical education in futuristic technologies with the perspective of innovative, industrial, and social applications for the betterment of humanity.

INSTITUTE MISSION

- To produce competent and disciplined high-quality professionals with the practical skills necessary to excel as innovative professionals and entrepreneurs for the benefit of society.
- To improve the quality of education through excellence in teaching and learning, research, leadership, and by promoting the principles of scientific analysis, and creative thinking.
- To provide excellent infrastructure, serene, and stimulating environment that is most conducive to learning.
- To strive for productive partnership between the Industry and the Institute for research and development in the emerging fields and creating opportunities for employability.
- To serve the global community by instilling ethics, values, and life skills among the students needed to enrich their lives.

Department Vision:

The department will serve as a centre of excellence in practicing, training and implementing AI and AI associated techniques that will enable /support innovative thoughts and ideas across industries and society.

Department Mission:

M1: To collaborate with industry and provide the state of the art infrastructural facilities to meet the global requirements and societal needs for AI.

M2: Promote learning and development of students in Artificial Intelligence thought leadership ,by providing them a suitable infrastructure and environment, enabling them to grow into successful entrepreneurs.

M3: To encourage students to pursue higher education and research in the field of AI.

M4: To impart moral and ethical values in their profession.

PEO'S

PEO1: To equip students with required skills and competencies in AI & Data Science - Data Toolkit - Python & Data Visualization in Python, Regression, Natural Language Processing, Deep Learning and

PEO2: Reinforcement Learning including basic electronics and applied mathematics to further inspire the student's creative and innovative thought process.

PEO3: Enabled students competent and employable by providing excellent Infrastructure to learn and contribute to the industry and for the welfare of the society.

PEO4: To produce professionally ethical individuals with multi-disciplinary skills that will meet the industry expectations

PEO5: Graduates will adapt to evolving technologies through life-long learning.

.

PROGRAM SPECIFIC OUTCOMES (PSOs)

Graduates should be able to:

PSO1: Evolve AI based efficient domain specific processes for effective decision making in several domains such as business and governance domains.

PSO 2: Arrive at actionable Foresight, Insight, hindsight from data for solving business and engineering problems

PSO 3: Create, select and apply the theoretical knowledge of AI and Data Analytics along with practical industrial tools and techniques to manage and solve wicked societal problems

PSO 4: Develop data analytics and data visualization skills, skills pertaining to knowledge acquisition, knowledge representation and knowledge engineering, and hence be capable of coordinating complex projects.

PSO 5: Able to carry out fundamental research to cater the critical needs of the society through cutting edge technologies of AI.

PROGRAM OUTCOMES (PO'S):

Engineering Graduates will be able to:

1. Engineering knowledge:(K3) Apply the knowledge of mathematics, science, engineering fundamentals, and an engineering specialization to the solution of complex engineering problems.

2. Problem analysis: (K4) 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: (K4) 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: (K5) 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: (K3, K5, K6) Create, select, and apply appropriate techniques, resources, and modern engineering and IT tools including prediction and modeling to complex engineering activities with an understanding of the limitations.

6. The engineer and society: (A3) 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: (A2) 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: (A3) Apply ethical principles and commit to professional ethics and responsibilities and norms of the engineering practice.

9. Individual and teamwork: (A3) Function effectively as an individual, and as a member or leader in diverse teams, and in multidisciplinary settings.

10. Communication: (A3) 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: (A3) 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: (A2) 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.

ANNA UNIVERSITY SYLLABUS

CCS341- DATA WAREHOUSING LABORATORY

L	T	P	C
2	0	2	3

1. Data exploration and integration with WEKA
2. Apply weka tool for data validation
3. Plan the architecture for real time application
4. Write the query for schema definition
5. Design data ware house for real time applications
6. Analyse the dimensional Modeling
7. Case study using OLAP
8. Case study using OTLP
9. Implementation of warehouse testing.

TOTAL : 30 PERIODS

CCS341- DATA WAREHOUSING LABORATORY

INDEX

S.No.	Date of Submission	Name of the Experiment	Marks	Signature

Signature of the Staff

CO-PO MAPPING

Course Outcome No.	Course Outcome	Highest Cognitive Level
CO305.1	Design Data Warehouse Architecture for various Problems	K6
CO305.2	Apply the OLAP technology	K3
CO305.3	Analyzing the partitioning strategy	K4
CO305.4	Critically analyze the differentiation of various schema for given problem	K4
CO306.5	Frame roles of process manager and system manager	K4

Course No.	Level of CO	Program Outcomes												Program Specific Outcomes		
		K2	K3	K3	K2	K3	K3	A3	A3	A3	A3	A3	A2	PSO-1	PSO-2	PSO-3
		PO-1	PO-2	PO-3	PO-4	PO-5	PO-6	PO-7	PO-8	PO-9	PO-10	PO-11	PO-12			
CO207.1	K2	3	1	3	3	-	-	-	-	1	1	1	3	2	2	1
CO207.2	K4	2	2	1	3	1	-	-	-	3	2	3	1	2	1	2
CO207.3	A3	2	1	3	1	-	-	-	-	3	3	1	1	2	1	1
CO207.4	A3	2	2	3	1	-	-	-	-	2	3	2	1	2	1	2
CO207.5	K4	3	3	1	3	1	-	-	-	1	3	2	3	3	3	2

CONTENT

DATE:

EX.NO:1

LISTING APPLICATIONS FOR MINING

AIM:

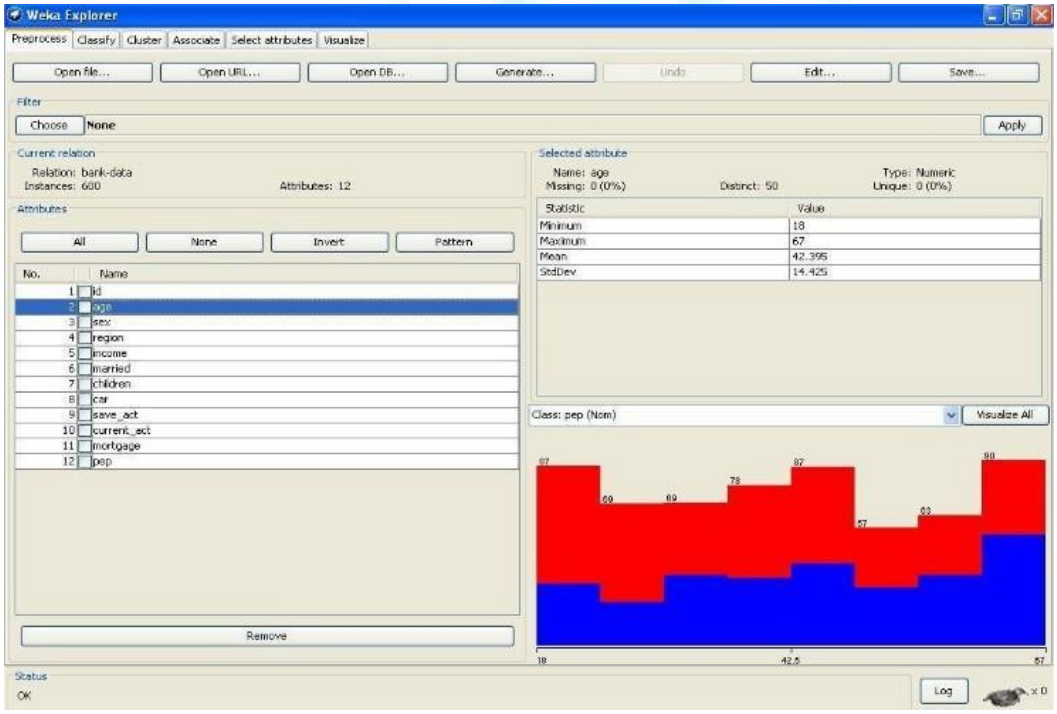
To list all the categorical (or nominal) attributes and the real-valued attributes separately.

RESOURCES: Weka mining tool.

PROCEDURE:

1) Open the Weka GUIChooser.
2) Select EXPLORER present in Applications.
3) Select PreprocessTab.
4) Go to OPEN file and browse the file that is already stored in the system"bank.csv".
5) Clicking on any attribute in the left panel will show the basic statistics on that selectedattribute.1.4

OUTPUT:



RESULT

Thus the listing applications for the data mining was studied.

DATE:

EX.NO:2

FILE FORMAT FOR DATA MINING

AIM: To study the file formats for the data mining.

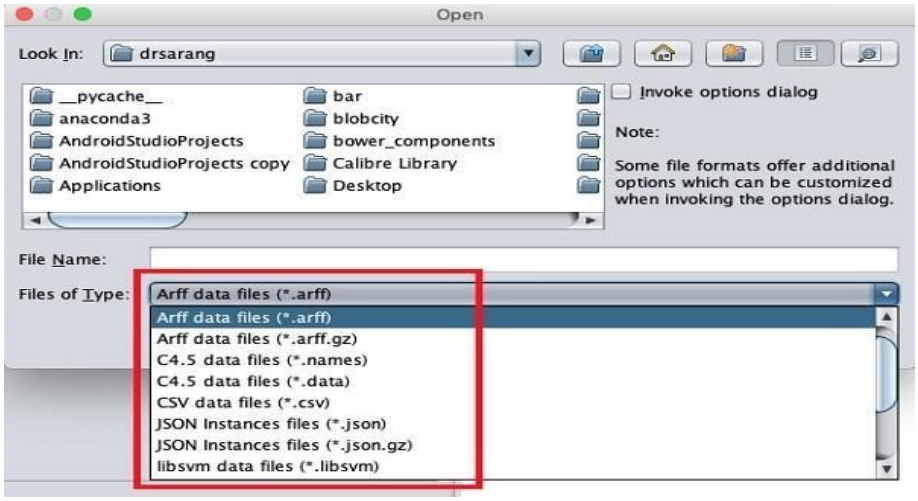
INTRODUCTION:

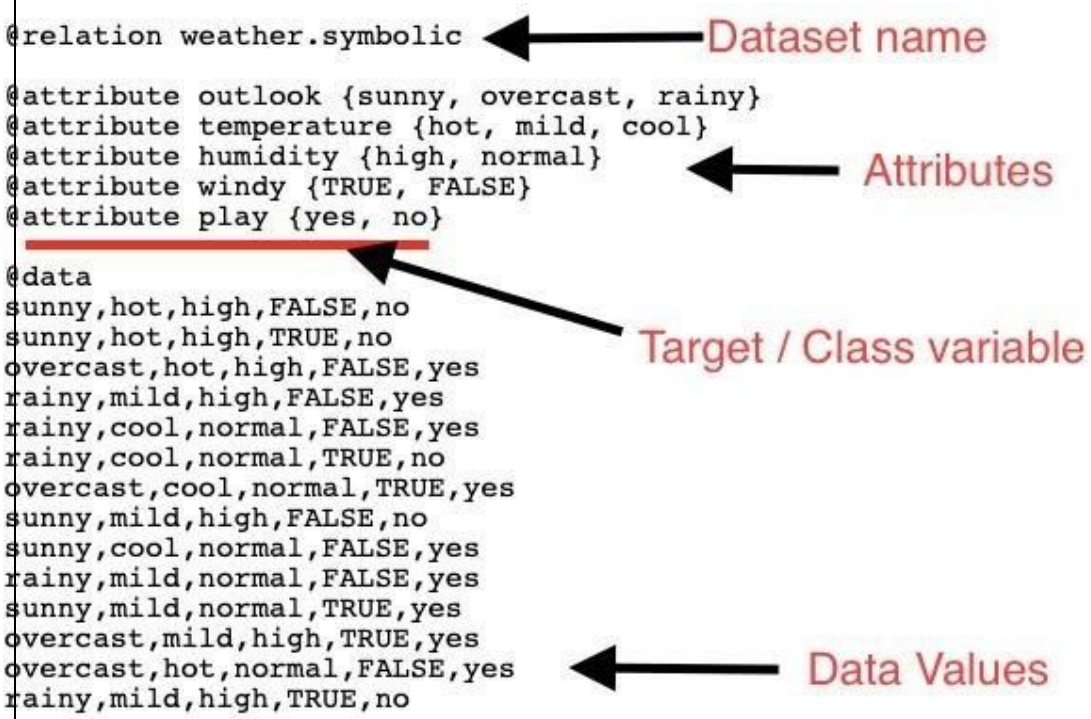
WEKA supports a large number of file formats for the data. The complete list of file formats are given here:

1. arff
2. arff.gz
3. bsi
4. csv
5. dat
6. data
7. json
8. json.gz
9. libsvm
10. m
11. names
12. xrff
13. xrff.gz

The types of files that it supports are listed in the drop-down list box at the bottom of the screen.

OUTPUT:




@relation weather.symbolic
Dataset name
@attribute outlook {sunny, overcast, rainy}
@attribute temperature {hot, mild, cool}
@attribute humidity {high, normal}
@attribute windy {TRUE, FALSE}
@attribute play {yes, no}
Attributes
@data
sunny,hot,high,FALSE,no
sunny,hot,high,TRUE,no
overcast,hot,high,FALSE,yes
rainy,mild,high,FALSE,yes
rainy,cool,normal,FALSE,yes
rainy,cool,normal,TRUE,no
overcast,cool,normal,TRUE,yes
sunny,mild,high,FALSE,no
sunny,cool,normal,FALSE,yes
rainy,mild,normal,FALSE,yes
sunny,mild,normal,TRUE,yes
overcast,mild,high,TRUE,yes
overcast,hot,normal,FALSE,yes
rainy,mild,high,TRUE,no
Target / Class variable
Data Values

RESULT:

Thus the different file formats for the data mining was studied.

DATE:

EX.NO:3 CONVERSION OF TEXT FILE INTO ARFFFILE

AIM:

To convert a text file to ARFF(Attribute-Relation File Format) using Weka3.8.2 tool.

OBJECTIVES:

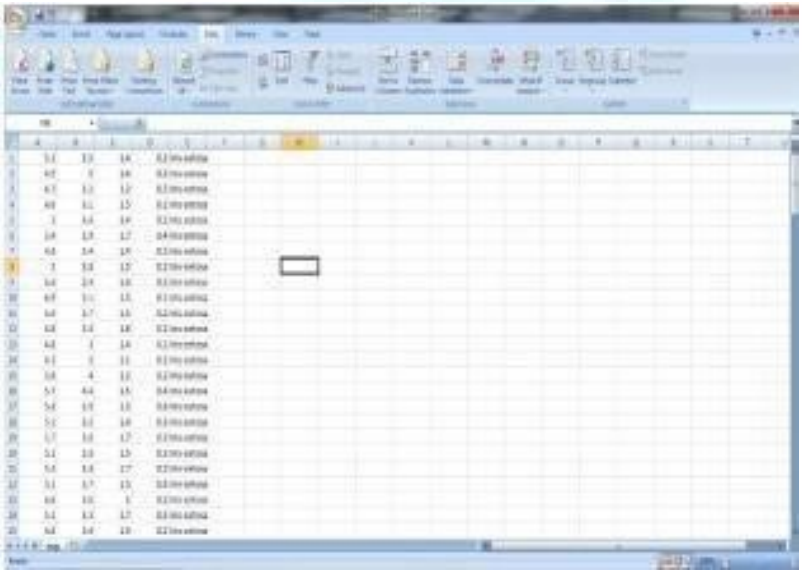
Most of the data that we have collected from public forum is in the text format that cannot be read by Weka tool. Since Weka (Data Mining tool) recognizes the data in ARFF format only we have to convert the text file into ARFF file.

ALGORITHM:

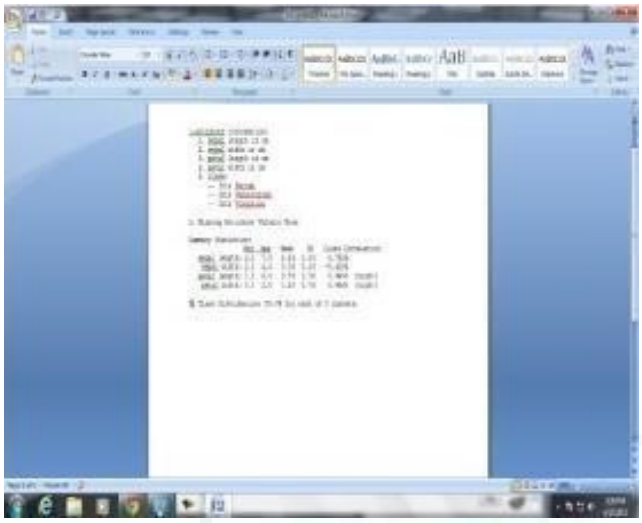
1. Download any data set from UCI datarepository.
2. Open the same data file from excel. It will ask for delimiter (which produce column) inexcel.
3. Add one row at the top of thedata.
4. Enter header for eachcolumn.
5. Save file as .CSV (Comma Separated Values)format.
6. Open Weka tool and open the CSVfile.
7. Save it as ARFFformat.

OUTPUT:

DATA TEXT FILE



DATA ARFF FILE



RESULT:

Thus, conversion of a text file to ARFF(Attribute-Relation File Format) using Weka3.8.2 tool is implemented.

DATE:

EX.No: 4 TRAINING THE GIVEN DATASET FOR ANAPPLICATION

AIM:

To apply the concept of Linear Regression for training the given dataset.

ALGORITHM:

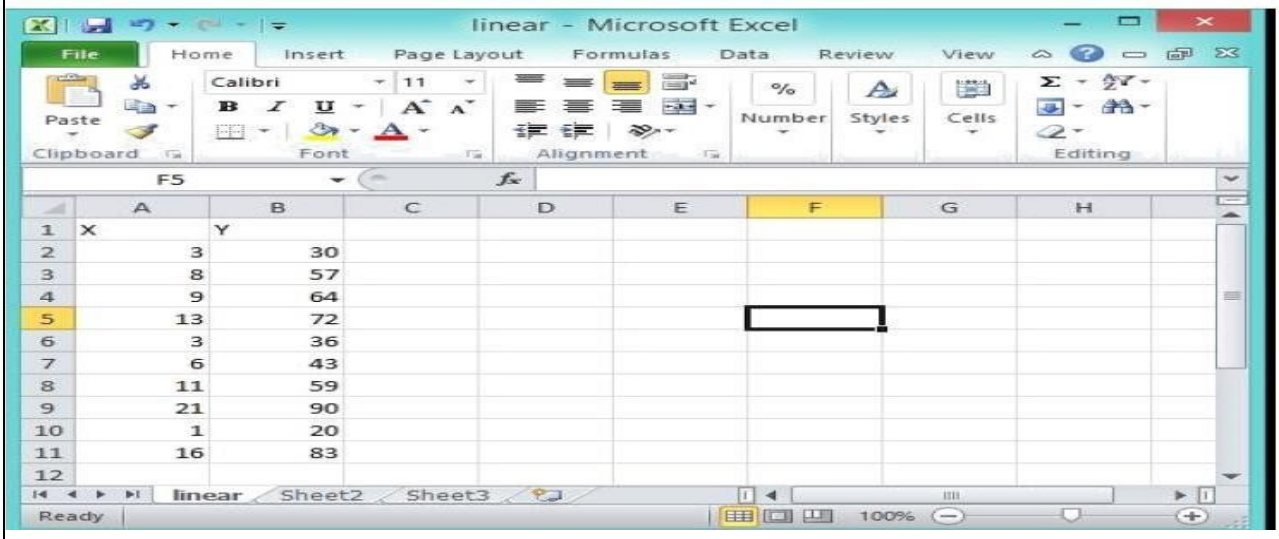
1. Open the wekatool.
2. Download a dataset by usingUCI.
3. Apply replace missingvalues.
4. Apply normalizefilter.
5. Click the ClassifyTab.
6. Choose the Simple Linear Regressionoption.
7. Select the training set ofdata.
8. Start the validationprocess.
9. Note theoutput.

LINEAR REGRESSION:

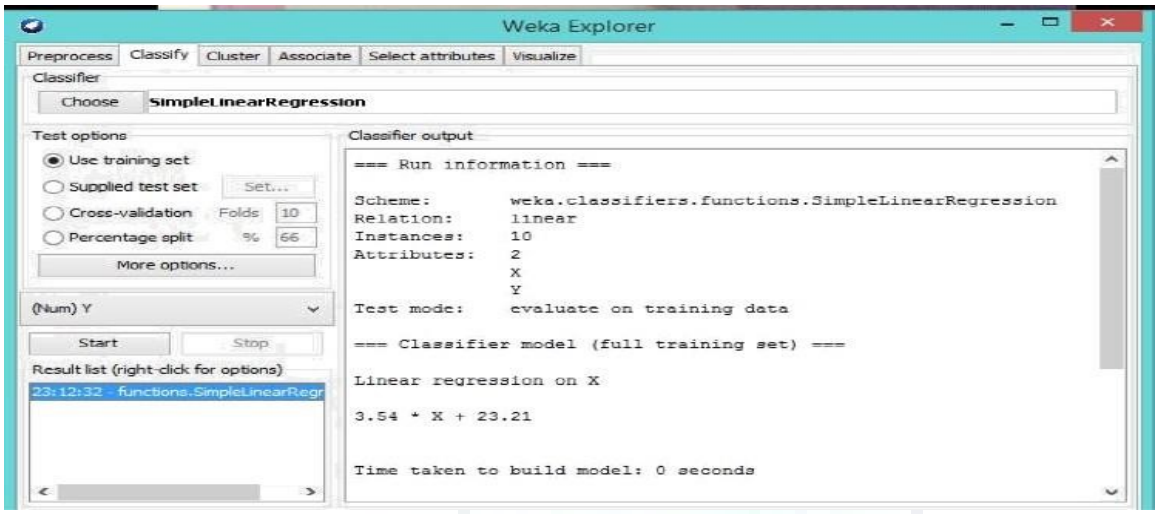
In statistics, Linear Regression is an approach for modeling a relationship between a scalar dependent variable Y and one or more explanatory variables denoted X.the case of explanatory variable is called Simple Linear Regression.
Coefficient of Linear Regression is given by: **Y=ax+b**

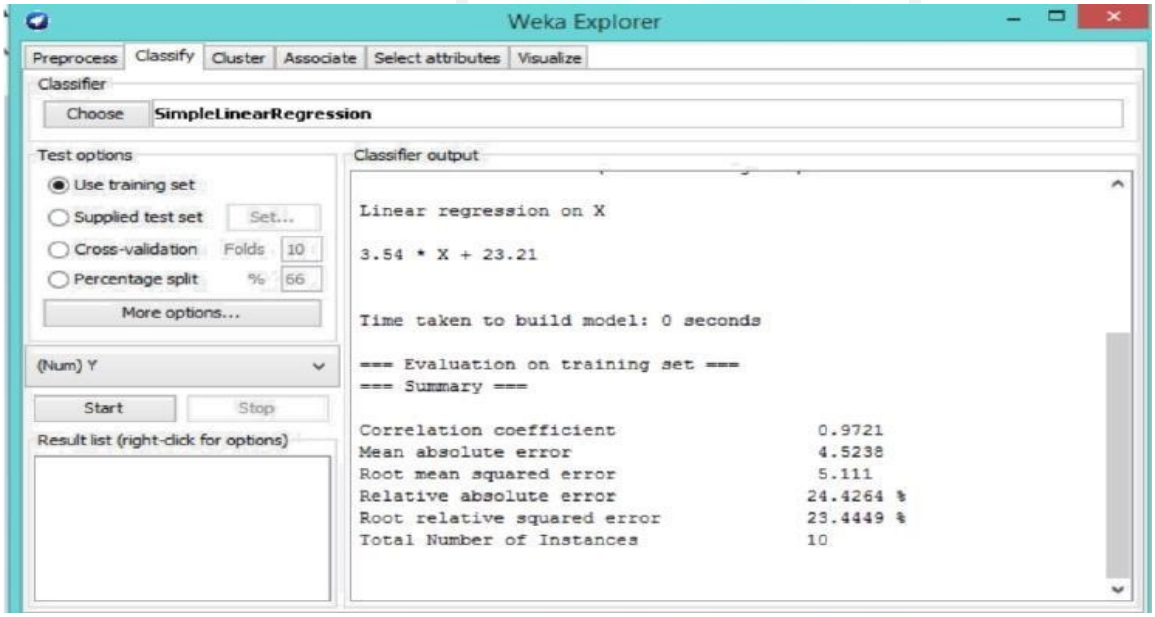
OUTPUT:

INPUT DATA



OUTPUT





RESULT:

Thus the concept of Linear Regression for training the given dataset is applied and implemented.

DATE:

EX.No: 5 TESTING THE GIVEN DATASET FOR ANAPPLICATION

AIM:

To apply the Navie Bayes Classification for testing the given dataset.

ALGORITHM:

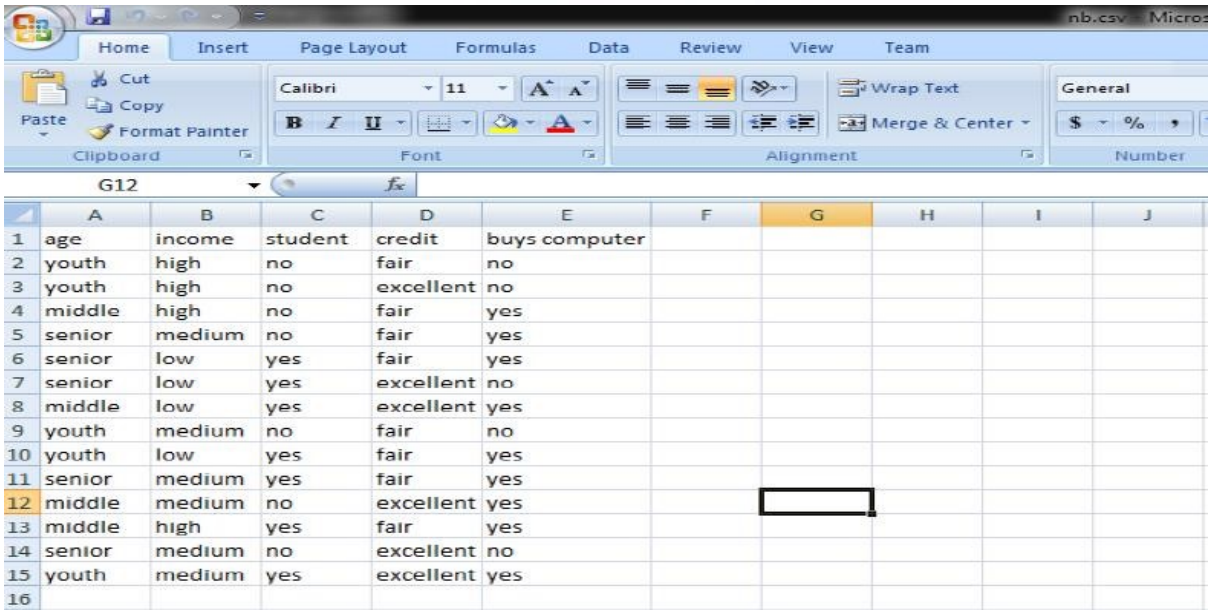
1. Open the wekatool.
2. Download a dataset by usingUCI.
3. Apply replace missingvalues.
4. Apply normalizefilter.
5. Click the ClassificationTab.
6. Apply Navie BayesClassification.
7. Find the ClassifiedValue.
8. Note theoutput.

Bayes' Theorem In the Classification Context:

X is a data tuple. In Bayesian term it is considered "evidence".H is some hypothesis that X belongs to a specified class C .P(H|X) is the posterior probability of H conditioned on X .

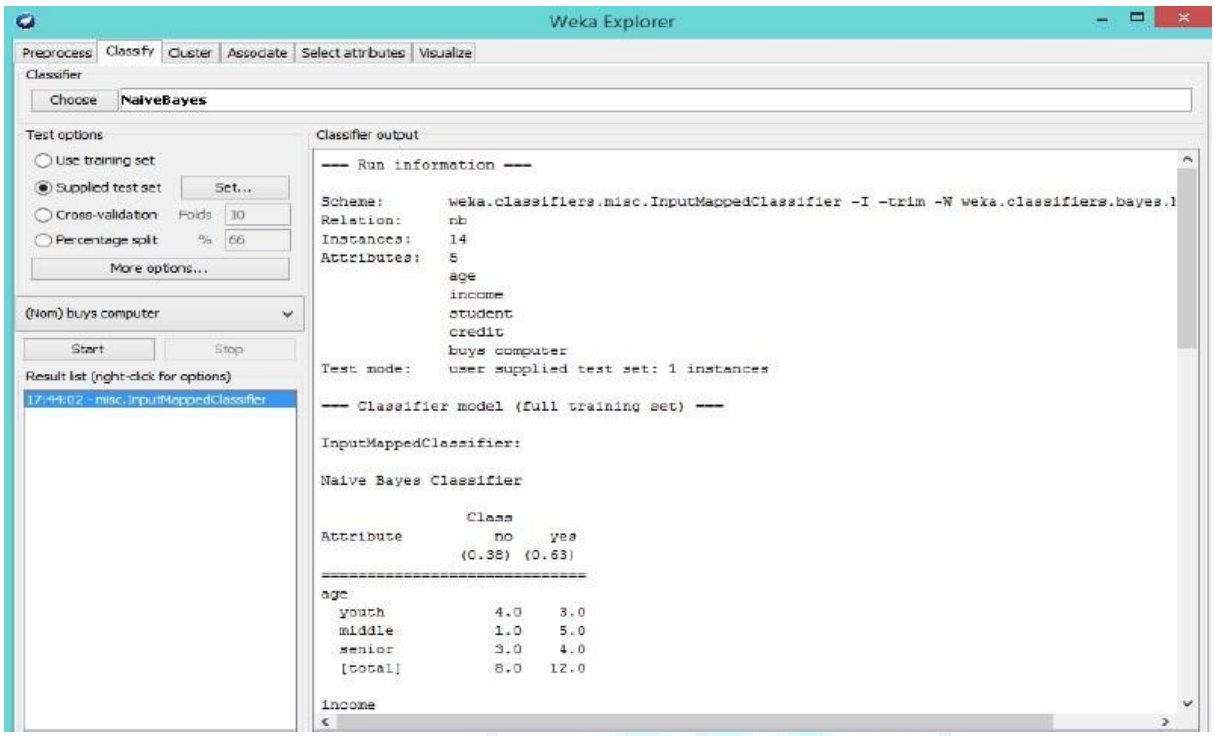
Example: predict whether a costumer will buy a computer or not " Costumers are described by two attributes: age and income " X is a 35 years-old costumer with an income of 40k " H is the hypothesis that the costumer will buy a computer " P(H|X) reflects the probability that costumer X will buy a computer given that we know the costumers' age and income

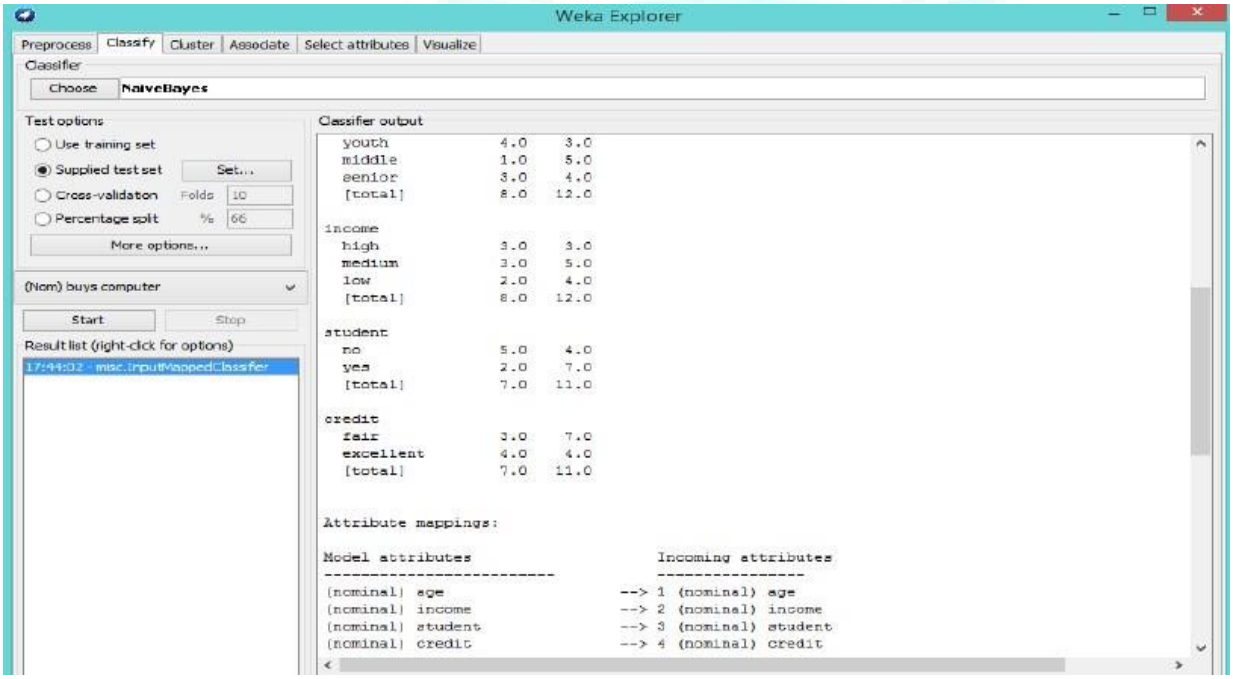
INPUT DATA

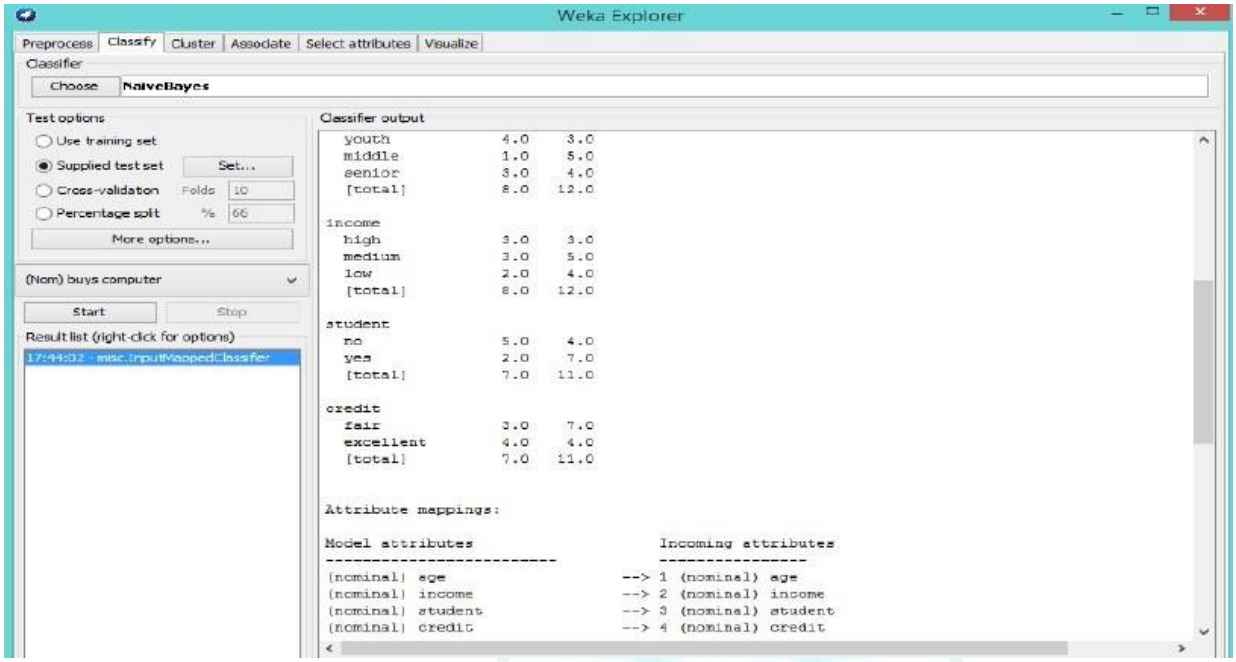


	A	B	C	D	E	F	G	H	I	J
1	age	income	student	credit	buys computer					
2	youth	high	no	fair	no					
3	youth	high	no	excellent	no					
4	middle	high	no	fair	yes					
5	senior	medium	no	fair	yes					
6	senior	low	yes	fair	yes					
7	senior	low	yes	excellent	no					
8	middle	low	yes	excellent	yes					
9	youth	medium	no	fair	no					
10	youth	low	yes	fair	yes					
11	senior	medium	yes	fair	yes					
12	middle	medium	no	excellent	yes					
13	middle	high	yes	fair	yes					
14	senior	medium	no	excellent	no					
15	youth	medium	yes	excellent	yes					
16										

OUTPUT:


Weka Explorer
Preprocess
Classify
Cluster
Associate
Select attributes
Visualize
Classifier
Choose
NaiveBayes
Test options
Use training set
Supplied test set
Set...
Cross-validation
Folds
10
Percentage split
%
66
More options...
(Nom) buys computer
Start
Stop
Result list (right-click for options)
17:44:02 - misc.InputMappedClassifier
Classifier output
=== Run information ===
Scheme: weka.classifiers.misc.InputMappedClassifier -I -trim -W weka.classifiers.bayes.
Relation: nb
Instances: 14
Attributes: 5
age
income
student
credit
buys computer
Test mode: user supplied test set: 1 instances
=== Classifier model (full training set) ===
InputMappedClassifier:
Naive Bayes Classifier
Class
Attribute no yes
(0.38) (0.63)
age
youth 4.0 3.0
middle 1.0 5.0
senior 3.0 4.0
[total] 8.0 12.0
income


Weka Explorer
Preprocess
Classify
Cluster
Associate
Select attributes
Visualize
Classifier
Choose
NaiveBayes
Test options
Use training set
Supplied test set
Set...
Cross-validation
Folds
10
Percentage split
%
66
More options...
(Nom) buys computer
Start
Stop
Result list (right-click for options)
17:44:02 - misc.InputMappedClassifier
Classifier output
youth 4.0 3.0
middle 1.0 5.0
senior 3.0 4.0
[total] 8.0 12.0
income
high 3.0 3.0
medium 3.0 5.0
low 2.0 4.0
[total] 8.0 12.0
student
no 5.0 4.0
yes 2.0 7.0
[total] 7.0 11.0
credit
fair 3.0 7.0
excellent 4.0 4.0
[total] 7.0 11.0
Attribute mappings:
Model attributes Incoming attributes
(nominal) age --> 1 (nominal) age
(nominal) income --> 2 (nominal) income
(nominal) student --> 3 (nominal) student
(nominal) credit --> 4 (nominal) credit



RESULT:

Thus the Navie Bayes Classification for testing the given dataset is implemented

DATE:

EX.No: 6 GENERATE ACCURATE MODEL

AIM:

To find the good result (by improving the performance) using the training set and testing data set for numerical values.

OBJECTIVES:

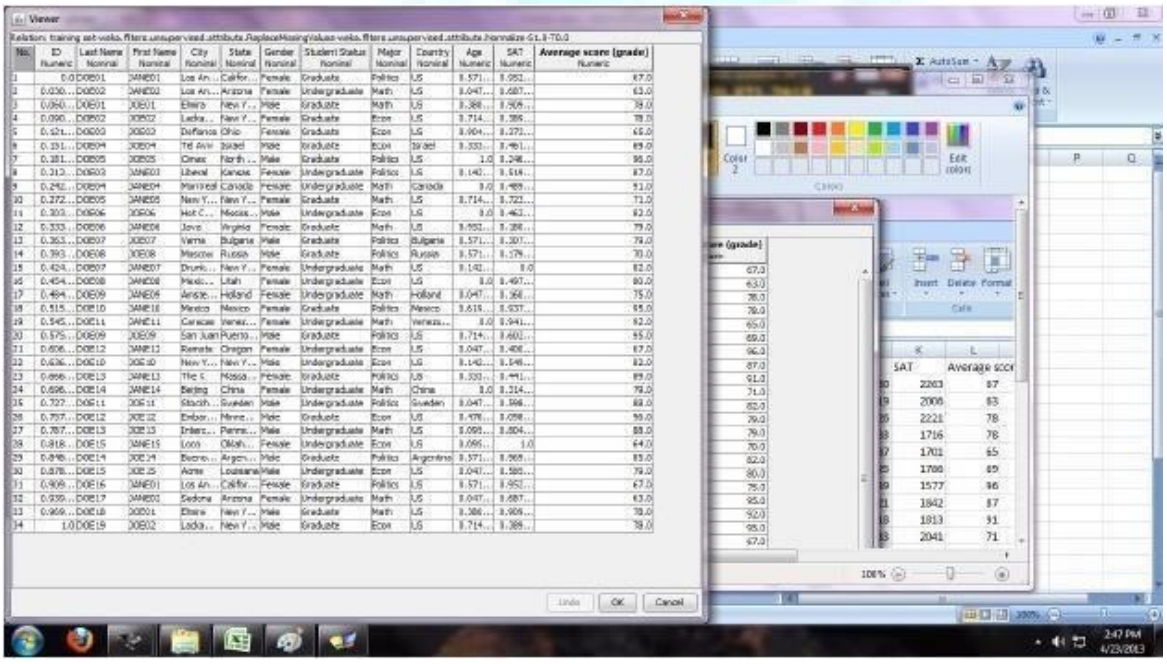
To develop training and testing data using numerical data set in order to get accurate model for classification.

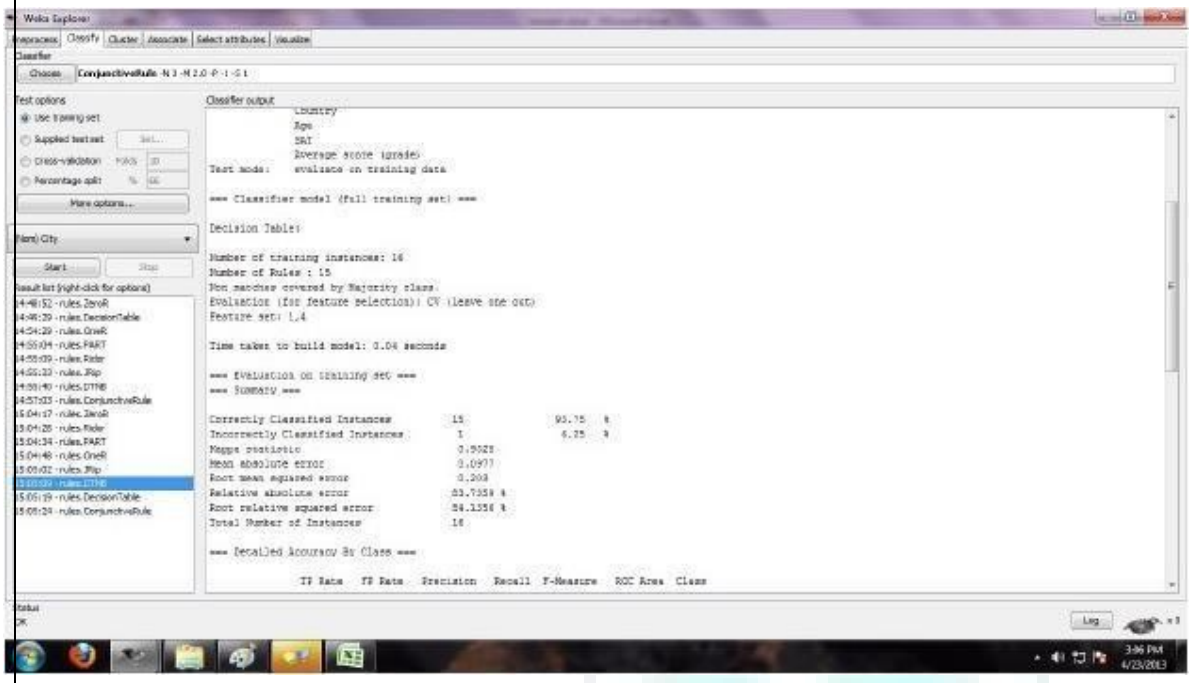
ALGORITHM:

1. Download any dataset.
2. Save the file with .ARFFformat.
3. Apply 'Replace Missing Values'filter.
4. Normalize the values by applying normalizeflter.
5. Go to unsupervised instance removepercentage
6. Right click on that (show properties) option then select 70% true and save it astraining.arff
7. Select the original data set then right click on show properties then give 70% false and save it astesting.arff
8. Select classification and apply variousalgorithms.

OUTPUT :

TRAINING DATA SET





RESULT

Thus, the good result (by improving the performance) using the training set and testing data set is done.

DATE:

EX.No: 7 DATA PRE-PROCESSING – DATAFILTERS

AIM:

To perform the data pre-processing by applying filter.

OBJECTIVES:

The data collected from public fourms have plenty of noise or missing data. Weka provides filter to replace the missing values and to remove the noisy data. So that the result will be more accurate.

ALGORITHM:

1. Download a complete data set (numeric) fromUCI.
2. Open the data set in Wekatool.
3. Save the data set with missingvalues.
4. Apply replace missing valuefilter.
5. Calculate the accuracy using theformula

$$\text{Accuracy}=\sqrt{\sum (\text{old-new})^2}$$

$$\text{Percentage of accuracy}=\frac{\text{Accuracy}}{\sum \text{old value}} \times 100$$

OUTPUT:

STUDENT DETAILS TABLE: MISSING VALUE

Viewer

Relation: weather

No.	1: outlook Nominal	2: temperature Numeric	3: humidity Numeric	4: windy Nominal	5: play Nominal
1	sunny	85.0	85.0	FALSE	no
2	sunny	80.0	90.0	TRUE	no
3	overcast	83.0	86.0	FALSE	yes
4	rainy		96.0	FALSE	yes
5	rainy	68.0	80.0	FALSE	yes
6	rainy	65.0		TRUE	no
7	overcast	64.0	65.0	TRUE	yes
8	sunny	72.0	95.0	FALSE	no
9	sunny			FALSE	yes
10	rainy	75.0	80.0	FALSE	yes
11	sunny	75.0	70.0	TRUE	yes
12	overcast			TRUE	yes
13	overcast	81.0	75.0	FALSE	yes
14	rainy		91.0	TRUE	no

STUDENT DETAILS TABLE: REPLACE MISSING VALUE

Viewer

Relation: weather-weka.filters.unsupervised.attribute.Replace

No.	1: outlook	2: temperature	3: humidity	4: windy	5: play
	Nominal	Numeric	Numeric	Nominal	Nominal
1	sunny	85.0	85.0	FALSE	no
2	sunny	80.0	90.0	TRUE	no
3	overcast	83.0	86.0	FALSE	yes
4	rainy	74.8	96.0	FALSE	yes
5	rainy	68.0	80.0	FALSE	yes
6	rainy	65.0	83.0	TRUE	no
7	overcast	64.0	65.0	TRUE	yes
8	sunny	72.0	95.0	FALSE	no
9	sunny	74.8	83.0	FALSE	yes
10	rainy	75.0	80.0	FALSE	yes
11	sunny	75.0	70.0	TRUE	yes
12	overcast	74.8	83.0	TRUE	yes
13	overcast	81.0	75.0	FALSE	yes
14	rainy	74.8	91.0	TRUE	no

RESULT:

Thus the data pre-processing by applying filter is performed.

DATE:

EX.No: 8 FEATURE SELECTION

AIM:

To find the good results by feature selection.

OBJECTIVES:

Any classifier/model has internal feature, those feature gives more accurate and optimal result.

ALGORITHM:

1. Download any dataset with nominalvalues.
2. Save it as text.arff.
3. Split it into training and testing dataset.
4. Go to unsupervised instance removepercentage.
5. Right click on that show properties then select 70% true and save it astraining.arff
6. Right click on that show properties then select 70% false and save it as testing.arff using original dataset.
7. Open the parameter for classifying.
8. Fix the set of changingvalues.
9. Look at theperformance.
10. Go to step 3 until the expected values of maximum value isreached

OUTPUT:

TRAINING SET		
CLASSIFIER	PARAMETER SETTING	PERFORMANCE
JRip	Seed=1	Root Mean Squared Error=0.1707 Mean Absolute Error=0.0583
JRip	Seed =2	Root Mean Squared Error=0.1764 Mean Absolute Error=0.0622
JRip	Seed =3	Root Mean Squared Error=0.1764 Mean Absolute Error=0.0622
Ridor	Seed =1	Root Mean Squared Error=0.2508 Mean Absolute Error=0.0629
Ridor	Seed=2	Root Mean Squared Error=0.2508 Mean Absolute Error=0.0629

TEST SET		
CLASSIFIER	PARAMETER SETTING	PERFORMANCE
JRip	Seed=1	Root Mean Squared Error=0.2431 Mean Absolute Error=0.1172
JRip	Seed =2	Root Mean Squared Error=0.2431 Mean Absolute Error=0.1172
JRip	Seed =3	Root Mean Squared Error=0.2431 Mean Absolute Error=0.1172
Ridor	Seed =1	Root Mean Squared Error=0.3423 Mean Absolute Error=0.1172
Ridor	Seed=2	Root Mean Squared Error=0.3423 Mean Absolute Error=0.1172

TRAINING		
JRip	Seed=1	Root Mean Squared Error=0.1707 Mean Absolute Error=0.0583
Ridor	Seed =1	Root Mean Squared Error=0.2508 Mean Absolute Error=0.0629

TEST		
JRip	Seed=1	Root Mean Squared Error=0.2431 Mean Absolute Error=0.1172
Rider	Seed =1	Root Mean Squared Error=0.3423 Mean Absolute Error=0.1172

RESULT:

Thus the good results by feature selection were found

DATE:

EX.No: 9 **WEB MINING**

AIM:

To apply the web mining technique clustering algorithm for the given dataset.

Introduction to Web Mining:

Web mining is an application of data mining techniques to find information patterns from the web data. Web mining helps to improve the power of web search engine by identifying the web pages and classifying the web documents.Web mining is very useful to e-commerce websites and e-services.

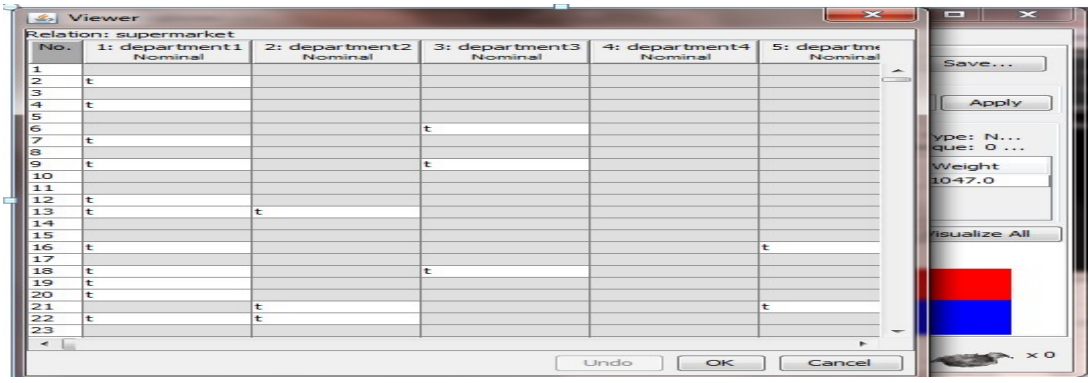
Web Content Mining :

Web content mining can be used for mining of useful data, information and knowledge from web page content. Web structure mining helps to find useful knowledge or information pattern from the structure of hyperlinks. Due to heterogeneity and absence of structure in web data, automated discovery of new knowledge pattern can be challenging to some extent.Web content mining performs scanning and mining of the text, images and groups of web pages according to the content of the input (query), by displaying the list in search engines. For example: If an user wants to search for a particular book, then search engine provides the list ofsuggestions.

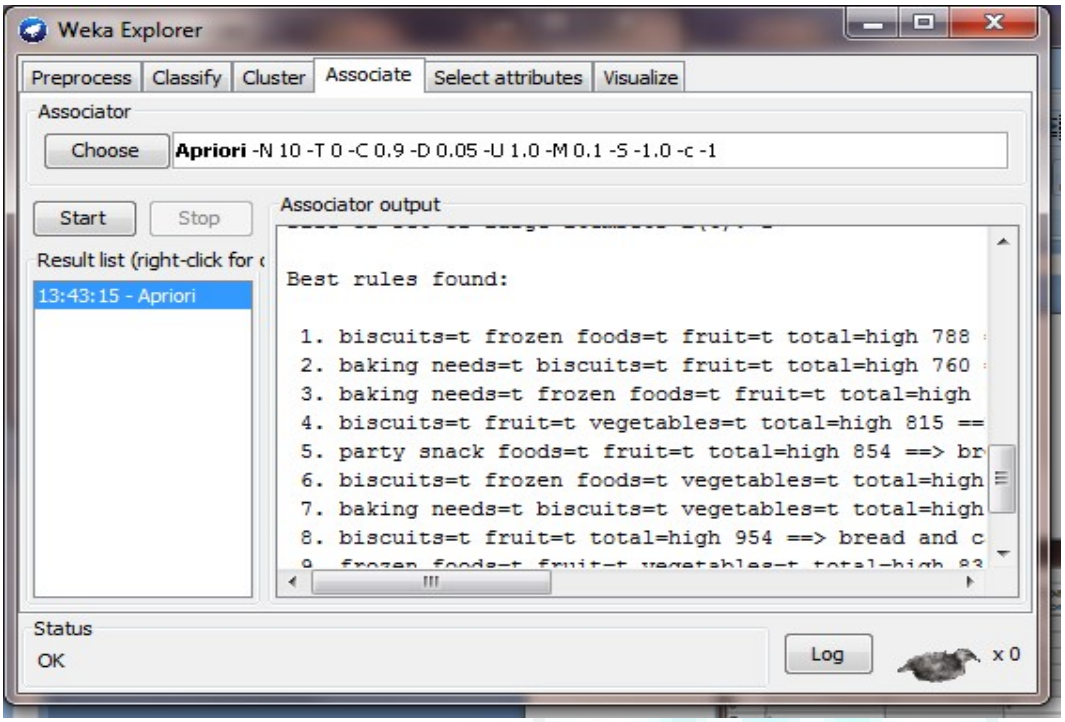
ALGORITHM:

1. Open the wekatool.
2. Download a dataset by usingUCI.
3. Apply replace missingvalues.
4. Apply normalizefilter.
5. Click the clustertab.
6. Apply all algorithms one byone.
7. Find the no of clusters that areformed
8. Note theoutput.

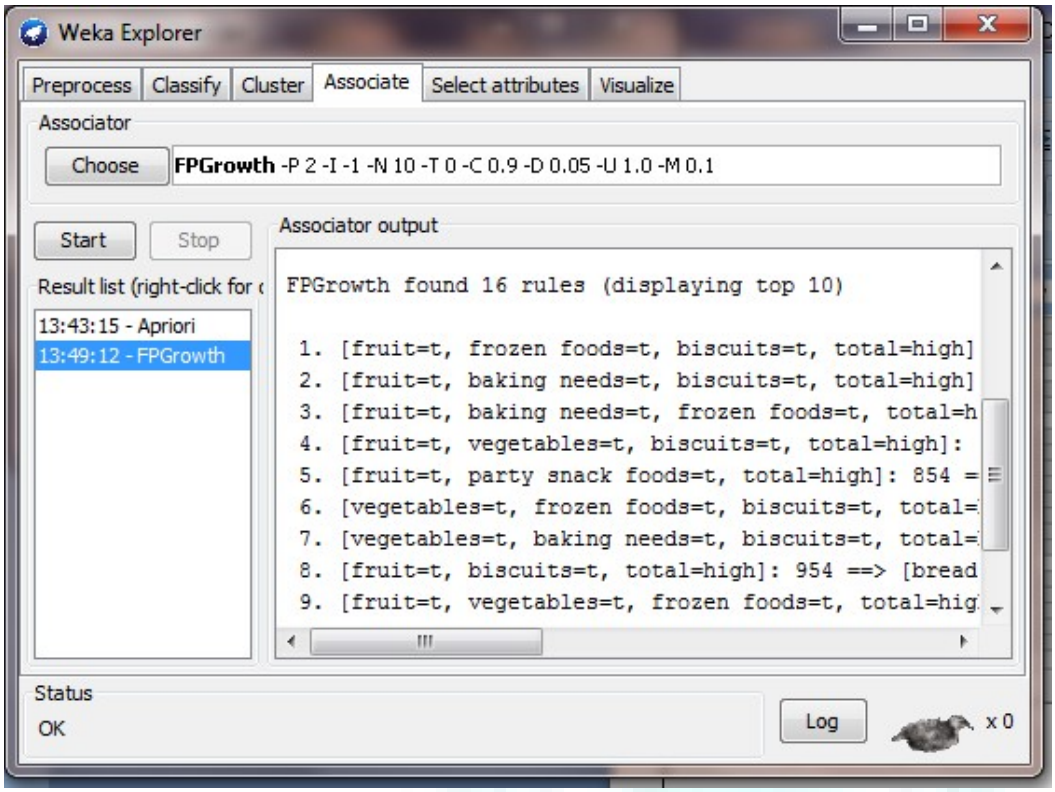
OUTPUT:



APRIORI


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
Result list (right-click for
13:43:15 - Apriori
Associator output
Best rules found:
1. biscuits=t frozen foods=t fruit=t total=high 788
2. baking needs=t biscuits=t fruit=t total=high 760
3. baking needs=t frozen foods=t fruit=t total=high
4. biscuits=t fruit=t vegetables=t total=high 815 ==
5. party snack foods=t fruit=t total=high 854 ==> br
6. biscuits=t frozen foods=t vegetables=t total=high
7. baking needs=t biscuits=t vegetables=t total=high
8. biscuits=t fruit=t total=high 954 ==> bread and c
Status
OK
Log
x 0

FP GROWTH ALGORITHM



RESULT:

Thusto apply the web mining technique clustering algorithm for the given dataset has been done.

DATE:

EX.No: 10 **TEXT MINING**

AIM:

To find association between data and to find the frequent item set for text mining.

Text Data Mining

Text data mining can be described as the process of extracting essential data from standard language text. All the data that we generate via text messages, documents, emails, files are written in common language text. Text mining is primarily used to draw useful insights or patterns from such data.The purchasing of one product when another product is purchased represents an association rule. Association rules are frequently used by retail store to assist in marketing, advertising, floor placement, and inventory control. Association rules are used to show the relationship between data items.

Keyword-based Association Analysis in text mining:

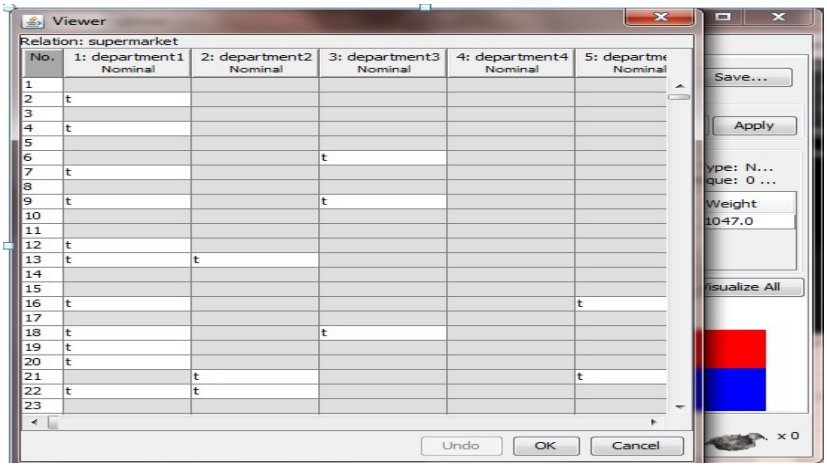
It collects sets of keywords or terms that often happen together and afterward discover the association relationship among them. First, it preprocesses the text data by parsing, stemming, removing stop words, etc. Once it pre-processed the data, then it induces association mining algorithms. Here, human effort is not required, so the number of unwanted results and the execution time is reduced.

ALGORITHM:

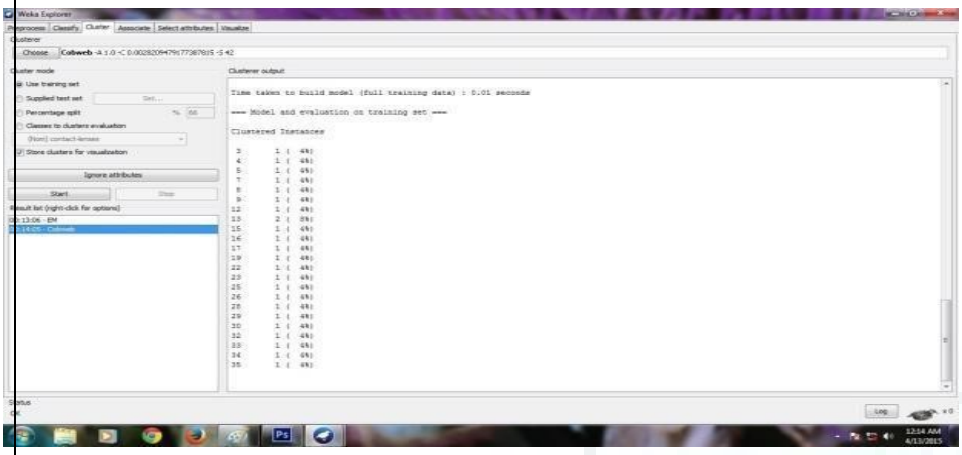
1. Opendataset
2. Selectassociate
3. Choose different algorithm forassociation
4. Observe theperformance
5. Select the association rule with the maximum confidencerule.

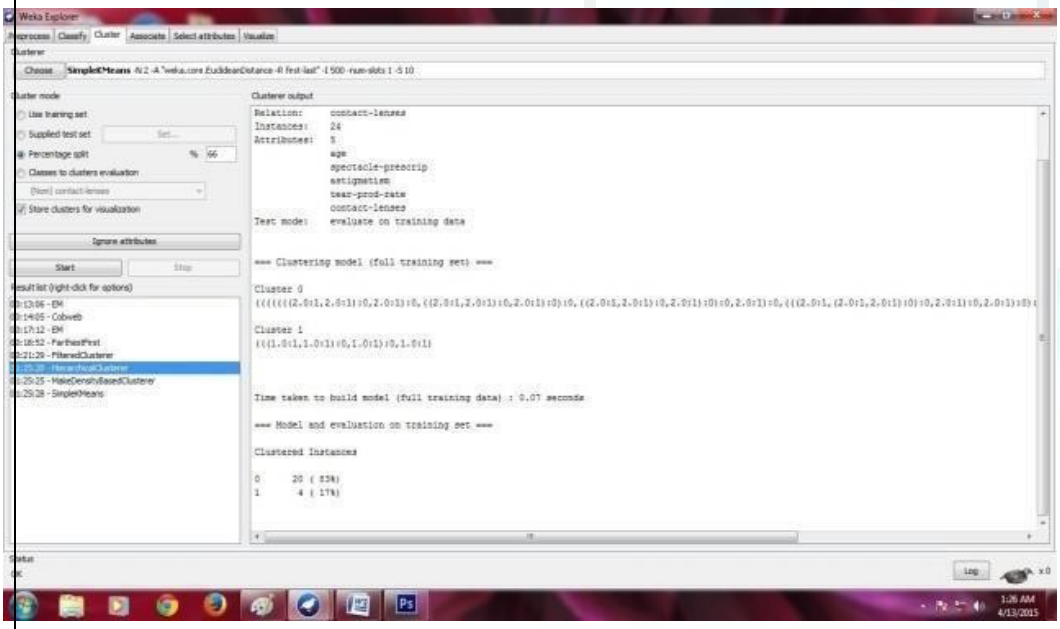
INPUT:

SuperMarket data set



OUTPUT:


Weka Explorer
Preprocess Classify Cluster Associate Select attributes Visualize
Clusterer
Choose Cobweb -A 1.0 -C 0.0028209479177387815 -S 42
Cluster mode
Use training set
Supplied test set
Percentage split
Classes to clusters evaluation
(Nom) contact-lenses
Store clusters for visualization
Ignore attributes
Start
Stop
Result list (right-click for options)
03:13:06 - EM
03:14:05 - Cobweb
Clusterer output
Time taken to build model (full training data) : 0.01 seconds
=== Model and evaluation on training set ===
Clustered Instances
Status
OK
Log
x 0


Weka Explorer
Preprocess Classify Cluster Associate Select attributes Visualize
Clusterer
Choose SimpleKMeans -N 2 -A "weka.core.EuclideanDistance -R first-last" -I 500 -num-slots 1 -S 10
Cluster mode
Use training set
Supplied test set
Percentage split
% 66
Classes to clusters evaluation
(Nom) contact-lenses
Store clusters for visualization
Ignore attributes
Start
Stop
Result list (right-click for options)
03:13:06 - EM
03:14:05 - Cobweb
03:17:12 - EM
03:18:52 - FarthestFirst
03:21:29 - FilteredClusterer
03:25:25 - MakeDensityBasedClusterer
03:25:28 - SimpleKMeans
Clusterer output
Relation: contact-lenses
Instances: 24
Attributes: 5
age
spectacle-prescrip
astigmatism
tear-prod-rate
contact-lenses
Test mode: evaluate on training data
=== Clustering model (full training set) ===
Cluster 0
Cluster 1
Time taken to build model (full training data) : 0.07 seconds
=== Model and evaluation on training set ===
Clustered Instances
0 20 (83%)
1 4 (17%)
Status
OK
Log
x 0

RESULT

Thus association between data and to find the frequent item set for text mining was found.

DATE:

EX.No:11 **DESIGN OF FACT AND DIMENSIONTABLES**

AIM:

To design fact and dimension tables.

Fact Table :

A fact table is used in the dimensional model in data warehouse design. A fact table is found at the center of a star schema or snowflake schema surrounded by dimension tables.A fact table consists of facts of a particular business process e.g., sales revenue by month by product. Facts are also known as measurements or metrics. A fact table record captures a measurement or a metric.

PROCEDURE:

1. **Choosing business process to model** – The first step is to decide what business process to model by gathering and understanding business needs and availabledata
2. **Declare the grain** – by declaring a grain means describing exactly what a fact table recordrepresents
3. **Choose the dimensions** – once grain of fact table is stated clearly, it is time to determine dimensions for the facttable.
4. **Identify facts** – identify carefully which facts will appear in the facttable.

RESULT:

Thus design fact and dimension tables are create.