LINUX PROGRAMMING AND DATA MINING

LAB MANUAL

COMPUTER SCIENCE AND ENGINEERING

Program Outcomes	
PO1	**Engineering knowledge**: Apply the knowledge of mathematics, science, engineering fundamentals, and an engineering specialization to the solution of complex engineering problems.
PO2	**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.
PO3	**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.
PO4	**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.
PO5	**Modern tool usage**: 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.
PO6	**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.
PO7	**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.
PO8	**Ethics**: Apply ethical principles and commit to professional ethics and responsibilities and norms of the engineering practice.
PO9	**Individual and team work**: Function effectively as an individual, and as a member or leader in diverse teams, and in multidisciplinary settings.
PO10	**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.
PO11	**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.
PO12	**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.
Program Specific Outcomes	
PSO1	**Professional Skills:** The ability to research, understand and implement computer programs in the areas related to algorithms, system software, multimedia, web design, big data analytics, and networking for efficient analysis and design of computer-based systems of varying complexity.
PSO2	**Problem-Solving Skills:** The ability to apply standard practices and strategies in software project development using open-ended programming environments to deliver a quality product for business success.
PSO3	**Successful Career and Entrepreneurship:** The ability to employ modern computer languages, environments, and platforms in creating innovative career paths, to be an entrepreneur, and a zest for higher studies.

LINUX PROGRAMMING AND DATA MINING LAB SYLLABUS

*Content beyond the university prescribed syllabi

ATTAINMENT OF PROGRAM OUTCOMES & PROGRAM SPECIFIC OUTCOMES

Exp. No.	Experiment	Program Outcomes Attained	Program Specific Outcomes Attained
	LINUX PROGRAMMING		
1	c) Write a shell script that accepts a file name, starting and ending line numbers as arguments and displays all the lines between the given line numbers. d) *Illustrate by writing script that will print, message "Hello World, in Bold and Blink effect, and in different colors like red, brown etc using echo commands?	PO1, PO2	PSO1
2	c) Write a shell script that deletes all lines containing a specified word in one or more files supplied as arguments to it. d) *Illustrate by writing script using for loop to print the following patterns? i) * * * * * * * * * * * * * * * ii) 1 2 2 3 3 3 4 4 4 4 5 5 5 5 5	PO1	PSO1
3	c) Write a shell script that displays a list of all the files in the current directory to which the user has read, write and execute permissions. d) * Illustrate to redirect the standard input (stdin) and the standard output (stdout) of a process, so that scanf () reads from the pipe and printf () writes into the pipe?	PO1, PO2	PSO1
4	c) Write a shell script that receives any number of file names as arguments checks if every argument supplied is a file or a directory and reports accordingly. Whenever the argument is a file, the number of lines on it is also reported. d) *Illustrate by writing c program where process forks to a child, and create a child process by using forks and suddenly terminates itself?	PO1	PSO1
5	Write a shell script that accepts a list of file names as its arguments, counts and reports the occurrence of each word that is present in the first argument file on other argument files.	PO1	PSO1
6	Write a shell script to list all of the directory files in a directory.	PO1	PSO1
7	Write a shell script to find factorial of a given integer.	PO1	PSO1
8	Write an awk script to count the number of lines in a file that do not contain vowels.	PO1	PSO1
9	Write an awk script to find the number of characters, words and lines in a file.	PO1	PSO1
10	Write a c program that makes a copy of a file using standard I/O and system calls	PO1	PSO1

Exp. No.	Experiment	Program Outcomes Attained	Program Specific Outcomes Attained
11	Implement in C the following UNIX commands using System calls a) cat b) ls c) mv	PO1, PO2	PSO1
12	Write a program that takes one or more file/directory names as command line input and reports the following information on the file. a) File type b) Number of links c) Time of last access d) Read Write and Execute permissions	PO1, PO2	PSO1
13	Write a C program to emulate the UNIX ls –l command.	PO1	PSO1
14	Write a C program to list for every file in a directory, its inode number and file name.	PO1	PSO1
15	Write a C program that demonstrates redirection of standard output to a file.Ex: ls > f1.	PO1	PSO1
16	Write a C program to create a child process and allow the parent to display “parent” and the child to display “child” on the screen.	PO1	PSO1
17	Write a C program to create a Zombie process.	PO1	PSO1
18	Write a C program that illustrates how an orphan is created.	PO1	PSO1
19	Write a C program that illustrates how to execute two commands concurrently with a command pipe. Ex: - ls –l | sort	PO1	PSO1
20	Write C programs that illustrate communication between two unrelated processes using named pipe.	PO1	PSO1
21	Write a C program to create a message queue with read and write permissions to write 3 messages to it with different priority numbers.	PO1	PSO1
22	Write a C program that receives the messages (from the above message queue as specified in (21)) and displays them.	PO1	PSO1
23	Write a C program to allow cooperating processes to lock a resource for exclusive use, using a) Semaphores b) flock or lockf system calls.	PO1, PO2	PSO1
24	Write a C program that illustrates suspending and resuming processes using signals	PO1, PO2	PSO1
25	Write a C program that implements a producer-consumer system with two processes.	PO1, PO2, PO4	PSO1, PSO2
26	Write client and server programs (using c) for interaction between server and client processes using Unix Domain sockets. (Using Semaphores).	PO1, PO2, PO3, PO4	PSO1, PSO2
27	Write client and server programs (using c) for interaction between server and client processes using Internet Domain sockets.	PO1, PO2, PO3, PO4	PSO1, PSO2
28	Write a C program that illustrates two processes communicating using shared memory.	PO1, PO2, PO3, PO4	PSO1, PSO2
	DATA MINING		
18	List all the categorical (or nominal) attributes and the real-valued attributes separately.	PO1, PO2	PSO1
19	What attributes do you think might be crucial in making the credit assessment? Come up with some simple rules in plain English using your selected attributes.	PO1, PO2	PSO1, PSO2

Exp. No.	Experiment	Program Outcomes Attained	Program Specific Outcomes Attained
20	*What attributes do you think might be crucial in making the bank assessment?	PO1, PO2, PO12	PSO1, PSO2
21	One type of model that you can create is a Decision Tree -train a Decision Tree using the complete dataset as the training data. Report the model obtained after training.	PO1, PO2, PO5	PSO1
22	Suppose you use your above model trained on the complete dataset, and classify credit good/bad for each of the examples in the dataset. What % of examples can you classify correctly? (This is also called testing on the training set) Why do you think you cannot get 100 % training accuracy?	PO1, PO2	PSO1, PSO2
23	*Find out the correctly classified instances, root mean squared error, kappa statistics, and mean absolute error for weather data set?	PO1, PO2, PO5	PSO1
24	Is testing on the training set as you did above a good idea? Why or Why not?	PO1, PO2, PO5, PO12	PSO1
25	One approach for solving the problem encountered in the previous question is using cross-validation? Describe what is cross -validation briefly. Train a Decision Tree again using cross - validation and report your results. Does your accuracy increase/decrease? Why?	PO1, PO2, PO5	PSO1
26	Check to see if the data shows a bias against "foreign workers" (attribute 20), or "personal -status" (attribute 9). One way to do this (perhaps rather simple minded) is to remove these attributes from the dataset and see if the decision tree created in those cases is significantly different from the full dataset case which you have already done. To remove an attribute you can use the preprocess tab in Weka's GUI Explorer. Did removing these attributes have any significant effect? Discuss.	PO1, PO2, PO4, PO5	PSO1
27	*Load the 'weather.arff' dataset in Weka and run the ID3 classification algorithm. What problem do you have and what is the solution?	PO1, PO2, PO5	PSO1, PSO2
28	Another question might be, do you really need to input so many attributes to get good results? Maybe only a few would do. For example, you could try just having attributes 2, 3, 5, 7, 10, 17 (and 21, the class attribute (naturally)). Try out some combinations. (You had removed two attributes in problem 7. Remember to reload the arff data file to get all the attributes initially before you start selecting the ones you want.)	PO1, PO2, PO4, PO12	PSO1
29	Sometimes, the cost of rejecting an applicant who actually has a good credit (case 1) might be higher than accepting an applicant who has bad credit (case 2). Instead of counting the misclassifications equally in both cases, give a higher cost to the first case (say cost 5) and lower cost to the second case. You can do this by using a cost matrix in Weka. Train your Decision Tree again and report the Decision Tree and cross - validation results. Are they significantly different from results obtained in problem 6 (using equal cost)?	PO1, PO2, PO5, PO12	PSO1, PSO2
30	Do you think it is a good idea to prefer simple decision trees instead of having long complex decision trees? How does the complexity of a Decision Tree relate to the bias of the model?	PO1, PO2, PO12	PSO1, PSO2
31	*Run the J48 and 1Bk classifiers using-the cross-validation strategy with various fold levels. Compare the accuracy results. Holdout strategy with three percentage levels. Compare the accuracy results.	PO1, PO2, PO4, PO5	PSO1, PSO2

Exp. No.	Experiment	Program Outcomes Attained	Program Specific Outcomes Attained
32	You can make your Decision Trees simpler by pruning the nodes. one approach is to use Reduced Error Pruning -Explain this idea briefly. Try reduced error pruning for training your Decision Trees using cross - validation (you can do this in Weka) and report the Decision Tree you obtain? Also, report your accuracy using the pruned model. Does your accuracy increase?	PO1, PO2, PO4, PO5, PO12	PSO1, PSO2
33	(Extra Credit): How can you convert a Decision Trees into "if –then - else rules". Make up your own small Decision Tree consisting of 2 - 3 levels and convert it into a set of rules. There also exist different classifiers that output the model in the form of rules -one such classifier in Weka is rules. PART, train this model and report the set of rules obtained. Sometimes just one attribute can be good enough in making the decision, yes, just one! Can you predict what attribute that might be in this dataset? OneR classifier uses a single attribute to make decisions (it chooses the attribute based on minimum error). Report the rule obtained by training a one R classifier. Rank the performance of j48, PART and oneR.	PO1, PO2, PO4, PO5, PO12	PSO1, PSO2
34	*Run J48 and Naïve Bayes classifiers on the following datasets and determine the accuracy: 1.vehicle.arff 2.kr-vs-kp.arff 3.glass.arff 4.wave-form-5000.arff On which datasets does the Naïve Bayes perform better?	PO1, PO2, PO4, PO5, PO12	PSO1, PSO2

*Content beyond the university prescribed syllabi

LINUX PROGRAMMING AND DATA MINING LABORATORY

OBJECTIVE:

The Linux programming laboratory course covers major methods of Inter Process Communication (IPC), which is the basis of all client / server applications under Linux, Linux Utilities, working with the Bourne again shell (bash), files, process and signals. There will be extensive programming exercises in shell scripts. It also emphasizes various concepts in multithreaded programming and socket programming.

Data mining tools allow predicting future trends and behaviors, allowing businesses to make proactive, knowledge-driven decisions. The data mining laboratory course is designed to exercise the data mining techniques such as classification, clustering, pattern mining etc with varied datasets and dynamic parameters. Weka data mining tool is used for the purpose of acquainting the students with the basic environment of the data mining tools.

OUTCOMES:

Upon the completion of Linux Programming and Data Mining practical course, the student will be able to:

1. **Understand** and implement basic system functionalities of Linux operating system.
2. **Write** shell scripts to automate different tasks.
3. **Demonstrate** Inter process Communication techniques.
4. **Design** and implement advanced features such as threads, IPC, pipes, FIFOs.
5. **Demonstrate** client server technology using socket programming.
6. **Understand** synchronized programs using shared memory,
7. **Implement** and manage client server technology using TCP and UDP.
8. **Learn** to build a data warehouse and query it using open source tools.
9. **Learn** to execute data mining tasks using a data mining toolkit (such as WEKA) and visualize the results.
10. **Demonstrate** the working of algorithms for data mining tasks such association rule mining, classification, clustering and regression.

LINUX PROGRAMMING

EXPERIMENT 1

1.1 OBJECTIVE

a) Write a shell script that accepts a file name, starting and ending line numbers as arguments and displays all the lines between the given line numbers.

b) *Illustrate by writing script that will print, message "Hello World, in Bold and Blink effect, and in different colors like red, brown etc using echo commands?

1.2 RESOURCE/REQUIREMENTS

Linux operating system ,vi-editor, shell-interpreter

1.3 PROGRAM LOGIC

1. Read a filename, starting and ending line numbers as arguments
2. Find difference between starting and ending line numbers
3. Test given filename exists or not
4. If exists display between lines to output stream else display file not exists

1.4.a DESCRIPTION / PROCEDURE

```
echo " Enter the file name"
read  fname
echo "enter starting line number"
read sl
echo "enter ending line number"
read el
d=`expr $el - $sl`
if [ -f $fname ]
   then
   echo "the lines between $sl and $el of given file are"
   head -$el $fname | tail -$d
   else
   echo "file doesnt exist"
fi
```

INPUT:

```
sh prog1.sh
enter the file name
file1
enter starting line number
15
enter ending line number
20
```

OUTPUT:

```
# It displays 15 to 20 between lines
```

1.4.b DESCRIPTION / PROCEDURE

```
# clear the screen
clear
echo -e "\033[1m Hello World"
# print bold effect
echo -e "\033[5m Blink"
 # blink effect
echo -e "\033[0m Hello World"
# print back to normal
echo -e "\033[31m Hello World"
#print in Red color
```

```
echo -e "\033[32m Hello World"
# Green color
echo -e "\033[33m Hello World"
# See remains on screen
echo -e "\033[34m Hello World"
echo -e "\033[35m Hello World"
echo -e "\033[36m Hello World"
echo -e -n "\033[0m "
 # print back to normal
echo -e "\033[41m Hello World"
echo -e "\033[42m Hello World"
echo -e "\033[43m Hello World"
echo -e "\033[44m Hello World"
echo -e "\033[45m Hello World"
echo -e "\033[46m Hello World"
echo -e "\033[0m Hello World"
# Print back to normal
```

1.5 PRE-LAB QUESTIONS

1. Define shell script? What is the difference between shell and kernel.
2. Name few file handling commands present in unix.

1.6 LAB ASSIGNMENT

1. Write a shell script to count number of words present in a file without using commands.
2. Write a menu driven shell script to execute a command as 1.for `ls` ,2 for grep and 3 for cat.

1.7 POST-LAB QUESTIONS

1. What is the purpose of case statement?
2. What the difference between break and exit statement?

EXPERIMENT 2

2.1 OBJECTIVE

a) Write a shell script that deletes all lines containing a specified word in one or more files supplied as arguments to it.

b) *Illustrate by writing script using for loop to print the following patterns?

2.2 RESOURCE/REQUIREMENTS

Linux operating system ,vi-editor, shell-interpreter

2.3 PROGRAM LOGIC

Read file name from command line arguments and display lines inverse of specified word.

2.4.a DESCRIPTION / PROCEDURE

```
if [ $# -ne 0 ]
then
     echo enter the word
     read word
     for fname in $*
        do
        if [ -f $fname ]
           then
           echo the given input filename is:$fname
           grep -v "$word" $fname
        else
          echo its not a file
        fi
       done
        else
        echo "enter atleast one argument as input"
fi
```

INPUT:

```
sh prog2.sh  3.sh
enter the word
echo
```

OUTPUT:

```
The given input filename is : 3.sh
It displays all the lines other than pattern matching
```

2.4.b.i DESCRIPTION / PROCEDURE

```
# do the following for loop
 echo "Stars"
# outer loop
 for (( i=1; i<=5; i++ ))
do
#inner loop

    for (( j=1; j<=i;  j++ ))
    do
     echo -n " *"
```

```
    done
    echo ""
done
 #
```

2.4.b.ii DESCRIPTION / PROCEDURE

```
# do the following for loop

echo "Can you see the following:"
# outer loop
for (( i=1; i<=5; i++ ))
do
#inner loop
    for (( j=1; j<=i;  j++ ))
    do
     echo -n "$i"
    done
    echo ""
done
#
```

2.5 PRE-LAB QUESTIONS

1. What are positional parameter and name any two.
2. Write down the syntax of `if` statement.

2.6 LAB ASSIGNMENT

1.Read two string str1 and str2 and check
  i) Compare two strings
  ii) Palindrome or not

2.7 POST-LAB QUESTIONS

1. What is the purpose of the variable $? What are the various output it has?

EXPERIMENT 3

3.1 OBJECTIVE

a) Write a shell script that displays a list of all the files in the current directory to which the user has read, write and execute permissions.

b) Illustrate to redirect the standard input (stdin) and the standard output (stdout) of a process, so that scanf () reads from the pipe and printf () writes into the pipe?

3.2 RESOURCE/REQUIREMENTS

Linux operating system ,vi-editor, shell-interpreter

3.3 PROGRAM LOGIC

Read a list of files from current directory and display the file names to output stream whose files has read, write, execute permissions.

3.4.a DESCRIPTION / PROCEDURE

```
echo "List of Files which have Read, Write and Execute Permissions in Current Directory"
for file in *
 do
 if [ -r $file -a -w $file -a -x $file ]
   then
   echo $file
 fi
done
```

INPUT:

sh prog3.sh

OUTPUT:

List of Files which have Read, Write and Execute Permissions in Current Directory
pp2.txt

3.4.b DESCRIPTION / PROCEDURE

```
#include <stdio.h>
#include <unistd.h>
#include <sys/ipc.h>
main()
{
int fd[2];int n=0, i;
pipe(fd);
if (fork() == 0) {
/* create Child process */
close(1) ; dup(fd[1]) ;
close(fd[0]);
/* try not  read   from the pipe in this example.So  close fd[0].    */
for (i=0; i < 10; i++) {printf("%d\n",n);
/* Now that stdout has been     redirected, printf automatically writes into the pipe.  */   n++;  }
} else {/* Parent process */close(0) ; dup(fd[0]) ;
/* Redirect the stdin of this   process to the pipe*/
close(fd[1]);
/*     will not write into the pipe.So  we close fd[1]. */

for (i=0; i < 10; i++) {scanf("%d",&n);
/* Now that stdin has been      redirected, scanf automatically reads from the pipe. */
printf("n = %d\n",n);
/* try stdout of this has not changed . So this will be shown in the terminal.      */   sleep(1);
```

```
}}}
```

3.5 PRE-LAB QUESTIONS

1. What is the difference between $* and $@.
2. How to read a variable ,assign ,and access it

3.6 LAB ASSIGNMENT

1. Read a file name from command line and check it's a file or not.
2. Read a file name from command line and check if it read and write permission or not.

3.7 POST-LAB QUESTIONS

1. How to check if file is existing, it has read, write and execution permission.

EXPERIMENT 4

4.1 OBJECTIVE

a) Write a shell script that receives any number of file names as arguments checks if every argument supplied is a file or a directory and reports accordingly. Whenever the argument is a file, the number of lines on it is also reported.

b) *Illustrate by writing c program where process forks to a child, and create a child process by using forks and suddenly terminates itself?

4.2 RESOURCE/REQUIREMENTS

Linux operating system ,vi-editor, shell-interpreter

4.3 PROGRAM LOGIC

Read a list of files from current directory and display the file names to output along with number of lines of each file

4.4.a DESCRIPTION / PROCEDURE

```
echo enter the name
for fname in *
 do
 if test -f $fname
   then
   echo "file" $fname
   echo "number of lines" `cat $fname | wc -l`
 else if test -d $fname
   then
   echo "dir" $fname
 fi
fi
done
```

INPUT:

```
sh prog4.sh
```

OUTPUT:

```
enter the name
file 3.sh
number of lines 9
```

4.4.b DESCRIPTION / PROCEDURE

```
#include <stdio.h>
 /* for fork() */
#include <sys/types.h>
#include <unistd.h>
 /* for wait*() */
#include <sys/wait.h>
int main() {
   pid_t mypid, childpid;
   int status;
   mypid = getpid();
   printf("Hi. I'm the parent process. My pid is %d.\n", mypid);
   childpid = fork();
   if ( childpid == -1 ) {
      perror("Cannot proceed. fork() error");
```

```
        return 1;
    }
    if (childpid  == 0) {
      printf("Child 1: I inherited my parent's pid as %d.\n", mypid);
       mypid = getpid();
       printf("Child 1: getppid() tells my parent is %d. My own pid instead is %d.\n", getppid(), mypid);
       /* forks another child */
       childpid = fork();
       if ( childpid == -1 ) {
          perror("Cannot proceed. fork() error");
          return 1;
       }
       if (childpid == 0) {
      /* this is the child of the first child, thus "Child 2" */
          printf("Child 2: I hinerited my parent's PID as %d.\n", mypid);
          mypid = getpid();
          printf("Child 2: getppid() tells my parent is %d. My own pid  instead is %d.\n", getppid(),
                                                        mypid);
          childpid = fork();
          if ( childpid == -1 ) {
             perror("Cannot proceed. fork() error");
             return 1;
          }
          if (childpid == 0) {
           /* "Child 3" sleeps 30 seconds then terminates 12, hopefully before its   parent "Child 2" */
             printf("Child 3: I hinerited my parent's PID as %d.\n", mypid);
             mypid = getpid();
             printf("Child 3: getppid() tells my parent is %d. My own pid instead is %d.\n", getppid(),
                     mypid);
             sleep(30);
             return 12;
          } else   /* the parent "Child 2" suddendly returns 15 */  return 15;
       } else {
           /* this is still "Child 1", which waits for its child to exit */
          while ( waitpid(childpid, &status, WNOHANG) == 0 ) sleep(1);
          if ( WIFEXITED(status) ) printf("Child1: Child 2 exited with exit status %d.\n",
                                                                WEXITSTATUS(status));
          else printf("Child 1: child has not terminated correctly.\n");
       }

    } else {
       /* then we're the parent process, "Parent" */
       printf("Parent: fork() went ok. My child's PID is %d\n", childpid);
       /* wait for the child to terminate and report about that */
       wait(&status);
       if ( WIFEXITED(status) ) printf("Parent: child has exited with status %d.\n",
                                            WEXITSTATUS(status));
       else printf("Parent: child has not terminated normally.\n");
    }
    return 0;
}
```

4.5 PRE-LAB QUESTIONS

1. How to write arithmetic multiplication operator in shell.
2. Write down the syntax for nested if statement.

4.6 LAB ASSIGNMENT

1. Write a shell script to count number of txt,c and shell programs present in current directory.
2. Write a shell script to count number of only files present in current directory.

4.7 POST-LAB QUESTIONS

1. What is means by relation operator , name any three relation operators present in shell.
2. What is meant by logic operator, and Explain about each operator.

EXPERIMENT 5

5.1 OBJECTIVE

Write a shell script that accepts a list of file names as its arguments, counts and reports the occurrence of each word that is present in the first argument file on other argument files.

5.2 RESOURCE/REQUIREMENTS

Linux operating system ,vi-editor, shell-interpreter

5.3 PROGRAM LOGIC

Read list of file names and counts and report the occurrence of each word that is present in the first argument file on other argument files using comm and grep commands.

5.4 DESCRIPTION / PROCEDURE

```
echo Enter file name:
read file1
read file2
a=`comm -2 $file1 $file2`
b=`grep -c $a $file2`
echo Words contained in file one occurred in file two $b times
grep -n $a $file2
```

INPUT:

```
sh prog5.sh
Enter file name:
f1
myfile
```

OUTPUT:

```
Words contained in file one occured in file two 3 times
1:myfile contains
5:myfile
8:myfile
```

5.5 PRE-LAB QUESTIONS

1. What is difference between comm and diff commands
2. What is the significance of using single and double quotation in echo statement.
3. What are environmental variables.

5.6 LAB ASSIGNMENT

1. Write a script to reverse a given string.
2. Write a script to copy list of file into specified directory.

5.7 POST-LAB QUESTIONS

1. What is difference between grep, egrep and fgrep commands
2. How to copy and paste lines in a vi-editor
3. What are different types of shell and how to move from one shell to other shell

EXPERIMENT 6

6.1 OBJECTIVE

Write a shell script to list all of the directory files in a directory.

6.2 RESOURCE/REQUIREMENTS

Linux operating system ,vi-editor, shell-interpreter

6.3 PROGRAM LOGIC

1. Read a directory
2. Test given directory is directory file and exist using test options -d
3. If its directory and exist display all the sub directories and files to output stream
4. else display its not directory or not exists.

6.4 DESCRIPTION / PROCEDURE

```
echo " Enter dir name "
read dir
if [ -d $dir ]
then
   echo " Files in $dir are "
   ls $dir
else
   echo " Dir does not exist"
fi
```

INPUT:

```
sh Lp6.sh
Enter dir name
Presanna
```

OUTPUT:

```
Files in prasanna are
3.sh
4.sh
pp2.txt
```

6.5 PRE-LAB QUESTIONS

1. Name few test commands present in unix.
2. Write down the syntax case statement

6.6 LAB ASSIGNMENT

1. Write a shell script to count number of words present in a file without using commands.
2. Write a menu driven shell script to execute a command as 1.for `ls` ,2 for grep and 3 for cat.

6.7 POST-LAB QUESTIONS

1. What is the purpose of shift statement.
2. What the difference is between break and exit statement.

EXPERIMENT 7

7.1 OBJECTIVE
Write a shell script to find factorial of a given number.

7.2 RESOURCE/REQUIREMENTS
Linux operating system ,vi-editor, shell-interpreter

7.3 PROFRAM LOGIC
1. Read a filename
2. take a variable i for count and execute expression in for loop ,increment the variable till loop ends, 3. Display the factorial of given number to output stream.

7.4 DESCRIPTION / PROCEDURE
```
echo Factorial
echo Enter number:
read n
fact=1
i=1
for((i=1;i<=n;i++))
do
   fact=`expr $fact \* $i`
done
echo Factorial of $n is $fact
```
INPUT:
```
sh p7.sh
Factorial
Enter number:5
```
OUTPUT:
```
Factorial of 5 is 120
```

7.5 PRE-LAB QUESTIONS
1. What is the use of while loop.
2. How you do command substitution in shell script.
3. How do you access command line arguments from within a shell script?

7.6 LAB ASSIGNMENT
1. Write a shell script for menu driven to execute a command like 1 for ls,2 for ls –l, etc..
2. Write a shell script to find the sum of digits.

7.7 POST-LAB QUESTIONS
1. Illustrate the difference between while and for loop.
2. Which operator is used to check string is Null .
3. What is the name of the variable which counts number of arguments passed?

EXPERIMENT 8

8.1 OBJECTIVE

Write an awk script to count the number of lines in a file that do not contain vowels.

8.2 RESOURCE/REQUIREMENTS

Linux operating system ,vi-editor, shell-interpreter

8.3 PROGRAM LOGIC

1. Initialize total=0 in begin part
2. Using if command check each line to count the number of lines in a file that do not contain vowels in body part.
3. Display total lines to output stream that do not contain vowels in end part

8.4 DESCRIPTION / PROCEDURE

```
BEGIN{ print Displaying number of lines in a file that do not contain vowels
        total=0}
{if($0!~/[aeiouAEIOU]/)
total=total + 1}
END{print "The total lines in a file that do not contain vowels:",total}
```

INPUT:

```
awk prog8.awk  lp1.sh
Displaying number of lines in a file that do not contain vowels
```

OUTPUT:

```
The total lines in a file that do not contain vowels:1
```

8.5 PRE-LAB QUESTIONS

1. Define awk? What are the features of awk script?
2. What is the syntax for awk script?

8.6 LAB ASSIGNMENT

1. Write a awk script to display between lines eg.2 to 8 lines to output stream.
2. Write a awk script to find sum of total salary of all employees of given file.

8.7 POST-LAB QUESTIONS

1. Illustrate difference between awk and sed.
2. Write a awk command to search given pattern in file if found display to output entire line.

EXPERIMENT 9

9.1 OBJECTIVE
Write an awk script to find the number of characters, words and lines in a file.

9.2 RESOURCE/REQUIREMENTS
Linux operating system ,vi-editor, shell-interpreter

9.3 PROGRAM LOGIC
Write awk script to find the number of characters, words and lines in a file

9.4 DESCRIPTION / PROCEDURE
```
BEGIN{ print Displaying number of characters, words and lines in a file}
{word=words + NF}
{len = length($0)}
{charcount=charcount + len}
END{print The total number of characters, words and lines in a file is:
print("Words:\t",words)
print("Lines:\t",NR)
print("Chars:\t",len) }
```

INPUT:
```
awk prog9.awk lp5.sh
```

OUTPUT:
```
The total number of characters, words and lines in a file is:
Words:12
Lines:3
Chars:39
```

9.5 PRE-LAB QUESTIONS
1. How input file is given to awk script?
2. What is the use of next, getline and exit control actions in awk

9.6 LAB ASSIGNMENT
1. Write a awk script to convert lower case characters to upper case of given file.

9.7 POST-LAB QUESTIONS
1. Explain about associative arrays in awk.

EXPERIMENT 10

10.1 OBJECTIVE

Write a C program that makes a copy of a file using Systems calls

10.2 RESOURCE/REQUIREMENTS

Linux operating system, vi –editor, shell interpreter

10.3 PROGRAM LOGIC

1. Read source filename, destination filename
2. Open given source filename
3. Read the content from file and write to destination file
4. Repeat step 3 till end of file reaches
5. close open files

10.4 DESCRIPTION / PROCEDURE

```
#include<stdio.h>
#include<unistd.h>
#include<sys/types.h>
#include<string.h>
void main()  {
  char src[10], dest[10], buff;
  int fd,fd1;
  printf("enter the source file name \n");
  scanf("%s\n",src);
  fd=open("src",O_RDONLY);
  printf("enter the destination file name\n"
  scanf("%s\n",dest);
  fd1=open("dest",O_WRONLY|O_CREAT|O_TRUNC|S_IRUSR|S_IWUSR);
  while(read(fd,&buff,1));
  wirte(fd1,&buff,1);
  printf("The copy of a file is successed");
  close(fd);
  close(fd1);
}
```

INPUT:

```
cc prog10.c
./a.out
entr the source file name:
file1
enter the destination file name:
file2
```

OUTPUT:

The copy of a file is successes

10.5 PRE-LAB QUESTIONS

1. What is meant by file descriptor and user file descriptor starts from which number

10.6 LAB ASSIGNMENT

1. Write a c-program to count number lines in a file.

10.7 POST-LAB QUESTIONS

1. What are the file descriptors values of keyword, monitor, error.
2. What is the use of lseek() function

EXPERIMENT 11

11.A.1 OBJECTIVE

Write a C Program to Implement the Unix command cat using system calls.

11.A.2 RESOURCE/REQUIREMENTS

Linux operating system, vi –editor, shell interpreter

11. A.3 PROGRAM LOGIC

1. Open file which is input given by command line arguments
2. Read content from opened file
3. Display content to output stream
4. Repeat step 2 and step 3 till end of file reach.

11. A.4 DESCRIPTION / PROCEDURE

```
#include<fcntl.h>
#include<sys/stat.h>
#define BUFSIZE 1
int main(int argc, char **argv)
{
  int fd1;
  int n;
  char buf;
  fd1=open(argv[1],O_RDONLY);
 printf("Displaying content of file\n");
  while((n=read(fd1,&buf,1))>0)
  {
   printf("%c",buf);
/*          or
    write(1,&buf,1);  */
 }
  return (0);
}
```

INPUT:

cc prog11a.c unit1

OUTPUT:

Displays content of file

11. A.5 PRE-LAB QUESTIONS

1. What is the difference between open() and fopen()?

11. A.6 LAB ASSIGNMENT

1. Write a c-program to count number words in a file.

11. A.7 POST-LAB QUESTIONS

1. What is the difference between read(), write() and scanf, printf respectively

11.B.1 OBJECTIVE

Write a C Program to Implement the Unix command ls using system calls

11. B.2 RESOURCE/REQUIREMENTS

Linux operating system, vi –editor, c-compiler

11. B.3 PROGRAM LOGIC

1. Open a directory of given directory name
2. Scan directory and Read file and display filename to output stream
3. Repeat step 2 till eof directory reach.

11. B.4 DESCRIPTION / PROCEDURE

```
#include <sys/types.h>
#include <sys/dir.h>
#include <sys/param.h>
#include <stdio.h>
#define FALSE 0
#define TRUE 1
extern int alphasort();
char pathname[MAXPATHLEN];
main() {
int count,i;
struct dirent **files;
int file_select();
if (getcwd(pathname) == NULL )
{ printf("Error getting pathn");
exit(0);
}
printf("Current Working Directory = %sn",pathname);
count = scandir(pathname, &files, file_select, alphasort);
if (count <= 0)
{ printf("No files in this directoryn");
exit(0);
}
printf("Number of files = %d \n",count);
for (i=1;i<count+1;++i)
printf("%s \n",files[i-1]->d_name);
}
int file_select(struct direct *entry)
{
if ((strcmp(entry->d_name, ".") == 0) ||(strcmp(entry->d_name, "..") == 0))
return (FALSE);
else
return (TRUE);
 }
```

INPUT:

Cc prog11c.c

OUTPUT:

Current Working Directory =/home/prasanna
Number of files:2
Lp1.sh
lp2.sh

11. B.5 PRE-LAB QUESTIONS

1. What is the difference between lseek() and seekdir();

11. B.6 LAB ASSIGNMENT

1. Write a c-program to reverse a file using lseek();

11. B.7 POST-LAB QUESTIONS

1. What is the difference between system call and library functions

11.C.1 OBJECTIVE

Write a C Program to Implement the Unix command mv using system calls.

11. C.2 RESOURCE/REQUIREMENTS

Linux operating system, vi –editor, c-compiler

11. C.3 PROGRAM LOGIC

1. Open two files. Read and write mode respectively.
2. Using rename function move file content to another file
3. Remove source file using unlink() function.

11. C.4 DESCRIPTION / PROCEDURE

```
#include<fcntl.h>
#include<stdio.h>
#include<unistd.h>
#include<sys/stat.h>
int main(int argc, char **argv)
{
  int fd1,fd2;
  int n,count=0;
  fd1=open(argv[1],O_RDONLY);
fd2=creat(argv[2],S_IWUSR);
rename(fd1,fd2);
unlink(argv[1]);
return (0);
}
```

INPUT:

cc mv.c file1 file2

OUTPUT:

creates file2 and copies the content of file1 to file2 and removes file1

11. C.5 PRE-LAB QUESTIONS

1. What is the difference between cp and mv commands?

11. C.6 LAB ASSIGNMENT

1. Write a C program to move file content to another file without using functions rename and unlink.
2. Write a C program to copy content from source to destination file

11. C.7 POST-LAB QUESTIONS

1. What is difference between file descriptor and file pointer.

EXPERIMENT 12

12.1 OBJECTIVE

Write a C program that takes one or more file or directory names as command line input and reports the following information on the file.

1. file type
2. number of links
3. read, write and execute permissions
4. time of last access

(Note: use /fstat system calls)

12.2 RESOURCE/REQUIREMENTS

Linux operating system, vi –editor, c-compiler

12.3 PROGRAM LOGIC

1. Open a file using fopen() function
2. Read a file and display a file properties to output stream.

12.4 DESCRIPTION / PROCEDURE

```
#include<stdio.h>
main()
{
FILE *stream;
int buffer_character;
stream=fopen("test","r");
if(stream==(FILE*)0)
{
fprintf(stderr,"Error opening file(printed to standard error)\n");
fclose(stream);
exit(1);
}}
if(fclose(stream))==EOF)
{
fprintf(stderr,"Error closing stream.(printed to standard error)\n);
exit(1);
}
return();
}
```

12.5 PRE-LAB QUESTIONS

1. What is the difference between stat( ), fstat() and lstat() functions

12.6 LAB ASSIGNMENT

1.Write a C program to count number of words, lines and characters using system calls

12.7 POST-LAB QUESTIONS

1.List properties of files and different types of files in Linux

EXPERIMENT 13

13.1 OBJECTIVE

Write a C program to emulate the Unix ls – l command**.**

13.2 RESOURCE/REQUIREMENTS

Linux operating system, vi –editor, c-compiler

13.3 PROGRAM LOGIC

1. Declare and initialize required objects.
2. Read the directory name form the user.
3. Open the directory using opendir() system call and report error if the directory is not available.
4. Read the entry available in the directory.
5. Display the directory entry ie., name of the file or sub directory.
6. Repeat the step 6 and 7 until all the entries were read.

13.4 DESCRIPTION / PROCEDURE

```
#include <stddef.h>
#include <stdio.h>
#include <sys/types.h>
#include <dirent.h>
int main (void) {
DIR *dp;
struct dirent *ep;
dp = opendir (".");
if (dp != NULL) {
while (ep = readdir (dp))
printf("%s\n", ep->d_name);
closedir (dp);
}
else
perror ("Couldn't open the directory");
return 0;
}
```

INPUT: cc 13.c
./a.out
OUTPUT: 2 3 a.out 1 4 …. . 13.c

13.5 PRE-LAB QUESTIONS

1. What is the purpose of 0_CREAT and O_SYNC.

13.6 LAB ASSIGNMENT

1. Write a C-program to simulate `nl` command.

13.7 POST-LAB QUESTIONS

1. What is the use of -> operator?

EXPERIMENT 14

14.1 OBJECTIVE

Write a C program to list for every file in a directory, its inode number and file name

14.2 RESOURCE/REQUIREMENTS

Linux operating system, vi –editor, c-compiler

14.3 PROGRAM LOGIC

1. Read Directory name as input
2. Scan entire directory and display its directory file name and inode number till eof of the directory

14.4 DESCRIPTION / PROCEDURE

```
#include<stddef.h>
#include<stdio.h>
#include<sys/types.h>
#include<dirent.h>
main()
{
DIR *dp;
Struct dirent *p;
char dname[20];
Struct stat x;
Printf("Enter the directory name:");
Scanf("%s",dname);
Dp=opendir(dname);
Printf("\n FILE NAME \t INODE NUMBER \n");
While((p=readdir(dp))!=NULL)
{
        Printf("%s\t %/d\n",p->d_name,x.st_ino);
}
}
```

INPUT:

```
cc inode.c –o inode
./inode
```

OUTPUT:

```
  FILE NAME   INODE NUMBER
……..          4195164
File2.c          4195164
….
File1.c          4195164
```

14.5 PRE-LAB QUESTIONS

1. What is the system call that change the directory.
2. List all directory systems calls with example.

14.6 LAB ASSIGNMENT

1.Write a C-program to display all files using system calls

14.7 POST-LAB QUESTIONS

1. What flag is used to append the data /value in the file.

EXPERIMENT 15

15.1 OBJECTIVE

Write a C program that demonstrates redirection of standard output to a file. Ex: ls >f1.
/ freopen example: redirecting stdout */*

15.2 RESOURCE/REQUIREMENTS

Linux operating system, vi –editor, c-compiler

15.3 PROGRAM LOGIC

1. Open file using freopen() function
2. Redirect of standard to out put to a new file

15.4 DESCRIPTION / PROCEDURE

```
#include <stdio.h>
int main ()
{
freopen ("myfile.txt","w",stdout);
printf ("This sentence is redirected to a file which is given at output.");
fclose (stdout);
return 0;
}
```

INPUT:

```
cc 15.c –o file1
./file1
```

OUTPUT:

This sentence is redirected to a file which is given at output

15.5 PRE-LAB QUESTIONS

1.What is the purpose of fopen(),fread(),fwrite() functions

15.6 LAB ASSIGNMENT

1.Write a C-program to append content of file to another file

15.7 POST-LAB QUESTIONS

1. Explain different streams in linux
2. Illustrate different types of files in linux

EXPERIMENT 16

16.1 OBJECTIVE

Write a C program to create a child process and allow the parent to display "parent" and the child to display "child" on the screen.

16.2 RESOURCE/REQUIREMENTS

Linux operating system, vi –editor, c-compiler

16.3 PROGRAM LOGIC

1. Create child process using fork() function
2. If pid is not equal to zero execute parent block
3. else if pid is equal to zero execute child block
4. Display process id using getpid() and parent process id getppid() functions.

16.4 DESCRIPTION / PROCEDURE

```
#include <stdio.h>
main()
{
   int pid;
   printf("I'm the original process with PID %d and PPID %d.\n",getpid(),getppid());
   pid=fork();  /* Duplicate. Child and parent continue from here.*/
   if (pid!=0)  {/* pid is non-zero, so I must be the parent  */
         printf("I'm the parent process with PID %d and PPID %d.\n",getpid(),getppid());
         printf("My child's PID is %d.\n", pid);
      }
   else { /* pid is zero, so I must be the child. */
         printf("I'm the child process with PID %d and PPID %d.\n",getpid(),getppid());
      }
   printf("PID %d terminates.\n",pid);  /* Both processes execute this */
}
```

INPUT:

```
$cc fork.c              ... run the program.
./a.out
```

OUTPUT:

```
I'm the original process with PID 13292 and PPID 13273.
I'm the parent process with PID 13292 and PPID 13273.
My child's PID is 13293.
I'm the child process with PID 13293 and PPID 13292.
PID 13293 terminates.     ... child terminates.
PID 13292 terminates.     ... parent terminates.
```

16.5 PRE-LAB QUESTIONS

1. What are process identifiers in linux programming.
2. What is process, how you create new process?

16.6 LAB ASSIGNMENT

1. Write a program to find sum of odd numbers of parent process and sum of even numbers by child process.

16.7 POST-LAB QUESTIONS

1. Illustrate difference between fork() and vfork() functions.
2. What are different process ids in Linux programming?

EXPERIMENT 17

17.1 OBJECTIVE

Write a C program to create a Zombie process.

17.2 RESOURCE/REQUIREMENTS

Linux operating system, vi –editor, c-compiler

17.3 PROGRAM LOGIC

1. Create a new child process using fork()
2. If pid is not equal to zero execute parent process and block the process 1000 seconds
3. If pid is equal to zero executer child process and terminate immediately
4. To see already terminated process i.e zombie process open new command prompt and type ps -a it shows status **z** means that process is zombie process.

17.4 DESCRIPTION / PROCEDURE

```
#include <stdio.h>
main(){
   int pid;
   pid=fork();  /* Duplicate  */
   if (pid!=0) /* Branch based on return value from fork() */      {
        while (1) /* never terminate, and never execute a wait() */
          sleep(1000);       }
   else       {
        exit(42); /* Exit with a silly number */
     }}
```

INPUT:

```
$ cc prog17.c
./a.out&   ... execute the program in the background.
[1] 13545
```

OUTPUT:

```
$ ps
  PID TT STAT TIME COMMAND
13535 p2 s    0:00 -ksh(ksh) ...the shell
13545 p2 s    0:00 aombie.exe...the parent process
13536 p2 z    0:00 <defunct> ...the zombie child process
13537 p2 R    0:00 ps
$ kill 13545                ... kill the parent process.
[1]  Terminated   zombie.exe
$ ps                        ... notice the zombie is gone now.
  PID TT STAT TIME COMMAND
13535 p2 s    0:00 -csh(csh)
13548 p2 R    0:00 ps
```

17.5 PRE-LAB QUESTIONS

1. Define Zombie process.
2. How to know the status of process using commands.

17.6 LAB ASSIGNMENT

1. Write a C-program to create a child process to convert lower to upper case letters.

17.7 POST-LAB QUESTIONS

1. Explain different ways to terminate the process?

EXPERIMENT 18

18.1 OBJECTIVE

Write a C program that illustrates how an orphan is created.

18.2 RESOURCE/REQUIREMENTS

Linux operating system, vi –editor, c-compiler

18.3 PROGRAM LOGIC

1. Create a new child process using fork()
2. If pid is not equal to zero execute parent process block of code and terminates
3. If pid is equal to zero executer child process and block process for 5 seconds in mean time parent process terminates.
4. To see child process is became orphan process open new command prompt and type ps -a it shows status **O** means that process is orphan process.

18.4 DESCRIPTION / PROCEDURE

```
#include <stdio.h>
main()
{
   int pid;
   printf("I'm the original process with PID %d and PPID %d.\n",
         getpid(),getppid());
   pid=fork();  /* Duplicate. Child and parent continue from here.*/
   if (pid!=0)  /* Branch based on return value from fork() */
      {  /* pid is non-zero, so I must be the parent  */
         printf("I'm the parent process with PID %d and PPID %d.\n",
               getpid(),getppid());
         printf("My child's PID is %d.\n", pid);
      }
   else  {  /* pid is zero, so I must be the child.  */
         sleep(5); /*Make sure that the parent terminates first. */
         printf("I'm the child process with PID %d and PPID %d.\n",
               getpid(),getppid());
      }
   printf("PID %d terminates.\n",pid);  /* Both processes execute this */}
```

INPUT:

```
$cc prog18.c
./a.out              ... run the program.
```

OUTPUT:

```
I'm the original process with PID 13364 and PPID 13346.
I'm the parent process with PID 13364 and PPID 13346.
PID 13364 terminates.
I'm the child process with PID 13365 and PPID 1.   ...orphaned!
PID 13365 terminates.     ... child terminates.
```

18.5 PRE-LAB QUESTIONS

1. Illustrate difference between zombie process and orphan process.

18.6 LAB ASSIGNMENT

1. Write a C program to find given number is Armstrong number or not by parent process.

18.7 POST-LAB QUESTIONS

1. What is the child parent process id if child process status is orphan process?

EXPERIMENT 19

19.1 OBJECTIVE

Write a C program that illustrates how to execute two commands concurrently with a command pipe. Eg. ls-l|sort.

19.2 RESOURCE/REQUIREMENTS

Linux operating system, vi –editor, c-compiler

19.3 PROGRAM LOGIC

1. Open a directory of given directory name
2. Scan directory and Read file and display filename to output stream
3. Repeat step 2 till eof directory reach.
4. Using stat function test type of file, its permissions and file properties and display to output stream.

19.4 DESCRIPTION / PROCEDURE

```
#include<stdio.h>
#include<unistd.h>
#include<sys/stat.h>
#include<fcntl.h>
#include<stdlib.h>
#include<dirent.h>
main(int argc,char *argv[])
{
      struct dirent *p;
      DIR *dp;
      struct stat b;
      dp=opendir(".");
      if(dp==NULL)
      {
            printf("\nDirectory opening problem\n");
            return;
      }
      while((p=readdir(dp))!=NULL)
      {
           if(stat(p->d_name,&b)<0)
           {
                 printf("\nStat failure \n");
                 exit(0);
           }
          switch(b.st_mode &  S_IFMT )
           {
                 case S_IFREG: printf(" -");
                        break;
                 case S_IFDIR: printf(" d");
                               break;
                 case S_IFCHR: printf(" c");
                               break;
                 case S_IFBLK: printf(" b");
                               break;
                 case S_IFLNK: printf(" l");
                               break;
                 case S_IFSOCK: printf(" s");
```

```
                                        break;
                        case S_IFIFO: printf(" p");
                                        break;
                }
                if(S_IRUSR & b.st_mode)
                        printf(" r");
                else
                        printf(" -");
if(S_IWUSR & b.st_mode)
                        printf(" w");
                else
                        printf(" -");
                if(S_IXUSR & b.st_mode)
                        printf(" x");
                else
                        printf(" -");
                if(S_IRGRP & b.st_mode)
                        printf(" r");
                else
                        printf(" -");
                if(S_IWGRP & b.st_mode)
                        printf(" w");
                else
                        printf(" -");
                if(S_IXGRP & b.st_mode)
                        printf(" x");
                else
                        printf(" -");
                if(S_IROTH & b.st_mode)
                        printf(" r");
                else
printf(" -");
if(S_IWOTH & b.st_mode)
                        printf(" w");
                else
                        printf(" -");
                if(S_IXOTH & b.st_mode)
                        printf(" x");
                else
                        printf(" -");
                printf("%3d ",b.st_nlink);
                printf("%4d ",b.st_uid);
                printf("%4d ",b.st_gid);
                printf("%6d  ",b.st_size);
                printf("%9ld",b.st_ctime);
                printf("  %s\n",p->d_name);
        }
}
```

INPUT:

```
 vi 19.c
cc 19.c
 ./a.out
```

OUTPUT:

```
-rw-rw-r--  1  500  500   1506  1380610351  19.c
-rw-rw-r--  1  500  500      0  1380523478  2
-rw-rw-r--  1  500  500      0  1380523478  3
-rwxrwxr-x  1  500  500   6038  1380610357  a.out
-rw-rw-r--  1  500  500      0  1380523478  1
-rw-rw-r--  1  500  500    421  1380524812  12.c
-rw-rw-r--  1  500  500      0  1380523478  4
drwx------ 15  500  500   4096  1380609957  ..
drwxrwxr-x  2  500  500   4096  1380610357  .
-rw-rw-r--  1  500  500    347  1380523684  13.c
```

19.5 PRE-LAB QUESTIONS

1. What is the purpose of stat() functions.
2. Define file system?

19.6 LAB ASSIGNMENT

1. Write a C program to find number of sub directories in given directory.

19.7 POST-LAB QUESTIONS

1. Which system call is used to remove any type of file.

EXPERIMENT 20

20.1 OBJECTIVE

Write a C program in which a parent writes a message to a pipe and the child reads the message.

20.2 RESOURCE/REQUIREMENTS

Linux operating system, vi –editor, c-compiler

20.3 PROGRAM LOGIC

1. Create pipe using pipe() function and new process using fork() function
2. If parent process executing write a message to pipe at one end
3. If child process executing read a message from pipe at other end which is written by parent

20.4 DESCRIPTION / PROCEDURE

```
#include<stdio.h>
#include<stdlib.h>
#include<sys/types.h>
#include<unistd.h>
main() {
      int pfds[2];
      char buf[30];
    if(pipe(pfds) == -1)       {
            perror("Pipe is not created");
            exit(1);       }
      if(!fork())       {
            printf("CHILD: Writing to the pipe\n");
            write(pfds[1],"Hello World, I am child",23);
            printf("CHILD:Exiting");
            exit(0);       }
      else       {
            printf("PARENT:reading from the pipe\n");
            read(pfds[0],buf,23);
            printf("PARENT:Received Data is :%s",buf);
      }}
```

INPUT:

```
cc prog20.c
./a.out
```

OUTPUT:

```
CHILD: Writing to the pipe
CHILD:Exiting
PARENT:reading from the pipe
PARENT:Received Data is : Hello World, I am child
```

20.5 PRE-LAB QUESTIONS

1. What are the return values of pipe function?

20.6 LAB ASSIGNMENT

1. Implement two way communication using pipes.

20.7 POST-LAB QUESTIONS

1. What is the use of read and write system calls
2. How many processes will create if 4 fork functions executed?
3. Illustrate difference between pipes and named pipes?

EXPERIMENT 21

21.1 OBJECTIVE

Write a C program (sender.c) to create a message queue with read and write permissions to write 3 messages to it with different priority numbers.

21.2 RESOURCE/REQUIREMENTS

Linux operating system, vi –editor, c-compiler

21.3 PROGRAM LOGIC

1. Create process using msgget(), identify process using a key
2. Declare structure to store message and message type.
3. Using msgsnd() function write a message to message queue.

21.4 DESCRIPTION / PROCEDURE

```
#include<sys/types.h>
#include<sys/ipc.h>
#include<sys/msg.h>
#include<stdio.h>
#include<string.h>
#include<stdlib.h>
main()
{
int qid,len,i;
char s[15];
struct
{
long mtype;
char mtext[15];
}message,buff;
qid=msgget((key_t)10,IPC_CREAT|0666);
if(qid==-1)
{
perror("message queue create failed");
exit(1);
}
for(i=1;i<=3;i++)
{
printf("Enter the message to send\n");
scanf("%s",s);
strcpy(message.mtext,s);
message.mtype=i;
len=strlen(message.mtext);
if(msgsnd(qid,&message,len+1,0)==-1)
{
perror("message failed\n");
exit(1);
}
}
}
```

INPUT:

cc prog21.c

OUTPUT:

Enter the message to send: hi
Enter the message to send: hello, how are you
Enter the message to send: bye

21.5 PRE-LAB QUESTIONS

1. What is the purpose of msgget(),msgsnd(),msgrcv().
2. What is structure of message queue.

21.6 LAB ASSIGNMENT

1. Implement message queues like sender and receiver, where receiver can receive the message in un-order.

21.7 POST-LAB QUESTIONS

1. Describe use of pipe, fifos and message queues

EXPERIMENT 22

22.1 OBJECTIVE

Write a C program (receiver.c) that receives the messages (from the above message queue as specified in (22)) and displays them.

22.2 RESOURCE/REQUIREMENTS

Linux operating system, vi –editor, c-compiler

22.3

i) Connect client and server with socket() fuction.
ii) Provide socket address (port,Ip-address).
iii) Use connect() and bind(),listen() in client and server respectively.
iv) If successful use accept() in server.
v) Use fprintf() and read() for sending and receiving the data for communication between server and client.
vi) Close the connectrion

22.4 DESCRIPTION / PROCEDURE

```
#include<sys/types.h>
#include<sys/ipc.h>
#include<sys/msg.h>
#include<stdio.h>
#include<string.h>
#include<stdlib.h>
main()
{
int qid,len,i;
char s[15];
struct
{
long mtype;
char mtext[15];
}buff;
qid=msgget((key_t)10,IPC_CREAT|0666);
if(qid==-1)
{
perror("message queue create failed");
exit(0);
}
for(i=1;i<=3;i++)
{
if(msgrcv(qid,&buff,15,i,0)==-1)
{
perror("message failed\n");
exit(0);
}
printf("Message received from sender is %s\n",buff.mtext);
}
}
```

INPUT:

```
cc prog22.c
./a.out
```

OUTPUT:

Message received from sender is: hi
Message received from sender is: hello, how are you
Message received from sender is: bye

22.5 PRE-LAB QUESTIONS

1. How many message queues can create in linux programming and what are limitations for message queues?

22.6 LAB ASSIGNMENT

1. Implement message queues like sender and receiver, where sender sends number and receiver can receive the message and find square of received number.

22.7 POST-LAB QUESTIONS

1. Which system call is used to remove or delete a message queue

EXPERIMENT 23

23.1 OBJECTIVE

Write a C program to allow cooperating processes to lock a resource for exclusive use, using
a) Semaphores b) flock or lockf system calls.

23.2 RESOURCE/REQUIREMENTS

Linux operating system, vi –editor, c-compiler

23.3 PROGRAM LOGIC

1. Create semaphore and create key using ftok() function
2. Use semctl() to control the process.

23.4 DESCRIPTION / PROCEDURE

```
#include<stdio.h>
#include<stdlib.h>
#include<error.h>
#include<sys/types.h>
#include<sys/ipc.h>
#include<sys/sem.h>
int main(void)
{
key_t key;
int semid;
union semun arg;
if((key==ftok("sem demo.c","j"))== -1)
{
perror("ftok");
exit(1);
}
if(semid=semget(key,1,0666|IPC_CREAT))== -1)
{
perror("semget"):
exit(1);
}
arg.val=1;
if(semctl(semid,0,SETVAL,arg)== -1)
{
perror("smctl");
exit(1);
}
return 0;
}
```

23.5 PRE-LAB QUESTIONS

1. How to control the process in shared segment

23.6 LAB ASSIGNMENT

1. **Write** a program for shared memory form of IPC using producer consumer relation in such a way that consumer should read only after the producer has written some text to the shared memory.

23.7 POST-LAB QUESTIONS

1. Which system call is used to remove or delete a semaphore?

EXPERIMENT 24

24.1 OBJECTIVE

Write a C program that illustrates suspending and resuming processes using signals.

24.2 RESOURCE/REQUIREMENTS

Linux operating system, vi –editor, c-compiler

24.3 PROGRAM LOGIC

1. Use kill() to suspend and resume process

24.4 DESCRIPTION / PROCEDURE

```
#include<sys/types.h>
#include<signal.h>
//suspend the process(same as hitting crtl+z)
kill(pid,SIGSTOP);

//continue the process
kill(pid,SIGCONT);
```

24.5 PRE-LAB QUESTIONS

1. Define signal? Explain how to handle signals.

24.6 LAB ASSIGNMENT

1. Write a program to handle the signals like SIGINT , SIGFPE. SIGQUIT

24.7 POST-LAB QUESTIONS

1. Which signals cannot catch using signal handler functions.

EXPERIMENT 25

25.1 OBJECTIVE

Write a C program that implements a producer-consumer system with two processes. (using Semaphores).

25.2 RESOURCE/REQUIREMENTS

Linux operating system, vi –editor, c-compiler

25.3 PROGRAM LOGIC

1. create semaphore using semget( ) system call
2. if successful it returns positive value
3. create two new processes
4. first process will produce
5. until first process produces second process cannot consume

25.4 DESCRIPTION / PROCEDURE

```
#include<stdio.h>
#include<stdlib.h>
#include<sys/types.h>
#include<sys/ipc.h>
#include<sys/sem.h>
#include<unistd.h>
#define num_loops 2
int main(int argc,char* argv[])
{
int sem_set_id;
int child_pid,i,sem_val;
struct sembuf sem_op;
int rc;
struct timespec delay;
clrscr();
sem_set_id=semget(ipc_private,2,0600);
if(sem_set_id==-1)
{
perror("main:semget");
exit(1);
}
printf("semaphore set created,semaphore setid'%d'\n ",
sem_set_id);
child_pid=fork();
switch(child_pid)
{
case -1:
perror("fork");
exit(1);
case 0:
for(i=0;i<num_loops;i++)
{
sem_op.sem_num=0;
sem_op.sem_op=-1;
sem_op.sem_flg=0;
```

```
semop(sem_set_id,&sem_op,1);
printf("producer:'%d'\n",i);
fflush(stdout);
}
break;
default:
for(i=0;i<num_loops;i++)
{
printf("consumer:'%d'\n",i);
fflush(stdout);
sem_op.sem_num=0;
sem_op.sem_op=1;
sem_op.sem_flg=0;
semop(sem_set_id,&sem_op,1);
if(rand()>3*(rano_max14));
{
delay.tv_sec=0;
delay.tv_nsec=10;
nanosleep(&delay,null);
}
}
break;
}
return 0;
}
```

INPUT AND OUTPUT:

```
semaphore set created
semaphore set id '327690'
producer: '0'
consumer:'0'
producer:'1'
  consumer:'1'
```

25.5 PRE-LAB QUESTIONS

1. Define semaphores? What is their purpose? List and explain the APIs used to create and control the semaphores.

25.6 LAB ASSIGNMENT

1. Write a C program to remove semaphores using semaphore functions.

25.7 POST-LAB QUESTIONS

1. Which system call is used to remove or delete a semaphores?

EXPERIMENT 26

26.1 OBJECTIVE

Write client and server programs (using c) for interaction between server and client processes using Unix Domain sockets.

26.2 RESOURCE/REQUIREMENTS

Linux operating system, vi –editor, c-compiler

26.3 PROGRAM LOGIC

1. Create socket at client side and server side for communication.
2. Server side execute bind and accept function to accept client
3. Client side execute connect function to establish connection between client and server
4. Exchange information through unix domain sockets.
5. Close socket file descriptor after completion of communication.

26.4 DESCRIPTION / PROCEDURE

```
#include <stdio.h>
#include <sys/socket.h>
#include <sys/un.h>
#include <sys/types.h>
#include <unistd.h>
#include <string.h>

int connection_handler(int connection_fd)
{
 int nbytes;
 char buffer[256];

 nbytes = read(connection_fd, buffer, 256);
 buffer[nbytes] = 0;

 printf("MESSAGE FROM CLIENT: %s\n", buffer);
 nbytes = snprintf(buffer, 256, "hello from the server");
 write(connection_fd, buffer, nbytes);

 close(connection_fd);
 return 0;
}

int main(void)
{
 struct sockaddr_un address;
 int socket_fd, connection_fd;
 socklen_t address_length;
 pid_t child;

 socket_fd = socket(PF_UNIX, SOCK_STREAM, 0);
 if(socket_fd < 0)
 {
  printf("socket() failed\n");
  return 1;
 }
```

```
unlink("./demo_socket");

/* start with a clean address structure */
memset(&address, 0, sizeof(struct sockaddr_un));

address.sun_family = AF_UNIX;
snprintf(address.sun_path, UNIX_PATH_MAX, "./demo_socket");

if(bind(socket_fd,
        (struct sockaddr *) &address,
        sizeof(struct sockaddr_un)) != 0)
{
 printf("bind() failed\n");
 return 1;
}

if(listen(socket_fd, 5) != 0)
{
 printf("listen() failed\n");
 return 1;
}

while((connection_fd = accept(socket_fd,
                        (struct sockaddr *) &address,
                        &address_length)) > -1)
{
 child = fork();
 if(child == 0)
 {
  /* now inside newly created connection handling process */
  return connection_handler(connection_fd);
 }

 /* still inside server process */
 close(connection_fd);
}

close(socket_fd);
unlink("./demo_socket");
return 0;
}
```

Client.c

```
#include <stdio.h>
#include <sys/socket.h>
#include <sys/un.h>
#include <unistd.h>
#include <string.h>

int main(void)
{
 struct sockaddr_un address;
 int  socket_fd, nbytes;
 char buffer[256];
```

```
socket_fd = socket(PF_UNIX, SOCK_STREAM, 0);
if(socket_fd < 0)
{
 printf("socket() failed\n");
 return 1;
}

/* start with a clean address structure */
memset(&address, 0, sizeof(struct sockaddr_un));

address.sun_family = AF_UNIX;
snprintf(address.sun_path, UNIX_PATH_MAX, "./demo_socket");

if(connect(socket_fd,
        (struct sockaddr *) &address,
        sizeof(struct sockaddr_un)) != 0)
{
 printf("connect() failed\n");
 return 1;
}

nbytes = snprintf(buffer, 256, "hello from a client");
write(socket_fd, buffer, nbytes);

nbytes = read(socket_fd, buffer, 256);
buffer[nbytes] = 0;
printf("MESSAGE FROM SERVER: %s\n", buffer);
close(socket_fd);
return 0;
}
```

26.5 PRE-LAB QUESTIONS

1. Define socket?
2. Differentiate between connection oriented and connection less communication.

26.6 LAB ASSIGNMENT

1. Write a program to design a TCP client – server application which takes IP address, Port number and string to be echoed as command line inputs in client application and implements echo service.

26.7 POST-LAB QUESTIONS

1. Explain about IPV6 socket address structure and compare it with IPV4 and unix socket address structures.

EXPERIMENT 27

27.1 OBJECTIVE

Write client and server programs (using c) for interaction between server and client processes using Internet Domain sockets.

27.2 RESOURCE/REQUIREMENTS

Linux operating system, vi –editor, c-compiler

27.3 PROGRAM LOGIC

1. Connect client and server with socket() fuction.
2. Provide socket address (port, Ip-address) of Internet Domain Sockets.
3. Use connect() and bind(),listen() in client and server respectively.
4. If successful use accept() in server.
5. Use fprintf() and read() for sending and receiving the data for communication between server and client.
6. Close the connection

27.4 DESCRIPTION / PROCEDURE

Server.c

```
#include <sys/socket.h>
#include <netinet/in.h>
#include <arpa/inet.h>
#include <stdio.h>
#include <stdlib.h>
#include <unistd.h>
#include <errno.h>
#include <string.h>
#include <sys/types.h>
#include <time.h>

int main(int argc, char *argv[])
{
   int listenfd = 0, connfd = 0;
   struct sockaddr_in serv_addr;

   char sendBuff[1025];
   time_t ticks;

   listenfd = socket(AF_INET, SOCK_STREAM, 0);
   memset(&serv_addr, '0', sizeof(serv_addr));
   memset(sendBuff, '0', sizeof(sendBuff));

   serv_addr.sin_family = AF_INET;
   serv_addr.sin_addr.s_addr = htonl(INADDR_ANY);
   serv_addr.sin_port = htons(5000);

   bind(listenfd, (struct sockaddr*)&serv_addr, sizeof(serv_addr));

   listen(listenfd, 10);
```

```
    while(1)
    {
       connfd = accept(listenfd, (struct sockaddr*)NULL, NULL);

       ticks = time(NULL);
       snprintf(sendBuff, sizeof(sendBuff), "%.24s\r\n", ctime(&ticks));
       write(connfd, sendBuff, strlen(sendBuff));

       close(connfd);
       sleep(1);
     }
}
```

Client.c

```
#include <sys/socket.h>
#include <sys/types.h>
#include <netinet/in.h>
#include <netdb.h>
#include <stdio.h>
#include <string.h>
#include <stdlib.h>
#include <unistd.h>
#include <errno.h>
#include <arpa/inet.h>

int main(int argc, char *argv[])
{
    int sockfd = 0, n = 0;
    char recvBuff[1024];
    struct sockaddr_in serv_addr;

    if(argc != 2)
    {
       printf("\n Usage: %s <ip of server> \n",argv[0]);
       return 1;
    }

    memset(recvBuff, '0',sizeof(recvBuff));
    if((sockfd = socket(AF_INET, SOCK_STREAM, 0)) < 0)
    {
       printf("\n Error : Could not create socket \n");
       return 1;
    }

    memset(&serv_addr, '0', sizeof(serv_addr));

    serv_addr.sin_family = AF_INET;
    serv_addr.sin_port = htons(5000);
```

```
    if(inet_pton(AF_INET, argv[1], &serv_addr.sin_addr)<=0)
    {
       printf("\n inet_pton error occured\n");
       return 1;
    }

    if( connect(sockfd, (struct sockaddr *)&serv_addr, sizeof(serv_addr)) < 0)
    {
      printf("\n Error : Connect Failed \n");
      return 1;
    }

    while ( (n = read(sockfd, recvBuff, sizeof(recvBuff)-1)) > 0)
    {
       recvBuff[n] = 0;
       if(fputs(recvBuff, stdout) == EOF)
       {
          printf("\n Error : Fputs error\n");
       }
    }

    if(n < 0)
    {
       printf("\n Read error \n");
    }

    return 0;
}
```

27.5 PRE-LAB QUESTIONS

1. Explain about IPV6 socket address structure and compare it with IPV4 and unix socket address structures.

27.6 LAB ASSIGNMENT

1. Write a program to implement UDP client server application in which client takes an file name from the command line and sends to the server. Server returns the content of received file to the client.

27.7 POST-LAB QUESTIONS

1. Which system call is used to UDP Client Server Communication.

EXPERIMENT 28

28.1 OBJECTIVE

Implement shared memory form of IPC

28.2 RESOURCE/REQUIREMENTS

Linux operating system, vi –editor, c-compiler

28.3 PROGRAM LOGIC

i) Create two process ,
ii) Use shmget() to connect the processes.
iii) Take a memory and attach this two process using shmat()
iv) Now both the process can access the common memory.

28.4 DESCRIPTION / PROCEDURE

```
#include <sys/types.h>
#include <sys/ipc.h>
#include <sys/shm.h>
#include <stdio.h>
#define SHMSZ     27
main()
{    char c;
   int shmid;
   key_t key;
   char *shm, *s;
   key = 5678;
   if ((shmid = shmget(key, SHMSZ, IPC_CREAT | 0666)) < 0)
 {       perror("shmget");
      exit(1);     }
   if ((shm = shmat(shmid, NULL, 0)) == (char *) -1)
 {       perror("shmat");
      exit(1);     }
   s = shm;
   for (c = 'a'; c <= 'z'; c++)
      *s++ = c;
   *s = NULL;
   while (*shm Linux operating  system, vi –editor, c-compiler= '*')
     sleep(1);
   exit(0);
}
```

shm_client.c

```
#include <sys/types.h>
#include <sys/ipc.h>
#include <sys/shm.h>
#include <stdio.h>
#define SHMSZ     27
main()
{
   int shmid;
   key_t key;
   char *shm, *s;
    key = 5678;
```

```
    if ((shmid = shmget(key, SHMSZ, 0666)) < 0) {
        perror("shmget");
        exit(1);    }
    if ((shm = shmat(shmid, NULL, 0)) == (char *) -1) {
        perror("shmat");
        exit(1);    }
    for (s = shm; *s != NULL; s++)
        putchar(*s);
    putchar('\n');

    *shm = '*';

    exit(0);
}
```

28.5 PRE-LAB QUESTIONS

1. What are the advantages of using shared memory?
2. shmget() has how many parameters.

28.6 LAB ASSIGNMENT

1. Write a c-program to implement shred memory among more than one process with effecting critical region (i.e. synchronization).

28.7 POST-LAB QUESTIONS

1. Which system call used to assign or allocate memory to the process?

DATA MINING LAB

Credit Risk Assessment

Description: The business of banks is making loans. Assessing the credit worthiness of an applicant is of crucial importance. You have to develop a system to help a loan officer decide whether the credit of a customer is good, or bad. A bank's business rules regarding loans must consider two opposing factors. On the one hand, a bank wants to make as many loans as possible. Interest on these loans is the ban's profit source. On the other hand, a bank cannot afford to make too many bad loans. Too many bad loans could lead to the collapse of the bank. The bank's loan policy must involve a compromise not too strict, and not too lenient.

To do the assignment, you first and foremost need some knowledge about the world of credit . You can acquire such knowledge in a number of ways.

1. Knowledge Engineering. Find a loan officer who is willing to talk. Interview her and try to represent her knowledge in the form of production rules.
2. Books. Find some training manuals for loan officers or perhaps a suitable textbook on finance. Translate this knowledge from text form to production rule form.
3. Common sense. Imagine yourself as a loan officer and make up reasonable rules which can be used to judge the credit worthiness of a loan applicant.
4. Case histories. Find records of actual cases where competent loan officers correctly judged when not to, approve a loan application.

The German Credit Data :

Actual historical credit data is not always easy to come by because of confidentiality rules. Here is one such dataset (original) Excel spreadsheet version of the German credit data (download from web).

In spite of the fact that the data is German, you should probably make use of it for this assignment, (Unless you really can consult a real loan officer !)

A few notes on the German dataset :

- DM stands for Deutsche Mark, the unit of currency, worth about 90 cents Canadian (but looks and acts like a quarter).
- Owns_telephone. German phone rates are much higher than in Canada so fewer people own telephones.
- Foreign_worker. There are millions of these in Germany (many from Turkey). It is very hard to get German citizenship if you were not born of German parents.
- There are 20 attributes used in judging a loan applicant. The goal is the classify the applicant into one of two categories, good or bad.

Subtasks : (Turn in your answers to the following tasks)

EXPERIMENT-1

1.1 OBJECTIVE:

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

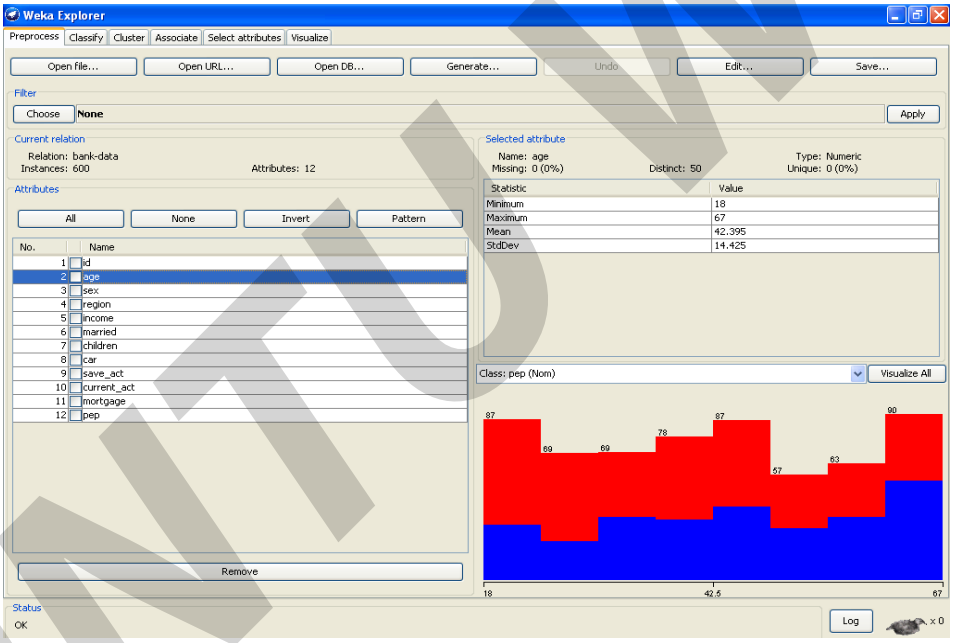
1.2 RESOURCES:

Weka mining tool

1.3 PROCEDURE:

1) Open the Weka GUI Chooser.
2) Select EXPLORER present in Applications.
3) Select Preprocess Tab.
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 selected attribute.

1.4 OUTPUT:



EXPERIMENT-2

2.1 OBJECTIVE:

Which attributes do you think might be crucial in making the credit assessment? Come up with some simple rules in plain English using your selected attributes.

2.2 RESOURCES:

Weka mining tool

2.3 THEORY:

Association rule mining is defined as: Let be a set of *n* binary attributes called items. Let be a set of transactions called the database. Each transaction in *D* has a unique transaction ID and contains a subset of the items in *I*. A rule is defined as an implication of the form X=>Y where X,Y C I and X Π Y=Φ . The sets of items (for short itemsets) X and Y are called antecedent (left hand side or LHS) and consequen*t* (right hand side or RHS) of the rule respectively.

To illustrate the concepts, we use a small example from the supermarket domain.

The set of items is *I* = {milk,bread,butter,beer} and a small database containing the items (1 codes presence and 0 absence of an item in a transaction) is shown in the table to the right. An example rule for the supermarket could be meaning that if milk and bread is bought, customers also buy butter.

Note: this example is extremely small. In practical applications, a rule needs a support of several hundred transactions before it can be considered statistically significant, and datasets often contain thousands or millions of transactions.

To select interesting rules from the set of all possible rules, constraints on various measures of significance and interest can be used. The best known constraints are minimum thresholds on support and confidence. The support supp(*X*) of an itemset *X* is defined as the proportion of transactions in the data set which contain the itemset. In the example database, the itemset {milk, bread} has a support of 2 / 5 = 0.4 since it occurs in 40% of all transactions (2 out of 5 transactions).

The confidence of a rule is defined. For example, the rule has a confidence of 0.2 / 0.4 = 0.5 in the database, which means that for 50% of the transactions containing milk and bread the rule is correct. Confidence can be interpreted as an estimate of the probability $P(Y \mid X)$, the probability of finding the RHS of the rule in transactions under the condition that these transactions also contain the LHS .

ALGORITHM:

Association rule mining is to find out association rules that satisfy the predefined minimum support and confidence from a given database. The problem is usually decomposed into two sub problems. One is to find those itemsets whose occurrences exceed a predefined threshold in the database; those itemsets are called frequent or large itemsets. The second problem is to generate association rules from those large itemsets with the constraints of minimal confidence.

Suppose one of the large itemsets is Lk, Lk = {I1, I2, … , Ik}, association rules with this itemsets are generated in the following way: the first rule is {I1, I2, … , Ik1} and {Ik}, by checking the confidence this rule can be determined as interesting or not. Then other rule are generated by deleting the last items in the antecedent and inserting it to the consequent, further the confidences of the new rules are checked to determine the interestingness of them. Those processes iterated until the antecedent becomes empty.

Since the second subproblem is quite straight forward, most of the researches focus on the first subproblem. The Apriori algorithm finds the frequent sets *L* In Database *D*.

· Find frequent set *Lk* − 1.

· Join Step.

o *Ck* is generated by joining *Lk* − 1with itself

· Prune Step.

o Any (*k* − 1) itemset that is not frequent cannot be a subset of a

frequent *k* itemset, hence should be removed.

Where · (*Ck*: Candidate itemset of size *k*)

· (*Lk*: frequent itemset of size *k*)

Apriori Pseudocode

Apriori (T,£)

L<{ Large 1itemsets that appear in more than transactions }

K<2

while L(k1)≠ Φ

C(k)<Generate(*Lk* − 1)

for transactions t Є T

C(t)Subset(*Ck*,*t*)

for candidates c Є C(t)

count[c]<count[c]+1

L(k)<{ c Є C(k)| count[c] ≥ *£*

K<K+ 1

return Ụ L(k) k

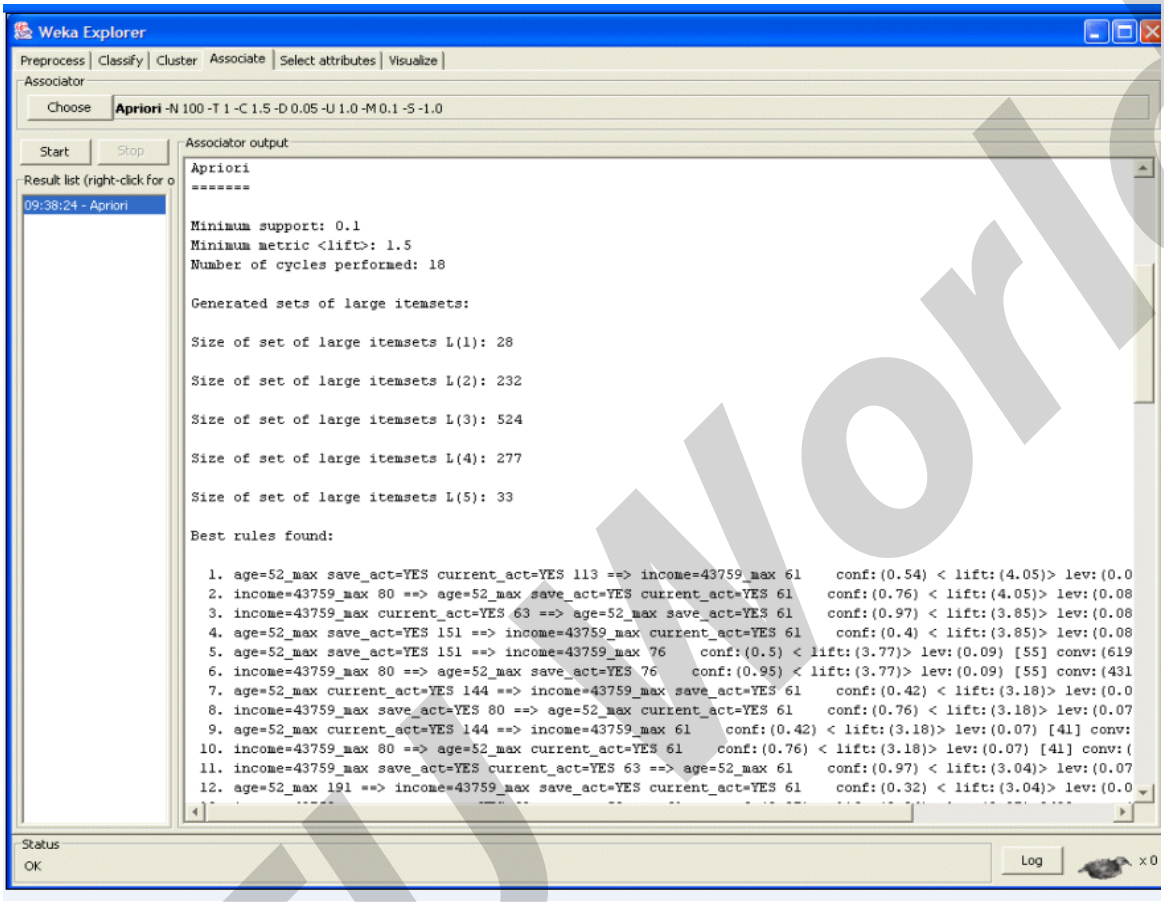
2.4 PROCEDURE:

1) Given the Bank database for mining.
2) Select EXPLORER in WEKA GUI Chooser.
3) Load “Bank.csv” in Weka by Open file in Preprocess tab.
4) Select only Nominal values.
5) Go to Associate Tab.
6) Select Apriori algorithm from “Choose “ button present in Associator
weka.associations.Apriori -N 10 -T 0 -C 0.9 -D 0.05 -U 1.0 -M 0.1 -S -1.0 -c -1

7) Select Start button

8) Now we can see the sample rules.

2.5 OUTPUT:



EXPERIMENT-3

3.1 OBJECTIVE:

*What attributes do you think might be crucial in making the bank assessment?

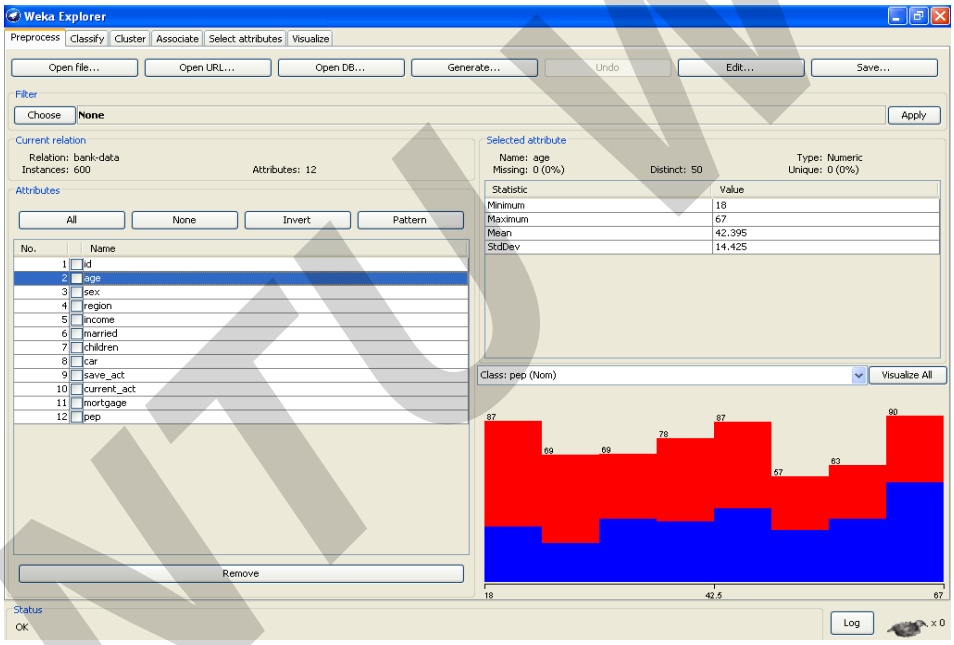
3.2 RESOURCES:

Weka mining tool

3.3 PROCEDURE:

1) Open the Weka GUI Chooser.
2) Select EXPLORER present in Applications.
3) Select Preprocess Tab.
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 selected attribute.

3.4 OUTPUT:



EXPERIMENT-4

4.1 OBJECTIVE:

One type of model that you can create is a decision tree. Train a decision tree using the complete dataset as the training data. Report the model obtained after training.

4.2 RESOURCES:

Weka mining tool

4.3 THEORY:

Classification is a data mining function that assigns items in a collection to target categories or classes. The goal of classification is to accurately predict the target class for each case in the data. For example, a classification model could be used to identify loan applicants as low, medium, or high credit risks. A classification task begins with a data set in which the class assignments are known. For example, a classification model that predicts credit risk could be developed based on observed data for many loan applicants over a period of time.

In addition to the historical credit rating, the data might track employment history, home ownership or rental, years of residence, number and type of investments, and so on. Credit rating would be the target, the other attributes would be the predictors, and the data for each customer would constitute a case.

Classifications are discrete and do not imply order. Continuous, floating point values would indicate a numerical, rather than a categorical, target. A predictive model with a numerical target uses a regression algorithm, not a classification algorithm. The simplest type of classification problem is binary classification. In binary classification, the target attribute has only two possible values: for example, high credit rating or low credit rating. Multiclass targets have more than two values: for example, low, medium, high, or unknown credit rating. In the model build (training) process, a classification algorithm finds relationships between the values of the predictors and the values of the target. Different classification algorithms use different techniques for finding relationships. These relationships are summarized in a model, which can then be applied to a different data set in which the class assignments are unknown.

Classification models are tested by comparing the predicted values to known target values in a set of test data. The historical data for a classification project is typically divided into two data sets: one for building the model; the other for testing the model. Scoring a classification model results in class assignments and probabilities for each case. For example, a model that classifies customers as low, medium, or high value would also predict the probability of each classification for each customer. Classification has many applications in customer segmentation, business modeling, marketing, credit analysis, and biomedical and drug response modeling.

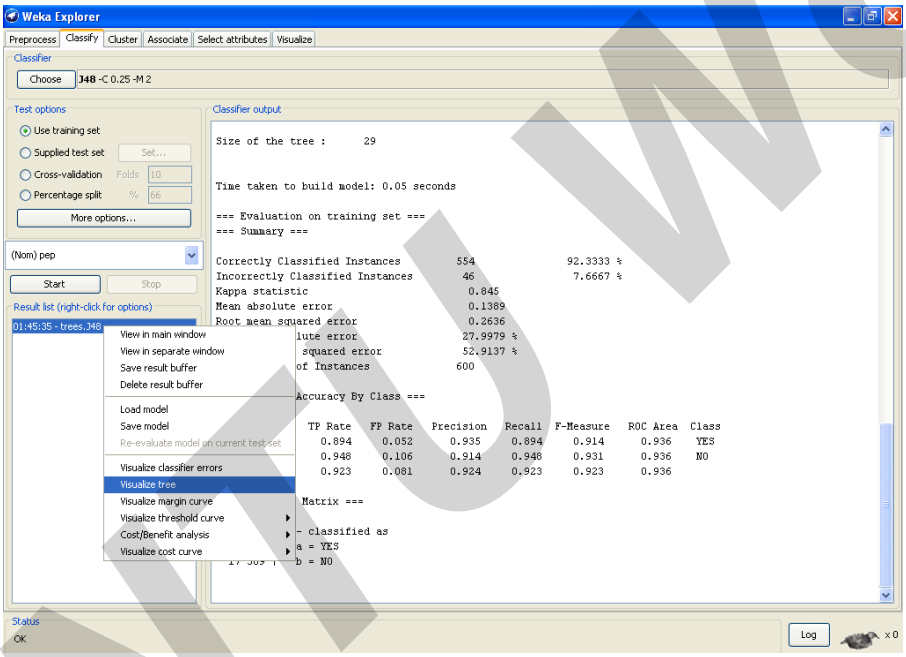
Different Classification Algorithms: Oracle Data Mining provides the following algorithms for classification:

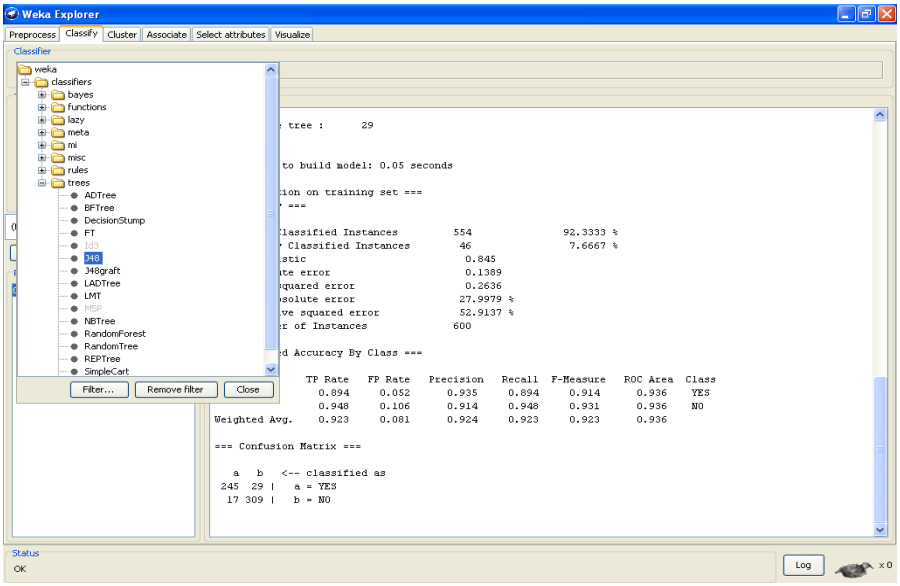
- Decision Tree - Decision trees automatically generate rules, which are conditional statements that reveal the logic used to build the tree.
- Naive Bayes - Naive Bayes uses Bayes' Theorem, a formula that calculates a probability by counting the frequency of values and combinations of values in the historical data.

4.4 PROCEDURE:

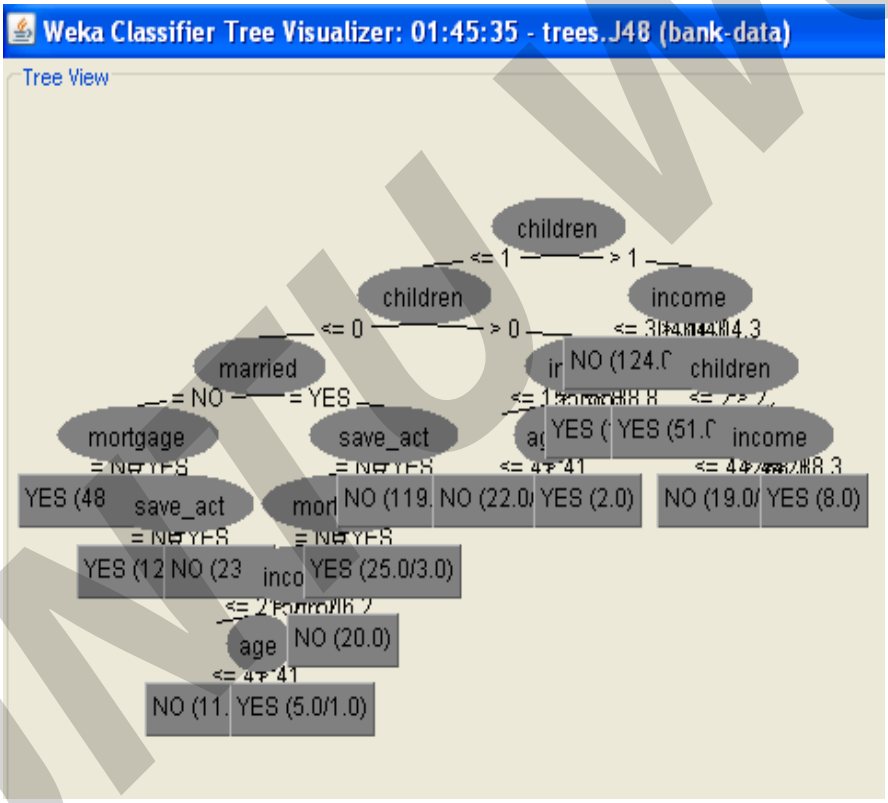
1) Open Weka GUI Chooser.
2) Select EXPLORER present in Applications.
3) Select Preprocess Tab.
4) Go to OPEN file and browse the file that is already stored in the system “bank.csv”.
5) Go to Classify tab.
6) Here the c4.5 algorithm has been chosen which is entitled as j48 in Java and can be selected by clicking the button choose and select tree j48
7) Select Test options “Use training set”
8) if need select attribute.
9) Click Start.
10) Now we can see the output details in the Classifier output.
11) Right click on the result list and select “visualize tree”option .

4.5 OUTPUT:





The decision tree constructed by using the implemented C4.5 algorithm



EXPERIMENT-5

5.1 OBJECTIVE:

Suppose you use your above model trained on the complete dataset, and classify credit good/bad for each of the examples in the dataset. What % of examples can you classify correctly? (This is also called testing on the training set) Why do you think you cannot get 100 % training accuracy?

5.2 RESOURCES:

Weka mining tool

5.3 THEORY:

Naive Bayes classifier assumes that the presence (or absence) of a particular feature of a class is unrelated to the presence (or absence) of any other feature. For example, a fruit may be considered to be an apple if it is red, round, and about 4" in diameter. Even though these features depend on the existence of the other features, a naive Bayes classifier considers all of these properties to independently contribute to the probability that this fruit is an apple.

An advantage of the naive Bayes classifier is that it requires a small amount of training data to estimate the parameters (means and variances of the variables) necessary for classification. Because independent variables are assumed, only the variances of the variables for each class need to be determined and not the entire covariance matrix The naive Bayes probabilistic model :

The probability model for a classifier is a conditional model

P(C|F1 .....Fn) over a dependent class variable *C* with a small number of outcomes or *classes*, conditional on several feature variables *F*1 through *Fn*. The problem is that if the number of features *n* is large or when a feature can take on a large number of values, then basing such a model on probability tables is infeasible. We therefore reformulate the model to make it more tractable.

Using Bayes' theorem, we write

P(C|F1.....Fn)=[{p(C)p(F1.....Fn|C)}/p(F1,.....Fn)]

In plain English the above equation can be written as

Posterior= [(prior *likehood)/evidence]

In practice we are only interested in the numerator of that fraction, since the denominator does not depend on *C* and the values of the features *Fi* are given, so that the denominator is effectively constant. The numerator is equivalent to the joint probability model p(C,F1.....Fn) which can be rewritten as follows, using repeated applications of the definition of conditional probability:

p(C,F1.....Fn) =p(C) p(F1.....Fn|C) =p(C)p(F1|C) p(F2.....Fn|C,F1,F2)

=p(C)p(F1|C) p(F2|C,F1)p(F3.....Fn|C,F1,F2)

= p(C)p(F1|C) p(F2|C,F1)p(F3.....Fn|C,F1,F2).....p(Fn|C,F1,F2,F3.....Fn1)

Now the "naive" conditional independence assumptions come into play: assume that each feature *Fi* is conditionally independent of every other feature *Fj* for j≠i .

This means that p(Fi|C,Fj)=p(Fi|C) and so the joint model can be expressed as
p(C,F1,.....Fn)=p(C)p(F1|C)p(F2|C).....= p(C)π p(Fi|C)

This means that under the above independence assumptions, the conditional distribution over the class variable *C* can be expressed like this:

p(C|F1.....Fn)= p(C) πp(Fi|C) Z

where *Z* is a scaling factor dependent only on F1.....Fn, i.e., a constant if the values of the feature variables are known.

Models of this form are much more manageable, since they factor into a so called *class prior p*(*C*) and independent probability distributions p(Fi|C). If there are *k* classes and if a model for each p(Fi|C=c) can be expressed in terms of *r* parameters, then the corresponding naive Bayes model has $(k - 1) + n\ r\ k$ parameters. In practice, often $k = 2$ (binary classification) and $r = 1$ (Bernoulli variables as features) are common, and so the total number of parameters of the naive Bayes model is $2n + 1$, where n is the number of binary features used for prediction

P(h/D)= P(D/h) P(h) P(D)

• P(h) : Prior probability of hypothesis h

• P(D) : Prior probability of training data D

• P(h/D) : Probability of h given D

• P(D/h) : Probability of D given h

Naïve Bayes Classifier : Derivation

• D : Set of tuples

– Each Tuple is an 'n' dimensional attribute vector

– X : (x1,x2,x3,…. xn)

• Let there me 'm' Classes : C1,C2,C3…Cm

• NB classifier predicts X belongs to Class Ci iff

– P (Ci/X) > P(Cj/X) for 1<= j <= m , j <> i

• Maximum Posteriori Hypothesis

– P(Ci/X) = P(X/Ci) P(Ci) / P(X)

– Maximize P(X/Ci) P(Ci) as P(X) is constant

Naïve Bayes Classifier : Derivation

• With many attributes, it is computationally expensive to evaluate P(X/Ci)

• Naïve Assumption of "class conditional independence"

• P(X/Ci) = n P(xk/ Ci)

k = 1

• P(X/Ci) = P(x1/Ci) * P(x2/Ci) *…* P(xn/ Ci)

5.4 PROCEDURE:

1) Given the Bank database for mining.
2) Use the Weka GUI Chooser.
3) Select EXPLORER present in Applications.
4) Select Preprocess Tab.
5) Go to OPEN file and browse the file that is already stored in the system "bank.csv".
6) Go to Classify tab.
7) Choose Classifier "Tree"
8) Select "NBTree" i.e., Navie Baysiean tree.
9) Select Test options "Use training set"
10) if need select attribute.
11) Now start weka.
12) Now we can see the output details in the Classifier output.

5.5 OUTPUT:

=== Evaluation on training set ===
=== Summary ===

Correctly Classified Instances	554	92.3333 %
Incorrectly Classified Instances	46	7.6667 %
Kappa statistic	0.845	
Mean absolute error	0.1389	
Root mean squared error	0.2636	
Relative absolute error	27.9979 %	
Root relative squared error	52.9137 %	
Total Number of Instances	600	

=== Detailed Accuracy By Class ===

TP Rate	FP Rate	Precision	Recall	F-Measure	ROC Area	Class
0.894	0.052	0.935	0.894	0.914	0.936	YES
0.948	0.106	0.914	0.948	0.931	0.936	NO
			Weighted Avg.			
0.923	0.081	0.924	0.923	0.923	0.936	

=== Confusion Matrix ===

A	b
245	29
17	309

<-- classified as
a = YES , b = NO

EXPERIMENT-6

6.1 OBJECTIVE:

*Find out the correctly classified instances, root mean squared error, kappa statistics, and mean absolute error for weather data set?

6.2 RESOURCES:

Weka mining tool

6.3 THEORY:

Naive Bayes classifier assumes that the presence (or absence) of a particular feature of a class is unrelated to the presence (or absence) of any other feature. For example, a fruit may be considered to be an apple if it is red, round, and about 4" in diameter. Even though these features depend on the existence of the other features, a naive Bayes classifier considers all of these properties to independently contribute to the probability that this fruit is an apple.

An advantage of the naive Bayes classifier is that it requires a small amount of training data to estimate the parameters (means and variances of the variables) necessary for classification. Because independent variables are assumed, only the variances of the variables for each class need to be determined and not the entire covariance matrix The naive Bayes probabilistic model :

The probability model for a classifier is a conditional model

P(C|F1 .....Fn) over a dependent class variable *C* with a small number of outcomes or *classes*, conditional on several feature variables *F*1 through *Fn*. The problem is that if the number of features *n* is large or when a feature can take on a large number of values, then basing such a model on probability tables is infeasible. We therefore reformulate the model to make it more tractable.

Using Bayes' theorem, we write

P(C|F1.....Fn)=[{p(C)p(F1.....Fn|C)}/p(F1,.....Fn)]

In plain English the above equation can be written as

Posterior= [(prior *likehood)/evidence]

In practice we are only interested in the numerator of that fraction, since the denominator does not depend on *C* and the values of the features *Fi* are given, so that the denominator is effectively constant. The numerator is equivalent to the joint probability model p(C,F1.....Fn) which can be rewritten as follows, using repeated applications of the definition of conditional probability:

p(C,F1.....Fn) =p(C) p(F1.....Fn|C) =p(C)p(F1|C) p(F2.....Fn|C,F1,F2)

=p(C)p(F1|C) p(F2|C,F1)p(F3.....Fn|C,F1,F2)

= p(C)p(F1|C) p(F2|C,F1)p(F3.....Fn|C,F1,F2).....p(Fn|C,F1,F2,F3.....Fn1)

Now the "naive" conditional independence assumptions come into play: assume that each feature *Fi* is conditionally independent of every other feature *Fj* for j≠i .

This means that p(Fi|C,Fj)=p(Fi|C) and so the joint model can be expressed as
p(C,F1,.....Fn)=p(C)p(F1|C)p(F2|C).....= p(C)π p(Fi|C)

This means that under the above independence assumptions, the conditional distribution over the class variable *C* can be expressed like this:

p(C|F1.....Fn)= p(C) πp(Fi|C) Z

where *Z* is a scaling factor dependent only on F1.....Fn, i.e., a constant if the values of the feature variables are known.

Models of this form are much more manageable, since they factor into a so called *class prior p*(*C*) and independent probability distributions p(Fi|C). If there are *k* classes and if a model for each p(Fi|C=c) can be expressed in terms of *r* parameters, then the corresponding naive Bayes model has $(k - 1) + n\ r\ k$ parameters. In practice, often $k = 2$ (binary classification) and $r = 1$ (Bernoulli variables as features) are common, and so the total number of parameters of the naive Bayes model is $2n + 1$, where n is the number of binary features used for prediction

P(h/D)= P(D/h) P(h) P(D)

• P(h) : Prior probability of hypothesis h

• P(D) : Prior probability of training data D

• P(h/D) : Probability of h given D

• P(D/h) : Probability of D given h

Naïve Bayes Classifier : Derivation

• D : Set of tuples

– Each Tuple is an 'n' dimensional attribute vector

– X : (x1,x2,x3,…. xn)

• Let there me 'm' Classes : C1,C2,C3…Cm

• NB classifier predicts X belongs to Class Ci iff

– P (Ci/X) > P(Cj/X) for 1<= j <= m , j <> i

• Maximum Posteriori Hypothesis

– P(Ci/X) = P(X/Ci) P(Ci) / P(X)

– Maximize P(X/Ci) P(Ci) as P(X) is constant

Naïve Bayes Classifier : Derivation

• With many attributes, it is computationally expensive to evaluate P(X/Ci)

• Naïve Assumption of "class conditional independence"

• P(X/Ci) = n P(xk/ Ci)

k = 1

• P(X/Ci) = P(x1/Ci) * P(x2/Ci) *…* P(xn/ Ci)

6.4 PROCEDURE:

1) Given the Bank database for mining.
2) Use the Weka GUI Chooser.
3) Select EXPLORER present in Applications.
4) Select Preprocess Tab.
5) Go to OPEN file and browse the file that is already stored in the system "bank.csv".
6) Go to Classify tab.
7) Choose Classifier "Tree"
8) Select "NBTree" i.e., Navie Baysiean tree.
9) Select Test options "Use training set"
10) if need select attribute.
11) Now start weka.
12) Now we can see the output details in the Classifier output.

6.5 OUTPUT:

=== Evaluation on training set ===
=== Summary ===

Correctly Classified Instances	554	92.3333 %
Incorrectly Classified Instances	46	7.6667 %
Kappa statistic	0.845	
Mean absolute error	0.1389	
Root mean squared error	0.2636	
Relative absolute error	27.9979 %	
Root relative squared error	52.9137 %	
Total Number of Instances	600	

=== Detailed Accuracy By Class ===

TP Rate	FP Rate	Precision	Recall	F-Measure	ROC Area	Class
0.894	0.052	0.935	0.894	0.914	0.936	YES
0.948	0.106	0.914	0.948	0.931	0.936	NO
			Weighted Avg.			
0.923	0.081	0.924	0.923	0.923	0.936	

=== Confusion Matrix ===

A	b
245	29
17	309

<-- classified as
a = YES , b = NO

EXPERIMENT-7

7.1 OBJECTIVE:

Is testing on the training set as you did above a good idea? Why or Why not?

7.2 RESOURCES:

Weka Mining tool

7.3 PROCEDURE:

1) In Test options, select the Supplied test set radio button
2) Click Set
3) Choose the file which contains records that were not in the training set we used to create the model.
4) Click Start(WEKA will run this test data set through the model we already created.)
5) Compare the output results with that of the 4th experiment

7.4 OUTPUT:

This can be experienced by the different problem solutions while doing practice.

The important numbers to focus on here are the numbers next to the "Correctly Classified Instances" (92.3 percent) and the "Incorrectly Classified Instances" (7.6 percent). Other important numbers are in the "ROC Area" column, in the first row (the 0.936); Finally, in the "Confusion Matrix," it shows the number of false positives and false negatives. The false positives are 29, and the false negatives are 17 in this matrix.

Based on our accuracy rate of 92.3 percent, we say that upon initial analysis, this is a good model.

One final step to validating our classification tree, which is to run our test set through the model and ensure that accuracy of the model

Comparing the "Correctly Classified Instances" from this test set with the "Correctly Classified Instances" from the training set, we see the accuracy of the model, which indicates that the model will not break down with unknown data, or when future data is applied to it.

EXPERIMENT-8

8.1 OBJECTIVE:

One approach for solving the problem encountered in the previous question is using cross-validation? Describe what is cross -validation briefly. Train a Decision Tree again using cross - validation and report your results. Does your accuracy increase/decrease? Why?

8.2 RESOURCES:

Weka mining tool

8.3 THEORY:

Decision tree learning, used in data mining and machine learning, uses a decision tree as a predictive model which maps observations about an item to conclusions about the item's target value In these tree structures, leaves represent classifications and branches represent conjunctions of features that lead to those classifications. In decision analysis, a decision tree can be used to visually and explicitly represent decisions and decision making. In data mining, a decision tree describes data but not decisions; rather the resulting classification tree can be an input for decision making. This page deals with decision trees in data mining.

Decision tree learning is a common method used in data mining. The goal is to create a model that predicts the value of a target variable based on several input variables. Each interior node corresponds to one of the input variables; there are edges to children for each of the possible values of that input variable. Each leaf represents a value of the target variable given the values of the input variables represented by the path from the root to the leaf.

A tree can be "learned" by splitting the source set into subsets based on an attribute value test. This process is repeated on each derived subset in a recursive manner called recursive partitioning. The recursion is completed when the subset at a node all has the same value of the target variable, or when splitting no longer adds value to the predictions. In data mining, trees can be described also as the combination of mathematical and computational techniques to aid the description, categorization and generalization of a given set of data.

Data comes in records of the form:

(x, y) = (x1, x2, x3..., xk, y)

The dependent variable, Y, is the target variable that we are trying to understand, classify or generalize. The vector x is comprised of the input variables, x1, x2, x3 etc., that are used for that task.

8.4 PROCEDURE:

1) Given the Bank database for mining.
2) Use the Weka GUI Chooser.
3) Select EXPLORER present in Applications.
4) Select Preprocess Tab.
5) Go to OPEN file and browse the file that is already stored in the system “bank.csv”.

6) Go to Classify tab.

7) Choose Classifier "Tree"

8) Select J48

9) Select Test options "Cross-validation".

10) Set "Folds" Ex:10

11) if need select attribute.

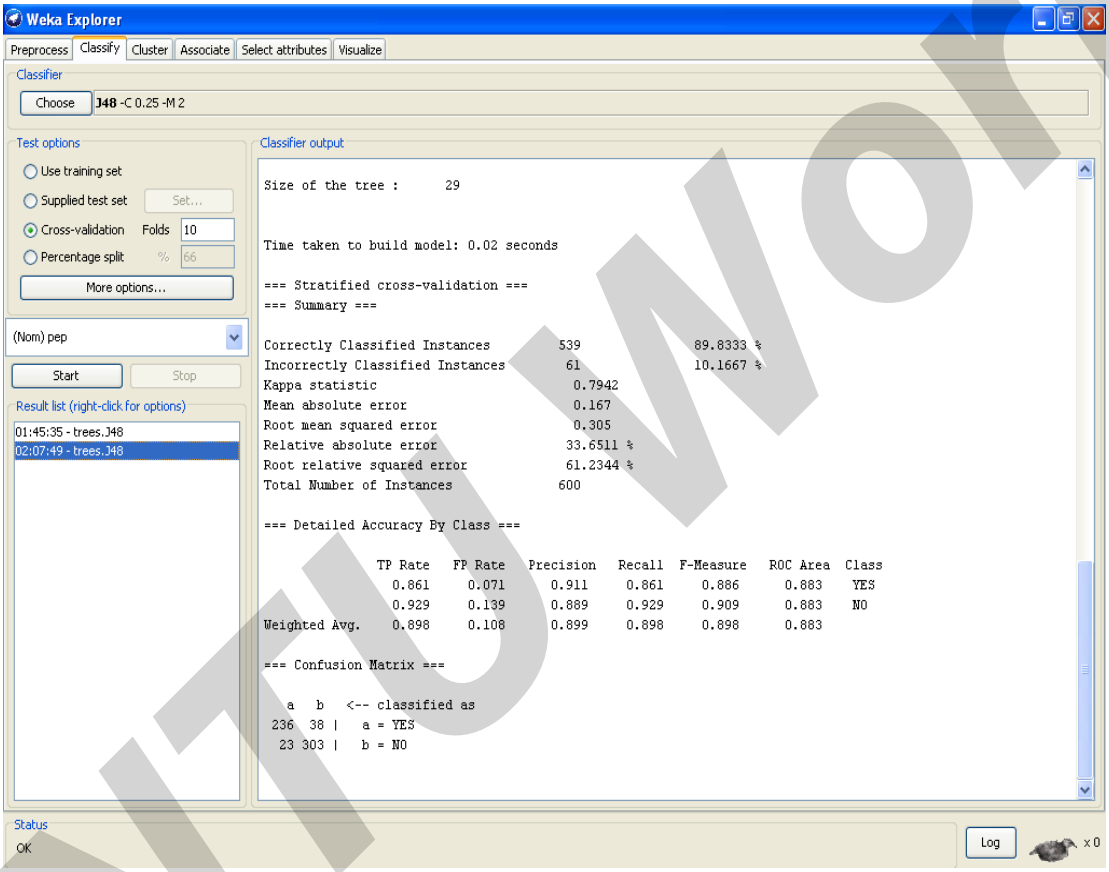
12) now Start weka.

13)now we can see the output details in the Classifier output.

14)Compare the output results with that of the 4th experiment

15) check whether the accuracy increased or decreased?

8.5 OUTPUT:



=== Stratified cross-validation ===

=== Summary ===

Correctly Classified Instances 539 89.8333 %

Incorrectly Classified Instances 61 10.1667 %

Kappa statistic 0.7942

Mean absolute error 0.167

Root mean squared error 0.305

Relative absolute error 33.6511 %

Root relative squared error 61.2344 %

Total Number of Instances 600

=== Detailed Accuracy By Class ===

TP Rate	FP Rate	Precision	Recall	F-Measure	ROC Area	Class
0.861	0.071	0.911	0.861	0.886	0.883	YES
0.929	0.139	0.889	0.929	0.909	0.883	NO
			Weighted Avg.			
0.898	0.108	0.899	0.898	0.898	0.883	

=== Confusion Matrix ===

a	b	<-- classified as
236	38	a = YES
23	303	b = NO

EXPERIMENT-9

9.1 OBJECTIVE

Check to see if the data shows a bias against "foreign workers" (attribute 20), or "personal -status" (attribute 9). One way to do this (perhaps rather simple minded) is to remove these attributes from the dataset and see if the decision tree created in those cases is significantly different from the full dataset case which you have already done. To remove an attribute you can use the preprocess tab in Weka's GUI Explorer. Did removing these attributes have any significant effect? Discuss.

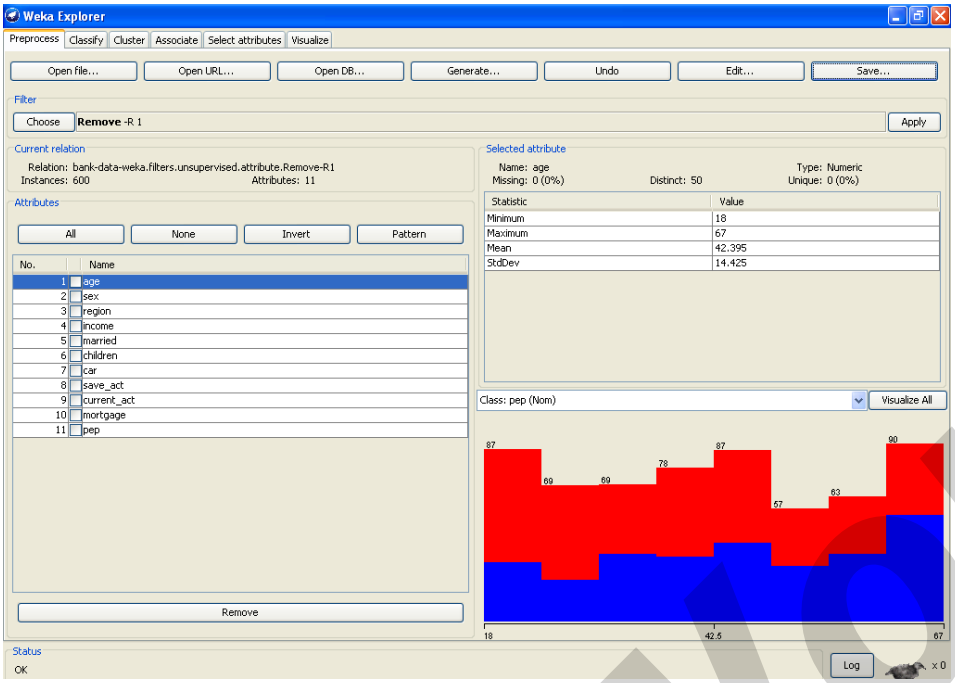
9.2 RESOURCES:

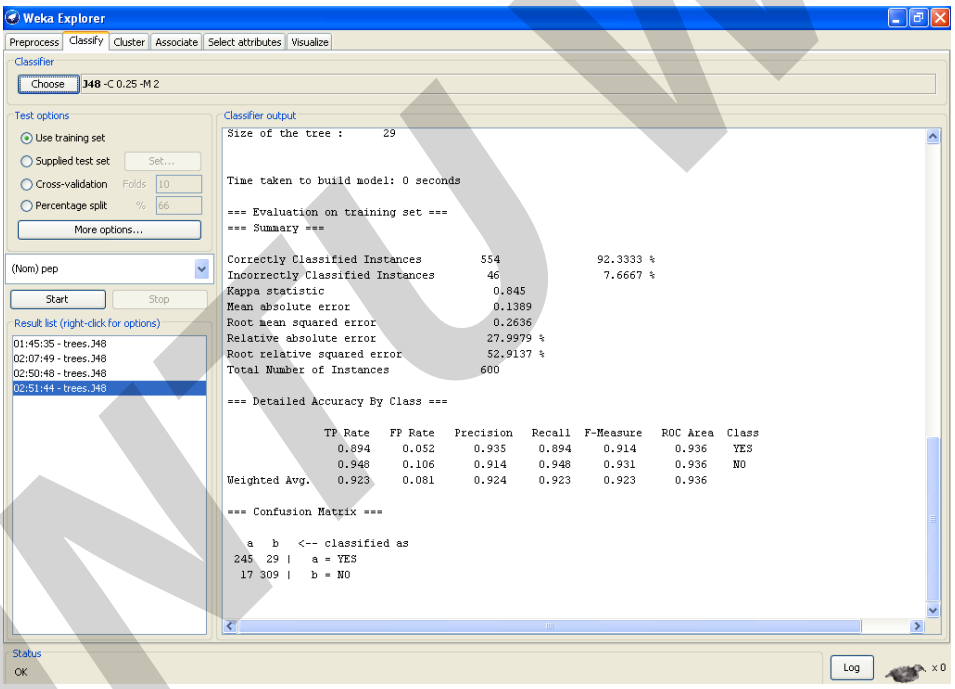
Weka mining tool

9.3 PROCEDURE:

1) Given the Bank database for mining.
2) Use the Weka GUI Chooser.
3) Select EXPLORER present in Applications.
4) Select Preprocess Tab.
5) Go to OPEN file and browse the file that is already stored in the system “bank.csv”.
6) In the "Filter" panel, click on the "Choose" button. This will show a popup window with list available filters.
7) Select “weka.filters.unsupervised.attribute.Remove”
8) Next, click on text box immediately to the right of the "Choose" button
9) In the resulting dialog box enter the index of the attribute to be filtered out (Make sure that the "invert Selection" option is set to false)
10) Then click "OK" . Now, in the filter box you will see "Remove -R 1"
11) Click the "Apply" button to apply this filter to the data. This will remove the "id" attribute and create a new working relation
12) To save the new working relation as an ARFF file, click on save button in the top panel.
13) Go to OPEN file and browse the file that is newly saved (attribute deleted file)
14) Go to Classify tab.
15) Choose Classifier “Tree”
16) Select j48 tree
17) Select Test options “Use training set”
18) If need select attribute.
19) Now start weka.
20) Now we can see the output details in the Classifier output.
21) Right click on the result list and select ” visualize tree “option .
22) Compare the output results with that of the 4th experiment
23) Check whether the accuracy increased or decreased.
24) Check whether removing these attributes have any significant effect.

9.4 OUTPUT:





EXPERIMENT-10

10.1 OBJECTIVE:

*Load the 'weather.arff' dataset in Weka and run the ID3 classification algorithm. What problem do you have and what is the solution?

10.2 RESOURCES:

Weka Mining tool

10.3 PROCEDURE:

1) In Test options, select the Supplied test set radio button
2) click Set
3) Choose the file which contains records that were not in the training set we used to create the model.
4) Click Start (WEKA will run this test data set through the model we already created.)
5) Compare the output results with that of the 4th experiment

10.4 OUTPUT:

This can be experienced by the different problem solutions while doing practice.

The important numbers to focus on here are the numbers next to the "Correctly Classified Instances" (92.3 percent) and the "Incorrectly Classified Instances" (7.6 percent). Other important numbers are in the "ROC Area" column, in the first row (the 0.936); finally, in the "Confusion Matrix," it shows the number of false positives and false negatives. The false positives are 29, and the false negatives are 17 in this matrix.

Based on our accuracy rate of 92.3 percent, we say that upon initial analysis, this is a good model.

One final step to validating our classification tree, which is to run our test set through the model and ensure that accuracy of the model

Comparing the "Correctly Classified Instances" from this test set with the "Correctly Classified Instances" from the training set, we see the accuracy of the model, which indicates that the model will not break down with unknown data, or when future data is applied to it.

EXPERIMENT-11

11.1 OBJECTIVE:

Another question might be, do you really need to input so many attributes to get good results? Maybe only a few would do. For example, you could try just having attributes 2, 3, 5, 7, 10, 17 (and 21, the class attribute (naturally)). Try out some combinations. (You had removed two attributes in problem 7. Remember to reload the arff data file to get all the attributes initially before you start selecting the ones you want).

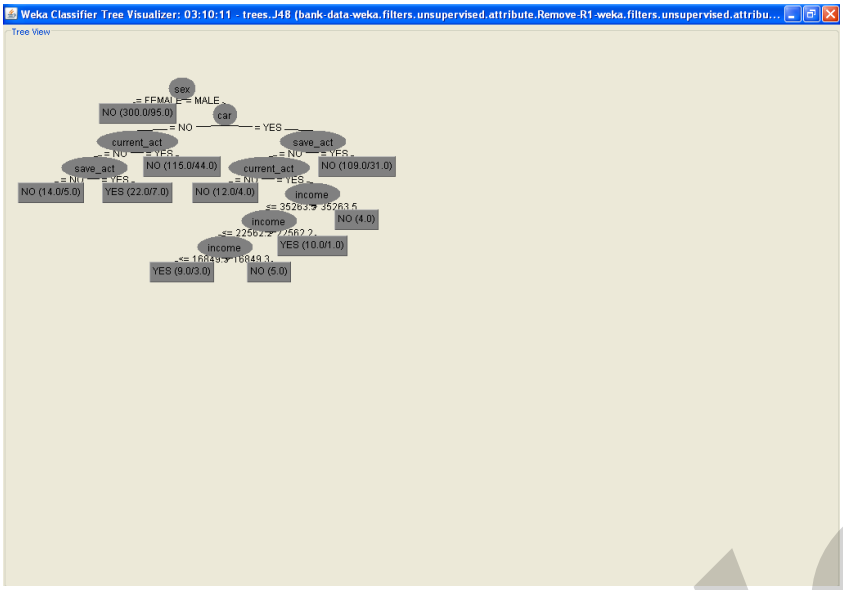
11.2 RESOURCES:

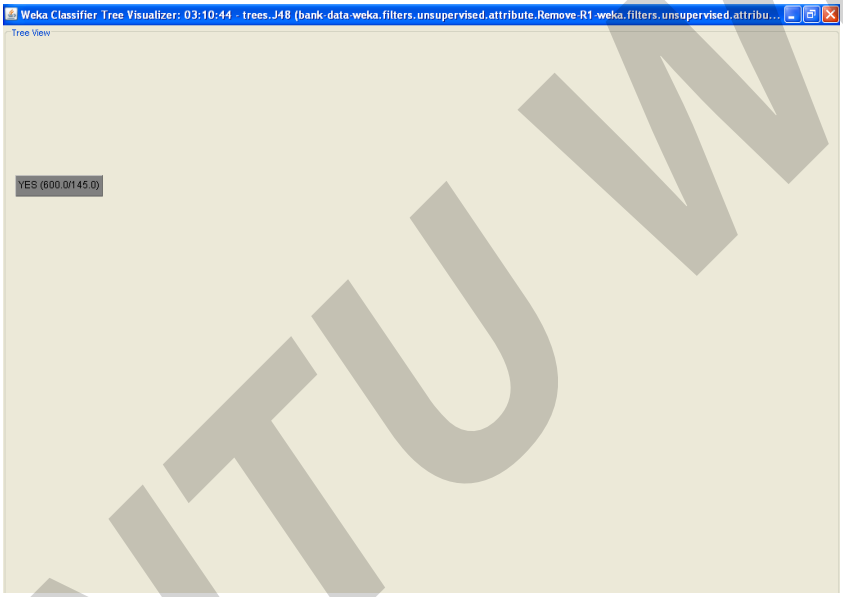
Weka mining tool.

11.3 PROCEDURE:

1) Given the Bank database for mining.
2) Use the Weka GUI Chooser.
3) Select EXPLORER present in Applications.
4) Select Preprocess Tab.
5) Go to OPEN file and browse the file that is already stored in the system "bank.csv".
6) Select some of the attributes from attributes list which are to be removed. With this step only the attributes necessary for classification are left in the attributes panel.
7) The go to Classify tab.
8) Choose Classifier "Tree"
9) Select j48
10) Select Test options "Use training set"
11) If need select attribute.
12) Now start Weka.
13) Now we can see the output details in the Classifier output.
14) Right click on the result list and select "visualize tree" option.
15) Compare the output results with that of the 4^{th} experiment.
16) Check whether the accuracy increased or decreased?
17) Check whether removing these attributes have any significant effect.

11.4 OUTPUT:





EXPERIMENT-12

12.1 OBJECTIVE:

Sometimes, the cost of rejecting an applicant who actually has a good credit (case 1) might be higher than accepting an applicant who has bad credit (case 2). Instead of counting the misclassifications equally in both cases, give a higher cost to the first case (say cost 5) and lower cost to the second case. You can do this by using a cost matrix in Weka. Train your Decision Tree again and report the Decision Tree and cross -validation results. Are they significantly different from results obtained in problem 6 (using equal cost)?

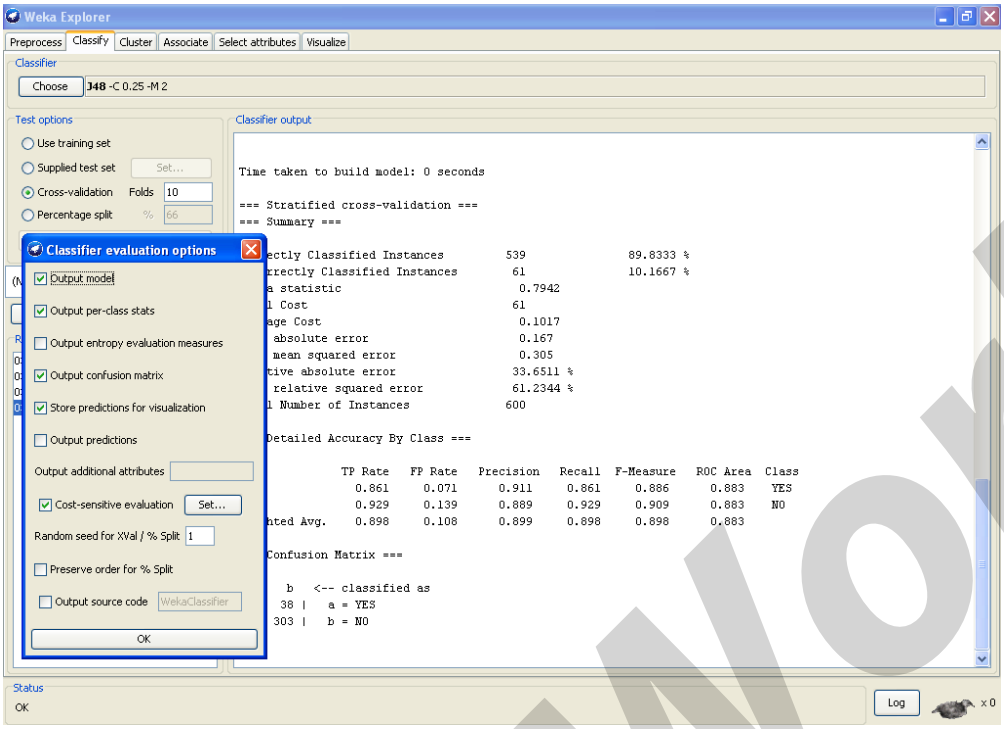
12.2 RESOURCES:

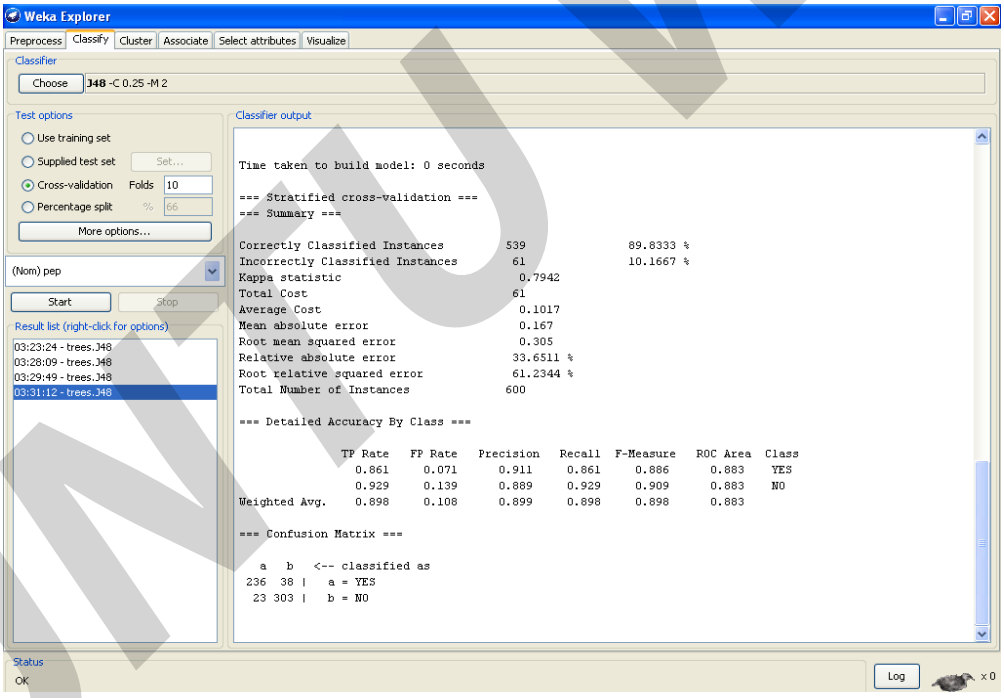
Weka mining tool

12.3 PROCEDURE:

1) Given the Bank database for mining.
2) Use the Weka GUI Chooser.
3) Select EXPLORER present in Applications.
4) Select Preprocess Tab.
5) Go to OPEN file and browse the file that is already stored in the system "bank.csv".
6) Go to Classify tab.
7) Choose Classifier "Tree"
8) Select j48
9) Select Test options "Training set".
10) Click on "more options".
11) Select cost sensitive evaluation and click on set button
12) Set the matrix values and click on resize. Then close the window.
13) Click Ok
14) Click start.
15) We can see the output details in the Classifier output
16) Select Test options "Cross-validation".
17) Set "Folds" Ex: 10
18) if need select attribute.
19) Now start weka.
20) Now we can see the output details in the Classifier output.
21) Compare results of 15^{th} and 20^{th} steps.
22) Compare the results with that of experiment 6.

12.4 OUTPUT:





EXPERIMENT-13

13.1 OBJECTIVE:

Do you think it is a good idea to prefer simple decision trees instead of having long complex decision trees? How does the complexity of a Decision Tree relate to the bias of the model?

13.2 RESOURCES:

Weka mining tool

13.3 PROCEDURE:

This will be based on the attribute set, and the requirement of relationship among attribute we want to study. This can be viewed based on the database and user requirement.

EXPERIMENT-14

14.1 OBJECTIVE:

*Run the J48 and 1Bk classifiers using-the cross-validation strategy with various fold levels. Compare the accuracy results. Hold out strategy with three percentage levels. Compare the accuracy results.

14.2 RESOURCES:

Weka mining tool

14.3 THEORY:

Reduced-error pruning

- Each node of the (over-fit) tree is examined for pruning
- A node is pruned (removed) only if the resulting pruned tree performs no worse than the original over the validation set
- Pruning a node consists of
  - Removing the sub-tree rooted at the pruned node
  - Making the pruned node a leaf node
  - Assigning the pruned node the most common classification of the training instances attached to that node
- Pruning nodes iteratively
  - Always select a node whose removal most increases the DT accuracy over the validation set
  - Stop when further pruning decreases the DT accuracy over the validation set

IF (Children=yes) Λ (income=>30000)

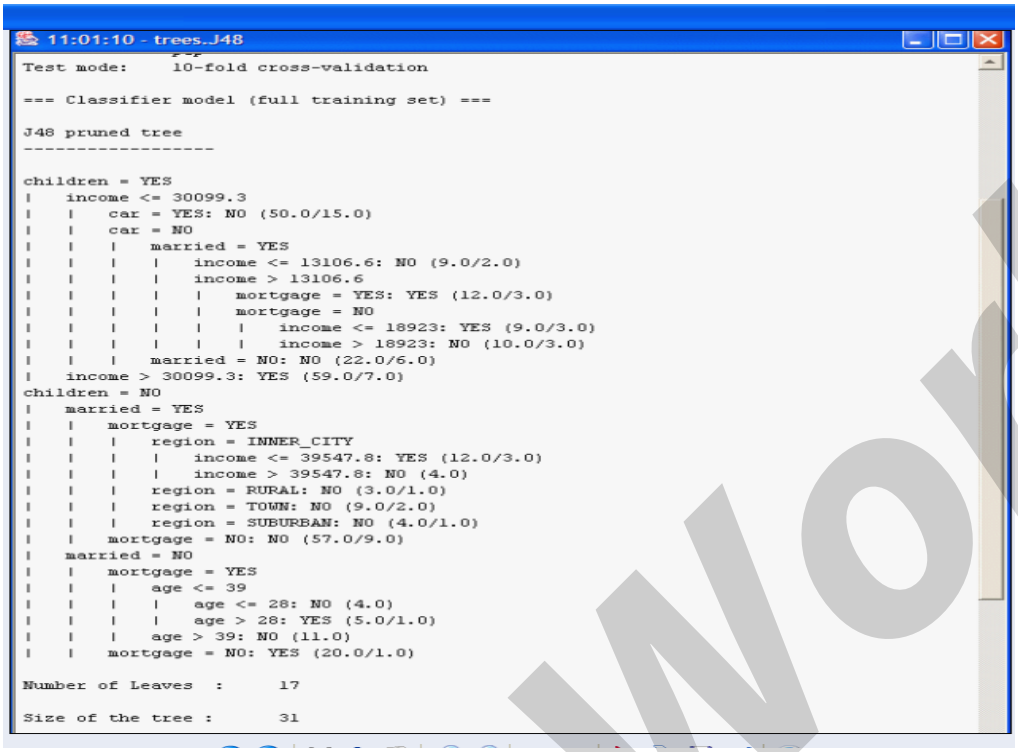
THEN (car=Yes)

14.4 PROCEDURE:

1) Given the Bank database for mining.
2) Use the Weka GUI Chooser.
3) Select EXPLORER present in Applications.
4) Select Preprocess Tab.
5) Go to OPEN file and browse the file that is already stored in the system “bank.csv”.
6) Select some of the attributes from attributes list
7) Go to Classify tab.
8) Choose Classifier “Tree”
9) Select “NBTree” i.e., Navie Bayesian tree.
10) Select Test options “Use training set”
11) Right click on the text box besides choose button, select show properties
12) Now change unprune mode “false” to “true”.
13) Change the reduced error pruning % as needed.
14) If need select attribute.
15) Now start Weka.
16) Now we can see the output details in the Classifier output.

17) Right click on the result list and select “visualize tree” option.

14.5 OUTPUT:



EXPERIMENT-15

15.1 OBJECTIVE:

You can make your Decision Trees simpler by pruning the nodes. one approach is to use Reduced Error Pruning -Explain this idea briefly. Try reduced error pruning for training your Decision Trees using cross -validation (you can do this in Weka) and report the Decision Tree you obtain? Also, report your accuracy using the pruned model. Does your accuracy increase?

15.2 RESOURCES:

Weka mining tool

15.3 THEORY:

Reduced-error pruning

- ❖ Each node of the (over-fit) tree is examined for pruning
- ❖ A node is pruned (removed) only if the resulting pruned tree

performs no worse than the original over the validation set

- ➢ Pruning a node consists of

• Removing the sub-tree rooted at the pruned node

• Making the pruned node a leaf node

• Assigning the pruned node the most common classification of the training instances attached to that node

- ➢ Pruning nodes iteratively

• Always select a node whose removal most increases the DT accuracy over the validation set

• Stop when further pruning decreases the DT accuracy over the validation set

IF (Children=yes) Λ (income=>30000)

THEN (car=Yes)

15.4 PROCEDURE:

1) Given the Bank database for mining.
2) Use the Weka GUI Chooser.
3) Select EXPLORER present in Applications.
4) Select Preprocess Tab.
5) Go to OPEN file and browse the file that is already stored in the system “bank.csv”.
6) Select some of the attributes from attributes list
7) Go to Classify tab.
8) Choose Classifier “Tree”
9) Select “NBTree” i.e., Navie Bayesian tree.
10) Select Test options “Use training set”
11) Right click on the text box besides choose button, select show properties
12) Now change unprone mode “false” to “true”.
13) Change the reduced error pruning % as needed.

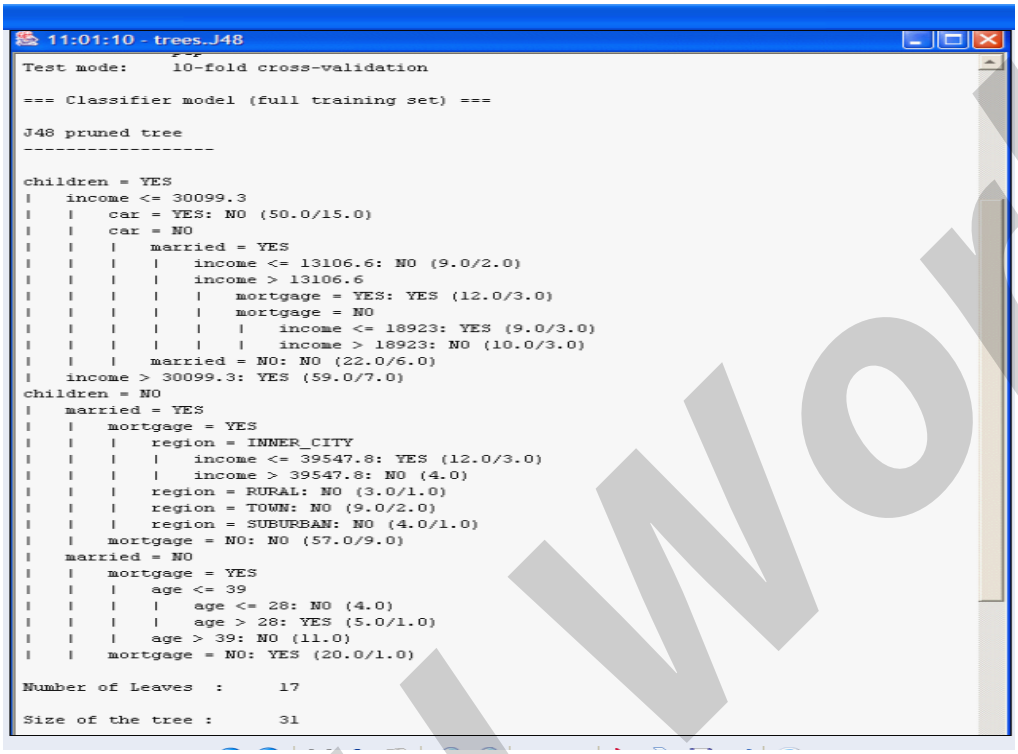
14) If need select attribute.

15) Now start weka.

16) Now we can see the output details in the Classifier output.

17) Right click on the result list and select ” visualize tree “option.

15.5 OUTPUT:



EXPERIMENT-16

16.1 OBJECTIVE:

(Extra Credit): How can you convert a Decision Trees into "if –then -else rules". Make up your own small Decision Tree consisting of 2 - 3 levels and convert it into a set of rules. There also exist different classifiers that output the model in the form of rules -one such classifier in Weka is rules. PART, train this model and report the set of rules obtained. Sometimes just one attribute can be good enough in making the decision, yes, just one! Can you predict what attribute that might be in this dataset? OneR classifier uses a single attribute to make decisions (it chooses the attribute based on minimum error). Report the rule obtained by training a one R classifier. Rank the performance of j48, PART and OneR.

16.2 RESOURCES:

Weka mining tool.

16.3 PROCEDURE:

1) Given the Bank database for mining.
2) Use the Weka GUI Chooser.
3) Select EXPLORER present in Applications.
4) Select Preprocess Tab.
5) Go to OPEN file and browse the file that is already stored in the system “bank.csv”.
6) Select some of the attributes from attributes list
7) Go to Classify tab.
8) Choose Classifier “Trees Rules”
9) Select “J48”.
10) Select Test options “Use training set”
11) If need select attribute.
12) Now start weka.
13) Now we can see the output details in the Classifier output.
14) Right click on the result list and select ” visualize tree “option .

(or)

- java weka.classifiers.trees.J48 -t c:\temp\bank.arff

Procedure for “OneR”:

1) Given the Bank database for mining.
2) Use the Weka GUI Chooser.
3) Select EXPLORER present in Applications.
4) Select Preprocess Tab.
5) Go to OPEN file and browse the file that is already stored in the system “bank.csv”.
6) Select some of the attributes from attributes list
7) Go to Classify tab.
8) Choose Classifier “Rules”
9) Select “OneR”.
10) Select Test options “Use training set”

11) if need select attribute.
12) Now start weka.
13) Now we can see the output details in the Classifier output.

Procedure for "PART":

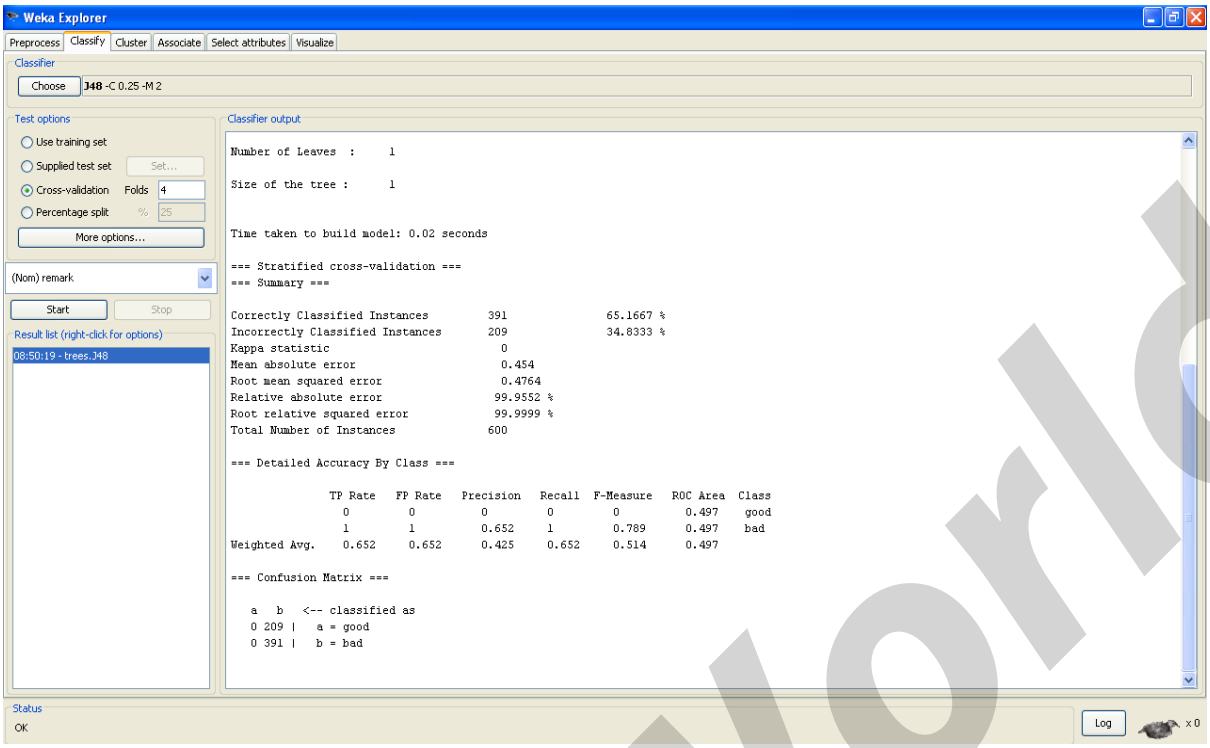
1) Given the Bank database for mining.
2) Use the Weka GUI Chooser.
3) Select EXPLORER present in Applications.
4) Select Preprocess Tab.
5) Go to OPEN file and browse the file that is already stored in the system "bank.csv".
6) Select some of the attributes from attributes list
7) Go to Classify tab.
8) Choose Classifier "Rules".
9) Select "PART".
10) Select Test options "Use training set"
11) if need select attribute.
12) Now start weka.
13) Now we can see the output details in the Classifier output.

Attribute relevance with respect to the class – relevant attribute (*science*)

IF accounting=1 THEN class=A (Error=0, Coverage = 7 instance)

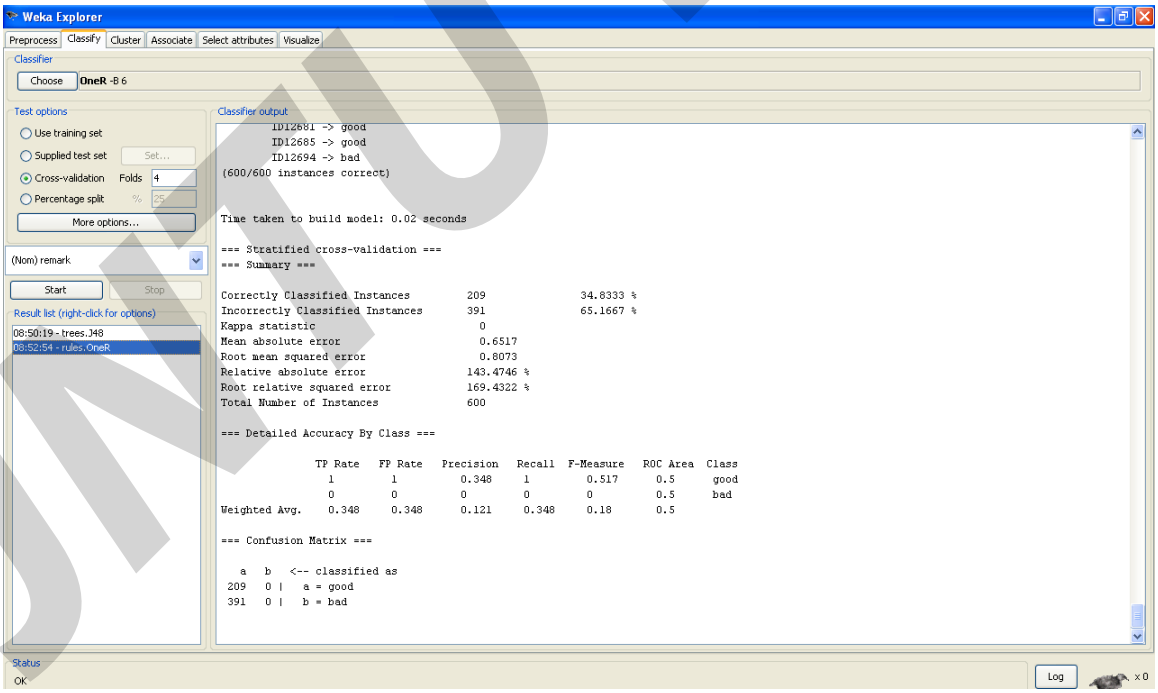
IF accounting=0 THEN class=B (Error=4/13, Coverage = 13 instances)

12.4 OUTPUT:

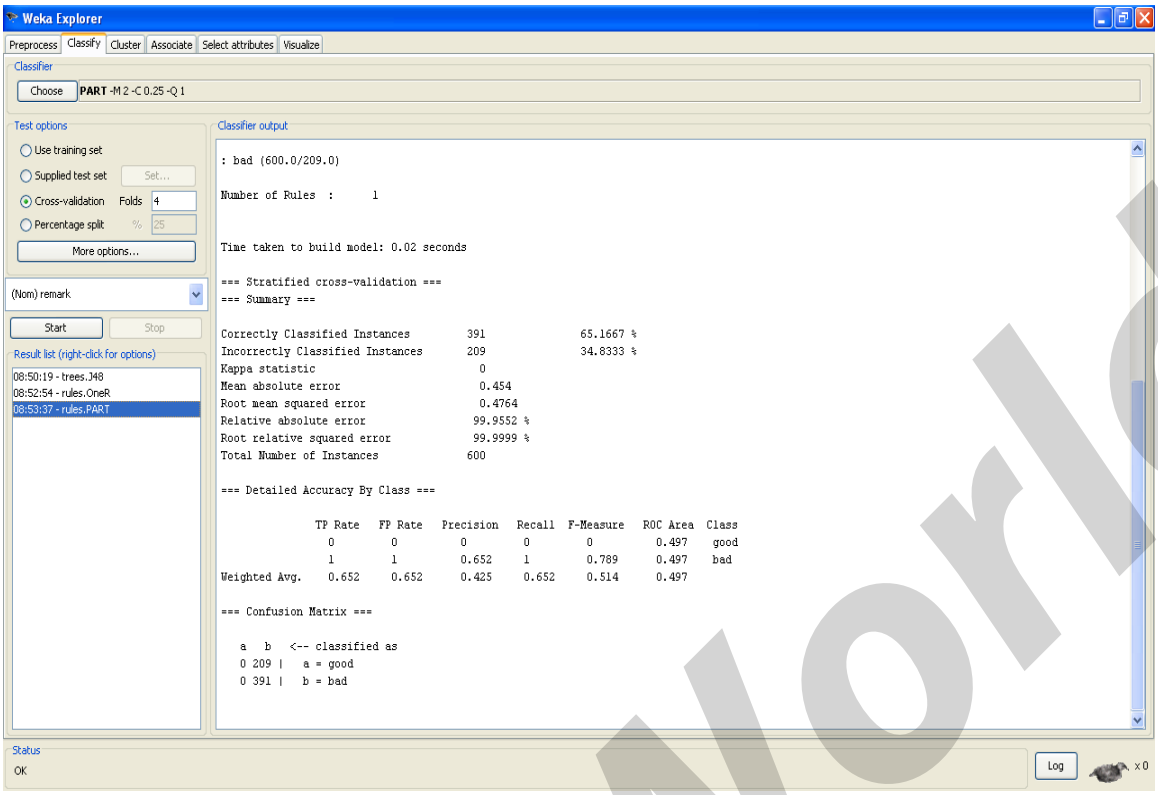


J48

One R



PART



EXPERIMENT-17

17.1 OBJECTIVE:

*Run J48 and Naïve Bayes classifiers on the following datasets and determine the accuracy:

1.vehicle.arff
2.kr-vs-kp.arff
3.glass.arff
4.wave-form-5000.arff

On which datasets does the Naïve Bayes perform better?

17.2 RESOURCES:

Weka mining tool.

17.3 PROCEDURE:

1) Given the Bank database for mining.
2) Use the Weka GUI Chooser.
3) Select EXPLORER present in Applications.
4) Select Preprocess Tab.
5) Go to OPEN file and browse the file that is already stored in the system "bank.csv".
6) Select some of the attributes from attributes list
7) Go to Classify tab.
8) Choose Classifier "Trees Rules"
9) Select "J48".
10) Select Test options "Use training set"
11) If need select attribute.
12) Now start weka.
13) Now we can see the output details in the Classifier output.
14) Right click on the result list and select " visualize tree "option .

(or)

- java weka.classifiers.trees.J48 -t c:\temp\bank.arff

Procedure for "OneR":

1) Given the Bank database for mining.
2) Use the Weka GUI Chooser.
3) Select EXPLORER present in Applications.
4) Select Preprocess Tab.
5) Go to OPEN file and browse the file that is already stored in the system "bank.csv".
6) Select some of the attributes from attributes list
7) Go to Classify tab.
8) Choose Classifier "Rules"
9) Select "OneR".
10) Select Test options "Use training set"
11) if need select attribute.

12) Now start weka.

13) Now we can see the output details in the Classifier output.

Procedure for "PART":

1) Given the Bank database for mining.

2) Use the Weka GUI Chooser.

3) Select EXPLORER present in Applications.

4) Select Preprocess Tab.

5) Go to OPEN file and browse the file that is already stored in the system "bank.csv".

6) Select some of the attributes from attributes list

7) Go to Classify tab.

8) Choose Classifier "Rules".

9) Select "PART".

10) Select Test options "Use training set"

11) if need select attribute.

12) Now start weka.

13) Now we can see the output details in the Classifier output.

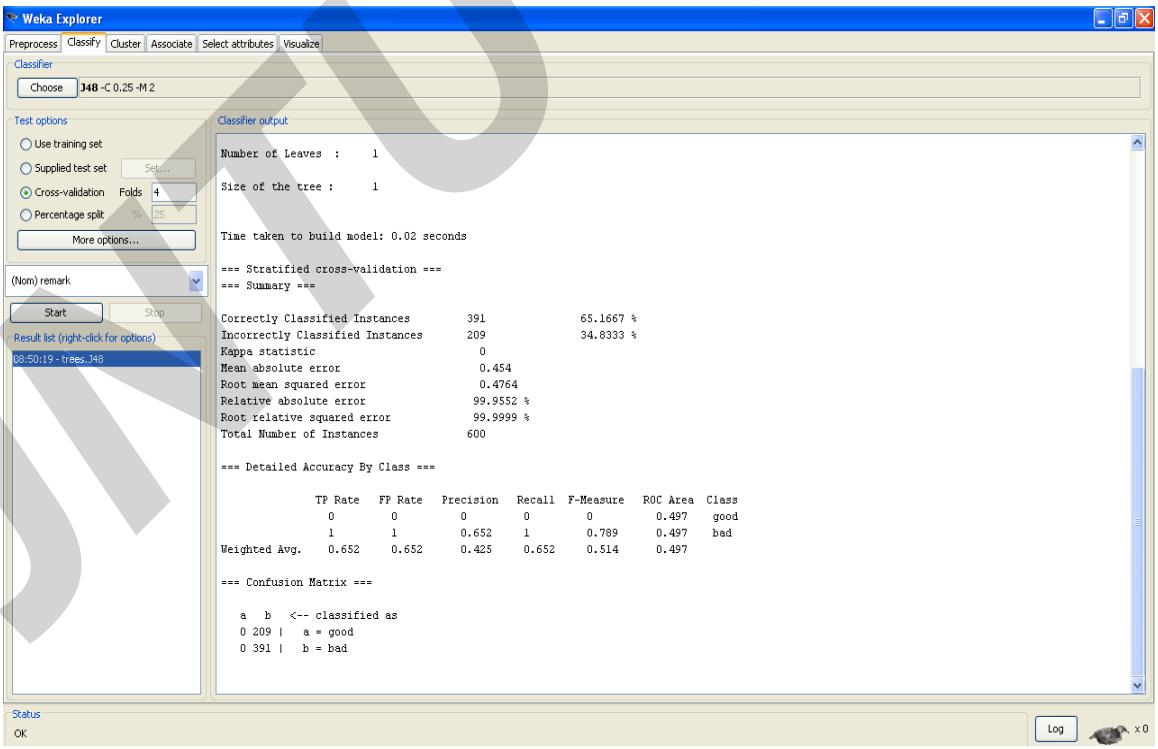
Attribute relevance with respect to the class – relevant attribute (*science*)

IF accounting=1 THEN class=A (Error=0, Coverage = 7 instance)

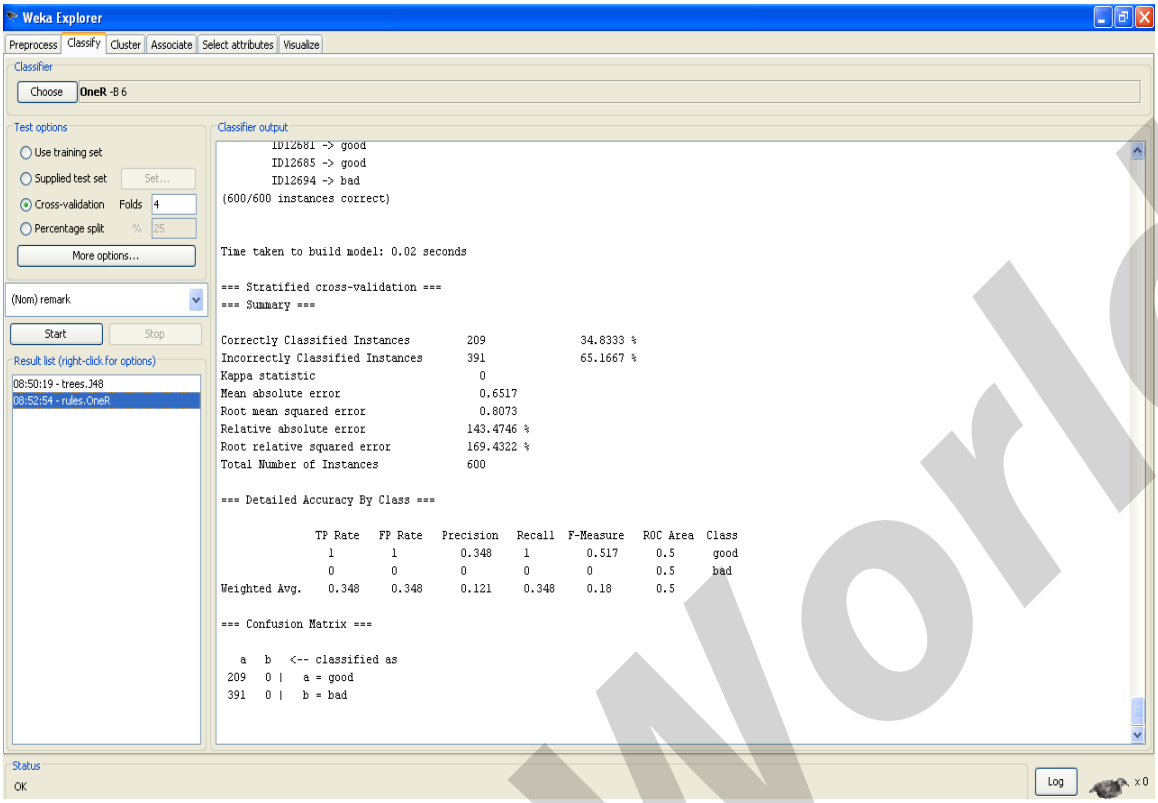
IF accounting=0 THEN class=B (Error=4/13, Coverage = 13 instances)

17.4 OUTPUT:

J48



One R



PART