NAWAB SHAH ALAM KHAN COLEEG OF ENGINEERING & TECHNOLOGY

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

LABORATORY IMPROVEMENT FOR FUTURE TRENDS (LIFT) - - -I

NAME OF THE LABORATORIES

S.No.	YEAR-SEM	NAME OF THE LAB
1	IV B.TECH-ISEM	DATA MINING

A Guide for execution of Lab Courses

VISION OF THE INSTITUTE:

To be a leading institute of world class quality technical education with strong ethical values, preparing students for leadership in their fields for the dynamic and global careers, developing breakthrough environment for professional education and research.

MISSION OF THE INSTITUTE:

M1: To enable the students to develop into outstanding professionals with high ethical standards capable of creating ,developing and managing local and global engineering enterprises

M2: To ensure quality assurance by fulfilling expectations of the society and industry with state of the art technology.

M3: To attract and retain knowledgeable, creative , motivated, and highly skilled individuals whose leadership and contributions uphold the college tenets of education through student–centric learning methodologies.

M4: To provide opportunities for deserving students of all communities.

M5: To promote all round personality development of the students through interactions with alumni and academia.

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

VISION:

To produce quality IT professionals, with an ability to adapt to ever changing IT needs of local, national and international arena, through effective teaching & learning, interactions with alumni and industry

MISSION:

- ➤ M1: To provide a holistic learning environment for students through ethical practices.

- ➤ M2: To provide quality infrastructure through practical exposure to the latest technology requirements.

- ➤ M3: To train the students in soft skills to excel in placements and competitive exams at higher level the industry ready.

- ➤ M4: To have a healthy Industry - Institute interaction through faculty development programs, student internships, guest lectures and using latest teaching learning methodologies.

- ➤ M5: To provide effective platform to meet the industrial requirement and provide research-oriented environment for the faculty to meet the continuous societal needs.

PROGRAM SPECIFIC OUTCOMES (PSO's)

- ➤ Develop efficient information management systems using latest development tools catering to the globally changing requirements in multi-disciplinary domains

- ➤ Manage real time IT projects with consideration of human, financial, ethical and environmental factors and an understanding of policy implications.

1. **AIM OF THE LIFT**:

  The main aim of the LIFT programme is to innovate, modify the existing facilities in labs, to create awareness among the students and develop Industry –Institution interactions and reach the standards in laboratories

2. **FUNCTIONS OF THE LIFT**:

  I. To create better understanding concepts of LIFT and other lab related activities among the staff and lab technicians for better improvement.

  II. To Arrange LIFT Presentations from each department about the lab activities by the staff handling the labs. (Lab Planners)

  III. To Prepare GAP ANALYSIS: This involves collection of requirements from each lab of every department, information about expansion of labs, repairs and maintenance of labs etc.

  IV. To arrange Industrial Visits/ Industrial training programs in coordination with concerned lab staff and Heads of the departments.

  V. A Report on Shadow Engineering: This involves arrangement of Industrial and Practical learning, Submission of Industrial Visit report, Technical Survey reports and Market Survey of a product for development in laboratories.

  VI. Verification of all the laboratories in every department by the LIFT Team along with the Principal and the concerned HODs, to check whether the activities are going according to LIFT guidelines, to check the Record Keeping, Lab Manuals and Viva sessions etc.

  VII. Check for LEAD Experiments and its follow up.

  VIII. Submission of proposals related to R&D, Project and Consultancy from lab staff to the Principal for further approvals.

LAB IMPROVEMENT FOR FUTURE TRENDS PROGRAMME (LIFT)

INDEX

CONTENTS:

1. OBJECTIVES AND RELEVANCE
2. SCOPE
3. PREREQUISITES
4. SYLLABUS AS PER JNTUH
5. LAB SCHEDULE
6. SUGGESTED BOOKS
7. WEBSITES (USEFUL LINKS)

1. OBJECTIVES AND RELEVANCE:

The main objective of the LIFT concept in lab course is to provide practical hands on experience for each student by providing them with good exposure to different experiments and to uplift the knowledge levels of the student, with different applications in various fields.

2. SCOPE:

The main scope of the LIFT lab course is to cover all the experiments as per the schedule given in the prescribed week wise periods. With this, a student can better understand the concepts and operating systems so that he could get better knowledge about each lab.

3. PREREQUISITES:

The basic level idea related to each experiment should be provided to the students before conducting main lab course. Following details are to be explained related to experiment:

1. Introduction to experiment – 30 min
2. The Operating of the equipment/instrument/software
3. Record of Experimental Results.
4. Sample Calculations / Executable Programs

4. SYLLABUS AS PER JNTUH:

The lab course should be planned as per the JNTUH syllabus. In this, LEAD experiments should also be included in the cycle of experiments.

5. (A) LAB SCHEDULE:

The lab schedule should be planned once in a week. The week wise scheduled experiment should be completed.

Batches	week-1	week-2	week-3	week-4	week-5	week-6
B1	Demo	Exp.1	Exp.2	Exp.10	Exp.9	Exp.8
B2	Demo	Exp.2	Exp.10	Exp.9	Exp.8	Exp.1
B3	Demo	Exp.10	Exp.9	Exp.8	Exp.1	Exp.2
B4	Demo	Exp.9	Exp.8	Exp.1	Exp.2	Exp.10
B5	Demo	Exp.8	Exp.1	Exp.2	Exp.10	Exp.9

(B) Scheme of Evaluation:

The scheme of evaluation for internal and external exams as follows:

LAB INTERNAL:

Day to Day Evalution-15					Internal Exam-10		
Uniform	Observation & Record	Performance Of the Experiment	Result	Viva	Write up	Execution & Results	Viva
Marks-3	Marks-3	Marks-3	Marks-3	Marks-3	Marks-4	Marks-3	Marks-3
Total Marks-25							

LAB EXTERNAL:

S.NO	Write up	Final Evaluation	Viva
1	1. **Aim** 2. **Procedure** 3. **Program** 4. **Expected output.**	**Based on correctness of the program and Results**	**Based on understanding of Experiment and theoretical questions in the related subjects**
Marks	20	20	10
Total Marks-50			

6. SUGGESTED BOOKS:

The suggested books should be recommended to the students as per the JNTUH syllabus prescribed.

7. WEBSITES (USEFUL LINKS):

The useful links should be provided to the students, where they can get an easy access to the knowledge of the experiment.

DATA MINING AND WARE HOUSING LAB

CONTENTS:

1. OBJECTIVES AND RELEVANCE

The main objective of the lab course is to gain practical hands on experience of data mining techniques using weka tool.

2. SCOPE

To make use of the weka tool for mining Credit Risk Assessment of an applicant in the German Credit Data set. Assessing the credit worthiness of an applicant is of crucial importance for banks. Using Weka data mining tool we have to develop a system to help a loan officer decide whether the credit of a customer is good or bad.

3. PREREQUISITES

Basic RDMS knowledge is required

PREAMBLE
This lab covers the experiments in DWDM subject. The JNTU has given 8 experiments in the syllabus.

4. SYLLABUS-JNTU

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 banks 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 a number of ways.

2. Books: Find some training manuals for loan officers or perhaps a suitable textbook on finance. Translate this knowledge from text 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 to and 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 data set consisting of 1000 actual cases collected in Germany. 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!) There are 20 attributes used in judging a loan applicant (ie, 7 Numerical attributes and 13 Categorical or Nominal attributes). The goal is to classify the applicant into one of two categories, Good or Bad.

Subtasks:

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

2. 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.

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

4. 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 we cannot get 100% training accuracy?

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

6. 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 accuracy increase/decrease? Why?

7. Check to see if the data shows a bias against "foreign workers" or "personal-status". One way to do this is to remove these attributes from the data set and see if the decision tree created in those cases is significantly different from the full data set case which you have already done. Did removing these attributes have any significant effect? Discuss.

8. Another question might be, do you really need to input so many attributes to get good results? May be only a few would do. For example, you could try just having attributes 2,3,5,7,10,17 and 21.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.)

9. Sometimes, The cost of rejecting an applicant who actually has good credit might be higher than accepting an applicant who has bad credit. Instead of counting the misclassification equally in both cases, give a higher cost to the first case (say cost 5) and lower cost to the second case. By using a cost matrix in weak. Train your decision tree and report the Decision Tree and cross validation results. Are they significantly different from results obtained in problem 6.

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

11. 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 and report the Decision Trees you obtain. Also Report your accuracy using the pruned model. Does your Accuracy increase?

12. How can you convert a Decision Tree into "if-then-else rules"? Make up your own small Decision Tree consisting 2-3 levels and convert 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 might be in this data set? 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, oneR.

The students will be given a demo on how to use the weka tool for solving the above tasks. The necessary manual and tutorial will be given.

5(A) LAB SCHEDULE:

CYCLE 1:

Batches	week-1	week-2	week-3	week-4	week-5	week-6	week-7	week-8
B1(501 to 512)	Demo & Basic commands	Exp.1,2&3	Exp.4 ,5&6	Exp.6,7&8	Exp.9&10	Exp.11 & 12	test	Exp.13&14
B2(513 to 524)	Demo & Basic commands	Exp.1&2	Exp.3 &4	Exp.5 &6	Exp.7 &8	Exp.9 & 10	test	Exp.13&14
B3(525 to 536)	Demo & Basic commands	Exp.1&2	Exp.3 &4	Exp.5 &6	Exp.7 &8	Exp.9 & 10	test	Exp.13&14
B4(537 to 548)	Demo & Basic commands	Exp.1&2	Exp.3 &4	Exp.5 &6	Exp.7 &8	Exp.9 & 10	test	Exp.13&14
B5(549 to 560)	Demo & Basic commands	Exp.1&2	Exp.3 &4	Exp.5 &6	Exp.7 &8	Exp.9 & 10	test	Exp.13&14

B) VIVA SCHEDULE: The viva schedule should be planned prior to the lab experiment.

CYCLE 1:   **ROUND -** 1

Batches	week-1	week-2	week-3	week-4	week-5	week-6	week-7	week-8	week-9	week-10	week-11	week-12	week-13	week-14
B1,B2,B3	viva					viva					viva			
B4,B5,B1		viva					viva					viva		
B2,B3,B4			viva					viva					viva	
B5,B1,B2				viva					viva					viva
B2,B3,B4					viva					viva				

(C) Scheme of Evaluation:

The scheme of evaluation for internal and external exams as follows:

LAB INTERNAL:

Day to Day Evalution-15					Internal Exam-10		
Uniform	Observation & Record	Performance Of the Experiment	Result	Viva	Write up	Execution & Results	Viva
Marks-3	Marks-3	Marks-3	Marks-3	Marks-3	Marks-4	Marks-3	Marks-3
Total Marks-25							

LAB EXTERNAL:

S.NO	Write up	Final Evaluation	Viva
1	1. **Aim** 2. **Procedure** 3. **Program** 4. **Expected output.**	**Based on correctness of the program and Results**	**Based on understanding of Experiment and theoretical questions in the related subjects**
Marks	20	20	10
Total Marks-50			

6. SUGGESTED BOOKS

Data mining : *practical* machine learning tools and techniques Ian H. Witten

7. WEB SITES

1. Weka Reference, http://www.gnu.org/copyleft/gpl.html

2. Prefuse Visualization Toolkit. See http:// prefuse.org/ for more information on the project

3. WekaWiki – http://weka.wikispaces.com/

4. Extensions for Weka's main GUI on WekaWiki –

5. http://weka.wikispaces.com/Extensions+for+Weka%27s+main+GUI

6. Adding tabs in the Explorer on WekaWiki –

   http://weka.wikispaces.com/Adding+tabs+in+the+Explorer

7. Explorer visualization plugins on WekaWiki –

   http://weka.wikispaces.com/Explorer+visualization+plugins

8. http://home.etf.rs/~vm/os/dmsw/Morgan.Kaufman.Publishers.Weka.2nd.Edition.2005.   Elsevier.pdf