



JAWAHARLAL NEHRU TECHNOLOGICAL UNIVERSITY HYDERABAD  
(Established by Andhra Pradesh Act No.30 of 2008)  
Kukatpally, Hyderabad – 500 085, TELANGANA (India)

ADD-ON COURSE STRUCTURE FOR III AND IV YEARS EFFECTIVE AFTER OCTOBER 2015  
B. TECH. COMPUTER SCIENCE AND ENGINEERING

**Note:**

1) Qualification pack for ASSOCIATE ANALYTICS includes the following three subjects.

1. Introduction to Analytics
2. Big Data Analytics
3. Predictive Analytics

2) Qualification pack for SECURITY ANALYST includes the following three subjects.

1. Information Security Management,
2. Information Security Assessments & Audits.
3. Information Security Incident Response & Management

**IV YEAR I SEMESTER**

Code	Subject	L	T/P/D	C
	Linux Programming	4	-	4
	Design Patterns	4	-	4
	Data Warehousing and Data Mining	4	-	4
	Cloud Computing	4	-	4
	ELECTIVE – I Software Project Management Image processing and Pattern Recognition Mobile Computing Computer Graphics Operations Research Big Data Analytics (Associate Analytics-2) * Information Security Assessments & Audits (Security Analyst -2) *	4	-	4
	ELECTIVE – II Machine Learning Soft Computing Information Retrieval Systems Artificial Intelligence Computer Forensics	4	-	4
	Linux Programming Lab	-	3	2
	Data Warehousing and Mining Lab	-	3	2
	Total	24	6	28

\* For registering for Big Data Analytics the prerequisite is Introduction to Analytics.

\* For registering for Information Security Assessments & Audits the prerequisite is Information Security Management.

**(A70517) LINUX PROGRAMMING****Objectives:**

- To understand and make effective use of Linux utilities and Shell scripting language (bash) to solve Problems.
- To implement in C some standard Linux utilities such as ls,mv,cp etc.using system calls.
- To develop the skills necessary for systems programming including file system programming, process and signal management, and interprocess communication.
- To develop the basic skills required to write network programs using Sockets.

**UNIT- I**

Linux Utilities-File handling utilities, Security by file permissions, Process utilities, Disk utilities, Networking commands, Filters, Text processing utilities and Backup utilities.

Sed-Scripts, Operation, Addresses, Commands, Applications, awk-Execution, Fields and Records, Scripts, Operation, Patterns, Actions, Associative Arrays, String and Mathematical functions, System commands in awk, Applications.

Shell programming with Bourne again shell(bash)- Introduction, shell responsibilities, pipes and Redirection, here documents, running a shell script, the shell as a programming language, shell meta characters, file name substitution, shell variables, command substitution, shell commands, the environment, quoting, test command, control structures, arithmetic in shell, shell script examples, interrupt processing, functions, debugging shell scripts.

**UNIT- II**

Files and Directories- File Concept, File types, File System Structure, file metadata-Inodes, kernel support for files, system calls for file I/O operations-open, creat, read, write, close, lseek, dup2, file status information-stat family, file and record locking- fcntl function, file permissions - chmod, fchmod, file ownership-chown, lchown, fchown, links-soft links and hard links – symlink, link, unlink.

Directories-Creating, removing and changing Directories-mkdir, rmdir, chdir, obtaining current working directory-getwd, Directory contents, Scanning Directories-opendir, readdir, closedir, rewaddir functions.

**UNIT- III**

Process – Process concept, Layout of a C program image in main

memory, Process environment-environment list, environment variables, getenv, setenv, Kernel support for process, process identification, process control - process creation, replacing a process image, waiting for a process, process termination, zombie process, orphan process, system call interface for process management-fork, vfork, exit, wait, waitpid, exec family, Process Groups, Sessions and Controlling Terminal, Differences between threads and processes.

Signals – Introduction to signals, Signal generation and handling, Kernel support for signals, Signal function, unreliable signals, reliable signals, kill, raise, alarm, pause, abort, sleep functions.

**UNIT- IV**

Interprocess Communication - Introduction to IPC, IPC between processes on a single computer system, IPC between processes on different systems, pipes-creation, IPC between related processes using unnamed pipes, FIFOs-creation, IPC between unrelated processes using FIFOs(Named pipes), differences between unnamed and named pipes, popen and pclose library functions.

Message Queues- Kernel support for messages, APIs for message queues, client/server example.

Semaphores-Kernel support for semaphores, APIs for semaphores, file locking with semaphores.

**UNIT- V**

Shared Memory- Kernel support for shared memory, APIs for shared memory, shared memory example.

Sockets- Introduction to Berkeley Sockets, IPC over a network, Client-Server model, Socket address structures (Unix domain and Internet domain), Socket system calls for connection oriented protocol and connectionless protocol, example-client/server programs-Single Server-Client connection, Multiple simultaneous clients, Socket options-setsockopt andfcntl system calls, Comparison of IPC mechanisms.

**TEXT BOOKS:**

1. Unix System Programming using C++, T.Chan, PHI.
2. Unix Concepts and Applications, 4th Edition, Sumitabha Das, TMH.
3. Unix Network Programming , W.R.Stevens, PHI.

**REFERENCE BOOKS:**

1. Beginning Linux Programming, 4<sup>th</sup> Edition, N.Matthew, R.Stones, Wrox, Wiley India Edition.
2. Unix for programmers and users, 3<sup>rd</sup> Edition, Graham Glass, King Ables, Pearson.

- System Programming with C and Unix, A.Hoover, Pearson.
- Unix System Programming, Communication, Concurrency and Threads, K.A.Robbins and S.Robbins, Pearson Education.
- Unix shell Programming, S.G.Kochan and P.Wood,3<sup>rd</sup> edition, Pearson Education.
- Shell Scripting, S.Parker, Wiley India Pvt. Ltd.
- Advanced Programming in the Unix Environment,2<sup>nd</sup> edition, W.R.Stevens and S.A.Rago, Pearson Education.
- Unix and Shell programming, B.A.Forouzan and R.F.Gilberg, Cengage Learning.
- Linux System Programming, Robert Love, O'Reilly, SPD.
- C Programming Language, Kernighan and Ritchie, PHI

#### Outcomes:

- Work confidently in Linux environment.
- Work with shell script to automate different tasks as Linux administration.

JAWAHARLAL NEHRU TECHNOLOGICAL UNIVERSITY HYDERABAD

IV Year B.Tech. CSE-I Sem

L	T/P/D	C
4	-/-	4

(A70530) DESIGN PATTERNS

#### Objectives:

- Understand the design patterns that are common in software applications.
- Understand how these patterns are related to Object Oriented design.

#### UNIT-I

Introduction: What Is a Design Pattern?, Design Patterns in Smalltalk MVC, Describing Design Patterns, The Catalog of Design Patterns, Organizing the Catalog, How Design Patterns Solve Design Problems, How to Select a Design Pattern, How to Use a Design Pattern.

#### UNIT-II

A Case Study: Designing a Document Editor: Design Problems, Document Structure, Formatting, Embellishing the User Interface, Supporting Multiple Look-and-Feel Standards, Supporting Multiple Window Systems, User Operations Spelling Checking and Hyphenation, Summary.

Creational Patterns: Abstract Factory, Builder, Factory Method, Prototype, Singleton, Discussion of Creational Patterns.

#### UNIT-III

Structural Pattern Part-I: Adapter, Bridge, Composite.

Structural Pattern Part-II: Decorator, acade, Flyweight, Proxy.

#### UNIT-IV

Behavioral Patterns Part-I: Chain of Responsibility, Command, Interpreter, Iterator.

Behavioral Patterns Part-II: Mediator, Memento, Observer.

#### UNIT-V

Behavioral Patterns Part-II (cont'd): State, Strategy, Template Method ,Visitor, Discussion of Behavioral Patterns.

What to Expect from Design Patterns, A Brief History, The Pattern Community An Invitation, A Parting Thought.

#### TEXT BOOK:

- Design Patterns By Erich Gamma, Pearson Education

#### REFERENCE BOOKS :

- Pattern's in JAVA Vol-I By Mark Grand, Wiley DreamTech.

2. Pattern's in JAVA Vol-II By Mark Grand, Wiley DreamTech.
3. JAVA Enterprise Design Patterns Vol-III By Mark Grand, Wiley DreamTech.
4. Head First Design Patterns By Eric Freeman-Oreilly-spd.
5. Peeling Design Patterns, Prof. Meda Srinivasa Rao, Narsimha Karumanchi, CareerMonk Publications.
6. Design Patterns Explained By Alan Shalloway, Pearson Education.
7. Pattern Oriented Software Architecture, F.Buschmann&others, John Wiley & Sons.

#### Outcomes:

- Ability to understand and apply common design patterns to incremental / iterative development.
- Ability to identify appropriate patterns for design of given problem.

## JAWAHARLAL NEHRU TECHNOLOGICAL UNIVERSITY HYDERABAD

### IV Year B.Tech. CSE-I Sem

L	T/P/D	C
4	--	4

### (A70520) DATA WAREHOUSING AND DATA MINING

#### Objectives:

Study data warehouse principles and its working learn data mining concepts understand association rules mining. Discuss classification algorithms learn how data is grouped using clustering techniques.

#### UNIT-I

**Data warehouse:** Introduction to Data warehouse, Difference between operational database systems and data warehouses, Data warehouse Characteristics, Data warehouse Architecture and its Components, Extraction-Transformation-Loading, Logical(Multi-Dimensional), Data Modeling, Schema Design, Star and Snow-Flake Schema, Fact Consultation, Fact Table, Fully Addictive, Semi-Addictive, Non Addictive Measures; Fact-Less-Facts, Dimension Table Characteristics; OLAP Cube, OLAP Operations, OLAP Server Architecture-ROLAP, MOLAP and HOLAP.

#### UNIT-II

**Introduction to Data Mining:** Introduction, What is Data Mining, Definition, KDD, Challenges, Data Mining Tasks, Data Preprocessing, Data Cleaning, Missing data, Dimensionality Reduction, Feature Subset Selection, Discretization and Binaryzation, Data Transformation; Measures of Similarity and Dissimilarity- Basics.

#### UNIT-III

**Association Rules:** Problem Definition, Frequent Item Set Generation, The APRIORI Principle, Support and Confidence Measures, Association Rule Generation; APRIORI Algorithm, The Partition Algorithms, FP-Growth Algorithms, Compact Representation of Frequent Item Set- Maximal Frequent Item Set, Closed Frequent Item Set.

#### UNIT-IV

**Classification:** Problem Definition, General Approaches to solving a classification problem , Evaluation of Classifiers , Classification techniques, Decision Trees-Decision tree Construction , Methods for Expressing attribute test conditions, Measures for Selecting the Best Split, Algorithm for Decision tree Induction ; Naive-Bayes Classifier, Bayesian Belief Networks; K- Nearest neighbor classification-Algorithm and Characteristics.

#### UNIT-V

**Clustering:** Problem Definition, Clustering Overview, Evaluation of Clustering Algorithms, Partitioning Clustering-K-Means Algorithm, K-Means Additional

issues, PAM Algorithm; Hierarchical Clustering-Agglomerative Methods and divisive methods, Basic Agglomerative Hierarchical Clustering Algorithm, Specific techniques, Key Issues in Hierarchical Clustering, Strengths and Weakness; Outlier Detection.

#### TEXT BOOKS:

- 1) Data Mining- Concepts and Techniques- Jiawei Han, Micheline Kamber, Morgan Kaufmann Publishers, Elsevier, 2 Edition, 2006.
- 2) Introduction to Data Mining, Pang-Ning Tan, Vipin Kumar, Michael Steinbach, Pearson Education.

#### REFERENCE BOOKS:

- 1) Data Mining Techniques, Arun K Pujari, 3<sup>rd</sup> Edition, Universities Press.
- 2) Data Warehousing Fundamentals, Pualraj Ponnaiah, Wiley Student Edition.
- 3) The Data Warehouse Life Cycle Toolkit – Ralph Kimball, Wiley Student Edition.
- 4) Data Mining, Vikaram Pudi, P Radha Krishna, Oxford University Press

#### Outcomes:

- Student should be able to understand why the data warehouse in addition to database systems.
- Ability to perform the preprocessing of data and apply mining techniques on it.
- Ability to identify the association rules, classification and clusters in large data sets.
- Ability to solve real world problems in business and scientific information using data mining

## JAWAHARLAL NEHRU TECHNOLOGICAL UNIVERSITY HYDERABAD

### IV Year B.Tech. CSE-I Sem

L	T/P/D	C
4	-/-	4

### (A70519) CLOUD COMPUTING

#### Objectives:

- To explain the evolving computer model called cloud computing.
- To introduce the various levels of services that can be achieved by cloud.
- To describe the security aspects in cloud.

#### UNIT- I

**Systems Modeling, Clustering and Virtualization:** Distributed System Models and Enabling Technologies, Computer Clusters for Scalable Parallel Computing, Virtual Machines and Virtualization of Clusters and Data centers.

#### UNIT- II

**Foundations:** Introduction to Cloud Computing, Migrating into a Cloud, Enriching the 'Integration as a Service' Paradigm for the Cloud Era, The Enterprise Cloud Computing Paradigm.

#### UNIT- III

**Infrastructure as a Service (IAAS) & Platform and Software as a Service (PAAS / SAAS):** Virtual machines provisioning and Migration services, On the Management of Virtual machines for Cloud Infrastructures, Enhancing Cloud Computing Environments using a cluster as a Service, Secure Distributed Data Storage in Cloud Computing.

Aneka, Comet Cloud, T-Systems', Workflow Engine for Clouds, Understanding Scientific Applications for Cloud Environments.

#### UNIT- IV

**Monitoring, Management and Applications:** An Architecture for Federated Cloud Computing, SLA Management in Cloud Computing, Performance Prediction for HPC on Clouds, Best Practices in Architecting Cloud Applications in the AWS cloud, Building Content Delivery networks using Clouds, Resource Cloud Mashups.

#### UNIT- V

**Governance and Case Studies:** Organizational Readiness and Change management in the Cloud age, Data Security in the Cloud, Legal Issues in Cloud computing, Achieving Production Readiness for Cloud Services

#### TEXT BOOKS:

1. Cloud Computing: Principles and Paradigms by Rajkumar Buyya, James Broberg and Andrzej M. Goscinski, Wiley, 2011.

2. Distributed and Cloud Computing, Kai Hwang, Geoffery C.Fox, Jack J.Dongarra, Elsevier, 2012.

#### REFERENCE BOOKS:

1. Cloud Computing : A Practical Approach, Anthony T.Velte, Toby J.Velte, Robert Elsenpeter, Tata McGraw Hill, rp2011.
2. Enterprise Cloud Computing, Gautam Shroff, Cambridge University Press, 2010.
3. Cloud Computing: Implementation, Management and Security, John W. Rittinghouse, James F.Ransome, CRC Press, rp2012.
4. Cloud Application Architectures: Building Applications and Infrastructure in the Cloud, George Reese, O'Reilly, SPD, rp2011.
5. Cloud Security and Privacy: An Enterprise Perspective on Risks and Compliance, Tim Mather, Subra Kumaraswamy, Shahed Latif, O'Reilly, SPD, rp2011.

#### Outcomes:

- Ability to understand the virtualization and cloud computing concepts.

## JAWAHARLAL NEHRU TECHNOLOGICAL UNIVERSITY HYDERABAD

### IV Year B.Tech. CSE-I Sem

L	T/P/D	C
4	--	4

### (A70540) SOFTWARE PROJECT MANAGEMENT

(Elective- I)

#### Objectives:

The main goal of software development projects is to create a software system with a predetermined functionality and quality in a given time frame and with given costs. For achieving this goal, models are required for determining target values and for continuously controlling these values. This course focuses on principles, techniques, methods & tools for model-based management of software projects, assurance of product quality and process adherence (quality assurance), as well as experience-based creation & improvement of models (process management). The goals of the course can be characterized as follows:

1. Understanding the specific roles within a software organization as related to project and process management
2. Understanding the basic infrastructure competences (e.g., process modeling and measurement)
3. Understanding the basic steps of project planning, project management, quality assurance, and process management and their relationships

#### UNIT- I

Conventional Software Management: The waterfall model, conventional software

Management performance. Evolution of Software Economics: Software Economics, pragmatic software cost estimation.

#### UNIT- II

Improving Software Economics: Reducing Software product size, improving software processes, improving team effectiveness, improving automation, Achieving required quality, peer inspections.

The old way and the new: The principles of conventional software engineering, principles of modern software management, transitioning to an iterative process.

#### UNIT- III

Life cycle phases: Engineering and production stages, inception, Elaboration, construction, transition phases.

Artifacts of the process: The artifact sets, Management artifacts, Engineering

artifacts, programmatic artifacts. Model based software architectures: A Management perspective and technical perspective.

#### UNIT- IV

Work Flows of the process: Software process workflows, Inter trans workflows. Checkpoints of the Process: Major Mile Stones, Minor Milestones, Periodic status assessments. Iterative Process Planning: Work breakdown structures, planning guidelines, cost and schedule estimating, Interaction planning process, Pragmatic planning.

Project Organizations and Responsibilities: Line-of-Business Organizations, Project Organizations, evolution of Organizations.

Process Automation : Automation Building Blocks, The Project Environment.

#### UNIT- V

Project Control and Process instrumentation: The server care Metrics, Management indicators, quality indicators, life cycle expectations pragmatic Software Metrics, Metrics automation. Tailoring the Process: Process discriminants, Example.

Future Software Project Management: Modern Project Profiles Next generation

Software economics, modern Process transitions.

Case Study : The Command Center Processing and Display System-Replacement(CCPDS-R).

#### TEXT BOOKS:

1. Software Project Management, Walker Royce, Pearson Education.
2. Software Project Management, Bob Hughes & Mike Cotterell, fourth edition, Tata McGraw Hill.

#### REFERENCE BOOKS:

1. Applied Software Project Management, Andrew Stellman & Jennifer Greene, O'Reilly, 2006
2. Head First PMP, Jennifer Greene & Andrew Stellman, O'Reilly, 2007
3. Software Engineering Project Management, Richard H. Thayer & Edward Yourdon, second edition, Wiley India, 2004.
4. Agile Project Management, Jim Highsmith, Pearson education, 2004
5. The art of Project management, Scott Berkun, O'Reilly, 2005.
6. Software Project Management in Practice, Pankaj Jalote, Pearson Education, 2002.

#### Outcomes:

- Describe and determine the purpose and importance of project management from the perspectives of planning, tracking and completion of project.
- Compare and differentiate organization structures and project structures.
- Implement a project to manage project schedule, expenses and resources with the application of suitable project management tools.

## JAWAHARLAL NEHRU TECHNOLOGICAL UNIVERSITY HYDERABAD

IV Year B.Tech. CSE -I Sem

L T/P/D C  
4 -/- 4

### BIG DATA ANALYTICS (ASSOCIATE ANALYTICS – II) (Elective-I)

#### Unit I:

##### Data Management (NOS 2101):

Design Data Architecture and manage the data for analysis, understand various sources of Data like Sensors/signal/GPS etc. Data Management, Data Quality (noise, outliers, missing values, duplicate data) and Data Preprocessing.

Export all the data onto Cloud ex. AWS/Rackspace etc.

##### Maintain Healthy, Safe & Secure Working Environment (NOS 9003):

Introduction, workplace safety, Report Accidents & Emergencies, Protect health & safety as your work, course conclusion, assessment

#### Unit II

##### Big Data Tools (NOS 2101):

Introduction to Big Data tools like Hadoop, Spark, Impala etc., Data ETL process, Identify gaps in the data and follow-up for decision making.

##### Provide Data/Information in Standard Formats (NOS 9004):

Introduction, Knowledge Management, Standardized reporting & compliances, Decision Models, course conclusion. Assessment.

#### Unit III

##### Big Data Analytics:

Run descriptives to understand the nature of the available data, collate all the data sources to suffice business requirement, Run descriptive statistics for all the variables and observe the data ranges, Outlier detection and elimination.

#### Unit IV

##### Machine Learning Algorithms (NOS 9003):

Hypothesis testing and determining the multiple analytical methodologies, Train Model on 2/3 sample data using various Statistical/Machine learning algorithms, Test model on 1/3 sample for prediction etc.

#### Unit V

##### (NOS 9004)

##### Data Visualization (NOS 2101):

Prepare the data for Visualization, Use tools like Tableau, QlickView and D3, Draw insights out of Visualization tool.

Product Implementation

#### TEXT BOOK:

1. Student's Handbook for Associate Analytics.

#### REFERENCE BOOKS:

1. Introduction to Data Mining, Tan, Steinbach and Kumar, Addison Wesley, 2006
2. Data Mining Analysis and Concepts, M. Zaki and W. Meira (the authors have kindly made an online version available): <http://www.dataminingbook.info/uploads/book.pdf>
3. Mining of Massive Datasets Jure Leskovec Stanford Univ. Anand RajaramanMilliway Labs Jeffrey D. Ullman Stanford Univ.
4. ([http://www.vistrails.org/index.php/Course:\\_Big\\_Data\\_Analysis](http://www.vistrails.org/index.php/Course:_Big_Data_Analysis))

JAWAHARLAL NEHRU TECHNOLOGICAL UNIVERSITY HYDERABAD

IV Year B.Tech. CSE -I Sem

L            T/P/D            C  
4            -/-            4

INFORMATION SECURITY ASSESSMENTS & AUDITS ( SECURITY ANALYST – II)  
(Elective-I)

**Unit I**

**Information Security Performance Metrics and Audit:**

Security Metrics and Reporting, Common Issues and Variances of Performance Metrics, Introduction to Security Audit, Servers and Storage devices, Infrastructure and Networks, Communication Routes, Information Security Methodologies (Black-box, White-box, Grey-box), Phases of Information Security Audit and Strategies, Ethics of an Information Security Auditor etc.

Maintain Healthy, Safe & Secure Working environment (NOS 9003).

**Unit II**

**Information Security Audit Tasks, Reports and Post Auditing Actions:**

Pre-audit checklist, Information Gathering, Vulnerability Analysis, External Security Audit, Internal Network Security Audit, Firewall Security Audit, IDS Security Auditing, Social Engineering Security Auditing, Web Application Security Auditing, Information Security Audit Deliverables & Writing Report, Result Analysis, Post Auditing Actions, Report Retention etc.

Provide Data/Information in Standard formats (NOS 9004).

**Unit III**

**Vulnerability Management:**

Information Security Vulnerabilities – Threats and Vulnerabilities, Human-based Social Engineering, Computer-based Social Engineering, Social Media Countermeasures, Vulnerability Management – Vulnerability Scanning, Testing, Threat management, Remediation etc.

**Unit IV**

**Information Security Assessments:**

Vulnerability Assessment, Classification, Types of Vulnerability Assessment, Vulnerability Assessment Phases, Vulnerability Analysis Stages, Characteristics of a Good Vulnerability Assessment Solutions & Considerations, Vulnerability Assessment Reports – Tools and choosing a right Tool, Information Security Risk Assessment, Risk Treatment, Residual Risk, Risk Acceptance, Risk Management Feedback Loops etc.

**Unit V**

**Configuration Reviews:**

Introduction to Configuration Management, Configuration Management Requirements-Plan-Control, Development of configuration Control Policies, Testing Configuration Management etc.

**TEXT BOOKS:**

Prescribed books:-

1. Assessing Information Security (strategies, tactics, logic and framework) by A Vladimirov, K.Gavrilko, and A.Michajlowski
2. "The Art of Computer Virus Research and Defense by Peter Szor."

**REFERENCES:**

1. <https://www.sans.org/reading-room/whitepapers/threats/implementing-vulnerability-management-process-34180>
2. <http://csrc.nist.gov/publications/nistpubs/800-40-Ver2/SP800-40v2.pdf>

**(A70532) IMAGE PROCESSING AND PATTERN RECOGNITION**

(Elective - I)

**Objectives:**

- Adequate background knowledge about image processing and pattern recognition
- Practical knowledge and skills about image processing and pattern recognition tools
- Necessary knowledge to design and implement a prototype of an image processing and pattern recognition *application*.

**UNIT – I**

Fundamental steps of image processing, components of an image processing system. The image model and image acquisition, sampling and quantization, relationship between pixels, distance functions, scanner.

Statistical and spatial operations, Intensity functions transformations, histogram processing, smoothing & sharpening – spatial filters Frequency domain filters, homomorphic filtering, image filtering & restoration. Inverse and weiner filtering, FIR weiner filter, Filtering using image transforms, smoothing splines and interpolation.

**UNIT – II**

Morphological and other area operations, basic morphological operations, opening and closing operations, dilation erosion, Hit or Miss transform, morphological algorithms, extension to grey scale images.

Segmentation and Edge detection region operations, basic edge detection, second order detection, crack edge detection, gradient operators, compass and Laplace operators, edge linking and boundary detection, thresholding, region based segmentation, segmentation by morphological watersheds.

**UNIT – III**

Image compression: Types and requirements, statistical compression, spatial compression, contour coding, quantizing compression, image data compression-predictive technique, pixel coding, transfer coding theory, lossy and lossless predictive type coding, Digital Image Water marking.

**UNIT – IV**

Representation and Description: Chain codes, Polygonal approximation, Signature Boundary Segments, Skeltons, Boundary Descriptors, Regional Descriptors, Relational Descriptors, Principal components for Description,

**Relational Descriptors****UNIT- V**

**Pattern Recognition Fundamentals:** Basic Concepts of pattern recognition, Fundamental problems in pattern recognition system, design concepts and methodologies, example of automatic pattern recognition systems, a simple automatic pattern recognition model.

**Pattern classification:** Pattern classification by distance function: Measures of similarity, Clustering criteria, K-means algorithm, Pattern classification by likelihood function: Pattern classification as a Statistical decision problem, Bayes classifier for normal patterns.

**TEXT BOOKS**

- Digital Image Processing Third edition, Pearson Education, Rafael C. Gonzalez, Richard E. Woods.
- Pattern recognition Principles: Julius T. Tou, and Rafael C. Gonzalez, Addison-Wesly Publishing Company.
- Digital Image Processing, M.Anji Reddy, Y.Hari Shankar, BS Publications.

**REFERENCE BOOKS:**

- Image Processing, Analysis and Machine Vision, Second Edition, Milan Sonka, Vaclav Hlavac and Roger Boyle. Thomson learning
- Digital Image Processing – William k. Pratl –John Wiley edition.
- Fundamentals of digital image processing – by A.K. Jain, PHI.
- Pattern classification, Richard Duda, Hart and David strok John Wiley publishers.
- Digital Image Processing, S.Jayaraman,S. Esakkirajan, T.Veerakumar, TMH.
- Pattern Recognition, R.Shinghal, Oxford University Press.

**Outcomes:**

- Ability to apply computer algorithms to practical problems.
- Ability to image segmentation, reconstruction and restoration.
- Ability to perform the classification of patterns

**(A70536) MOBILE COMPUTING****(Elective – I)****Objectives:**

- To make the student understand the concept of mobile computing paradigm, its novel applications and limitations.
- To understand the typical mobile networking infrastructure through a popular GSM protocol
- To understand the issues and solutions of various layers of mobile networks, namely MAC layer, Network Layer & Transport Layer
- To understand the database issues in mobile environments & data delivery models.
- To understand the ad hoc networks and related concepts.
- To understand the platforms and protocols used in mobile environment.

**UNIT- I**

**Introduction:** Mobile Communications, Mobile Computing – Paradigm, Promises/Novel Applications and Impediments and Architecture; Mobile and Handheld Devices, Limitations of Mobile and Handheld Devices.

**GSM – Services, System Architecture, Radio Interfaces, Protocols, Localization, Calling, Handover, Security, New Data Services, GPRS, CSHSD, DECT.**

**UNIT – II**

**(Wireless) Medium Access Control (MAC):** Motivation for a specialized MAC (Hidden and exposed terminals, Near and far terminals), SDMA, FDMA, TDMA, CDMA, Wireless LAN/(IEEE 802.11)

**Mobile Network Layer:** IP and Mobile IP Network Layers, Packet Delivery and Handover Management, Location Management, Registration, Tunneling and Encapsulation, Route Optimization, DHCP.

**UNIT – III**

**Mobile Transport Layer:** Conventional TCP/IP Protocols, Indirect TCP, Snooping TCP, Mobile TCP, Other Transport Layer Protocols for Mobile Networks.

**Database Issues:** Database Hoarding & Caching Techniques, Client-Server Computing & Adaptation, Transactional Models, Query processing, Data Recovery Process & QoS Issues.

**UNIT- IV**

**Data Dissemination and Synchronization:** Communications Asymmetry, Classification of Data Delivery Mechanisms, Data Dissemination, Broadcast Models, Selective Tuning and Indexing Methods, Data Synchronization – Introduction, Software, and Protocols

**UNIT- V**

**Mobile Ad hoc Networks (MANETs):** Introduction, Applications & Challenges of a MANET, Routing, Classification of Routing Algorithms, Algorithms such as DSR, AODV, DSDV, etc. , Mobile Agents, Service Discovery.

**Protocols and Platforms for Mobile Computing :**WAP, Bluetooth, XML, J2ME, JavaCard, PalmOS, Windows CE, SymbianOS, Linux for Mobile Devices, Android.

**TEXT BOOKS:**

1. Jochen Schiller, "Mobile Communications", Addison-Wesley, Second Edition, 2009.
2. Raj Kamal, "Mobile Computing", Oxford University Press, 2007, ISBN: 0195686772

**REFERENCE BOOKS:**

1. Jochen Schiller, "Mobile Communications", Addison-Wesley, Second Edition, 2004.
2. Stojmenovic and Cacute, "Handbook of Wireless Networks and Mobile Computing", Wiley, 2002, ISBN 0471419028.
3. Reza Behravanfar, "Mobile Computing Principles: Designing and Developing Mobile Applications with UML and XML", ISBN: 0521817331, Cambridge University Press, Oct 2004,

**Outcomes:**

- Able to think and develop new mobile application.
- Able to take any new technical issue related to this new paradigm and come up with a solution(s).
- Able to develop new ad hoc network applications and/or algorithms/ protocols.
- Able to understand & develop any existing or new protocol related to mobile environment

## (A70529) COMPUTER GRAPHICS

(Elective- I)

**Objectives:**

- To make students understand about fundamentals of Graphics to enable them to design animated scenes for virtual object creations.
- To make the student present the content graphically.

**UNIT- I**

**Introduction:** Application areas of Computer Graphics, overview of graphics systems, video-display devices, raster-scan systems, random scan systems, graphics monitors and work stations and input devices

**Output primitives:** Points and lines, line drawing algorithms, mid-point circle and ellipse algorithms. Filled area primitives: Scan line polygon fill algorithm, boundary-fill and flood-fill algorithms.

**UNIT- II**

**2-D Geometrical transforms:** Translation, scaling, rotation, reflection and shear transformations, matrix representations and homogeneous coordinates, composite transforms, transformations between coordinate systems.

**2-D Viewing:** The viewing pipeline, viewing coordinate reference frame, window to view-port coordinate transformation, viewing functions, Cohen-Sutherland and Cyrus-beck line clipping algorithms, Sutherland –Hodgeman polygon clipping algorithm.

**UNIT- III**

**3-D Object representation:** Polygon surfaces, quadric surfaces, spline representation, Hermite curve, Bezier curve and B-spline curves, Bezier and B-spline surfaces, sweep representations, octrees BSP Trees,

**3-D Geometric transformations:** Translation, rotation, scaling, reflection and shear transformations, composite transformations, 3-D viewing: Viewing pipeline, viewing coordinates, view volume and general projection transforms and clipping.

**UNIT- IV**

**Visible surface detection methods:** Classification, back-face detection, depth-buffer, scan-line, depth sorting, BSP-tree methods, area sub-division and octree methods

**Illumination Models and Surface rendering Methods:** Basic illumination

models, polygon rendering methods

**UNIT- V**

**Computer animation:** Design of animation sequence, general computer animation functions, raster animation, computer animation languages, key frame systems, motion specifications

**TEXT BOOKS:**

- "Computer Graphics C version", Donald Hearn and M. Pauline Baker, Pearson education.
- "Computer Graphics Second edition", Zhigand xiang, Roy Plastock, Schaum's outlines, Tata Mc Graw hill edition.

**REFERENCE BOOKS:**

- "Computer Graphics Principles & practice", second edition in C, Foley, VanDam, Feiner and Hughes, Pearson Education.
- "Procedural elements for Computer Graphics", David F Rogers, Tata Mc Graw hill, 2nd edition.
- "Principles of Interactive Computer Graphics", Neuman and Sproul, TMH.
- "Principles of Computer Graphics", Shalini, Govil-Pai, Springer.
- "Computer Graphics", Steven Harrington, TMH
- Computer Graphics, F.S.Hill, S.M.Kelley, PHI.
- Computer Graphics, P.Shirley, Steve Marschner & Others, Cengage Learning.
- Computer Graphics & Animation, M.C.Trivedi, Jaico Publishing House.
- An Integrated Introduction to Computer Graphics and Geometric Modelling, R.Goldman, CRC Press, Taylor&Francis Group.
- Computer Graphics, Rajesh K.Maurya, Wiley India.

**Outcomes:**

- Students can animate scenes entertainment.
- Will be able work in computer aided design for content presentation..
- Better analogy data with pictorial representation.

**(A70352) OPERATIONS RESEARCH**

(Elective-I)

**Objectives:**

- To introduce the methods of Operations Research.
- Emphasize the mathematical procedures of non linear programming search techniques.
- Introduce advanced topics such as Probabilistic models and dynamic programming.

**UNIT – I**

Development – Definition – Characteristics and Phases – Types of models – Operations Research models – applications.

**Allocation:** Linear Programming Problem Formulation – Graphical solution – Simplex method – Artificial variables techniques: Two-phase method, Big-M method.

**UNIT – II**

**Transportation Problem** – Formulation – Optimal solution, unbalanced transportation problem – Degeneracy.

**Assignment problem** – Formulation – Optimal solution - Variants of Assignment Problem- Traveling Salesman problem.

**UNIT – III**

**Sequencing** – Introduction – Flow –Shop sequencing – n jobs through two machines – n jobs through three machines – Job shop sequencing – two jobs through 'm' machines

**Replacement:** Introduction – Replacement of items that deteriorate with time – when money value is not counted and counted – Replacement of items that fail completely- Group Replacement.2

**UNIT – IV**

**Theory of Games:** Introduction –Terminology– Solution of games with saddle points and without saddle points- 2 x 2 games – dominance principle – m x 2 & 2 x n games -graphical method.

**Inventory:** Introduction – Single item, Deterministic models – Purchase inventory models with one price break and multiple price breaks –Stochastic models – demand may be discrete variable or continuous variable – Single Period model and no setup cost.

**UNIT – V**

**Waiting Lines:** Introduction –Terminology-Single Channel – Poisson arrivals and Exponential Service times – with infinite population and finite population models– Multichannel – Poisson arrivals and exponential service times with infinite population.

**Dynamic Programming:** Introduction – Terminology- Bellman's Principle of Optimality – Applications of dynamic programming- shortest path problem – linear programming problem.

**Simulation:** Introduction, Definition, types of simulation models, Steps involved in the simulation process- Advantages and disadvantages- applications of simulation to queuing and inventory.

**TEXT BOOK :**

- Operations Research /J.K.Sharma 4e. /MacMilan
- Introduction to O.R/Hillier & Libermann/TMH

**REFERENCE BOOKS :**

- Introduction to O.R /Taha/PHI
- Operations Research/ NVS Raju/ SMS Education/3<sup>rd</sup> Revised Edition
- Operations Research /A.M.Natarajan, P.Balasubramaniam, A. Tamilarasi/Pearson Education.
- Operations Research / Wagner/ PHI Publications.
- Operations Research/M.V. Durga Prasad, K. Vijaya Kumar Reddy, J. Suresh Kumar/ Cengage Learning.

**(A70534) MACHINE LEARNING**

**(Elective – II)**

**Objectives:**

- To be able to formulate machine learning problems corresponding to different applications.
- To understand a range of machine learning algorithms along with their strengths and weaknesses.
- To understand the basic theory underlying machine learning.

**UNIT – I**

**Introduction:** An illustrative learning task, and a few approaches to it. What is known from algorithms? Theory, Experiment. Biology. Psychology.

**Concept Learning:** Version spaces. Inductive Bias. Active queries. Mistake bound/PAC model. basic results. Overview of issues regarding data sources, success criteria.

**UNIT – II**

**Decision Tree Learning:** - Minimum Description Length Principle. Occam's razor. Learning with active queries

**Neural Network Learning:** Perceptions and gradient descent back propagation.

**UNIT – III**

**Sample Complexity and Over fitting:** Errors in estimating means. Cross Validation and jackknifing VC dimension. Irrelevant features: Multiplicative rules for weight tuning.

**Bayesian Approaches:** The basics Expectation Maximization. Hidden Markov Models

**UNIT – IV**

**Instance-based Techniques:** Lazy vs. eager generalization. K nearest neighbor, case- based reasoning.

**UNIT – V**

**Genetic Algorithms:** Different search methods for induction - Explanation-based Learning: using prior knowledge to reduce sample complexity.

**TEXT BOOKS:**

1. Tom Michel, Machine Learning, McGraw Hill, 1997

2. Trevor Has tie, Robert Tibshirani & Jerome Friedman. The Elements of Statically Learning, Springer Verlag, 2001

**REFERENCE BOOKS:**

1. Machine Learning Methods in the Environmental Sciences, Neural Networks, William W Hsieh, Cambridge Univ Press.
2. Richard o. Duda, Peter E. Hart and David G. Stork, pattern classification, John Wiley & Sons Inc.,2001
3. Chris Bishop, Neural Networks for Pattern Recognition, Oxford University Press, 1995

**Outcomes:**

- Student should be able to understand the basic concepts such as decision trees and neural networks.
- Ability to formulate machine learning techniques to respective problems.
- Apply machine learning algorithms to solve problems of moderate complexity

**(A70539) SOFT COMPUTING****(Elective – II)****Objectives:**

- To give students knowledge of soft computing theories fundamentals, i.e. Fundamentals of artificial and neural networks, fuzzy sets and fuzzy logic and genetic algorithms.

**UNIT-I**

**AI Problems and Search:** AI problems, Techniques, Problem Spaces and Search, Heuristic Search Techniques- Generate and Test, Hill Climbing, Best First Search Problem reduction, Constraint Satisfaction and Means End Analysis. Approaches to Knowledge Representation- Using Predicate Logic and Rules.

**UNIT-II**

**Artificial Neural Networks:** Introduction, Basic models of ANN, important terminologies, Supervised Learning Networks, Perceptron Networks, Adaptive Linear Neuron, Backpropagation Network. Associative Memory Networks. Training Algorithms for pattern association, BAM and Hopfield Networks.

**UNIT-III**

**Unsupervised Learning Network-** Introduction, Fixed Weight Competitive Nets, Maxnet, Hamming Network, Kohonen Self-Organizing Feature Maps, Learning Vector Quantization, Counter Propagation Networks, Adaptive Resonance Theory Networks. Special Networks-Introduction to various networks.

**UNIT-IV**

**Introduction to Classical Sets ( crisp Sets)and Fuzzy Sets-** operations and Fuzzy sets. Classical Relations -and Fuzzy Relations- Cardinality, Operations, Properties and composition. Tolerance and equivalence relations.

**Membership functions- Features, Fuzzification, membership value assignments, Defuzzification.**

**UNIT-V**

**Fuzzy Arithmetic and Fuzzy Measures, Fuzzy Rule Base and Approximate Reasoning Fuzzy Decision making**

**Fuzzy Logic Control Systems. Genetic Algorithm- Introduction and basic operators and terminology. Applications: Optimization of TSP, Internet Search Technique**

**TEXT BOOKS:**

- Principles of Soft Computing- S N Sivanandam, S N Deepa, Wiley India, 2007.
- Soft Computing and Intelligent System Design -Fakhreddine O Karray, Clarence D Silva, Pearson Edition, 2004.

**REFERENCE BOOKS:**

- Artificial Intelligence and Soft Computing- Behavioural and Cognitive Modelling of the Human Brain- Amit Konar, CRC press, Taylor and Francis Group.
- Artificial Intelligence – Elaine Rich and Kevin Knight, TMH, 1991, rp2008.
- Artificial Intelligence – Patric Henry Winston – Third Edition, Pearson Education.
- A first course in Fuzzy Logic-Hung T Nguyen and Elbert A Walker, CRC Press Taylor and Francis Group.
- Artificial Intelligence and Intelligent Systems, N.P.Padhy, Oxford Univ. Press.

**Outcomes:**

- Student can able to building intelligent systems through soft computing techniques.
- Student should be able to understand the concept of artificial neural networks, fuzzy arithmetic and fuzzy logic with their day to day applications.

**(A70533) INFORMATION RETRIEVAL SYSTEMS**  
**(Elective – II)**

**Objectives:**

- To learn the different models for information storage and retrieval
- To learn about the various retrieval utilities
- To understand indexing and querying in information retrieval systems
- To expose the students to the notions of structured and semi structured data
- To learn about web search

**UNIT-I**

## Introduction

Retrieval Strategies: Vector space model, Probabilistic retrieval strategies: Simple term weights, Non binary independence model Language Models.

**UNIT-II**

Retrieval Utilities: Relevance feedback, Clustering, N-grams, Regression analysis, Thesauri.

**UNIT-III**

Retrieval Utilities: Semantic networks, Parsing.

Cross-Language Information Retrieval: Introduction, Crossing the language barrier.

**UNIT-IV**

Efficiency: Inverted index, Query processing, Signature files, Duplicate document detection

**UNIT-V**

Integrating Structured Data and Text: A Historical progression, Information retrieval as a relational application, Semi-structured search using a relational schema.

Distributed Information Retrieval: A Theoretical model of distributed retrieval, Web search.

**TEXT BOOK:**

- David A. Grossman, Ophir Frieder, Information Retrieval – Algorithms and Heuristics, Springer, 2<sup>nd</sup> Edition (Distributed by Universities Press), 2004.

**REFERENCE BOOKS:**

- Gerald J Kowalski, Mark T Maybury, Information Storage and Retrieval Systems, Springer, 2000.
- Soumen Chakrabarti, Mining the Web : Discovering Knowledge from Hypertext Data, Morgan-Kaufmann Publishers, 2002.
- Christopher D. Manning, Prabhakar Raghavan, Hinrich Schütze, An Introduction to Information Retrieval, Cambridge University Press, Cambridge, England, 2009.

**Outcomes:**

- Possess the ability to store and retrieve textual documents using appropriate models.
- Possess the ability to use the various retrieval utilities for improving search.
- Possess an understanding of indexing and compressing documents to improve space and time efficiency.
- Possess the skill to formulate SQL like queries for unstructured data.
- Understand issues in web search.

L	T/P/D	C
4	-/-	4

**(A70526) ARTIFICIAL INTELLIGENCE  
(ELECTIVE- II)**

**Objectives:**

- To learn the difference between optimal reasoning vs human like reasoning
- To understand the notions of state space representation, exhaustive search, heuristic search along with the time and space complexities
- To learn different knowledge representation techniques
- To understand the applications of AI: namely Game Playing, Theorem Proving, Expert Systems, Machine Learning and Natural Language Processing

**UNIT-I**

Introduction, History, Intelligent Systems, Foundations of AI, Sub areas of AI, Applications.

Problem Solving - State-Space Search and Control Strategies: Introduction, General Problem Solving, Characteristics of Problem, Exhaustive Searches, Heuristic Search Techniques, Iterative-Deepening A\*, Constraint Satisfaction. Game Playing, Bounded Look-ahead Strategy and use of Evaluation Functions, Alpha-Beta Pruning

**UNIT-II**

Logic Concepts and Logic Programming: Introduction, Propositional Calculus, Propositional Logic, Natural Deduction System, Axiomatic System, Semantic Tableau System in Propositional Logic, Resolution Refutation in Propositional Logic, Predicate Logic, Logic Programming.

Knowledge Representation: Introduction, Approaches to Knowledge Representation, Knowledge Representation using Semantic Network, Extended Semantic Networks for KR, Knowledge Representation using Frames.

**UNIT-III**

Expert System and Applications: Introduction, Phases in Building Expert Systems, Expert System Architecture, Expert Systems Vs Traditional Systems, Truth Maintenance Systems, Application of Expert Systems, List of Shells and Tools.

Uncertainty Measure - Probability Theory: Introduction, Probability Theory, Bayesian Belief Networks, Certainty Factor Theory, Dempster-Shafer Theory.

**UNIT-IV**

Machine-Learning Paradigms: Introduction. Machine Learning Systems. Supervised and Unsupervised Learning. Inductive Learning. Learning Decision Trees (Text Book 2), Deductive Learning. Clustering, Support Vector Machines.

Artificial Neural Networks: Introduction, Artificial Neural Networks, Single-Layer Feed-Forward Networks, Multi-Layer Feed-Forward Networks, Radial-Basis Function Networks, Design Issues of Artificial Neural Networks, Recurrent Networks.

**UNIT-V**

Advanced Knowledge Representation Techniques: Case Grammars, Semantic Web

Natural Language Processing: Introduction, Sentence Analysis Phases, Grammars and Parsers, Types of Parsers, Semantic Analysis, Universal Networking Knowledge.

**TEXT BOOKS:**

1. Saroj Kaushik. Artificial Intelligence. Cengage Learning, 2011.
2. Russell, Norvig: Artificial intelligence, A Modern Approach, Pearson Education, Second Edition. 2004.

**REFERENCE BOOK:**

1. Rich, Knight, Nair: Artificial intelligence, Tata McGraw Hill, Third Edition 2009.

**Outcomes:**

- Possess the ability to formulate an efficient problem space for a problem expressed in English.
- Possess the ability to select a search algorithm for a problem and characterize its time and space complexities.
- Possess the skill for representing knowledge using the appropriate technique.
- Possess the ability to apply AI techniques to solve problems of Game Playing, Expert Systems, Machine Learning and Natural Language Processing.

**(A70528) COMPUTER FORENSICS****(Elective-II)****Objectives:**

- A brief explanation of the objective is to provide digital evidences which are obtained from digital media.
- In order to understand the objectives of computer forensics, first of all, people have to recognize the different roles computer plays in a certain crime.
- According to a snippet from the United States Security Service, the functions computer has in different kinds of crimes.

**UNIT – I**

**Computer Forensics Fundamentals:** What is Computer Forensics?, Use of Computer Forensics in Law Enforcement, Computer Forensics Assistance to Human Resources/Employment Proceedings, Computer Forensics Services, Benefits of Professional Forensics Methodology, Steps taken by Computer Forensics Specialists

**Types of Computer Forensics Technology:** Types of Military Computer Forensic Technology, Types of Law Enforcement – Computer Forensic Technology – Types of Business Computer Forensic Technology

**Computer Forensics Evidence and Capture:** Data Recovery Defined – Data Back-up and Recovery – The Role of Back-up in Data Recovery – The Data-Recovery Solution

**UNIT – II**

**Evidence Collection and Data Seizure:** Why Collect Evidence? Collection Options – Obstacles – Types of Evidence – The Rules of Evidence – Volatile Evidence – General Procedure – Collection and Archiving – Methods of Collection – Artifacts – Collection Steps – Controlling Contamination: The Chain of Custody

**Duplication and Preservation of Digital Evidence:** Preserving the Digital Crime Scene – Computer Evidence Processing Steps – Legal Aspects of Collecting and Preserving Computer Forensic Evidence

**Computer Image Verification and Authentication:** Special Needs of Evidential Authentication – Practical Consideration – Practical Implementation

**UNIT – III**

**Computer Forensics analysis and validation:** Determining what data to collect and analyze, validating forensic data, addressing data-hiding techniques, performing remote acquisitions

**Network Forensics:** Network forensics overview, performing live acquisitions, developing standard procedures for network forensics, using

network tools, examining the honeynet project.

**Processing Crime and Incident Scenes:** Identifying digital evidence, collecting evidence in private-sector incident scenes, processing law enforcement crime scenes, preparing for a search, securing a computer incident or crime scene, seizing digital evidence at the scene, storing digital evidence, obtaining a digital hash, reviewing a case

**UNIT – IV**

**Current Computer Forensic tools:** evaluating computer forensic tool needs, computer forensics software tools, computer forensics hardware tools, validating and testing forensics software

**E-Mail Investigations:** Exploring the role of e-mail in investigation, exploring the roles of the client and server in e-mail, investigating e-mail crimes and violations, understanding e-mail servers, using specialized e-mail forensic tools

**Cell phone and mobile device forensics:** Understanding mobile device forensics, understanding acquisition procedures for cell phones and mobile devices.

**UNIT – V**

**Working with Windows and DOS Systems:** understanding file systems, exploring Microsoft File Structures, Examining NTFS disks, Understanding whole disk encryption, windows registry, Microsoft startup tasks, MS-DOS startup tasks, virtual machines.

**TEXT BOOKS:**

1. Computer Forensics, Computer Crime Investigation by John R. Vacca, Firewall Media, New Delhi.
2. Computer Forensics and Investigations by Nelson, Phillips Enfinger, Steuart, CENGAGE Learning

**REFERENCE BOOKS:**

1. Real Digital Forensics by Keith J. Jones, Richard Bejtlich, Curtis W. Rose, Addison- Wesley Pearson Education
2. Forensic Compiling, A Practitioner's Guide by Tony Sammes and Brian Jenkinson, Springer International edition.
3. Computer Evidence Collection & Presentation by Christopher L.T. Brown, Firewall Media.
4. Homeland Security, Techniques & Technologies by Jesus Mena, Firewall Media.
5. Software Forensics Collecting Evidence from the Scene of a Digital Crime by Robert M. Slade, TMH 2005
6. Windows Forensics by Chad Steel, Wiley India Edition.

**Outcomes:**

- Students will understand the usage of computers in forensic, and how to use various forensic tools for a wide variety of investigations.
- It gives an opportunity to students to continue their zeal in research in computer forensics.

## (A70596) LINUX PROGRAMMING LAB

**Objectives:**

- To write shell scripts to solve problems.
- To implement some standard Linux utilities such as ls, cp etc using system calls.
- To develop network-based applications using C.

**List of sample problems:****Note: Use Bash for Shell scripts.**

1. 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.
2. Write a shell script that deletes all lines containing a specified word in one or more files supplied as arguments to it.
3. 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.
4. 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.
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.
6. Write a shell script to list all of the directory files in a directory.
7. Write a shell script to find factorial of a given integer.
8. Write an awk script to count the number of lines in a file that do not contain vowels.
9. Write an awk script to find the number of characters, words and lines in a file.
10. Write a C program that makes a copy of a file using standard I/O and system calls.
11. Implement in C the following Linux commands using System calls  
a). cat      b) mv
12. Write a C program to list files in a directory.
13. Write a C program to emulate the Unix ls -l command.
14. Write a C program to list for every file in a directory, its inode number and file name.
15. Write a C program that redirects standard output to a file.Ex: ls > f1.
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.

17. Write a C program to create a Zombie process.
18. Write a C program that illustrates how an orphan is created.
19. Write a C program that illustrates how to execute two commands concurrently with a command pipe. Ex:- ls -l | sort
20. Write C programs that illustrate communication between two unrelated processes using named pipe(FIFO File).
21. Write a C program in which a parent writes a message to a pipe and the child reads the message.
22. 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.
23. Write a C program (receiver.c) that receives the messages (from the above message queue as specified in (22)) and displays them.
24. Write a C program that illustrates suspending and resuming processes using signals.
25. Write Client and Server programs in C for connection oriented communication between Server and Client processes using Unix Domain sockets to perform the following:  
Client process sends a message to the Server Process. The Server receives the message, reverses it and sends it back to the Client. The Client will then display the message to the standard output device.
26. Write Client and Server programs in C for connection oriented communication between Server and Client processes using Internet Domain sockets to perform the following:  
Client process sends a message to the Server Process. The Server receives the message, reverses it and sends it back to the Client. The Client will then display the message to the standard output device.
27. Write C programs to perform the following:  
One process creates a shared memory segment and writes a message("Hello") into it. Another process opens the shared memory segment and reads the message(ie. "Hello"). It will then display the message("Hello") to standard output device.

**TEXT BOOKS:**

1. Beginning Linux Programming, 4<sup>th</sup> Edition, N. Matthew, R. Stones, Wrox, Wiley India Edition.
2. Advanced Unix Programming, N.B. Venkateswarulu, BS Publications.
3. Unix and Shell Programming, M.G. Venkatesh Murthy, Pearson Education.
4. Unix Shells by Example, 4th Edition, Ellie Quigley, Pearson Education.
5. Sed and Awk, O. Dougherty & A. Robbins, 2<sup>nd</sup> edition, SPD.

**Outcomes:**

- Ability to understand the Linux environment
- Ability to perform the file management and multiple tasks using shell scripts in Linux environment

## (A70595) DATA WAREHOUSING AND MINING LAB

**Objectives:**

Learn how to build a data warehouse and query it (using open source tools like Pentaho Data Integration and Pentaho Business Analytics), Learn to perform data mining tasks using a data mining toolkit (such as open source WEKA), Understand the data sets and data preprocessing, Demonstrate the working of algorithms for data mining tasks such association rule mining, classification, clustering and regression, Exercise the data mining techniques with varied input values for different parameters.

**UNIT-1. Build Data Warehouse and Explore WEKA**

- A. Build a Data Warehouse/Data Mart (using open source tools like Pentaho Data Integration tool, Pentoaho Business Analytics; or other data warehouse tools like Microsoft-SSIS, Informatica, Business Objects, etc.).

- Identify source tables and populate sample data
- Design multi-dimensional data models namely Star, snowflake and Fact constellation schemas for any one enterprise (ex. Banking, Insurance, Finance, Healthcare, Manufacturing, Automobile, etc.).
- Write ETL scripts and implement using data warehouse tools
- Perform various OLAP operations such slice, dice, roll up, drill up and pivot
- Explore visualization features of the tool for analysis like identifying trends etc.

## B. Explore WEKA Data Mining/Machine Learning Toolkit

- Downloading and/or installation of WEKA data mining toolkit,
- Understand the features of WEKA toolkit such as Explorer, Knowledge Flow interface, Experimenter, command-line interface.
- Navigate the options available in the WEKA (ex. Select attributes panel, Preprocess panel, Classify panel, Cluster panel, Associate panel and Visualize panel)
- Study the arff file format
- Explore the available data sets in WEKA.
- Load a data set (ex. Weather dataset, Iris dataset, etc.)
- Load each dataset and observe the following:
  - List the attribute names and their types

- Number of records in each dataset

- Identify the class attribute (if any)

- Plot Histogram

- Determine the number of records for each class.

- Visualize the data in various dimensions

**Unit 2 Perform data preprocessing tasks and Demonstrate performing association rule mining on data sets**

- Explore various options available in Weka for preprocessing data and apply (like Discretization Filters, Resample filter, etc.) on each dataset
- Load each dataset into Weka and run Apriori algorithm with different support and confidence values. Study the rules generated.
- Apply different discretization filters on numerical attributes and run the Apriori association rule algorithm. Study the rules generated. Derive interesting insights and observe the effect of discretization in the rule generation process.

**Unit 3 Demonstrate performing classification on data sets**

- Load each dataset into Weka and run Id3, J48 classification algorithm. Study the classifier output. Compute entropy values, Kappa statistic.
- Extract if-then rules from the decision tree generated by the classifier. Observe the confusion matrix and derive Accuracy, F-measure, TPR, FPR, Precision and Recall values. Apply cross-validation strategy with various fold levels and compare the accuracy results.
- Load each dataset into Weka and perform Naïve-bayes classification and k-Nearest Neighbour classification. Interpret the results obtained.
- Plot RoC Curves
- Compare classification results of ID3, J48, Naïve-Bayes and k-NN classifiers for each dataset, and deduce which classifier is performing best and poor for each dataset and justify.

**Unit 4 Demonstrate performing clustering on data sets**

- Load each dataset into Weka and run simple k-means clustering algorithm with different values of k (number of desired clusters). Study the clusters formed. Observe the sum of squared errors and centroids, and derive insights.
- Explore other clustering techniques available in Weka.
- Explore visualization features of Weka to visualize the clusters. Derive interesting insights and explain.

## **Unit 5 Demonstrate performing Regression on data sets**

- A. Load each dataset into Weka and build Linear Regression model. Study the clusters formed. Use Training set option. Interpret the regression model and derive patterns and conclusions from the regression results.
- B. Use options cross-validation and percentage split and repeat running the Linear Regression Model. Observe the results and derive meaningful results.
- C. Explore Simple linear regression technique that only looks at one variable.

### **Resource Sites:**

1. <http://www.pentaho.com/>
2. <http://www.cs.waikato.ac.nz/ml/weka/>

### **Outcomes:**

- o Ability to understand the various kinds of tools.
- o Demonstrate the classification, clusters and etc. in large data sets

## **DATA MINING LAB**

### **Objectives:**

- To obtain practical experience using data mining techniques on real world data sets.
- Emphasize hands-on experience working with all real data sets.

### **List of Sample Problems:**

#### **Task 1: 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 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, 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 dataset, consisting of 1000 actual cases collected in Germany. credit dataset (original) Excel spreadsheet version of the German credit data.

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)**

1. List all the categorical (or nominal) attributes and the real-valued attributes separately. (5 marks)
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. (5 marks)
3. 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. (10 marks)
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 you cannot get 100 % training accuracy? (10 marks)

5. Is testing on the training set as you did above a good idea? Why or Why not ? (10 marks)
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 your accuracy increase/decrease? Why? (10 marks)
7. 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. (10 marks)
8. 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.) (10 marks)
9. 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)? (10 marks)
10. 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? (10 marks)
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 (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? (10 marks)
12. (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. (10 marks)

#### Task Resources:

- Mentor lecture on Decision Trees
- Andrew Moore's Data Mining Tutorials (See tutorials on Decision Trees and Cross Validation)
- Decision Trees (Source: Tan, MSU)
- Tom Mitchell's book slides (See slides on Concept Learning and Decision Trees)
- Weka resources:
  - Introduction to Weka (html version) (download ppt version)
  - Download Weka
  - Weka Tutorial
  - ARFF format
  - Using Weka from command line

#### Task 2: Hospital Management System

Data Warehouse consists Dimension Table and Fact Table.

REMEMBER The following

Dimension

The dimension object (Dimension):

- \_ Name
- \_ Attributes (Levels) , with one primary key
- \_ Hierarchies

One time dimension is must.

About Levels and Hierarchies

Dimension objects (dimension) consist of a set of levels and a set of hierarchies defined over those levels. The levels represent levels of aggregation. Hierarchies describe parent-child relationships among a set of levels.

For example, a typical calendar dimension could contain five levels. Two

hierarchies can be defined on these levels:

H1: YearL > QuarterL > MonthL > WeekL > DayL

H2: YearL > WeekL > DayL

The hierarchies are described from parent to child, so that Year is the parent of Quarter, Quarter the parent of Month, and so forth.

#### About Unique Key Constraints

When you create a definition for a hierarchy, Warehouse Builder creates an identifier key for each level of the hierarchy and a unique key constraint on the lowest level (Base Level)

Design a Hospital Management system data warehouse (TARGET) consists of Dimensions Patient, Medicine, Supplier, Time. Where measures are 'NO UNITS', UNIT PRICE.

Assume the Relational database (SOURCE) table schemas as follows

TIME (day, month, year),

PATIENT (patient\_name, Age, Address, etc.,)

MEDICINE ( Medicine\_Brand\_name, Drug\_name, Supplier, no\_units, Uinit\_Price, etc.,)

SUPPLIER :( Supplier\_name, Medicine\_Brand\_name, Address, etc.,)

If each Dimension has 6 levels, decide the levels and hierarchies, Assume the level names suitably.

Design the Hospital Management system data warehouse using all schemas. Give the example 4-D cube with assumption names.

#### Outcomes:

- Ability to add mining algorithms as a component to the exiting tools
- Ability to apply mining techniques for realistic data.