Hindi Vidya Prachar Samiti's

Ramniranjan Jhunjhunwala College

of

Arts, Science & Commerce

(Autonomous College)

Affiliated to

UNIVERSITY OF MUMBAI

Syllabus for Semester I

Program: M.Sc. Information Technology

Program Code: RJSPIT

(CBCS 2018-19)

M.Sc. Information Technology Semester I Syllabus

SEMESTER I

Course Code	Unit	Topic Headings	Credits	L / Week
RJSPIT101	Paper Title: Data Mining			
	I	• Introduction • Data • Knowledge Representation	4	4
	II	• Data Preprocessing • Mining Frequent Patterns, Associations, and Correlations		4
	III	• Classification and Prediction • Support Vector Machines, Associative Classification		4
	IV	• Cluster Analysis		4
	V	• Graph Mining, Social Network Analysis, and Multirelational Data Mining • Mining Object, Spatial, Multimedia, Text, and Web Data		4

RJSPIT102	Paper Title: Distributed Systems			
	I	• Characterization of Distributed Systems • System Models	4	4
	II	• Networking And Internetworking • Interprocess Communication		4
	III	• Remote Invocation • Indirect Communication • Web Services		4
	IV	• Coordination And Agreement • Name Services • Time And Global States		4
	V	• Distributed Transactions • Replication • Mobile And Ubiquitous Computing		4

RJSPIT103	Paper Title: Data Analysis Tools			
	I	*PART I : COMPUTING* • Statistics in Modern day Using C • Databases	4	4
	II	• Matrices and models • Graphics		4
	III	• More coding tools		4

		PART II : STATISTICS • Distributions for description		
	IV	• Maximum likelihood estimation • Monte Carlo		4
	V	• Maximum likelihood estimation • Monte Carlo		4

RJSPIT104	Paper Title: Software Testing			
	I	• Test Basics • Testing Processes	4	4
	II	• Test Management		4
	III	• Test Techniques		4
	IV	• Tests of Software characteristics • Reviews • Incident Management		4
	V	• Standards and Test Process Improvement • Test Techniques • People Skills and Team Composition		4

RJSPITP101	Data Mining	2	4
RJSPITP102	Distributed Systems	2	4
RJSPITP103	Data Analysis Tools	2	4
RJSPITP104	Software Testing	2	4

Theory semester I

Course Code	Title	Credits
RJSPIT101	Data Mining	4

Unit	Description
I	**Introduction:** Basics of data mining, related concepts, Data mining techniques. **Data:** Introduction, Attributes, Data Sets, and Data Storage, Issues Concerning the Amount and Quality of Data **Knowledge Representation:** Data Representation and their Categories: General Insights, Categories of Knowledge Representation, Granularity of Data and Knowledge Representation Schemes, Sets and Interval Analysis, Fuzzy Sets as Human-Centric Information Granules, Shadowed Sets, Rough Sets, Characterization of Knowledge Representation Schemes, Levels of Granularity and Perception Perspectives, The Concept of Granularity in Rules.
II	**Data Preprocessing:** Descriptive Data Summarization, Data Cleaning, Data Integration and Transformation, Data Reduction, Data Discretization and Concept Hierarchy Generation. **Mining Frequent Patterns, Associations, and Correlations:** Basic Concepts, Efficient and Scalable Frequent Item set Mining Methods, Mining Various Kinds of Association Rules, Constraint-Based Association Mining
III	**Classification and Prediction:** What Is Classification?, What Is Prediction?, Issues Regarding Classification and Prediction, Classification by Decision Tree Induction, Bayesian Classification, Rule-Based Classification, Classification by Back-propagation, **Support Vector Machines, Associative Classification:** Classification by Association Rule Analysis, Lazy Learners, Other Classification Methods, Prediction, Accuracy and Error Measures, Evaluating the Accuracy of a Classifier or Predictor, Ensemble Methods Increasing the Accuracy, Model Selection.
IV	**Cluster Analysis:** What Is Cluster Analysis?, Types of Data in Cluster Analysis, A Categorization of Major Clustering Methods, Partitioning Methods, Hierarchical Methods, Density-Based Methods, Grid-Based Methods, Model-Based Clustering Methods, Clustering High-Dimensional Data, Constraint-Based Cluster Analysis, Outlier Analysis
V	**Graph Mining, Social Network Analysis, and Multirelational Data Mining:** Graph Mining, Social Network Analysis, Multirelational Data Mining. **Mining Object, Spatial, Multimedia, Text, and Web Data:** Multidimensional Analysis and Descriptive Mining of Complex Data Objects, Spatial Data Mining, Multimedia Data Mining, Text Mining, Mining the World Wide Web.

M.Sc. IT	Semester I Theory
RJSPIT101 Course I Data Mining	Course Outcomes 1.1 : This course: 1. Provide students with an overview of the methodologies and approaches to data mining. 2. Gain insight into the challenges and limitations of different data mining techniques. 3. Provide the students with practice on applying data mining solutions using common data mining software tool (e.g. WEKA, SPSS, Data Miner,…). 4. Prepare students for research in the area of data mining and related app. Learning outcomes : Students will be able to: ➢ Understand data mining principles and techniques: Introduce DM as a cutting edge business intelligence method and acquaint the students with the DM techniques for building competitive advantage through proactive analysis, predictive modeling, and identifying new trends and behaviors. ➢ Build basic terminology and knowledge representation. ➢ Learn how to gather and analyze large sets of data to gain useful business understanding. ➢ Learn how to preprocess (cleansing) data. ➢ Understand data mining functions. ➢ Learn how to produce a quantitative analysis report/memo with the necessary information to make decisions. ➢ Describe and demonstrate basic data mining algorithms, methods, and tools. ➢ Identify business applications of data mining. ➢ Overview of the developing areas - web mining, text mining, and ethical aspects of data mining.

Course Code	Title	Credits
RJSPIT102	Distributed Systems	4

Unit	Description
I	**Characterization Of Distributed Systems**: Introduction, Examples of Distributed Systems, Trends In Distributed Systems, Focus On Resource Sharing, Challenges, Case Study: The World Wide Web. **System Models:** Physical Models, Architectural Models, Fundamental Models
II	**Networking And Internetworking:**

	Types Of Network, Network Principles, Internet Protocols, Case Studies: Ethernet, Wi-Fi And Bluetooth. **Interprocess Communication:** The API For The Internet Protocols, External Data Representation And Marshalling, Multicast Communication, **Network Virtualization:** Overlay Networks, Case Study: MPI
III	**Remote Invocation:** Request-Reply Protocols, Remote Procedure Call, Remote Method Invocation, Case Study: Java RMI **Indirect Communication:** Group communication, Publish-subscribe systems, Message queues, Shared memory approaches **Web Services:** Web services, Service descriptions and IDL for web services, A directory service for use with web services, XML security, Coordination of web services, applications of web services.
IV	**Coordination And Agreement:** Distributed mutual exclusion Elections Coordination and agreement in group communication, Consensus and related problems **Name Services:** Name services and the Domain Name System, Directory services, Case study: The Global Name Service, Case study: The X.500 Directory Service. **Time And Global States:** Clocks, events and process states , Synchronizing physical clocks , Logical time and logical clocks, Global states, Distributed debugging
V	**Distributed Transactions:** Flat and nested distributed transactions, Atomic commit protocols, Concurrency control in distributed transactions, Distributed deadlocks. **Replication:** System model and the role of group communication, Fault-tolerant services, Case studies of highly available services: The gossip architecture, Bayou and Coda, Transactions with replicated data **Mobile And Ubiquitous Computing:** Association, Interoperation, Sensing and context awareness, Security and privacy, Adaptation, Case study: Cool town

M.Sc. IT	Semester I Theory
RJSPIT102 Course II Distributed Systems	Course Outcomes 1.2 : **1.** Given the knowledge of operating systems and sequential program design, the students will learn efficient distributed algorithms to solve large problems where data and control is distributed over different nodes. Learning outcomes : Students will be able to: ➢ Demonstrate the knowledge of the basic elements and the concepts related to distributed system technologies. ➢ Demonstrate the knowledge of core architectural aspects of distributed systems. ➢ Design and implement distributed applications. ➢ Demonstrate details of the main underlying components of

	distributed systems (such as RPC, file systems). ➢ Use and apply important methods in distributed systems to support scalability and fault tolerance. ➢ Demonstrate experience in building large-scale distributed applications.

Course Code	Title	Credits
RJSPIT103	Data Analysis Tools	4

Unit	Description
I	***PART I : COMPUTING*** **Statistics in Modern day**, **C :** Lines, Variables and their declarations, Functions, The debugger , Compiling and running, Pointers , Arrays and other pointer tricks, Strings **Databases** : Basic queries , Doing more with queries, Joins and sub queries, On database design , Folding queries into C code
II	**Matrices and models** : The GSL's matrices and vectors apo_dat, Shunting data, Linear algebra, Numbers, gsl_matrix and gsl_vector internals, Models, **Graphics:** plot , Some common settings, From arrays to plots, A sampling of special plots, Animation, On producing good plots, Graphs--nodes and flowcharts, Printing and LATEX
III	**More coding tools :** Function pointers , Data structures, Parameters, Syntactic sugar, More tools ***PART II : STATISTICS*** **Distributions for description** : Moments, Sample distributions, Using the sample distributions, Non-parametric description.
IV	**Linear projections**: Principal component analysis, OLS and friends, Discrete variables, Multilevel modeling. **Hypothesis testing with the CLT**: The Central Limit Theorem, Meet the Gaussian family, Testing a hypothesis, ANOVA, Regression, Goodness of fit.
V	**Maximum likelihood estimation:** Log likelihood and friends, Description: Maximum likelihood estimators, Missing data, Testing with likelihoods. **Monte Carlo** : Random number generation, Description: Finding statistics for a distribution, Inference: Finding statistics for a parameter, Drawing a distribution, Non-parametric testing.

M.Sc. IT	Semester I Theory
RJSPIT103 Course III Data Analysis Tools	Course Outcomes 1.3 : 1. The students will learn to build the statistical model based on another model. 2. The students will develop and test hypotheses for various data sets. 3. The students will learn a variety of statistical tests, as well as strategies to know how to apply the appropriate one to your specific data and question. 4. The students will learn to draw the functions using gnuplot. Learning outcomes : Students will be able to: ➢ Handle various data structures and SQLite database in C. ➢ Estimate statistical models. ➢ Use apophenia and gsl library functions for matrices and models. ➢ Use gnuplot package for plotting the functions and data. ➢ Identify potential sources of data and distinguish between quantitative and qualitative data. ➢ Demonstrate basic data analysis techniques and show how this analysis can contribute to a business' future growth. ➢ Understand the concept of distribution, linear projections, Central limit theorem, Maximum likelihood estimation and Monte Carlo methods.

Course Code	Title	Credits
RJSPIT104	Software Testing	4

Unit	Description
I	**Test Basics:** Introduction, Testing in the Software Lifecycle, Specific Systems, Metrics and Measurement, Ethics **Testing Processes:** Introduction, Test Process Models, Test Planning and Control, Test Analysis and Design, Non-functional Test Objectives, Identifying and Documenting Test Conditions, Test Oracles, Standards, Static Tests, Metrics, Test Implementation and Execution, Test Procedure Readiness, Test Environment Readiness, Blended Test Strategies, Starting Test Execution, Running a Single Test Procedure, Logging Test Results, Use of Amateur Testers, Standards, Metrics, Evaluating Exit Criteria and Reporting, Test Suite, Defect Breakdown, Confirmation Test Failure Rate, System Test Exit Review, Standards, Evaluating Exit Criteria and Reporting Exercise, System Test Exit Review, Test Closure Activities
II	**Test Management:** Introduction, Test Management Documentation, Test Plan Documentation Templates, Test Estimation, Scheduling and Test Planning, Test Progress Monitoring and Control, Business Value of Testing, Distributed, Outsourced, and

	Insourced Testing, Risk- Based Testing, Risk Management, Risk Identification, Risk Analysis or Risk Assessment, Risk Mitigation or Risk Control, Risk Identification and Assessment Results, Risk-Based Testing throughout the Lifecycle, Risk-Aware Testing Standards, Risk-Based Testing Exercise, Project Risk By-Products, Requirements Defect By-Products, Test Case Sequencing Guidelines, Failure Mode and Effects Analysis, Test Management Issues
III	**Test Techniques** Introduction, Specification-Based, Equivalence Partitioning, Avoiding Equivalence Partitioning Errors, Composing Test Cases with Equivalence Partitioning, Equivalence Partitioning Exercise, Boundary Value Analysis, Examples of Equivalence Partitioning and Boundary Values, Non-functional Boundaries, Functional Boundaries, Integers, Floating Point Numbers, Testing Floating Point Numbers, Number of Boundaries, Boundary Value Exercise, Decision Tables, Collapsing Columns in the, Combining Decision Table Testing with Other Techniques, Nonexclusive Rules in Decision Tables, 4 Decision Table Exercise, Decision Table Exercise Debrief, State-Based Testing and State Transition Diagrams, Super states and sub states, State Transition Tables, Switch Coverage, State Testing with Other Techniques, State Testing Exercise, State Testing Exercise Debrief, Requirements- Based Testing Exercise, Requirements-Based Testing Exercise Debrief, Structure-Based, Control-Flow Testing, Building Control-Flow Graphs, Statement Coverage, Decision Coverage, Loop Coverage, Hexadecimal Converter Exercise, Hexadecimal Converter Exercise Debrief, Condition Coverage, Decision/Condition Coverage, Modified Condition /Decision Coverage(MC/DC), Multiple Condition Coverage, Control-Flow Exercise, Control-Flow Exercise Debrief, Path Testing, LCSAJ, Basis Path/Cyclomatic Complexity Testing, Cyclomatic Complexity Exercise, Cyclomatic Complexity Exercise Debrief, Final Word on Structural Testing, Structure-Based Testing Exercise, Structure-Based Testing Exercise Debrief, Defect- and Experience-Based, Defect Taxonomies, Error Guessing, Checklist Testing, Exploratory Testing, Test Charters, Exploratory Testing Exercise, Software Attacks, An Example of Effective Attacks, Other Attacks, Software Attack Exercise, Software Attack Exercise Debrief, Specification-, Defect-, and Experience-Based Exercise, Specification-, Defect-,and Experience-Based Exercise Debrief, Common Themes, Static Analysis, Complexity Analysis, Code Parsing Tools, Standards and Guidelines, Data-Flow Analysis, Set-Use Pairs, Set-Use Pair Example, Data-Flow Exercise, Data-Flow Exercise Debrief, Data-Flow Strategies, Static Analysis for Integration Testing, Call-Graph Based Integration Testing, McCabe Design Predicate Approach to Integration Testing, Hex Converter Example, McCabe Design Predicate Exercise, McCabe Design Predicate Exercise Debrief, Dynamic Analysis, Memory Leak Detection, Wild Pointer Detection, API Misuse Detection.
IV	**Tests of Software Characteristics** Introduction, Quality Attributes for Domain Testing, Accuracy, Suitability, Interoperability, Usability, Usability Test Exercise, Usability Test Exercise Debrief, Quality Attributes for Technical Testing, Technical Security, Security Issues, Timely Information, Reliability, Efficiency, Multiple Flavors of Efficiency Testing, Modeling the System, Efficiency Measurements, Examples of Efficiency Bugs, Exercise: Security, Reliability and Efficiency, Exercise: Security, Reliability, and Efficiency Debrief, Maintainability, Sub characteristics of Maintainability, Portability, Maintainability and Portability Exercise.

	Reviews Introduction, The Principles of Reviews, Types of Reviews, Introducing Reviews, Success Factors for Reviews, Deutsch's Design Review Checklist, Marick's Code Review Checklist, The Open Laszlo Code Review Checklist, Code Review Exercise, Deutsch Checklist Review Exercise. **Incident Management** Introduction, When Can a Defect Be Detected? Defect Lifecycle, Defect Fields, Metrics and Incident Management, Communicating Incidents, Incident Management Exercise.
V	**Standards and Test Process Improvement** Introduction, Standards Considerations, Test Improvement Process, Improving the Test Process, Improving the Test Process with TMM, Improving the Test Process with TPI, Improving the Test Process with CTP, Improving the Test Process with STEP, Capability Maturity Model Integration, CMMI, Test Improvement Process Exercise. **Test Techniques** Introduction, Test Tool Concepts, The Business Case for Automation, General Test Automation Strategies, An Integrated Test System Example, Test Tool Categories, Test Management Tools, Test Execution Tools, Debugging, Troubleshooting, Fault Seeding, and Injection Tools, Static and Dynamic Analysis Tools, Performance Testing Tools, Monitoring Tools, Web Testing Tools, Simulators and Emulators, Keyword-Driven Test Automation, Capture/Replay Exercise, Capture/Replay Exercise Debrief, Evolving from Capture/Replay, The Simple Framework Architecture, Data-Driven Architecture, Keyword-Driven Architecture, Keyword Exercise, Performance Testing, Performance Testing Exercise. **People Skills and Team Composition** Introduction, Individual Skills, Test Team Dynamics, Fitting Testing within an Organization, Motivation, and Communication.

M.Sc. IT	Semester I Theory
RJSPIT104 Course IV Software Testing	Course Outcomes 1.4 : **1.** The students will learn advance software testing concepts as per ISQTB certification course. Learning outcomes : Students will be able to: ➢ Manual Testing concepts and techniques ➢ Testing types & Test methods ➢ Functional & Nonfunctional testing types ➢ Test execution, defect management and test case management

Course Code	Practical Title	Credits
RJSPITP101	Data Mining	2

1. Design the data mining model using SQL server / Oracle.
2. Show the implementation of Naïve Bayes algorithm.
3. Show the implementation of Decision Tree.
4. Show the implementation of Time Series Algorithm.
5. Show the implementation of Clustering Algorithm.
6. Show the implementation of k-nearest neighbor and Apriori algorithm.
7. Show the implementation of Association Algorithm.
8. Show the implementation of Text Mining.
9. Show the implementation of Multimedia Mining.
10. Show the implementation of Spatial Mining.

M.Sc. IT	Semester I Practical
RJSPITP101 Practical I Data Mining	Course Outcome: 1. The student will learn to implement the various algorithms of data processing and data mining. Learning outcomes : ➢ The student will be able to implement the various algorithms of data processing and data mining.

Course Code	Practical Title	Credits
RJSPITP102	Distributed Systems	2

1. Implement the concept for sharing the resources using distributed system.
2. Write a program for implementing Client Server communication model.
3. Write a program to show the object communication using RMI.
4. Show the implementation of Remote Procedure Call.
5. Show the implementation of web services.
6. Write a program to execute any one mutual exclusion algorithm.
7. Write a program to implement any one election algorithm.
8. Show the implementation of any one clock synchronization algorithm.
9. Write a program to implement two phase commit protocol.
10. Implement the concept of distributed file system architecture.

M.Sc. IT	Semester I Practical
RJSPITP102 Practical II Distributed System	Course Outcome: The students will learn to: 1. Implement the concept for sharing the resources using distributed system. 2. Implement Client Server communication model and show the object communication using RMI.

	3. Implement Remote Procedure Call, web services, mutual exclusion algorithm, election algorithm and clock synchronization algorithm. 4. Implement two phase commit protocol. 5. Implement the concept of distributed file system architecture. Learning outcomes : The students will be able to: ➢ Implement the concept for sharing the resources using distributed system. ➢ Implement Client Server communication model and show the object communication using RMI. ➢ Implement Remote Procedure Call, web services, mutual exclusion algorithm, election algorithm and clock synchronization algorithm. ➢ Implement two phase commit protocol. ➢ Implement the concept of distributed file system architecture.

Course Code	Practical Title	Credits
RJSPITP103	Data Analysis Tools	2

1. Some SQL queries based on the 1st Unit.
2. Implementing GSL matrix and vectors
3. Graph Plotting
4. Implement the statistical distributions
5. Implement regression and goodness of fit
6. Implement testing with likelihood
7. Generate random numbers using Monte Carlo method
8. Implementing Parametric testing
9. Drawing an Inference
10. Implement Non-parametric Testing

M.Sc. IT	Semester I Practical
RJSPITP103 Practical III Data Analysis Tools	Course Outcome: 1. The students will learn to use the various packages like SQLite, GNU Plot, Apophenia library, and GSL library in C. Learning outcomes : The students will be able to: ➢ Use the various packages like SQLite, GNU Plot, Apophenia library, and GSL library in C. ➢ Implement the statistical models. ➢ Build statistical model using another statistical model.

Course Code	Practical Title	Credits
RJSPITP104	Software Testing	2

1. Evaluating Test Exit Criteria and Reporting
2. Static and Dynamic Analysis
3. Rate Quality Attributes for Domain and Technical Testing
4. Perform Review
5. Incident Management
6. Path Testing and Equivalence Partitioning
7. Performance Testing
8. Using Testing Tool Selenium
9. Using Testing Tool QTP
10. Using Testing Tool WAPT
11. Using Testing Tool VTEST
12. Using Testing Tool AutoIT

M.Sc. IT	Semester I Practical
RJSPITP104 Practical IV Software Testing	Course Outcome: 1. The students will learn to test softwares through all phases of software testing lifecycle. Learning outcomes : The students will be able to: ➢ Demonstrate the knowledge of advanced software testing techniques. ➢ Demonstrate different tools of software testing like Selenium, QTP, VTEST, WAPT and AutoIT.

References

1. M. H. Dunham. Data Mining: Introductory and Advanced Topics. Pearson Education. 2010. (Unit I)
2. Krzysztof J. Cios,W. Pedrycz,R. W.Swiniarski,L.A. Kurgan," Data Mining" A Knowledge Discovery Approach", Springer (Unit I).
3. J. Han and M. Kamber, " Data Mining: Concepts and Techniques", Second Edition, Elsevier, Reprinted 2008(Unit II-Unit VI)
4. Dr.CarolynK.Hamm,"Oracle Data Mining",RampantTechPress, SPD.
5. C.Ballard, Dynamic Warehousing and Data Mining Made Easy,ReddBooks,IBM (SPD)
6. H.Witten and E. Frank. Data Mining: Practical Machine Learning Tools and Techniques. Morgan Kaufmann. 2005.
7. D. Hand, H. Mannila and P. Smyth. Principles of Data Mining. Prentice-Hall. 2001.
8. Z.Tang and J MacLennan, "Data Mining with SQL Server 2005",Wiley
9. George Coulouris, Jean Dollimore, Tim Kindberg, Gordon Blair , Distributed Systems - Concepts and Design (Unit I-Unit VI)

10. A. Taunenbaum, "Distributed Systems: Principles and Paradigms"
11. G. Coulouris, J. Dollimore, and T. Kindberg, "Distributed Systems: Concepts and Design", Pearson Education
12. Modeling with Data: Tools and Techniques for Scientific Computing Ben Klemens, Princeton University Press.
13. Computational Statistics, James E. Gentle, Springer.
14. Computational Statistics, Second Edition, Geof H. Givens and Jennifer A. Hoeting, Wiley Publications.
15. Advanced Software Testing—Vol. 3 by Rex Black and· Jamie L. Mitchell, Rocky Nook Publication,
16. Advanced Software Testing Vol. 2 by Rex Black, Rocky Nook Publication, 2008
17. W.E. Perry, "Effective Methods for Software Testing", John Wiley.
18. Kaner C., Nguyen H., Falk J., "Testing Computer Software", John Wiley.
19. Boris Beizer, "Software Testing Techniques", Dreamtech
    Louise Tamres, "Introducing Software Testing", Pearson Education.