

3.3.3 Number of books and chapters in edited volumes/books published and papers published in national/ international conference proceedings per teacher during last five years

3.3.3.1. Total number of books and chapters in edited volumes/books published and papers in national/ international conference proceedings year-wise during last five years

DVV Query :

1) Kindly provide the authenticated list of Number of books and chapters in edited volumes/books published and papers published in national/ international conference proceedings per teacher during last five years duly signed by competent authority.

2) Kindly provide Cover page, content page and first page of the selected publication. 3) Kindly provide Web-link of research papers

INDEX

Sr. No.	Name of the teacher	Title of the book/chapters published	Year of publication	Page No of E- copies
1	Dr. Bhushan Jadhav	Book - Parallel and Distributed Systems	July 2020-June 2021	78
2	ARUN KULKARNI	Book - Parallel and Distributed Systems	July 2020-June 2021	72
3	Shilpa Verma	Chapter	July 2019-June 2020	131
4	Dr. Seema Kolkur	Chapter	July 2018-June 2019	127
5	Dr. Arti Deshpande	Book- Data warehousing and mining	July 2018-June 2019	25
6	Dr. Ujwala Bharambe	Chapter	July 2018-June 2019	149



Dr. G. T. Thampi
PRINCIPAL
Thadomal Shahani Engineering College
Bandra (W), Mumbai - 400 050.



7	Dr Madhuri rao	Book- Advances in Data Sciences, Security and Applications	July 2018-June 2019	145
8	GAURAV MITTAL VINAYAK KULKARNI	Book - Advances in Fluid Mechanics and Solid Mechanics	July 2017- June 2018	135
9	Shilpa Ingoley	Book- Computer Network	July 2017-June 2018	119
10	Shilpa Ingoley	Book - Computer Network	July 2017-June 2018	12
11	Dr. Arti Deshpande	Book -Machine Learning	July 2017-June 2018	10
12	Dr. Arti Deshpande	Book - Data Warehousing and mining	July 2017-June 2018	23
13	Dr. Arti Deshpande	Book -Data Warehousing and mining techniques	July 2017-June 2018	30



Dr. G. T. Thampi
PRINCIPAL
Thadomal Shahani Engineering College
Bandra (W), Mumbai - 400 050.



14	Dr. Arti Deshpande	Book- Data mining	July 2017-June 2018	37
15	Dr. Arti Deshpande	Book- Data Warehousing and mining	July 2017-June 2018	19
16	Dr. Arti Deshpande	Book -Data mining and Warehousing	July 2017-June 2018	51
17	Dr. Arti Deshpande	Book - Data Mining and Business Intelligence	July 2017-June 2018	58
18	Dr. Arti Deshpande	Book - Data Mining and Business Intelligence	July 2017-June 2018	65
19	Bhushan Jadhav	Book - Distibuted Computing	July 2017-June 2018	85
20	Bhushan Jadhav	Book - Information Management	July 2017-June 2018	92



Dr. G. T. Thampi
PRINCIPAL
Thadomal Shahani Engineering College
Bandra (W), Mumbai - 400 050.



21	Bhushan Jadhav	Book - Internet of things	July 2017-June 2018	104
22	Dr. Arti Deshpande	Book -Machine Learning	July 2017-June 2018	10
23	Dr. Arti Deshpande	Book-Data Warehousing and mining techniques	July 2017-June 2018	32
24	Dr. Arti Deshpande	Book - Data Mining and Business Intelligence	July 2017-June 2018	58
25	Dr. Arti Deshpande	Book - Data Mining and Business Intelligence	July 2017-June 2018	60
26	Dr. Arti Deshpande	Book -Machine Learning	July 2017-June 2018	15
27	Dr. Arti Deshpande	Book – Data Warehousing with Mining Techniques	July 2017-June 2018	44



Dr. G. T. Thampi
PRINCIPAL
Thadomal Shahani Engineering College
Bandra (W), Mumbai - 400 050.



28	Dr. Arti Deshpande	Book – Data Mining	July 2017-June 2018	34
29	Dr. G.T. Thampi	Book - Project Management	July 2016-June 2017	9
30	Ashwini S. Kunte	Chapter	July 2016-June 2017	134
31	Bhushan Jadhav	Book	July 2016-June 2017	104
32	Bhushan Jadhav	Book – Distributed computing	July 2016-June 2017	83
33	Bhushan Jadhav	Book - Cloud Computing Services	July 2016-June 2017	111



Dr. G. T. Thampi
PRINCIPAL
Thadomal Shahani Engineering College
Bandra (W), Mumbai - 400 050.



34	Bhushan Jadhav	Book - Internet of things	July 2016-June 2017	102
35	Bhushan Jadhav	Book - Distributed Systems	July 2016-June 2017	67



Dr. G. T. Thampi
PRINCIPAL
 Thadomal Shahani Engineering College
 Bandra (W), Mumbai - 400 050.



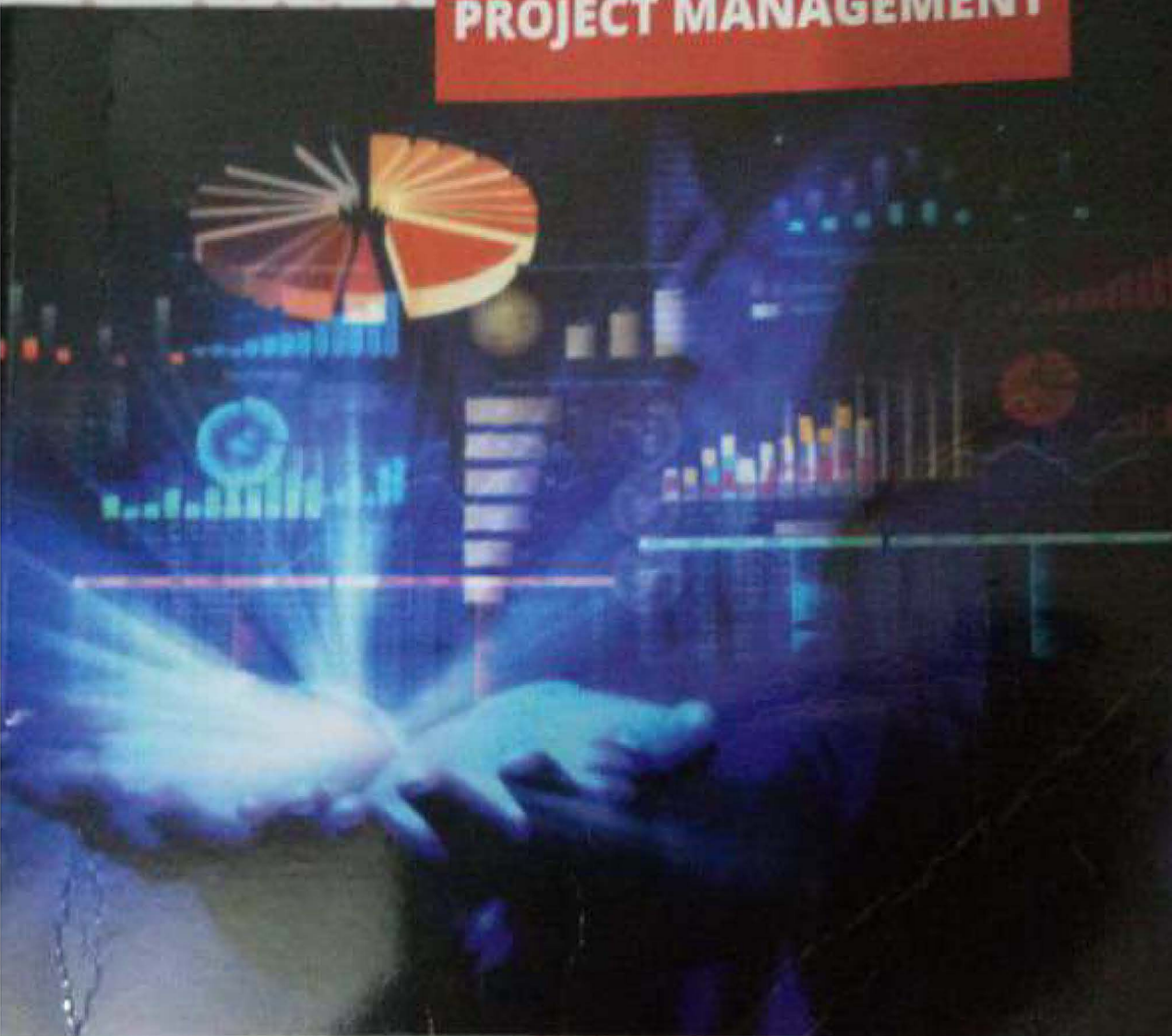
Books by Author – Dr G.T. Thampi



Dr. G. T. Thampi
PRINCIPAL
Thadomal Shahani Engineering College
Bandra (W), Mumbai - 400 050.



PROJECT MANAGEMENT



C. T. Upadhyaya

S. T. Tharopi

Dr. G. T. Thampi
PRINCIPAL
Thadomal Shahani Engineering College
Borivali (W), Mumbai - 400 075.



www.staredusolutions.org

Copyright ©StarEdu Solutions

All Rights Reserved. No part of this book may be reproduced, stored in a retrieval system, or transmitted in any form or by any means, electronic, mechanical, photocopying, recording or scanning without written permission of the publisher.

Limits of Liability:

While the publisher and the author have used their best efforts in preparing this book, StarEdu Solutions and the author make no representation warranties with respect to the accuracy or completeness of the contents of this book, and specifically disclaim any implied warranties of merchantability or fitness for any particular purpose. There are no warranties which extend beyond the descriptions contained in this paragraph. No warranty may be created or extended by sales representatives or written sales materials.

Disclaimer:

The contents of this book have been checked for accuracy. Since deviations cannot be precluded entirely, StarEdu Solutions or its author cannot guarantee full agreement. As the book is intended for educational purpose, StarEdu Solutions or its author shall not be responsible for any errors, omissions or damages arising out of the use of the information contained in the book. This publication is designed to provide accurate and authoritative information with regard to the subject matter covered. Our credit for all contributions, images and reference are given in the courseware. In case, we missed on anything, we request you to bring it to our knowledge and we will address the same in next edition of the book. We believe in protecting the intellectual property. If you are copyright owner and if you believe that content constitutes an infringement of copyright, please bring to our knowledge and in case of justified issue, we will address the same in next edition of the book.

This book may contain reference to other resources related to topics. These references are for training purpose only. However, StarEdu Solutions do not recommend or promoting such references, solutions software and products mentioned in reference or training content.

Images used in the book are for education purpose only and we are not promoting or advertising any product.

All Readers are requested to report errors, inaccuracies and omissions to StarEdu Solutions at support@staredusolutions.org, so that same can be addressed in the next edition of the the book.

ISBN: 978-93-86765-71-0

Edition: 2020



Dr. G. T. Thampi
PRINCIPAL
Thadomal Shahani Engineering College
Bandra (W), Mumbai - 400 050.



Dr. G. T. Thampi
PRINCIPAL
Thadomal Shahani Engineering College
Bandra (W), Mumbai - 400 050.



Preface

Project Management is a significant concept in the modern times. The business organisations have been established to earn profits by creating deliverables and selling them in the market. These deliverables are developed through projects. Often these projects are carried out without any management and supervision. This leads to increased costs, delayed projects and unachieved aims. Hence, the Project Management is the need of the hour.

The project completion is achieved through the coordinated efforts of the members of project teams led by the project manager. A project leader is responsible for managing all the stages of the project and garnering support from the sponsors and the management, so as to ensure that the goals of the project are achieved. Therefore, the project managers should be familiar with all the project management techniques to control the execution of the projects and deliver the intended results on time. These techniques should be revisited, reorganised and reframed, depending on the ever-changing business scenarios. Keeping this view in mind, the Project Management book has been composed, which serves as a powerful tool for enabling the students to get acquainted with the innovative concepts of project management, its techniques and its applications.

The **Project Management book** is designed to enlighten the students about the importance of project management in business organisations. The book covers all the major aspects of project management, including project life cycle, initiation of projects, project planning and scheduling, project risk management, project execution and monitoring, and project termination. After studying this book, the students will be able to comprehend and apply the various concepts of project management in real-world business scenarios, so as to gain hands-on experience.

The **Project Management book** is developed with an updated content about the different aspects of the project management. It is written in a simple, lucid and easy-to-understand language. In this book, various real-life examples have been provided in all the chapters to enable the students to gain insights into the application of project management concepts. The examples will also guide them on how to deal with the similar real-life situations and take the necessary measures.

The content of the **Project Management book** has been developed after a wide-scale research done on the subject. The book has been designed in such a way so as to enhance the understanding of the students regarding the complex issues, often faced while handling project management related concepts. The book will enable the students to gain a clear understanding and confidence of using project management techniques and lead the multitude of projects in their real lives.

About the Authors

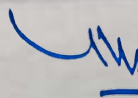


Dr. K. T. Upadhyaya has over 17 years experience in executing large scale projects initially in Cement, Petrol chemical, Food, and Dairy industry and currently in Software development, Data Analytics and Automation projects. Dr. K. T. Upadhyaya is engaged in training in Project Management, Risk Management, Data warehousing for past 16 years. He has trained over 10,000 participants from companies in different domains and students from top B-schools in India. He also trains on Data Analytics, Supply Chain Analytics, Business Process Management, Process Automation and AI.

Dr. K. T. Upadhyaya has keen interest in relating Indian Mythology, particularly Mahabharata, to management lessons. He is an avid reader of books relating to topics ranging from Data to Fiction and everything in between. Sachin Dev Burman and Sachin Tendulkar are his two favourites in their respective domains. Dr. K. T. Upadhyaya holds a Mechanical Engineering Degree from Sardar Patel College of Engineering, Mumbai and Masters and Ph. D. from BITS Pilani. He is a Certified Project Management Professional from PMI(R), Pennsylvania, USA.



Dr. G. T. Thampi is currently The Principal at Thadomal Shahani Engineering College, Bandra, Mumbai. He hold a Ph.D Degree in Technology and has more than 33 years of experience in renowned college. Dr. G.T. Thampi has been a part of some interesting researches and holds interests in Business Process and Re-Engineering in realm of Engineering Education. Apart from his own researches he has been a guide for multiple researches done in technology front. 17 research scholars awarded Ph.D. under his guidance. Dr. Thampi is also a co-author of more than 80 research publications and books



Dr. G. T. Thampi
PRINCIPAL
Thadomal Shahani Engineering College
Bandra (W), Mumbai - 400 050.



About the Book

Project Management book is designed to familiarise the students with the different aspects of successful project completion. This book provides comprehensive knowledge about the various concepts and strategies for managing projects effectively. It includes various examples to provide an understanding of the application of project management concepts in the real world. The content of the book is designed in a simple and student-friendly manner to help them apply project management techniques in real life.

Salient Features of the Book

Easy-to-understand language

Comprehensive coverage of all the relevant concepts

Numerous examples to enhance the understanding of students

Variety of diagrams and cases to impart the knowledge about the complexities faced during the management of projects

Several review questions after each chapter to appraise the performance of the students

Index terms at the end of the book to provide the definitions of project-related terms

Dr. G. T. Thampi
PRINCIPAL

Thadomal Shahani Engineering College
Bandra (W), Mumbai - 400 050.



StarEdu Solutions India Pvt. Ltd.

- 450, North Brand Blvd, Suite 600, Glendale, CA 91203
- support@staredusolutions.org
- www.staredusolutions.org

ISBN: 978-93-86765-71-0



9 789386 765710

Books by Author – Dr Arti Deshpande



Dr. G. T. Thampi
PRINCIPAL
Thadomal Shahani Engineering College
Bandra (W), Mumbai - 400 050.



Your Success is Our Goal 2020 SPPU

Semester VIII - Computer Engineering

- **Machine Learning**
Dr. Pallavi N. Halarankar, Dr. Arti Deshpande
- **Information and Cyber Security**
Pravin Goyal
- **COMPILERS** (Elective III)
Kavita Sultanpure, Pujashree S. Vidap, Jagadish K. Kamble
- **Embedded and Real Time Operating Systems** (Elective III)
Harish G. Narula, Sanjay B. Deshmukh, Priya Charles, Pratiksha P. Shevtekar, Dr. A. V. Patil, Dr. Manisha Bhende
- **Soft Computing and Optimization Algorithms** (Elective III)
Dipali Y. Koshti, Nikahat S. Mulla
- **Human Computer Interface** (Elective IV)
Dr. Jayshree Pansare
- **Cloud Computing** (Elective IV)
Pravin Goyal

coming soon.....

es easy-solutions now with **TechKnowledge** Publications

Paper Solutions Trusted by lakhs of students from more than 15 years

<p>Head Office : B/5, First Floor, Maniratna Complex, Taware Colony, Aranayeshwar Corner, Pune - 411009, Maharashtra, India. Tel. : 91-20-24221234, 91-20-24225678</p> <p>Distribution Address : Shop No. 8-9, Ground Floor, Shaan Bramha Complex, Appa Baiwani Chowk, Opp. Farakhana Police Station, Pune - 411002, Mobile : 74389 01812 / 074089 01812.</p> <p>Our Branches : Pune Mumbai Kolhapur Nagpur Solapur Nashik</p> <p>Email : info@techknowledgebooks.com Website : www.techknowledgebooks.com</p>	<p>ISBN : 978-93-85748-20-4</p> <p>Price : 150/-</p> <p>PE66A</p> <p>© TechKnowledgePublications</p>
--	--

MACHINE LEARNING

Semester VIII - Computer Engineering

Book Code: PE66A

Price: 150/-

ISBN

Includes
⇒ Solved University Question Papers of In-Sem (March 2019) & End-Sem (May 2019) Exams.

2020 SPPU

Strictly as per the new Credit System Syllabus (2015 Course)
Savitribai Phule Pune University
w.e.f. academic year 2018-2019

MACHINE LEARNING

Semester VIII - Computer Engineering

Same Subject, Same Authors with New Publication

Dr. Pallavi N. Halarankar
Dr. Arti Deshpande

TechKnowledge Publications

Dr. G. T. Thampi
PRINCIPAL
Thadomal Shahani Engineering College
Bandra (W), Mumbai - 400 050.



Machine Learning

(Code : 410250)

Semester VIII - Computer Engineering
(Savitribai Phule Pune University)

Strictly as per the New Credit System Syllabus (2015 Course)
Savitribai Phule Pune University w.e.f. Academic Year 2018-2019


Dr. Pallavi N. Halarnkar

Department of Computer Engineering
Thadomal Shahani Engineering College, Mumbai.
Maharashtra, India.

Dr. Arti Deshpande

Department of Computer Engineering
Thadomal Shahani Engineering College,
Mumbai.
Maharashtra, India.




Dr. G. T. Thampi
PRINCIPAL
Thadomal Shahani Engineering College
Bandra (W), Mumbai - 400 050.



PE66A Price ₹ 150/-



Machine Learning

Dr. Pallavi N. Halarnkar, Dr. Arti Deshpande

(Semester VIII - Computer Engineering) (Savitribai Phule Pune University)

Copyright © by Authors. All rights reserved. No part of this publication may be reproduced, copied, or stored in a retrieval system, distributed or transmitted in any form or by any means, including photocopy, recording, or other electronic or mechanical methods, without the prior written permission of the publisher.

This book is sold subject to the condition that it shall not, by the way of trade or otherwise, be lent, resold, hired out, or otherwise circulated without the publisher's prior written consent in any form of binding or cover other than which it is published and without a similar condition including this condition being imposed on the subsequent purchaser and without limiting the rights under copyright reserved above.

First Edition : **February 2019**

Second Revised Edition : **January 2020 (TechKnowledge Publications)**

This edition is for sale in India, Bangladesh, Bhutan, Maldives, Nepal, Pakistan, Sri Lanka and designated countries in South-East Asia. Sale and purchase of this book outside of these countries is unauthorized by the publisher.


Printed at : 37/2, Ashtavinayak Industrial Estate,
Near Pari Company,
Narhe, Pune, Maharashtra State, India.
Pune – 411041

ISBN : 978-93-89748-20-8

Published by

TechKnowledge Publications

Head Office : B/5, First floor, Maniratna Complex,
Taware Colony, Aranyeshwar Corner,
Pune - 411 009. Maharashtra State, India
Ph : 91-20-24221234, 91-20-24225678.
Email : info@techknowledgebooks.com,
Website : www.techknowledgebooks.com

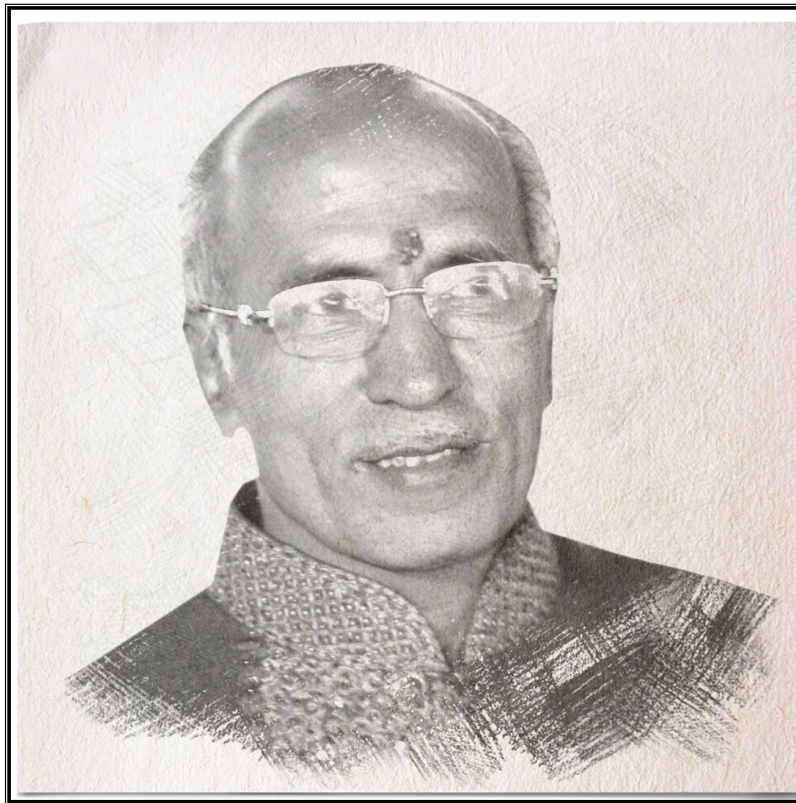

Dr. G. T. Thampi
PRINCIPAL
Thadomal Shahani Engineering College
Bandra (W), Mumbai - 400 050.



[410250] (FID : PE66) (Book Code : PE66A)

(Book Code : PE66A)

*We dedicate this Publication soulfully and wholeheartedly,
in loving memory of our beloved founder director,
Late Shri. Pradeepji Lalchandji Lunawat,
who will always be an inspiration, a positive force and strong support
behind us.*



“My work is my prayer to God”

- Lt. Shri. Pradeepji L. Lunawat

Dr. G. T. Thampi
PRINCIPAL
Thadomal Shahani Engineering College
Bandra (W), Mumbai - 400 050.



***Soulful Tribute and Gratitude for all Your
Sacrifices, Hardwork and 40 years of Strong Vision...***

Preface

Dear students,

We are extremely happy to present the book of “**Machine Learning**” for you. We have divided the subject into small chapters so that the topics can be arranged and understood properly. The topics within the chapters have been arranged in a proper sequence to ensure smooth flow of the subject.

We present this book in the loving memory of **Late Shri. Pradeepji Lunawat**, our source of inspiration and a strong foundation of “**TechKnowledge Publications**”. He will always be remembered in our heart and motivate us to achieve our milestone.

We are thankful to Mr. Shital Bhandari, Shri. Arunoday Kumar and Shri. Chandroday Kumar for the encouragement and support that they have extended. We are also thankful to the staff members of TechKnowledge Publications and others for their efforts to make this book as good as it is. We have jointly made every possible efforts to eliminate all the errors in this book. However if you find any, please let us know, because that will help us to improve further.

We are also thankful to my family members and friends for patience and encouragement.


Dr. G. T. Thampi
PRINCIPAL
Thadomal Shahani Engineering College
Bandra (W), Mumbai - 400 050.

 **Authors**


Syllabus

Savitribai Phule Pune University Fourth Year of Computer Engineering (2015 Course)

410250 : Machine Learning

Teaching Scheme : TH : 03 Hours/Week	Credit 03	Examination Scheme : In-Sem (Paper) : 30 Marks End-Sem (Paper) : 70 Marks
---	----------------------------	--

Prerequisite Courses : 207003 - Engineering Mathematics III

Companion Course:410254 - Laboratory Practice III

Course Objectives

- To understand human learning aspects and relate it with machine learning concepts.
- To understand nature of the problem and apply machine learning algorithm.
- To find optimized solution for given problem.

Course Outcomes

On completion of the course, student will be able to -

- Distinguish different learning based applications.
- Apply different preprocessing methods to prepare training data set for machine learning.
- Design and implement supervised and unsupervised machine learning algorithm.
- Implement different learning models.
- Learn Meta classifiers and deep learning concepts.


Course Contents

Unit I : Introduction to Machine Learning

(08 Hours)

Classic and adaptive machines, Machine learning matters, Beyond machine learning-deep learning and bio inspired adaptive systems, Machine learning and Big data.

Important Elements of Machine Learning - Data formats, Learnability, Statistical learning approaches, Elements of information theory.


Dr. G. T. Thampi
PRINCIPAL
Thadomal Shahani Engineering College
Bandra (W), Mumbai - 400 050.



(Refer chapter 1)

Unit II : Feature Selection

(08 Hours)

Scikit - learn Dataset, Creating training and test sets, managing categorical data, Managing missing features, Data scaling and normalization, Feature selection and Filtering, Principle Component Analysis(PCA) - non negative matrix factorization, Sparse PCA, Kernel PCA. Atom Extraction and Dictionary Learning.

(Refer chapter 2)

(Book Code : PE66A)

Unit III : Regression

(08 Hours)

Linear Regression : Linear models, A bi-dimensional example, Linear Regression and higher dimensionality, Ridge, Lasso and ElasticNet, Robust regression with random sample consensus, Polynomial regression, Isotonic regression.

Logistic Regression : Linear classification, Logistic regression, Implementation and Optimizations, Stochastic gradient descent algorithms, Finding the optimal hyper-parameters through grid search, Classification metric, ROC Curve.

(Refer chapter 3)

Unit IV : Naïve Bayes and Support Vector Machine

(08 Hours)

Bayes' Theorem, Naïve Bayes' Classifiers, Naïve Bayes in Scikit - learn- Bernoulli Naïve Bayes, Multinomial Naïve Bayes, and Gaussian Naïve Bayes.

Support Vector Machine(SVM) : Linear Support Vector Machines, Scikit- learn implementation-Linear Classification, Kernel based classification, Non-linear Examples. Controlled Support Vector Machines, Support Vector Regression.

(Refer chapter 4)

Unit V : Decision Trees and Ensemble Learning

(08 Hours)

Decision Trees : Impurity measures, Feature Importance. Decision Tree Classification with Scikit-learn, Ensemble Learning-Random Forest, AdaBoost, Gradient Tree Boosting, Voting Classifier.

Clustering Fundamentals: Basics, K-means: Finding optimal number of clusters, DBSCAN, Spectral Clustering. Evaluation methods based on Ground Truth- Homogeneity, Completeness, Adjusted Rand Index.

Introduction to Meta Classifier : Concepts of Weak and eager learner, Ensemble methods, Bagging, Boosting, Random Forests.

(Refer chapter 5)

Unit VI : Clustering Techniques

(08 Hours)


Hierarchical Clustering, Expectation maximization clustering, Agglomerative Clustering-Dendrograms, Agglomerative clustering in Scikit- learn, Connectivity Constraints.

Introduction to Recommendation Systems : Naïve User based systems, Content based Systems, Model free collaborative filtering-singular value decomposition, alternating least squares.

Fundamentals of Deep Networks: Defining Deep learning, common architectural principles of deep networks, building blocks of deep networks.

(Refer chapter 6)




Dr. G. T. Thampi
PRINCIPAL
Thadomal Shahani Engineering College
Bandra (W), Mumbai - 400 050.



eBooks are available on www.techknowledgebooks.com

Your Success is Our Goal
Semester V - Computer Engineering

Theoretical Computer Science	Dilip Kumar Sultania
Software Engineering	M. A. Ansari
Computer Network	J. S. Katre
Data Warehousing & Mining	Dr. Arti Deshpande, Dr. Pallavi N. Halarnkar
Department Level Optional Course - 1	
Internet Programming	Yogesh Prabhakar Pingle, Smita Jawale, Pooja Vartak
Adv. Data Management System	Mahesh Mali

coming soon.....



es
easy-solutions

Paper Solutions Trusted by lakhs of students from more than 15 years

Head Office :
 B/5, First Floor, Maniratra Complex, Taware Colony,
 Aranyeshwar Corner, Pune - 411009, Maharashtra, India.
 Tel. : 91-20-24221234, 91-20-24223678

ISBN : 978-93-50954-96-9


Price ₹ 295/-

MO175A

Our Branches : Pune | Mumbai | Kolhapur | Nagpur | Solapur | Nashik

For Library Orders Contact - Ved Book Distributors M : 80975 71421 / 92208 77214

Email : info@techknowledgebooks.com | Website : www.techknowledgebooks.com



DATA WAREHOUSING AND MINING
 Semester V - Computer Engineering

Dr. Arti Deshpande
 Dr. Pallavi N. Halarnkar

Book Code: K0175A
 Price: ₹ 295/-

Strictly as per the New Revised Syllabus (REV- 2019 'C' Scheme)
 of Mumbai University w.e.f. academic year 2021-22

**Data Warehousing
 and Mining**

(Code : CSC504)
 Semester V - Computer Engineering

Includes :

- Solved Latest University Question Papers.

Dr. Arti Deshpande
 Dr. Pallavi N. Halarnkar

Tech Knowledge
 Publications



Dr. G. T. Thampi
 PRINCIPAL
 Thadomal Shahani Engineering College
 Bandra (W), Mumbai-400 050.



Data Warehousing and Mining

(Code : CSC504)

Semester V : Computer Engineering (Mumbai University)

**Strictly as per the New Revised Syllabus (Rev-2019 'C' Scheme)
of Mumbai University w.e.f. academic year 2021-2022
(As per Choice Based Credit and Grading System)**

Dr. Arti Deshpande

Department of Computer Engineering
Thadomal Shahani Engineering College, Mumbai.
Maharashtra, India

Dr. Pallavi Halarnkar



Dr. G. T. Thampi
PRINCIPAL
Thadomal Shahani Engineering College
Bandra (W), Mumbai - 400 050.



 **Tech Knowledge**TM
Publications

MO175A Price ₹ 295/-



Data Warehousing and Mining (CSC504)

Dr. Arti Deshpande, Dr. Pallavi Halarnkar

Semester V : Computer Engineering (Mumbai University)

Copyright © Authors. All rights reserved. No part of this publication may be reproduced, copied, or stored in a retrieval system, distributed or transmitted in any form or by any means, including photocopy, recording, or other electronic or mechanical methods, without the prior written permission of the publisher.

This book is sold subject to the condition that it shall not, by the way of trade or otherwise, be lent, resold, hired out, or otherwise circulated without the publisher's prior written consent in any form of binding or cover other than which it is published and without a similar condition including this condition being imposed on the subsequent purchaser and without limiting the rights under copyright reserved above.

First Printed in India : February 2019

First Edition : July 2021

This edition is for sale in India, Bangladesh, Bhutan, Maldives, Nepal, Pakistan, Sri Lanka and designated countries in South-East Asia. Sale and purchase of this book outside of these countries is unauthorized by the publisher.

ISBN : 978-93-90694-96-9

Published By

TECHKNOWLEDGE PUBLICATIONS

Printed @

37/2, Ashtavinayak Industrial Estate,
Near Pari Company,
Narhe, Pune, Maharashtra State, India.
Pune - 411041

Head Office

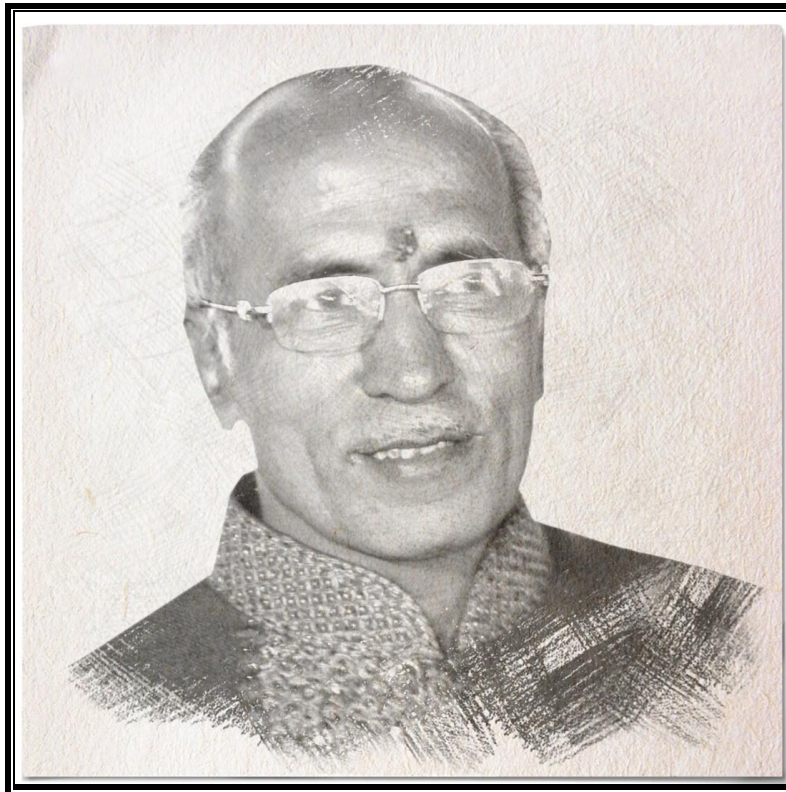
B/5, First floor, Maniratna Complex, Taware Colony,
Aranyeshwar Corner, Pune - 411009.
Maharashtra State, India

PRINCIPAL
Thadomal Shahani Engineering College
Bandra (W), Mumbai - 400 050.
Ph : 91-20-24221234, 91-20-24225678.
Email : info@techknowledgebooks.com,
Website : www.techknowledgebooks.com

Subject Code : CSC504

Book Code : MO175A

*We dedicate this Publication soulfully and wholeheartedly,
in loving memory of our beloved founder director,
Late Shri. Pradeepji Lalchandji Lunawat,
who will always be an inspiration, a positive force and strong support
behind us.*



“My work is my prayer to God”

- Lt. Shri Pradeepji L. Lunawat

Dr. G. T. Thampi
Thadomal Shahani Engineering College
Bandra (W), Mumbai - 400 050.



*Soulful Tribute and Gratitude for all Your
Sacrifices, Hardwork and 40 years of Strong Vision...*

Preface


My Dear Students,

We are extremely happy to come out with this book on “**Data Warehousing and Mining**” for you. The topics within the chapters have been arranged in a proper sequence to ensure smooth flow of the subject.

We present this book in the loving memory of Late Shri. Pradeepji Lunawat, our source of inspiration and a strong foundation of “TechKnowledge Publications”. He will always be remembered in our heart and motivate us to achieve our milestone.

We are thankful to Prof. J. S. Katre, Mr. Shital Bhandari, Prof. Arunoday Kumar and Shri. Chandroday Kumar for the encouragement and support that they have extended. We also thankful to Seema Lunawat for technology enhanced reading, E-books support and the staff members of TechKnowledge Publications for their efforts to make this book as good as it is. We have jointly made every possible efforts to eliminate all the errors in this book. However if you find any, please let me know, because that will help me to improve further.

We are thankful to my family members and friends for patience and encouragement.


Dr. G. T. Thampi
PRINCIPAL
Thadomal Shahani Engineering College
Bandra (W), Mumbai - 400 050.


Authors
□□□

SYLLABUS

Mumbai University Third Year of Computer Engineering (2019 Course)		
Subject Code	Subject Name	Credits
CSC504	Data Warehousing and Mining	03


Prerequisite : Database Concepts

Course Objectives :

1	To identify the significance of Data Warehousing and Mining.
2	To analyze data, choose relevant models and algorithms for respective applications.
3	To study web data mining.
4	To develop research interest towards advances in data mining.

Course Outcomes : At the end of the course, the student will be able to

1	Understand data warehouse fundamentals and design data warehouse with dimensional modelling and apply OLAP operations.
2	Understand data mining principles and perform Data preprocessing and Visualization.
3	Identify appropriate data mining algorithms to solve real world problems.
4	Compare and evaluate different data mining techniques like classification, prediction, clustering and association rule mining.
5.	Describe complex information and social networks with respect to web mining


Dr. C. T. Thampi
 PRINCIPAL
 Thadomal Shahani Engineering College
 Bandra, Mumbai - 400 050



Detailed Syllabus :

Module	Course Module / Contents	Periods
1	<p>Data Warehousing Fundamentals</p> <p>Introduction to Data Warehouse, Data warehouse architecture, Data warehouse versus Data Marts, E-R Modeling versus Dimensional Modeling, Information Package Diagram, Data Warehouse Schemas; Star Schema, Snowflake Schema, Factless Fact Table, Fact Constellation Schema. Update to the dimension tables. Major steps in ETL process, OLTP versus OLAP, OLAP operations: Slice, Dice, Rollup, Drilldown and Pivot.</p> <p style="text-align: right;">(Refer Chapter 1)</p>	08
2	<p>Introduction to Data Mining, Data Exploration and Data Pre-processing</p> <p>Data Mining Task Primitives, Architecture, KDD process, Issues in Data Mining, Applications of Data Mining, Data Exploration: Types of Attributes, Statistical Description of Data, Data Visualization, Data Preprocessing: Descriptive data summarization, Cleaning, Integration & transformation, Data reduction, Data Discretization and Concept hierarchy generation.</p> <p style="text-align: right;">(Refer Chapter 2)</p>	08
3	<p>Classification</p> <p>Basic Concepts, Decision Tree Induction, Naïve Bayesian Classification, Accuracy and Error measures, Evaluating the Accuracy of a Classifier: Holdout & Random Subsampling, Cross Validation, Bootstrap.</p> <p style="text-align: right;">(Refer Chapter 3)</p>	06
4	<p>Clustering</p> <p>Types of data in Cluster analysis, Partitioning Methods (k-Means, k-Medoids), Hierarchical Methods (Agglomerative, Divisive).</p> <p style="text-align: right;">(Refer Chapter 4)</p>	06
5	<p>Mining Frequent Patterns and Associations</p> <p>Market Basket Analysis, Frequent Item sets, Closed Item sets, and Association Rule, Frequent Pattern Mining, Apriori Algorithm, Association Rule Generation, Improving the Efficiency of Apriori, Mining Frequent Itemsets without candidate generation, Introduction to Mining Multilevel Association Rules and Mining Multidimensional Association Rules.</p> <p style="text-align: right;">(Refer Chapter 5)</p>	06
6	<p>Web Mining</p> <p>Introduction, Web Content Mining: Crawlers, Harvest System, Virtual Web View, Personalization, Web Structure Mining: Page Rank, Clever, Web Usage Mining</p> <p style="text-align: right;">(Refer Chapter 6)</p>	05

□□□

- Third Year Diploma - Semester VI
COMPUTER ENGINEERING GROUP
- **MANAGEMENT**
Virat V. Giri, Dr. Yogeshwari L. Giri
- **MOBILE APPLICATION DEVELOPMENT**
Virat V. Giri, Sagar Chavan, Ashwini Mane
- **PROGRAMMING WITH 'PYTHON'**
Ravi Majithia
- **WEB BASED APPLICATION DEVELOPMENT WITH PHP (Elective)**
Vijay T. Patil, Yogita N. Jore, Prasad J. Koyande
- **NETWORK AND INFORMATION SECURITY (Elective)**
Shital M. Mate
- **DATA WAREHOUSING WITH MINING TECHNIQUES (Elective)**
Dr. Arti Deshpande, Dr. Pallavi N. Halarnkar

coming soon.....


 Paper Solutions Trusted by lakhs of students from more than 15 years


 Head Office :
 B15, First Floor, Maniratna Complex, Taware Colony, Aranyeshwar Corner,
 Pune - 411009, Maharashtra, India. Tel. : 91-20-24221234, 91-20-24225678
 ISBN - 978-93-89684-45-2
 Price ₹ 145/-
 MDE57A
 Distribution Address :
 Shop No. 8-3, Ground Floor, Shan Braham Complex, Appa Bahawt Chowk, Opp. Faraskhana Police Station,
 Pune - 411002, Mobile : 74883 91613 / 974088 91912
 Our Branches : Pune | Mumbai | Kolhapur | Nagpur | Solapur | Nashik
 Email : info@techknowledgebooks.com | Website : www.techknowledgebooks.com
 Like us on :      TechKnowledgePublications



Strictly as per new revised 'I' Scheme
w.e.f. academic year 2019-2020

DATA WAREHOUSING WITH MINING TECHNIQUES

(Code : 22621) (Elective)
Semester VI – Computer Engineering Program Group (CO/CM/CW)

Same Subject, Same Authors with New Publication
Dr. Arti Deshpande Dr. Pallavi N. Halarnkar

Includes :
• Model Question Papers as per Bloom's Revised Taxonomy.



ISBN - 978-93-89684-45-2
 Price ₹ 145/-
 MDE57A
 Tech Knowledge Publications

Dr. G. T. Thampi
PRINCIPAL
Thadomal Shahani Engineering College
Bandra (W), Mumbai-400 050.



Data Warehousing and Mining

(Code - CSC603)

Semester VI - Computer Engineering

(Mumbai University)

Strictly as per the Choice Based Credit and Grading System
(Revise 2016) of Mumbai University w.e.f. academic year 2018-2019

Dr. Arti Deshpande

Department of Computer Engineering
Thadomal Shahani Engineering College,
Mumbai.
Maharashtra, India

Dr. Pallavi Halarnkar

Department of Computer Engineering
Thadomal Shahani Engineering College,
Mumbai.
Maharashtra, India.



(Book Code : ME40A)

Dr. G. T. Thampi
PRINCIPAL
Thadomal Shahani Engineering College
Bandra (W), Mumbai - 400 050.

Data Warehousing and Mining

Dr. Arti Deshpande, Dr. Pallavi Halarnkar
(Semester VI - Computer Engineering, Mumbai University)

Copyright © by Authors. All rights reserved. No part of this publication may be reproduced, copied, or stored in a retrieval system, distributed or transmitted in any form or by any means, including photocopy, recording, or other electronic or mechanical methods, without the prior written permission of the publisher.

This book is sold subject to the condition that it shall not, by the way of trade or otherwise, be lent, resold, hired out, or otherwise circulated without the publisher's prior written consent in any form of binding or cover other than which it is published and without a similar condition including this condition being imposed on the subsequent purchaser and without limiting the rights under copyright reserved above.

First Edition : February 2019
Second Revised Edition : January 2020 (**TechKnowledge Publications**)

This edition is for sale in India, Bangladesh, Bhutan, Maldives, Nepal, Pakistan, Sri Lanka and designated countries in South-East Asia. Sale and purchase of this book outside of these countries is unauthorized by the publisher.

Printed at : 37/2, Ashtavinayak Industrial Estate,
Near Pari Company, Narhe, Pune, Maharashtra State, India.
Pune – 411041.


ISBN : 978-93-89748-90-1


Published by

TechKnowledge Publications

Head Office : B/5, First floor, Maniratna Complex, Taware Colony, Aranyeshwar Corner,
Pune - 411 009. Maharashtra State, India
Ph : 91-20-24221234, 91-20-24225678.
Email : info@techknowledgebooks.com,
Website : www.techknowledgebooks.com

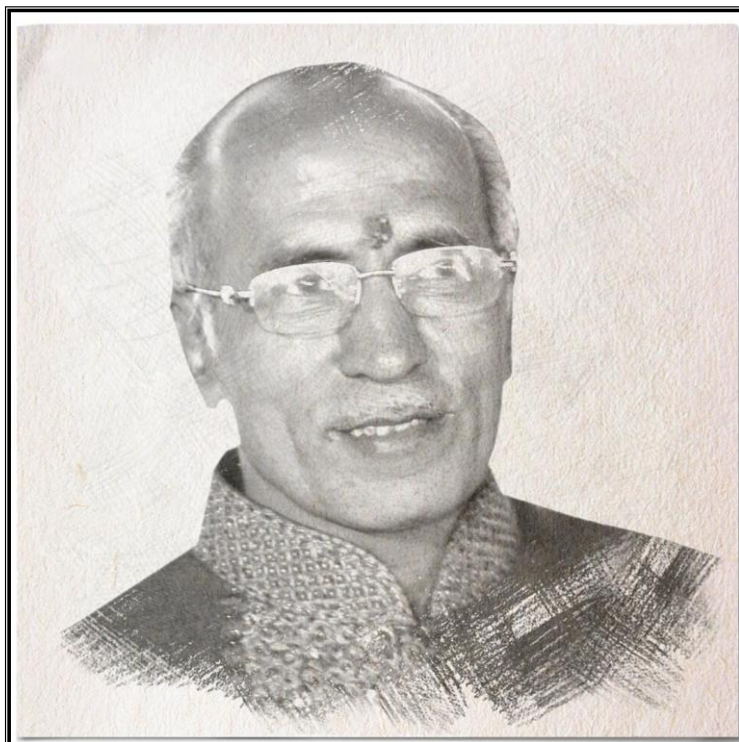
[CSC603] (FID : ME40) (Book Code : ME40A)


Dr. G. T. Thampi
PRINCIPAL
Thadomal Shahani Engineering College
Bandra (W), Mumbai - 400 050.



(Book Code : ME40A)

We dedicate this Publication soulfully and wholeheartedly, in loving memory of our beloved founder director, Late Shri. Pradeepji Lalchandji Lunawat, who will always be an inspiration, a positive force and strong support behind us.



“My work is my prayer to God”

- Lt. Shri. Pradeepji L. Lunawat

Soulful Tribute and Gratitude for all Your Sacrifices, Hardwork, and 40 years of Strong Vision...

Dr. G. T. Titampi
PRINCIPAL
Thadomal Shahani Engineering College
Bandra (W), Mumbai - 400 050.



(Book Code : ME40A)

Preface

Dear Students,

We are extremely happy to present the book of “**Data Warehousing and Mining**” for you. We have divided the subject into small chapters so that the topics can be arranged and understood properly. The topics within the chapters have been arranged in a proper sequence to ensure smooth flow of the subject.

We present this book in the loving memory of **Late. Shri. Pradeepji Lunawat**, our source of inspiration and a strong foundation of “**TechKnowledge Publications**”. He will always be remembered in our hearts and motivate us to achieve our new milestone.

We are thankful to Prof. Arunoday Kumar, Mr. Shital Bhandari and Shri. Chandroday Kumar for the encouragement and support that they have extended. We also thankful to the staff members of TechKnowledge Publications for their efforts to make this book as good as it is. We have made every possible efforts to eliminate all the errors in this book. However if you find any, please let us know, because that will help us to improve the book quality further.

We are thankful to my family members and friends for their patience and encouragement.


Dr. G. I. T. Jadhav
PRINCIPAL
Thadomal Shahani Engineering College
Bandra (W), Mumbai - 400 050.



(Book Code : ME40A)

Syllabus

Course Code	Course Name	Credits
CSC603	Data Warehousing and Mining	4

Course Objectives :

1. To identify the scope and essentiality of Data Warehousing and Mining.
2. To analyze data, choose relevant models and algorithms for respective applications.
3. To study spatial and web data mining.
4. To develop research interest towards advances in data mining.

Course Outcomes : On successful completion of course learner will be able to :

1. Understand Data Warehouse fundamentals, Data Mining Principles.
2. Design data warehouse with dimensional modelling and apply OLAP operations.
3. Identify appropriate data mining algorithms to solve real world problems.
4. Compare and evaluate different data mining techniques like classification, prediction, clustering and association rule mining.
5. Describe complex data types with respect to spatial and web mining.
6. Benefit the user experiences towards research and innovation.

Prerequisite : Basic database concepts, Concepts of algorithm design and analysis.

Module No.	Topics	Hrs.
1.0	Introduction to Data Warehouse and Dimensional modeling : Introduction to Strategic Information, Need for Strategic Information, Features of Data Warehouse, Data warehouses versus Data Marts, Top-down versus Bottom-up approach. Data warehouse architecture, metadata, E-R modelling versus Dimensional Modelling, Information Package Diagram, STAR schema, STAR schema keys, Snowflake Schema, Fact Constellation Schema, Factless Fact tables, Update to the dimension tables, Aggregate fact tables. (Refer chapter 1)	8
2.0	ETL Process and OLAP : Major steps in ETL process, Data extraction : Techniques, Data transformation : Basic tasks, Major transformation types, Data Loading : Applying Data, OLTP Vs OLAP, OLAP definition, Dimensional Analysis, Hypercubes, OLAP operations : Drill down, Roll up, Slice, Dice and Rotation, OLAP models : MOLAP, ROLA (Refer chapter 2)	8

(Book Code : ME40A)

Module No.	Topics	Hrs.
3.0	Introduction to Data Mining, Data Exploration and Preprocessing : Data Mining Task Primitives, Architecture, Techniques, KDD process, Issues in Data Mining, Applications of Data Mining, Data Exploration : Types of Attributes, Statistical Description of Data, Data Visualization, Data Preprocessing : Cleaning, Integration, Reduction : Attribute subset selection, Histograms, Clustering and Sampling, Data Transformation and Data Discretization : Normalization, Binning, Concept hierarchy generation, Concept Description : Attribute oriented Induction for Data Characterization. (Refer chapter 3)	10
4.0	Classification, Prediction and Clustering : Basic Concepts, Decision Tree using Information Gain, Induction : Attribute Selection Measures, Tree pruning, Bayesian Classification : Naive Bayes, Classifier Rule - Based Classification : Using IF-THEN Rules for classification, Prediction : Simple linear regression, Multiple linear regression Model Evaluation and Selection : Accuracy and Error measures, Holdout, Random Sampling, Cross Validation, Bootstrap, Clustering : Distance Measures, Partitioning Methods (k-Means, k-Medoids), Hierarchical Methods (Agglomerative, Divisive). (Refer chapter 4)	12
5.0	Mining Frequent Patterns and Association Rules : Market Basket Analysis, Frequent Item sets, Closed Item sets and Association Rule, Frequent Pattern Mining, Efficient and Scalable Frequent Item set Mining Methods : Apriori Algorithm, Association Rule Generation, Improving the Efficiency of Apriori, FP growth, Mining frequent Itemsets using Vertical Data Format, Introduction to Mining Multilevel Association Rules and Multidimensional Association Rules. (Refer chapter 5)	8
6.0	Spatial and Web Mining : Spatial Data, Spatial Vs. Classical Data Mining, Spatial Data Structures, Mining Spatial Association and Co-location Patterns, Spatial Clustering Techniques : CLARANS Extension, Web Mining : Web Content Mining, Web Structure Mining, Web Usage mining, Applications of Web Mining. (Refer chapter 6)	6
	Total	52



Dr. G. T. Thampi
PRINCIPAL
Thadomal Shahani Engineering College
Bandra (W), Mumbai - 400 050.



(Book Code : ME40A)


Your Success is Our Goal

Semester VI - Computer Engg. / CSE

- **Theory of Computation**
Dilip Kumar Sultana
- **Microprocessor and Interfacing**
Harish G. Narula, Khushboo Shah
- **Cryptography and Network Security**
Pravin Goyal
- **Data Mining (Elective-II)**
Arti Deshpande, Pallavi Halarankar
- **Advance Java Programming (Elective-III)**
Ravi Majethia
- **Image Processing (Elective-II)**
Dhananjay K. Thekedath
- **IOT & Applications (Open Elective-II)**
Pravin Goyal

eBooks are available on www.techknowledgebooks.com

coming soon.....



easy-solutions

Paper Solutions Trusted by lakhs of students from more than 15 years

Head Office :
B/S, First Floor, Mainitna Complex, Taware Colony,
Aayazpeth Corner, Pune - 411009, Maharashtra, India
Tel. : 91-20-24221234, 91-20-24225678


Our Distributors

Student's Agencies (I) Pvt. Ltd. Vidyarthi Sales Agencies
T. : (022) 40496161, 91072 90777 T. : (022) 23847279, 98197 76110

Our Branches : Pune | Mumbai | Kolhapur | Nagpur | Solapur | Nashik


Email : info@techknowledgebooks.com Website : www.techknowledgebooks.com

ISBN : 978-93-90694-65-1




Price ₹ 295/-

GERIA



1300 40 0000

TechKnowledgePublications



Strictly as per the New Revised Syllabus of
Gujarat Technological University
w.e.f. academic year 2020-2021

New Edition
GTU

Data Mining


(Code : 3160714) (Professional Elective - II)


Semester VI - Computer Engineering /
Computer Science and Engineering

Dr. Arti Deshpande
Dr. Pallavi Halarankar

TechKnowledge
Publications

Includes :
Solved Latest University Question Papers


Dr. G. T. Thampi
PRINCIPAL
 Thadomal Shahani Engineering College
 Bandra (W), Mumbai - 400 050.



Data Mining

(Code : 3160714) (Professional Elective-II)

Semester VI - Computer Engineering/Computer Science and Engineering,
(Gujarat Technological University)

Strictly as per the New Revised Syllabus of
Gujarat Technological University w.e.f. academic year 2020-2021

Dr. Arti Deshpande

Assistant Professor,
Department of Computer Engineering
Thadomal Shahani Engineering College,
Mumbai.
Maharashtra, India

Dr. Pallavi N. Halarnkar

Associate Professor,
Department of Computer Engineering
Thadomal Shahani Engineering College,
Mumbai.
Maharashtra, India.



Dr. G. T. Thampi
PRINCIPAL
Thadomal Shahani Engineering College
Bandra (W), Mumbai - 400 050.



Data Mining (3160714)

Dr. Arti Deshpande, Dr. Pallavi N. Halarankar

Semester VI – Computer Engineering/Computer Science and Engineering, (Gujarat Technological University)

Copyright © by Authors. All rights reserved. No part of this publication may be reproduced, copied, or stored in a retrieval system, distributed or transmitted in any form or by any means, including photocopy, recording, or other electronic or mechanical methods, without the prior written permission of the publisher.

This book is sold subject to the condition that it shall not, by the way of trade or otherwise, be lent, resold, hired out, or otherwise circulated without the publisher's prior written consent in any form of binding or cover other than which it is published and without a similar condition including this condition being imposed on the subsequent purchaser and without limiting the rights under copyright reserved above.

First Printed In India : April 2010 (Mumbai University)

First Edition : February 2021

This edition is for sale in India, Bangladesh, Bhutan, Maldives, Nepal, Pakistan, Sri Lanka and designated countries in South-East Asia. Sale and purchase of this book outside of these countries is unauthorized by the publisher.

ISBN : 978-93-90694-05-1

Published By

TechKnowledge Publications

Printed @

37/2, Ashtavinayak Industrial Estate,
Near Pari Company,
Narhe, Pune, Maharashtra State, India.
Pune - 411041

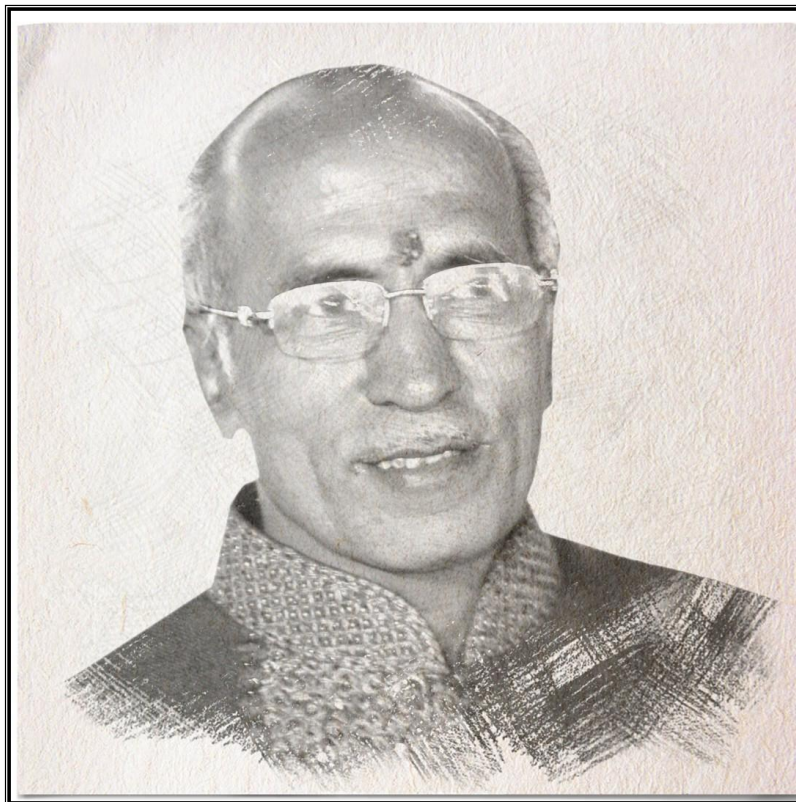
Head Office

B/5, First floor, Manvatna Complex
Taware Colony, Aranyeshwar Corner
Pune - 411 009, Maharashtra State, India
Ph : 91-20-24221234, 91-20-24225678.
Email : info@techknowledgebooks.com,
Website : www.techknowledgebooks.com

Subject Code : 3160714

Book Code : GE91A

*We dedicate this Publication soulfully and wholeheartedly,
in loving memory of our beloved founder director,
Late Shri. Pradeepji Lalchandji Lunawat,
who will always be an inspiration, a positive force and strong support
behind us.*



“My work is my prayer to God”

- Lt. Shri. Pradeepji L. Lunawat

Dr. G. T. Thampi
Thadomal Shamani Engineering College
Bandra (W), Mumbai - 400 050.



*Soulful Tribute and Gratitude for all Your
Sacrifices, Hardwork, and 40 years of Strong Vision...*



Dr. G. T. Thampi
PRINCIPAL
Thadomal Shahani Engineering College
Bandra (W), Mumbai - 400 050.



Preface

My Dear Students,

We are extremely happy to come out with this book on **Data Mining** for you. The topics within the chapters have been arranged in a proper sequence to ensure smooth flow of the subject.

We present this book in the loving memory of **Late Shri. Pradeepji Lunawat**, our source of inspiration and a strong foundation of “**TechKnowledge Publications**”. He will always be remembered in our heart and motivate us to achieve our milestone.

We are thankful to Shri. J. S. Katre, Shri. Shital Bhandari, Shri. Arunoday Kumar and Shri. Chandroday Kumar for the encouragement and support that they have extended. We are also thankful to Seema Lunawat for technology enhanced reading, E-books support and the staff members of TechKnowledge Publications for their efforts to make this book as good as it is. We have jointly made every possible efforts to eliminate all the errors in this book. However if you find any, please let us know, because that will help us to improve further.

We are also thankful to our family members and friends for their patience and encouragement.


Dr. G. T. Thampi
PRINCIPAL
Thadomal Shahani Engineering College
Bandra (W), Mumbai-400 050.


- Authors

SYLLABUS

Gujarat Technological University Sixth Semester of Computer Engineering / Computer Science and Engineering Data Mining (Code : 3160714)

Teaching and Examination Scheme

Teaching scheme			Credits C	Examination Marks				Total Marks
L	T	P		Theory Marks		Practical Marks		
				ESE (E)	PA (M)	ESE (V)	PA (I)	
3	0	2	4	70	30	30	20	150

Sr. No.	Content	Total Hours
1.	Introduction to data mining (DM) : Motivation for Data Mining - Data Mining-Definition and Functionalities – Classification of DM Systems - DM task primitives - Integration of a Data Mining system with a Database or a Data Warehouse - Issues in DM – KDD Process (Refer Chapter 1)	3
2.	Data Pre-processing : Data summarization, data cleaning, data integration and transformation, data reduction, data discretization and concept hierarchy generation, feature extraction, feature transformation, feature selection, introduction to Dimensionality Reduction, CUR decomposition (Refer Chapter 2)	4
3.	Concept Description, Mining Frequent Patterns, Associations and Correlations : What is concept description? - Data Generalization and summarization-based characterization - Attribute relevance - class comparisons, Basic concept, efficient and scalable frequent item-set mining methods, mining various kind of association rules, from association mining to correlation analysis, Advanced Association Rule Techniques, Measuring the Quality of Rules. (Refer Chapter 3)	10
4.	Classification and Prediction : Classification vs. prediction, Issues regarding classification and prediction, Statistical-Based Algorithms, Distance-Based Algorithms, Decision Tree-Based Algorithms, Neural Network-Based Algorithms, Rule-Based Algorithms, Combining Techniques, accuracy and error measures, evaluation of the accuracy of a classifier or predictor. Neural Network Prediction methods: Linear and nonlinear regression, Logistic Regression Introduction of tools such as DB Miner / WEKA / DTREG DM Tools (Refer Chapter 4)	10
5.	Cluster Analysis : 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, Strengths and Weakness; Outlier Detection, Clustering high dimensional data, clustering Graph and Network data. (Refer Chapter 5)	10
8.	Web mining and other Data Mining : Web Mining: Introduction to Web Mining, Web content mining, Web usage mining, Web Structure mining, Web log structure and issues regarding web logs, Spatial Data Mining, Temporal Mining, And Multimedia Mining. Applications of Distributed and parallel Data Mining. (Refer Chapter 6)	5

□□□

- Your Success is Our Goal** 2020 MU I
- Semester VI - Computer Engineering**
- SOFTWARE ENGINEERING
Dr. Pathan Mohd Shafi, Rashmi B. Kale, Arti A. Bhise, Prasad B. Chaudhari, M. A. Ansari
 - SYSTEM PROGRAMMING & COMPLIER CONSTRUCTION
Shiweta A. Loonkar
 - DATA WAREHOUSING & MINING
Dr. Arti Deshpande, Dr. Pallavi Halarnkar
 - CRYPTOGRAPHY & SYSTEM SECURITY
Pravin Goyal
 - ADVANCED DATABASE MANAGEMENT SYSTEM (Department Level Elective)
Mahesh Mali
 - ADVANCED COMPUTER NETWORK (Department Level Elective)
J. S. Katre

coming soon.....

easy-solutions now with **Tech Knowledge Publications**

Paper Solutions Trusted by lakhs of students from more than 15 years

Head Office :
B/S, First Floor, Maniratra Complex, Taware Colony,
Aranyeshwar Corner, Pune - 411009, Maharashtra, India.
Tel. : 91-20-2421234, 91-20-24229578

Distributors

Shree's Agencies (P) Pvt. Ltd. T. : (020) 4460165, 04152 80771	Vijayathi Sales Agency T. : (020) 2380729, 043119 7410	Bhorat Sales Agency T. : (020) 2381258, 040572 32787
---	---	---

Our Branches : Pune | Mumbai | Kolhapur | Nagpur | Solapur | Nashik

For Library Orders Contact - Ved Book Distributors M : 80975 71421 / 92208 77214

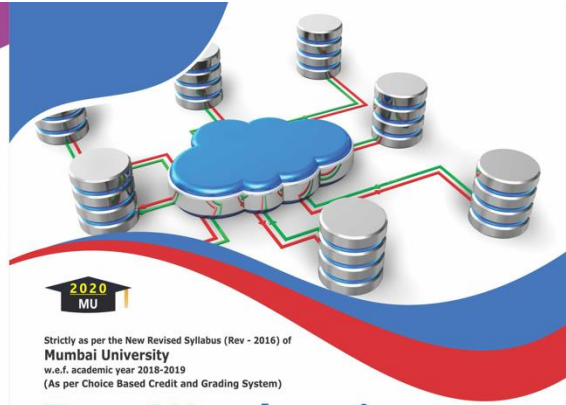
Email : info@techknowledgebooks.com Website : www.techknowledgebooks.com

ISBN : 978-93-93748-80-1
Price ₹ 325/-
ME40A

Recycled Paper with Recycled Ink

Tech Knowledge Publications

2020 MU I
DATA WAREHOUSING AND MINING
Semester VI - Computer Engineering
Dr. Arti Deshpande
Dr. Pallavi Halarnkar
Book Code ME40A
Price ₹ 325/-
B-15



2020 MU I

Strictly as per the New Revised Syllabus (Rev - 2016) of
Mumbai University
w.e.f. academic year 2018-2019
(As per Choice Based Credit and Grading System)

Data Warehousing and Mining

Semester VI - Computer Engineering

Same Subject, Same Authors with New Publication

Dr. Arti Deshpande Dr. Pallavi Halarnkar

With Solved Latest University Question Paper of May 2019.



Handwritten signature

Dr. G. T. Thampi
PRINCIPAL
Thadomal Shahani Engineering College
Bandra (W), Mumbai - 400 050.



Data Warehousing and Mining

(Code - CSC603)

Semester VI - Computer Engineering

(Mumbai University)

Strictly as per the Choice Based Credit and Grading System
(Revise 2016) of Mumbai University w.e.f. academic year 2018-2019

Dr. Arti Deshpande

Department of Computer Engineering
Thadomal Shahani Engineering College,
Mumbai.
Maharashtra, India

Dr. Pallavi Halarnkar

Department of Computer Engineering
Thadomal Shahani Engineering College,
Mumbai.
Maharashtra, India.



Dr. G. T. Thampi
PRINCIPAL
Thadomal Shahani Engineering College
Bandra (W), Mumbai - 400 050.



 **TechKnowledge**TM
Publications

ME40A Price ₹ 325/-



Thad

(Book Code : ME40A)

Data Warehousing and Mining

Dr. Arti Deshpande, Dr. Pallavi Halarnkar
(Semester VI - Computer Engineering, Mumbai University)

Copyright © by Authors. All rights reserved. No part of this publication may be reproduced, copied, or stored in a retrieval system, distributed or transmitted in any form or by any means, including photocopy, recording, or other electronic or mechanical methods, without the prior written permission of the publisher.

This book is sold subject to the condition that it shall not, by the way of trade or otherwise, be lent, resold, hired out, or otherwise circulated without the publisher's prior written consent in any form of binding or cover other than which it is published and without a similar condition including this condition being imposed on the subsequent purchaser and without limiting the rights under copyright reserved above.

First Edition : February 2019
Second Revised Edition : January 2020 (**TechKnowledge Publications**)

This edition is for sale in India, Bangladesh, Bhutan, Maldives, Nepal, Pakistan, Sri Lanka and designated countries in South-East Asia. Sale and purchase of this book outside of these countries is unauthorized by the publisher.

Printed at : 37/2, Ashtavinayak Industrial Estate,
Near Pari Company, Narhe, Pune, Maharashtra State, India.
Pune – 411041.


ISBN : 978-93-89748-90-1


Published by

TechKnowledge Publications

Head Office : B/5, First floor, Maniratna Complex, Taware Colony, Aranyeshwar Corner,
Pune - 411 009. Maharashtra State, India
Ph : 91-20-24221234, 91-20-24225678.
Email : info@techknowledgebooks.com,
Website : www.techknowledgebooks.com

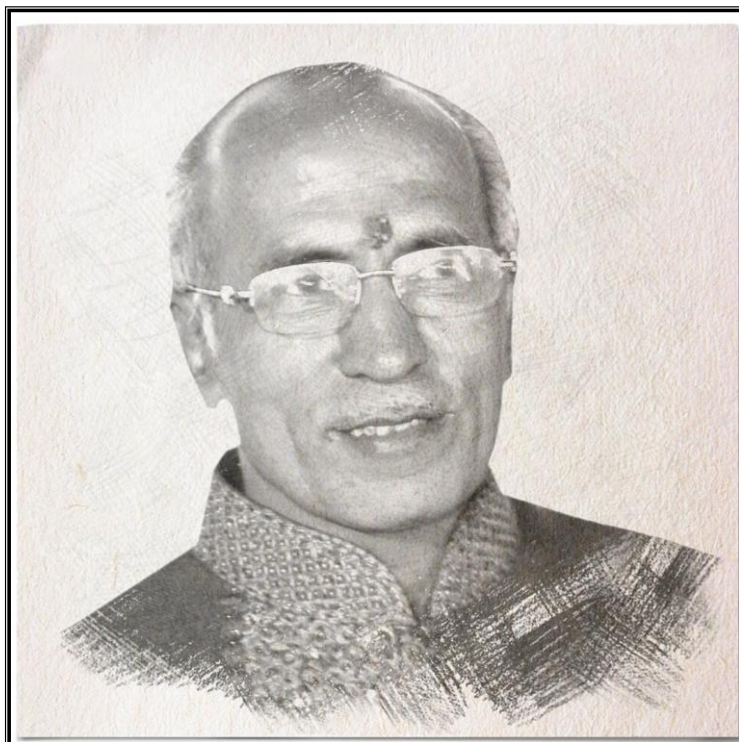
[CSC603] (FID : ME40) (Book Code : ME40A)


Dr. G. T. Thampi
PRINCIPAL
Thadomal Shahani Engineering College
Bandra (W), Mumbai - 400 050.



(Book Code : ME40A)

We dedicate this Publication soulfully and wholeheartedly, in loving memory of our beloved founder director, Late Shri. Pradeepji Lalchandji Lunawat, who will always be an inspiration, a positive force and strong support behind us.



“My work is my prayer to God”

- Lt. Shri. Pradeepji L. Lunawat

Soulful Tribute and Gratitude for all Your Sacrifices, Hardwork, and 40 years of Strong Vision...

Dr. G. T. Titampi
PRINCIPAL
Thadomal Shahani Engineering College
Bandra (W), Mumbai - 400 050.



(Book Code : ME40A)

Preface

Dear Students,

We are extremely happy to present the book of “**Data Warehousing and Mining**” for you. We have divided the subject into small chapters so that the topics can be arranged and understood properly. The topics within the chapters have been arranged in a proper sequence to ensure smooth flow of the subject.

We present this book in the loving memory of **Late. Shri. Pradeepji Lunawat**, our source of inspiration and a strong foundation of “**TechKnowledge Publications**”. He will always be remembered in our hearts and motivate us to achieve our new milestone.

We are thankful to Prof. Arunoday Kumar, Mr. Shital Bhandari and Shri. Chandroday Kumar for the encouragement and support that they have extended. We also thankful to the staff members of TechKnowledge Publications for their efforts to make this book as good as it is. We have made every possible efforts to eliminate all the errors in this book. However if you find any, please let us know, because that will help us to improve the book quality further.

We are thankful to my family members and friends for their patience and encouragement.


Dr. G. I. T. Jadhav
PRINCIPAL
Thadomal Shahani Engineering College
Bandra (W), Mumbai - 400 050.



(Book Code : ME40A)

Syllabus

Course Code	Course Name	Credits
CSC603	Data Warehousing and Mining	4

Course Objectives :

1. To identify the scope and essentiality of Data Warehousing and Mining.
2. To analyze data, choose relevant models and algorithms for respective applications.
3. To study spatial and web data mining.
4. To develop research interest towards advances in data mining.

Course Outcomes : On successful completion of course learner will be able to :

1. Understand Data Warehouse fundamentals, Data Mining Principles.
2. Design data warehouse with dimensional modelling and apply OLAP operations.
3. Identify appropriate data mining algorithms to solve real world problems.
4. Compare and evaluate different data mining techniques like classification, prediction, clustering and association rule mining.
5. Describe complex data types with respect to spatial and web mining.
6. Benefit the user experiences towards research and innovation.

Prerequisite : Basic database concepts, Concepts of algorithm design and analysis.

Module No.	Topics	Hrs.
1.0	Introduction to Data Warehouse and Dimensional modeling : Introduction to Strategic Information, Need for Strategic Information, Features of Data Warehouse, Data warehouses versus Data Marts, Top-down versus Bottom-up approach. Data warehouse architecture, metadata, E-R modelling versus Dimensional Modelling, Information Package Diagram, STAR schema, STAR schema keys, Snowflake Schema, Fact Constellation Schema, Factless Fact tables, Update to the dimension tables, Aggregate fact tables. (Refer chapter 1)	8
2.0	ETL Process and OLAP : Major steps in ETL process, Data extraction : Techniques, Data transformation : Basic tasks, Major transformation types, Data Loading : Applying Data, OLTP Vs OLAP, OLAP definition, Dimensional Analysis, Hypercubes, OLAP operations : Drill down, Roll up, Slice, Dice and Rotation, OLAP models : MOLAP, ROLAP. (Refer chapter 2)	8

(Book Code : ME40A)

Module No.	Topics	Hrs.
3.0	Introduction to Data Mining, Data Exploration and Preprocessing : Data Mining Task Primitives, Architecture, Techniques, KDD process, Issues in Data Mining, Applications of Data Mining, Data Exploration : Types of Attributes, Statistical Description of Data, Data Visualization, Data Preprocessing : Cleaning, Integration, Reduction : Attribute subset selection, Histograms, Clustering and Sampling, Data Transformation and Data Discretization : Normalization, Binning, Concept hierarchy generation, Concept Description : Attribute oriented Induction for Data Characterization. (Refer chapter 3)	10
4.0	Classification, Prediction and Clustering : Basic Concepts, Decision Tree using Information Gain, Induction : Attribute Selection Measures, Tree pruning, Bayesian Classification : Naive Bayes, Classifier Rule - Based Classification : Using IF-THEN Rules for classification, Prediction : Simple linear regression, Multiple linear regression Model Evaluation and Selection : Accuracy and Error measures, Holdout, Random Sampling, Cross Validation, Bootstrap, Clustering : Distance Measures, Partitioning Methods (k-Means, k-Medoids), Hierarchical Methods (Agglomerative, Divisive). (Refer chapter 4)	12
5.0	Mining Frequent Patterns and Association Rules : Market Basket Analysis, Frequent Item sets, Closed Item sets and Association Rule, Frequent Pattern Mining, Efficient and Scalable Frequent Item set Mining Methods : Apriori Algorithm, Association Rule Generation, Improving the Efficiency of Apriori, FP growth, Mining frequent Itemsets using Vertical Data Format, Introduction to Mining Multilevel Association Rules and Multidimensional Association Rules. (Refer chapter 5)	8
6.0	Spatial and Web Mining : Spatial Data, Spatial Vs. Classical Data Mining, Spatial Data Structures, Mining Spatial Association and Co-location Patterns, Spatial Clustering Techniques : CLARANS Extension, Web Mining : Web Content Mining, Web Structure Mining, Web Usage mining, Applications of Web Mining. (Refer chapter 6)	6
	Total	52



Dr. G. T. Thampi
PRINCIPAL
Thadomal Shahani Engineering College
Bandra (W), Mumbai - 400 050.



(Book Code : ME40A)

Your Success is Our Goal

Strictly as per the new Credit System Syllabus (2015 Course)
Savitribai Phule Pune University
 w.e.f. academic year 2018-2019

Semester VII - Computer Engineering

coming soon.....
es easy-solutions now with **Tech Knowledge Publications**

Head Office :
 B/S, First Floor, Maniratra Complex, Tawara Colony, Aranyeshwar Corner,
 Pune - 411009, Maharashtra, India. Tel. : 91-20-24221234, 91-20-24225678

Distribution Address :
 Shop No. 8-9, Ground Floor, Shaan Bramba Complex, Appa Baiwani Chowk, Opp. Farakhana Police Station,
 Pune - 411002. Mobile : 74609 91813 / 74609 91812.

Our Branches : Pune | Mumbai | Kolhapur | Nagpur | Solapur | Nashik

Email : info@techknowledgebooks.com | Website : www.techknowledgebooks.com

ISBN : 978-81-93209-36-6
 Price : 175/-
 9788193209366
 9078A

Dr. Arti Deshpande
 Dr. Pallavi N. Halarnkar

Book Code: P078A
 Price : 175/-

DATA MINING AND WAREHOUSING

Includes
 ⇒ Chapterwise Solved SPPU Question Papers Upto Dec. 2018.

Edition 2019

SPPU

Strictly as per the new Credit System Syllabus (2015 Course)
Savitribai Phule Pune University
 w.e.f. academic year 2018-2019

Data Mining and Warehousing

(Elective 1)
 Semester VII - Computer Engineering

Same Subject, Same Author with New Publication
Dr. Arti Deshpande
Dr. Pallavi N. Halarnkar

Tech Knowledge Publications

Dr. G. T. Thampi
 PRINCIPAL
 Thadomal Shahani Engineering College
 Bandra (W), Mumbai-400 050.



Data Mining and Warehousing

Elective I

(Code : 410244(D))

Semester VII - Computer Engineering

(Savitribai Phule Pune University)

**Strictly as per the New Credit System Syllabus (2015 Course)
Savitribai Phule Pune University w.e.f. academic year 2018-2019**

Dr. Arti Deshpande

Assistant Professor,
Department of Computer Engineering
Thadomal Shahani Engineering College, Mumbai.
Maharashtra, India.

Dr. Pallavi N. Halarnkar

Associate Professor,
Department of Computer Engineering
Thadomal Shahani Engineering College, Mumbai.
Maharashtra, India.



Dr. G. T. Thampi
PRINCIPAL
Thadomal Shahani Engineering College
Bandra (W), Mumbai - 400 050.



 **TechKnowledge**TM
Publications

P078A Price ₹ 175/-



Data Mining and Warehousing

Dr. Arti Deshpande, Dr. Pallavi N. Halarnkar
(Semester VII - Computer Engineering) (Savitribai Phule Pune University)

Copyright © by Authors. All rights reserved. No part of this publication may be reproduced, copied, or stored in a retrieval system, distributed or transmitted in any form or by any means, including photocopy, recording, or other electronic or mechanical methods, without the prior written permission of the publisher.

This book is sold subject to the condition that it shall not, by the way of trade or otherwise, be lent, resold, hired out, or otherwise circulated without the publisher's prior written consent in any form of binding or cover other than which it is published and without a similar condition including this condition being imposed on the subsequent purchaser and without limiting the rights under copyright reserved above.

First Edition : July 2018
Second Revised Edition : July 2019 (TechKnowledge Publications)

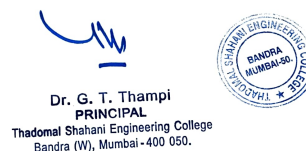
This edition is for sale in India, Bangladesh, Bhutan, Maldives, Nepal, Pakistan, Sri Lanka and designated countries in South-East Asia. Sale and purchase of this book outside of these countries is unauthorized by the publisher.

Printed at :: 37/2, Ashtvinayak Industrial Estate, Near Pari Company,
Narhe, Pune, Maharashtra State India, Pune - 411041

ISBN 978-93-89299-36-6

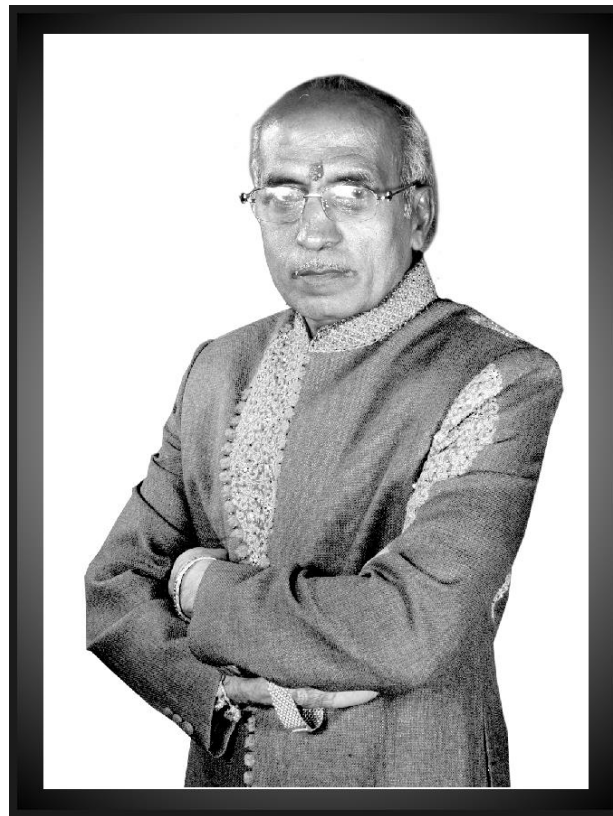
Published by
TechKnowledge Publications

Head Office : B/5, First floor, Maniratna Complex,
Taware Colony, Aranyeshwar Corner,
Pune - 411 009. Maharashtra State,
India. Ph : 91-20-24221234, 91-20-24225678.



[410244(D)] (FID : PO78) (Book Code : PO78A)

*We dedicate this Publication soulfully and wholeheartedly,
in loving memory of our beloved founder director
Late. Shri. Pradeepsheth Lalchandji Lunawat,
who will always be an inspiration, a positive force and strong support
behind us.*



Lt. Shri. Pradeepji L. Lunawat



Dr. G. T. Thampi
PRINCIPAL
Thadomal Shamji Engineering College
Bandra, Mumbai



*Soulful Tribute and Gratitude for all Your
Sacrifices, Hardwork and 40 years of Strong Vision.....*

Preface

Dear Students,

We are extremely happy to present the book of “**Data Mining and Warehousing**” for you. We have divided the subject into small chapters so that the topics can be arranged and understood properly. The topics within the chapters have been arranged in a proper sequence to ensure smooth flow of the subject.

We present this book in the loving memory of **Late. Shri. Pradeepji Lunawat**, our source of inspiration and a strong foundation of “**TechKnowledge Publications**”. He will always be remembered in our heart and motivate us to achieve our milestone.

We are thankful to Shri. J. S. Katre, Shri. Shital Bhandari, Shri. Arunoday Kumar and Shri. Chandroday Kumar for the encouragement and support that they have extended. We are also thankful to the staff members of TechKnowledge Publications and others for their efforts to make this book as good as it is.

We have jointly made every possible effort to eliminate all the errors in this book. However if you find any, please let us know, because that will help us to improve further.


Dr. Arti Deshpande

Dr. Pallavi Halarnkar
Dr. G. T. Deshpande
PRINCIPAL
Thadomal Shahani Engineering College
Bandra (W), Mumbai - 400 050.

□□□

Syllabus

Savitribai Phule Pune University
Fourth Year of Computer Engineering (2015 Course)
Elective I

410244(D) : Data Mining and Warehousing

Teaching Scheme : TH : 03 Hours/Week	Credit 03	Examination Scheme : In-Sem (Paper) : 30 Marks End-Sem (Paper) : 70 Marks
--	---------------------	--

Pre-requisites Courses

310242-Database Management Systems, 310244 - Information Systems and Engineering Economics

Companion Course : 410247- Laboratory Practice II

Course Objectives

- To understand the fundamentals of Data Mining.
- To identify the appropriateness and need of mining the data.
- To learn the preprocessing, mining and post processing of the data.
- To understand various methods, techniques and algorithms in data mining.

Course Outcomes

On completion of the course the student should be able to :

- Apply basic, intermediate and advanced techniques to mine the data.
- Analyze the output generated by the process of data mining.
- Explore the hidden patterns in the data.
- Optimize the mining process by choosing best data mining technique.

Course Contents

Unit I : Introduction

(08 Hours)

Data Mining, Data Mining Task Primitives, Data : Data, Information and Knowledge; Attribute Types : Nominal, Binary, Ordinal and Numeric attributes, Discrete versus Continuous Attributes; Introduction to Data Preprocessing, Data Cleaning : Missing values, Noisy data; Data integration : Correlation analysis; transformation : Min-max normalization, z-score normalization and decimal scaling; data reduction : Data Cube Aggregation, Attribute Subset Selection, sampling; and Data Discretization : Binning, Histogram Analysis

Dr. G. T. Thampi
Thadomal Shahani Engineering College
Bandra (W), Mumbai - 400 050.

(Refer chapter 1)

Unit II : Data Warehouse

(08 Hours)

Data Warehouse, Operational Database Systems and Data Warehouses (OLTP Vs OLAP), A Multidimensional Data Model: Data Cubes, Stars, Snowflakes, and Fact Constellations Schemas; OLAP Operations in the Multidimensional Data Model, Concept Hierarchies, Data Warehouse Architecture, The Process of Data Warehouse Design, A three-tier data warehousing architecture, Types of OLAP Servers : ROLAP versus MOLAP versus HOLAP. **(Refer chapter 2)**

Unit III : Measuring Data Similarity and Dissimilarity**(08 Hours)**

Measuring Data Similarity and Dissimilarity, Proximity Measures for Nominal Attributes and Binary Attributes, interval scaled; Dissimilarity of Numeric Data : Minkowski Distance, Euclidean distance and Manhattan distance; Proximity Measures for Categorical, Ordinal Attributes, Ratio scaled variables; Dissimilarity for Attributes of Mixed Types, Cosine Similarity. **(Refer chapter 3)**

Unit IV : Association Rules Mining**(08 Hours)**

Market basket Analysis, Frequent item set, Closed item set, Association Rules, a-priori Algorithm, Generating Association Rules from Frequent Item sets, Improving the Efficiency of a-priori, Mining Frequent Item sets without Candidate Generation : FP Growth Algorithm; Mining Various Kinds of Association Rules : Mining multilevel association rules, constraint based association rule mining, Meta rule-Guided Mining of Association Rules. **(Refer chapter 4)**

Unit V : Classification**(08 Hours)**

Introduction to : Classification and Regression for Predictive Analysis, Decision Tree Induction, Rule-Based Classification : using IF-THEN Rules for Classification, Rule Induction Using a Sequential Covering Algorithm. Bayesian Belief Networks, Training Bayesian Belief Networks, Classification Using Frequent Patterns, Associative Classification, Lazy Learners-k-Nearest-Neighbor Classifiers, Case-Based Reasoning. **(Refer chapter 5)**

Unit VI : Multiclass Classification**(08 Hours)**

Multiclass Classification, Semi-Supervised Classification, Reinforcement learning, Systematic Learning, Wholistic learning and multi-perspective learning. Metrics for Evaluating Classifier Performance : Accuracy, Error Rate, precision, Recall, Sensitivity, Specificity; Evaluating the Accuracy of a Classifier : Holdout Method, Random Sub sampling and Cross-Validation. **(Refer chapter 6)**

□□□



Dr. G. T. Thampi
PRINCIPAL
Thadomal Shahani Engineering College
Bandra (W), Mumbai - 400 050.



"Your Success is Our Goal"

es
easy-solutions

now with

TechKnowledge
Publications

Study, Learn & Excel in Examinations

Includes:

- University Paper Solved by Eminent Faculty Members.
- Unit-wise distribution of solved questions.
- Statistical Analysis of each unit.
- Up-to-date and Quality contents.
- In-line with the books of TechKnowledge Publications.

<p>Head Office : B/5, First Floor, Maniratra Complex, Taware Colony, Aranyeshwar Corner, Pune - 411009, Maharashtra, India. Tel. : 91-20-24221234, 91-20-24225678</p> <p>Our Distributors</p> <table border="0"> <tr> <td style="width: 50%;"> <p>Student's Agencies (I) Pvt. Ltd. T. : (022) 40408101, 919172 90777</p> </td> <td style="width: 50%;"> <p>Vidyarthi Sales Agencies T. : (022) 23887279, 918197 78110</p> </td> </tr> </table> <p>Our Branches : Pune Mumbai Kolhapur Nagpur Solapur Nashik</p> <p>Email : info@techknowledgebooks.com Website : www.techknowledgebooks.com</p>	<p>Student's Agencies (I) Pvt. Ltd. T. : (022) 40408101, 919172 90777</p>	<p>Vidyarthi Sales Agencies T. : (022) 23887279, 918197 78110</p>	<p>ISBN : 978-93-89503-32-6</p>  <p>Price ₹ 175/-</p> <p>GO46A</p>  <p>Use us on</p>  <p>TechKnowledgePublications</p>
<p>Student's Agencies (I) Pvt. Ltd. T. : (022) 40408101, 919172 90777</p>	<p>Vidyarthi Sales Agencies T. : (022) 23887279, 918197 78110</p>		

DATA MINING AND BUSINESS INTELLIGENCE

Dr. Arti Deshpande, Dr. Pallavi N. Halarnkar

Tech Knowledge Publications

B-20



Strictly as per the New Revised syllabus of Gujarat Technological University w. e. f. academic year 2016-2017

Data Mining and Business Intelligence

(Departmental Elective II)

Semester VII - Information Technology / Computer Engineering & Computer Science Engineering

Same Subject, Same Authors with New Publication

Arti Deshpande **Dr. Pallavi N. Halarnkar**

Chapterwise Solved University Question Papers Upto Dec. 2018.



Dr. G. T. Thampi
PRINCIPAL
Thadomal Shahani Engineering College
Bandra (W), Mumbai - 400 050.



Data Mining and Business Intelligence

(Code - 2170715)

Semester VII - Information Technology / Computer Engineering & Computer Science Engineering (Departmental Elective-II)
(Gujarat Technological University)

Strictly as per the New Revised Syllabus of
Gujarat Technological University w.e.f. academic year 2016-2017

Mrs. Arti Deshpande

ME (Comp. Engg.)
Thadomal Shahani Engineering College , Mumbai.

Dr. Pallavi N. Halarnkar

ME (Comp. Engg.)
Mukesh Patel School of Technology, Management and Engineering, Mumbai.



Dr. G. T. Thampi
PRINCIPAL
Thadomal Shahani Engineering College
Bandra (W), Mumbai - 400 050.



G046A Price ₹ 175/-



Data Mining and Business Intelligence

Mrs. Arti Deshpande, Dr. Pallavi N. Halarnkar

(Semester VII - Information Technology/ Computer Engineering &
Computer Science Engineering : Departmental Elective-II, GTU)

Copyright © by Authors. All rights reserved. No part of this publication may be reproduced, copied, or stored in a retrieval system, distributed or transmitted in any form or by any means, including photocopy, recording, or other electronic or mechanical methods, without the prior written permission of the publisher.

This book is sold subject to the condition that it shall not, by the way of trade or otherwise, be lent, resold, hired out, or otherwise circulated without the publisher's prior written consent in any form of binding or cover other than which it is published and without a similar condition including this condition being imposed on the subsequent purchaser and without limiting the rights under copyright reserved above.

- First Printed In India** : April 2010 (Mumbai University)
First Edition : July 2016
Second Revised Edition : June 2017
Third Revised Edition : June 2018
Fourth Revised Edition : October 2019 (TechKnowledge Publications)

This edition is for sale in India, Bangladesh, Bhutan, Maldives, Nepal, Pakistan, Sri Lanka and designated countries in South-East Asia. Sale and purchase of this book outside of these countries is unauthorized by the publisher.

Printed at : 37/2, Ashtvinayak Industrial Estate, Near Pari Company,
Narhe, Pune, Maharashtra State India,
Pune – 411041

ISBN 978-93-89503-32-6

Published by
TechKnowledge Publications

Head Office : B/5, First floor, Maniratna Complex, Taware Colony, Aranyeshwar
Pune - 411 009. Maharashtra State, India
Ph : 91-20-24221234, 91-20-24225678.



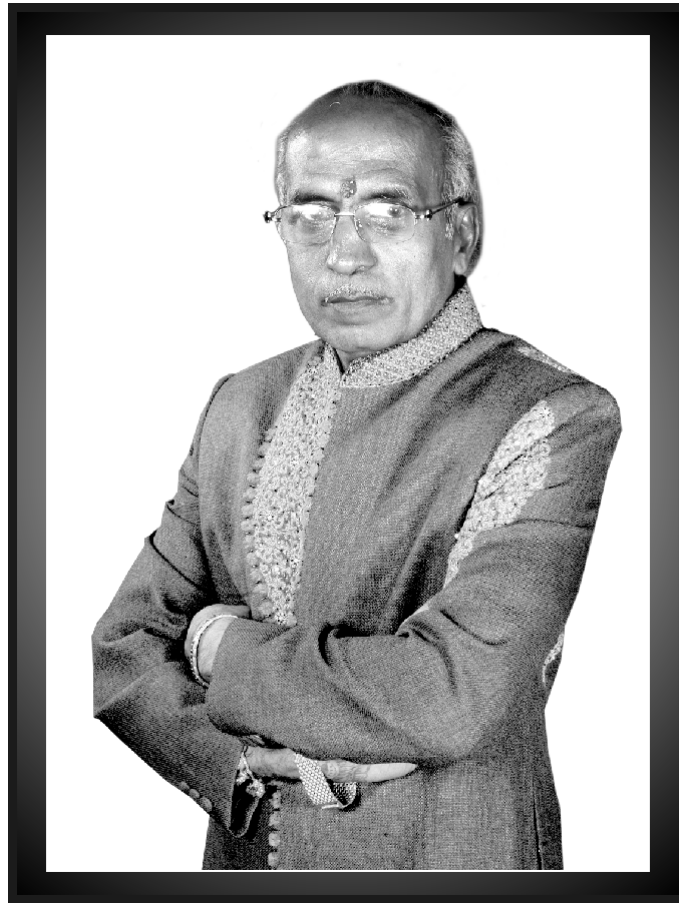
Dr. G. T. Thampi
PRINCIPAL
Thadomal Shahani Engineering College
Bandra (W), Mumbai - 400 050.



[2K16GTUCOMP29] (FID : GO46) (Book Code : GO46A)

(Book Code : GO46A)

*We dedicate this Publication soulfully and wholeheartedly,
in loving memory of our beloved founder director,
Late Shri. Pradeepsheth Lalchandji Lunawat,
who will always be an inspiration, a positive force and strong support
behind us.*



Lt. Shri. Pradeepji L. Lunawat

Dr. G. T. Thampi
PRINCIPAL
Thadomal Shahani Engineering College
Bandra (W), Mumbai - 400 050.



*Soulful Tribute and Gratitude for all Your
Sacrifices, Hardwork and 40 years of Strong Vision...*

Preface

My Dear Students,

We are extremely happy to come out with this book on **“Data Mining & Business Intelligence”** for you. The topics within the chapters have been arranged in a proper sequence to ensure smooth flow of the subject.

We present this book in the loving memory of **Late Shri. Pradeepji Lunawat**, our source of inspiration and a strong foundation of **“TechKnowledge Publications”**. He will always be remembered in our heart and motivate us to achieve our milestone.

We are thankful to Shri. J. S. Katre, Mr. Shital Bhandari, Shri. Arunoday Kumar and Shri. Chandroday Kumar for the encouragement and support that they have extended. We are also thankful to the staff members of TechKnowledge Publications and others for their efforts to make this book as good as it is. We have jointly made every possible efforts to eliminate all the errors in this book. However if you find any, please let us know, because that will help us to improve further.

We are also thankful to my family members and friends for patience and encouragement.


Dr. G. T. Thadomal
PRINCIPAL
Thadomal Shahani Engineering College
Bandra (W), Mumbai - 400 058

Arti Deshpande
Pallavi N. Halankar



Syllabus

Unit 1 : Overview and concepts Data Warehousing and Business Intelligence :

Why reporting and Analysing data, Raw data to valuable information-Lifecycle of Data - What is Business Intelligence - BI and DW in today's perspective - What is data warehousing - The building Blocks : Defining Features - Data warehouses and data Imarts - Overview of the components - Metadata in the data warehouse - Need for data warehousing - Basic elements of data warehousing - trends in data warehousing. **(Refer Chapter 1)**

Unit 2 : The Architecture of BI and DW :

BI and DW architectures and its types - Relation between BI and DW - OLAP (Online analytical processing) definitions - Difference between OLAP and OLTP - Dimensional analysis - What are cubes? Drill-down and roll-up - slice and dice or rotation - OLAP models - ROLAP versus MOLAP - defining schemas : Stars, snowflakes and fact constellations. **(Refer Chapter 2)**

Unit 3 : Introduction to data mining (DM) :

Motivation for Data Mining - Data Mining-Definition and Functionalities – Classification of DM Systems - DM task primitives - Integration of a Data Mining system with a Database or a Data Warehouse - Issues in DM – KDD Process. **(Refer Chapter 3)**

Unit 4 : Data Pre-processing :

Why to pre-process data? - Data cleaning: Missing Values, Noisy Data - Data Integration and transformation - Data Reduction : Data cube aggregation, Dimensionality reduction - Data Compression - Numerosity Reduction - Data Mining Primitives - Languages and System Architectures : Task relevant data - Kind of Knowledge to be mined - Discretization and Concept Hierarchy. **(Refer Chapter 4)**

Unit 5 : Concept Description and Association Rule Mining :

What is concept description? - Data Generalization and summarization-based characterization - Attribute relevance - class comparisons Association Rule Mining: Market basket analysis – basic concepts - Finding frequent item sets: Apriori algorithm - generating rules – Improved Apriori algorithm – Incremental ARM – Associative Classification – Rule Mining. **(Refer Chapter 5)**

Unit 6 : Classification and Prediction :

What is classification and prediction? – Issues regarding Classification and prediction :

Classification methods : Decision tree, Bayesian Classification, Rule based, **ART Neural Network**

Prediction methods : Linear and nonlinear regression, Logistic Regression

Introduction of tools such as DB Miner /WEKA/DTREG DM Tools.

(Refer Chapter 6)



Dr. T. Ramoji
Principal
Thadomal Shahani Engineering College
Bandra (W), Mumbai - 400 050.

Unit 7 : Data Mining for Business Intelligence Applications :

Data mining for business Applications like Balanced Scorecard, Fraud Detection, Click stream Mining, Market Segmentation, retail industry, telecommunications industry, banking & finance and CRM etc.

Data Analytics Life Cycle : Introduction to Big data Business Analytics - State of the practice in analytics role of data scientists

Key roles for successful analytic project : Main phases of life cycle - Developing core deliverables for stakeholders. **(Refer Chapter 7)**

Unit 8 : Advance topics :

Introduction and basic concepts of following topics.

Clustering, Spatial mining, web mining, text mining,

Big Data : Introduction to big data : distributed file system – Big Data and its importance, Four Vs, Drivers for Big data, Big data analytics, Big data applications. Algorithms using map reduce, Matrix-Vector Multiplication by Map Reduce. Introduction to Hadoop architecture: Hadoop Architecture, Hadoop Storage: HDFS, Common Hadoop Shell commands , Anatomy of File Write and Read., NameNode, Secondary NameNode, and DataNode, Hadoop MapReduce paradigm, Map and Reduce tasks, Job, Task trackers – Cluster Setup – SSH & Hadoop Configuration – HDFS Administering – Monitoring & Maintenance. **(Refer Chapter 8)**

□□□



Dr. G. T. Thampi
PRINCIPAL
Thadomal Shahani Engineering College
Bandra (W), Mumbai - 400 050.



- **Your Success is Our Goal**
- **Semester VI - Information Technology**
- **SOFTWARE ENGINEERING WITH PROJECT MANAGEMENT**
M. A. Ansari
- **DATA MINING AND BUSINESS INTELLIGENCE**
Dr. Arti Deshpande, Dr. Pallavi Halarnkar
- **CLOUD COMPUTING & SERVICES**
Pravin Goyal
- **WIRELESS NETWORKS**
J. S. Katre
- **DIGITAL FORENSICS** (Department Level Elective)
Suvama Shirke-Pansambal
- **MULTIMEDIA SYSTEMS** (Department Level Elective)
Tushiram Sule

coming soon.....

es easy-solutions **now with** **TechKnowledge** Publications

Paper Solutions Trusted by lakhs of students from more than 15 years

Head Office :
B/S, First Floor, Maniratra Complex, Taware Colony,
Aranyeshwar Corner, Pune - 411009, Maharashtra, India.
Tel : 91-20-24221234, 91-20-24221879

Distributors
Student's Agencies (I) Pvt. Ltd., Vidyanth Sales Agencies, Bharat Sales Agency
T : (022) 40486161, (0)91872 80777 T : (022) 23872729, (0)98197 78110 T : (022) 23810309, (0)98072 82797

Our Branches : Pune | Mumbai | Kolhapur | Nagpur | Solapur | Nashik

For Library Orders Contact - Ved Book Distributors M : 90975 71421 / 92208 77214

Email : info@techknowledgebooks.com Website : www.techknowledgebooks.com

ISSN : 978-93-88628-77-1
Price ₹ 245/-
ME59A

2020 MU
DATA MINING AND BUSINESS INTELLIGENCE
Semester VI - Information Technology

Book Code: ME59A
Price: ₹245/-



2020 MU

Data Mining and Business Intelligence

Semester VI - Information Technology

Same Subject, Same Authors with New Publication

Dr. Arti Deshpande Dr. Pallavi Halarnkar

With Solved Latest University Question Paper of May 2019.



Strictly as per the New Revised Syllabus (Rev - 2016) of
Mumbai University
w.e.f. academic year 2018-2019
(As per Choice Based Credit and Grading System)

Dr. G. T. Thampi
PRINCIPAL
Thadomal Shahani Engineering College
Bandra (W), Mumbai - 400 050.



Data Mining and Business Intelligence

(Code - ITC602)

Semester VI - Information Technology

(Mumbai University)

Strictly as per the Choice Based Credit and Grading System
(Revise 2016) of Mumbai University w.e.f. academic year 2018-2019

Dr. Arti Deshpande

Department of Computer Engineering
Thadomal Shahani Engineering College,
Mumbai.
Maharashtra, India.

Dr. Pallavi Halarnkar

Ph.D (Computer Engineering)



Dr. G. T. Thampi
PRINCIPAL

Thadomal Shahani Engineering College
Bandra (W), Mumbai - 400 050.



 **TechKnowledge**
Publications

ME69A Price ₹ 245/-



(Book Code : ME69A)

Data Mining and Business Intelligence

Dr. Arti Deshpande, Dr. Pallavi Halarnkar

(Semester VI - Information Technology, Mumbai University)

Copyright © by Authors. All rights reserved. No part of this publication may be reproduced, copied, or stored in a retrieval system, distributed or transmitted in any form or by any means, including photocopy, recording, or other electronic or mechanical methods, without the prior written permission of the publisher.

This book is sold subject to the condition that it shall not, by the way of trade or otherwise, be lent, resold, hired out, or otherwise circulated without the publisher's prior written consent in any form of binding or cover other than which it is published and without a similar condition including this condition being imposed on the subsequent purchaser and without limiting the rights under copyright reserved above.

First Edition : February 2019

Second Revised Edition : February 2020 (**TechKnowledge Publications**)


This edition is for sale in India, Bangladesh, Bhutan, Maldives, Nepal, Pakistan, Sri Lanka and designated countries in South-East Asia. Sale and purchase of this book outside of these countries is unauthorized by the publisher.


Printed at : 37/2, Ashtavinayak Industrial Estate,
Near Pari Company, Narhe, Pune, Maharashtra State, India.
Pune – 411041.

ISBN : 978-93-89828-77-1

Published by : TechKnowledge Publications

Head Office : B/5, First floor, Maniratna Complex, Taware Colony, Aranyeshwar Corner,
Pune - 411 009. Maharashtra State, India
Ph : 91-20-24221234, 91-20-24225678.
Email : info@techknowledgebooks.com,
Website : www.techknowledgebooks.com

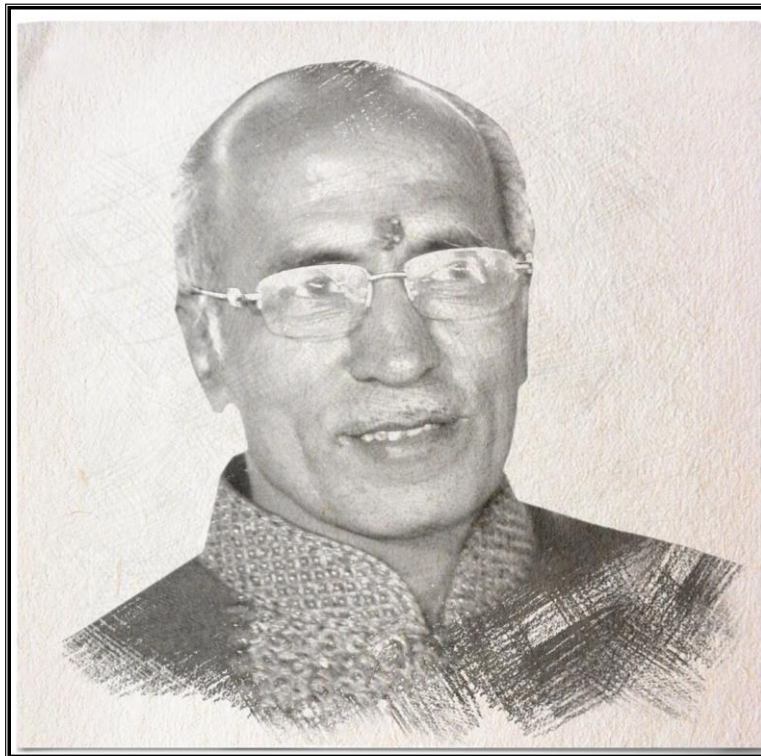

Dr. G. T. Thampi
PRINCIPAL
Thadomal Shahani Engineering College
Bandra (W), Mumbai - 400 050.



[ITC602] (FID : ME69) (Book Code : ME69A)

(Book Code : ME69A)

*We dedicate this Publication soulfully and
wholeheartedly,
in loving memory of our beloved founder director,
Late Shri. Pradeepji Lalchandji Lunawat,
who will always be an inspiration, a positive force and
strong support behind us.*



“My work is my prayer to God”

- Lt. Shri. Pradeepji L. Lunawat

*Soulful Tribute and Gratitude for all Your
Sacrifices, Hardwork and 40 years of
Strong Vision...*



PRINCIPAL
Thadomal Shahani Engineering College
Bandra (W), Mumbai - 400 050.



Dr. G. T. Thampi
PRINCIPAL
Thadomal Shahani Engineering College
Bandra (W), Mumbai - 400 050.



Preface



Dear students,


We are extremely happy to present the book on “**Data Mining and Business Intelligence**” for you. We have divided the subject into small chapters so that the topics can be arranged and understood properly. The topics within the chapters have been arranged in a proper sequence to ensure smooth flow of the subject.

We present this book in the loving memory of **Late. Shri. Pradeepji Lunawat**, our source of inspiration and a strong foundation of “**TechKnowledge Publications**”. He will always be remembered in our hearts and motivate us to achieve our new milestone.

We are thankful to Mr. Arunoday Kumar, Mr. Shital Bhandari and Mr. Chandroday Kumar for the encouragement and support that they have extended. We also thankful to the staff members of TechKnowledge Publications for their efforts to make this book as good as it is. We have made every possible efforts to eliminate all the errors in this book. However if you find any, please let us know, because that will help us to improve the book quality further.

We are also thankful to our family members and friends for their patience and encouragement.

 - **Authors**
Dr. G. T. Thampi
PRINCIPAL
Thadomal Shahani Engineering College
Bandra (W), Mumbai - 400 050. 



(Book Code : ME69A)

Books by Author – Dr Arun Kulkarni



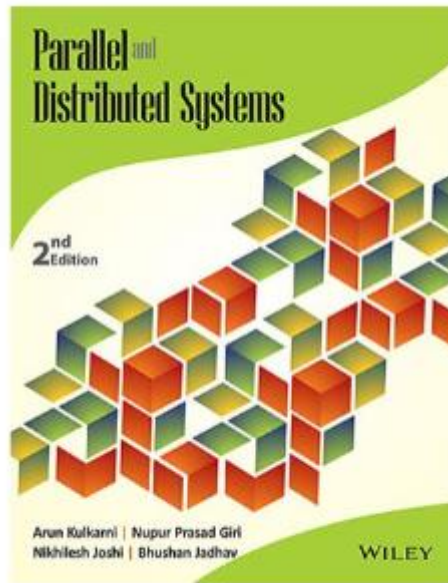
Dr. G. T. Thampi
PRINCIPAL
Thadomal Shahani Engineering College
Bandra (W), Mumbai - 400 050.



Parallel and Distributed Systems

by [Bhushan Jadhav](#), [Arun Kulkarni](#), [Nupur Prasad Giri](#), [Nikhilesh Joshi](#) (Author)

ISBN: 978-8126558674



Preface

Acknowledgement

About the Authors

Chapter 1 Introduction to Parallel Computing

1.1 Introduction


1.2 Computing


1.3 Parallel Architecture

1.4 Classification Based on Architectural Schemes

1.5 Classification Based on Memory Access

1.6 Classification Based on Interconnections between PEs and Memory Modules


Dr. G. T. Thampi
PRINCIPAL
Thadomal Shahani Engineering College
Bandra (W), Mumbai - 400 050.



1.7 Classification Based on Characteristic Nature of Processing Elements

1.8 Performance Metrics

1.9 Parallel Programming Models

1.10 Serial and Parallel Algorithms

1.11 Parallelism

Chapter 2 Pipelining

2.1 Introduction

2.2 Pipeline Performance

2.3 Types of Pipeline

2.4 Pipeline Stage Design

2.5 Pipeline Hazards

2.6 Instruction Scheduling

Chapter 3 Synchronous Parallel Processing

3.1 Introduction

3.2 SIMD Architecture and Its Programming Principle

3.3 Single Instruction Multiple Data (SIMD) Parallel Algorithms

3.4 Data Mapping and Memory in Array Processor

3.5 Case Studies of SIMD Parallel Processors

Chapter 4 Introduction to Distributed Systems

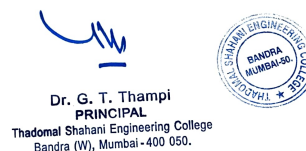
4.1 Introduction

4.2 Definition

4.3 Goals of the Distributed System

4.4 Issues Related to the Distributed System

4.5 Types of Distributed System



- 4.6 Distributed System Models
- 4.7 Hardware Concept
- 4.8 Software Concept
- 4.9 Models of Middleware
- 4.10 Services Offered by Middleware System
- 4.11 Client–Server Model

Chapter 5 Communication

- 5.1 Introduction
- 5.2 Layered Protocols
- 5.3 Remote Procedure Call
- 5.4 Remote Object Invocation
- 5.5 Remote Method Invocation
- 5.6 Message-Oriented Communication
- 5.7 Stream-Oriented Communication

Chapter 6 Resource and Process Management

- 6.1 Resource Management in Distributed System
- 6.2 Desirable Features of Global Scheduling Algorithm
- 6.3 Scheduling in the Distributed System
- 6.4 Taxonomy of the Distributed Scheduling
- 6.5 Task Assignment Approach
- 6.6 Load Balancing Approach
- 6.7 Issues in Designing Load Balancing Algorithm
- 6.8 Load Sharing Approach
- 6.9 Introduction to Process Management



Dr. G. T. Thampi
PRINCIPAL
Thadomal Shahani Engineering College
Bandra (W), Mumbai - 400 050.



Chapter 7 Synchronization

7.1 Introduction: Clock Synchronization

7.2 Physical Clock

7.3 Logical Clock

7.4 Election Algorithms

7.5 Mutual Exclusion

7.6 Centralized Algorithm

7.7 Distributed Mutual Exclusion

Chapter 8 Replication, Consistency and Distributed File System

8.1 Introduction

8.2 Replication and Consistency

8.3 Replication Management

8.4 Distributed File Systems

8.5 Case Studies

Summary

Multiple Choice Questions

Short Answer Questions

Long Answer Questions

Answers

Practical No. 1

Practical No. 2

Practical No. 3

Practical No. 4

Practical No. 5



Dr. G. T. Thampi
PRINCIPAL
Thadomal Shahani Engineering College
Bandra (W), Mumbai - 400 050.



Practical No. 6

Practical No. 7

Practical No. 8

Practical No. 9

Practical No. 10

Practical No. 11

Practical No. 12

Index



Dr. G. T. Thampi
PRINCIPAL
Thadomal Shahani Engineering College
Bandra (W), Mumbai - 400 050.



Books by Author – Dr Bhushan Jadhav



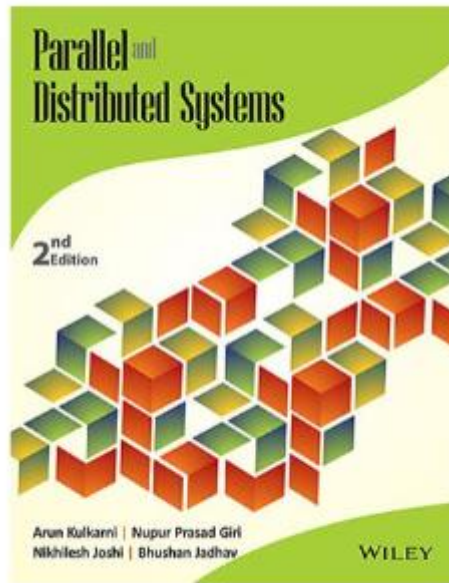
Dr. G. T. Thampi
PRINCIPAL
Thadomal Shahani Engineering College
Bandra (W), Mumbai - 400 050.



Parallel and Distributed Systems

by [Bhushan Jadhav](#), [Arun Kulkarni](#), [Nupur Prasad Giri](#), [Nikhilesh Joshi](#) (Author)

ISBN: 978-8126558674



Preface

Acknowledgement

About the Authors

Chapter 1 Introduction to Parallel Computing

1.1 Introduction

1.2 Computing

1.3 Parallel Architecture

1.4 Classification Based on Architectural Schemes

1.5 Classification Based on Memory Access

1.6 Classification Based on Interconnections between PEs and Memory Modules



Dr. G. T. Thampi
PRINCIPAL
Thadomal Shahani Engineering College
Bandra (W), Mumbai - 400 050.



1.7 Classification Based on Characteristic Nature of Processing Elements

1.8 Performance Metrics

1.9 Parallel Programming Models

1.10 Serial and Parallel Algorithms

1.11 Parallelism

Chapter 2 Pipelining

2.1 Introduction

2.2 Pipeline Performance

2.3 Types of Pipeline

2.4 Pipeline Stage Design

2.5 Pipeline Hazards

2.6 Instruction Scheduling

Chapter 3 Synchronous Parallel Processing

3.1 Introduction

3.2 SIMD Architecture and Its Programming Principle

3.3 Single Instruction Multiple Data (SIMD) Parallel Algorithms

3.4 Data Mapping and Memory in Array Processor

3.5 Case Studies of SIMD Parallel Processors

Chapter 4 Introduction to Distributed Systems

4.1 Introduction

4.2 Definition

4.3 Goals of the Distributed System

4.4 Issues Related to the Distributed System

4.5 Types of Distributed System



Dr. G. T. Thampi
PRINCIPAL
Thadomal Shahani Engineering College
Bandra (W), Mumbai - 400 050.




- 4.6 Distributed System Models
- 4.7 Hardware Concept
- 4.8 Software Concept
- 4.9 Models of Middleware
- 4.10 Services Offered by Middleware System
- 4.11 Client–Server Model


Chapter 5 Communication

- 5.1 Introduction
- 5.2 Layered Protocols
- 5.3 Remote Procedure Call
- 5.4 Remote Object Invocation
- 5.5 Remote Method Invocation
- 5.6 Message-Oriented Communication
- 5.7 Stream-Oriented Communication

Chapter 6 Resource and Process Management

- 6.1 Resource Management in Distributed System
- 6.2 Desirable Features of Global Scheduling Algorithm
- 6.3 Scheduling in the Distributed System
- 6.4 Taxonomy of the Distributed Scheduling
- 6.5 Task Assignment Approach
- 6.6 Load Balancing Approach
- 6.7 Issues in Designing Load Balancing Algorithm
- 6.8 Load Sharing Approach
- 6.9 Introduction to Process Management


Dr. G. T. Thampi
PRINCIPAL
Thadomal Shahani Engineering College
Bandra (W), Mumbai - 400 050.



Chapter 7 Synchronization

7.1 Introduction: Clock Synchronization

7.2 Physical Clock

7.3 Logical Clock

7.4 Election Algorithms

7.5 Mutual Exclusion

7.6 Centralized Algorithm

7.7 Distributed Mutual Exclusion

Chapter 8 Replication, Consistency and Distributed File System

8.1 Introduction

8.2 Replication and Consistency

8.3 Replication Management

8.4 Distributed File Systems

8.5 Case Studies

Summary

Multiple Choice Questions

Short Answer Questions

Long Answer Questions

Answers

Practical No. 1

Practical No. 2

Practical No. 3

Practical No. 4

Practical No. 5



Dr. G. T. Thampi
PRINCIPAL
Thadomal Shahani Engineering College
Bandra (W), Mumbai - 400 050.



Practical No. 6

Practical No. 7

Practical No. 8

Practical No. 9

Practical No. 10

Practical No. 11

Practical No. 12

Index



Dr. G. T. Thampi
PRINCIPAL
Thadomal Shahani Engineering College
Bandra (W), Mumbai - 400 050.



DISTRIBUTED COMPUTING

Includes Labs



Nupur Giri

Lata Ragha

Bhushan Jadhav

Dr. G. T. Thampi
PRINCIPAL

Thadomal Shahani Engineering College
Bandra (W), Mumbai - 400 050.



www.staredusolutions.org

DISTRIBUTED COMPUTING

Nupur Prasad Giri

Lata Ragh

Bhushan Jadhav



Dr. G. T. Thampi
PRINCIPAL

Thadomal Shahani Engineering College
Bandra (W), Mumbai - 400 050.



Author Profile



Dr. Nupur Prasad Giri is Professor and Head of Department, Computer Engineering at Vivekanand Education Society Institute of Technology, Mumbai. She has more than 25 years of industry and teaching experiences. Her doctoral research was in the multidisciplinary field of Multi Agent System and Mobile Computing, typically Cellular Networks. Her contributions in the field of Artificial Intelligence, Mobile and Distributed Computing have been published in many journals and International conferences. She has filed many patents.

She is also a committee member of many international conferences. She is recognized PhD guide in University of Mumbai. She has represented India as an Expert in Worldskills-2011, London as well as in Worldskills-2013, Germany in the category of Web Design. She has also been awarded Microsoft's "AI for Earth" Grant in the year 2018.



Dr. Lata L. Ragha is Professor and Head, Department of Computer Engineering at Fr. C. Rodrigues Institute of Technology, Vashi, Navi-Mumbai. She has received her Ph. D. degree from Jadavpur University, Kolkata in 2011. Her research interests include in the areas of Networking, Security, Internet Routing, and Data Mining. She has more than 100 research publications in International Journals and conferences. She is having 32 years of teaching experience. She is a member of Board of Studies, Mumbai University. She is associated with two colleges as member of Department Advisory Board. Two members have received their PhD degree under her guidance and two more have submitted their thesis under her guidance.

Dr. G. T. Thampi
PRINCIPAL

Thadomal Shahani Engineering College
Bandra (W), Mumbai-400 050.



Dr. Bhushan A Jadhav is an Assistant Professor at Thadomal Sahani Engineering College, Bandra, Mumbai. He has more than 11 years of teaching experience and has completed his Ph.D in area of Cloud Computing & Big Data Analytics. His area of interest and research includes cutting edge technologies such as DevOps, Internet of Things (IOT), Cloud Computing, Big Data Analytics, Python Programming, R Programming and Advance Information Security. He has published more than 8 research papers in National and International journals and conferences He has published six books with various publishers. He is certified trainer for Star Certification's Python Programming, Cloud Computing and DevOps Courses.

Contents

1. INTRODUCTION TO DISTRIBUTED SYSTEMS	1
1.1. Introduction	2
1.2. Definition of Distributed Systems	3
1.3. Significant Consequences of Distributed Systems	5
1.4. Goals of Distributed Systems	5
1.5. Issues of Distributed Systems	10
1.6. Types of Distributed Systems	13
1.6.1. Distributed Computing system	13
1.6.2. Distributed information systems	17
1.6.3. Distributed pervasive system:	20
1.7. Distributed System Models	23
1.7.1. Architectural Model:	23
1.7.2. Interaction Model:	25
1.7.3. Fault Models	27
1.7.4. Security Model	29
1.8. Hardware Concept	30
1.8.1. Multiprocessor System	30
1.8.2. Multicomputer System	32


Dr. G. T. Thampi
PRINCIPAL
Thadomal Shahani Engineering College
Bandra (W), Mumbai-400 050.



1.9. Software Concept	33
1.9.1. Distributed Operating System	33
1.9.2. Network Operating System	34
1.9.3. Middleware	35
1.10. Services offered by Middleware	36
1.11. Models of Middleware.	36
1.11.1. Remote Procedure Call (RPC)	36
1.11.2. Message Oriented Middleware (MOM).	37
1.11.3. Distributed Object Technology	37
1.12. Client-Server Model	38
1.12.1. Client	38
1.12.2. Server.	39
1.12.3. Network	39
Summary	39
Review Questions	40
2. COMMUNICATION	41
2.1. Introduction.	42
2.2. Layered Protocols	43
2.2.1. Seven Layers of OSI	49
2.3. Remote Procedure call.	50
2.3.1. The RPC Model	51
2.3.2. Implementing RPC Mechanism	53
2.3.3. Parameter Passing in RPC	53
2.3.4. Extended RPC Models.	56
2.3.5. Example of RPC.	56
2.4. Remote Object Invocation	57
2.4.1. Types of Object and Binding	58
2.5. Remote Method Invocation	58
2.5.1. Architecture of Remote Method Invocation	60
2.5.2. Remote Method Invocation Process	62
2.6. Message-Oriented Communication	




2.6.1. Types of Communication	63
2.6.2. Message-Oriented Transient Communication	65
2.6.3. Message-Oriented Persistent Communication	69
2.7. Group Communication	71
2.7.1. One – to - many communication	71
2.7.2. Many – to – one Communication	73
2.7.3. Many –to – many	74
2.8. Stream-Oriented Communication	77
Summary	79
Review Questions	80
3. SYNCHRONIZATION	81
3.1 Introduction: Clock Synchronization	82
3.2 Physical clock	83
3.2.1 Cristian's Algorithm	83
3.2.2 Berkeley Algorithm	85
3.2.3 Network Time Protocol	87
3.3 Logical clock	88
3.1.1 Lamport's Scalar Clock	89
3.3.2 Vector Timestamp Ordering	91
3.4 Election algorithms	93
3.4.1 Bully Algorithm	93
3.4.2 Ring Algorithm	97
3.5 Mutual exclusion	99
3.6 Centralized algorithm	101
3.7 Distributed Mutual exclusion	104
3.7.1 Non-Token Based Algorithms	104
3.7.2 Token-Based Algorithms	113
3.7.3 Singhal's Heuristic Algorithm	118
3.7.4 Raymond's Tree-Based Algorithm	121
Summary	127
Review Questions	128

Dr. G. T. Thampi
PRINCIPAL
Thadomal Shahani Engineering College
Bandra (W), Mumbai-400 050.



4. RESOURCE AND PROCESS MANAGEMENT	128
4.1 Resource Management in Distributed system.	130
4.2 Desirable Features of global scheduling algorithm.	130
4.3 Scheduling in the Distributed system	132
4.4 Taxonomy of the Distributed scheduling	134
4.4.1 Hierarchical Classification.	134
4.4.2 Dynamic Scheduling Algorithm	137
4.5 Task assignment approach	139
4.6 Load balancing approach	142
4.6.1 Benefits of Load Balancing	142
4.6.2 Static Load Balancing	143
4.6.3 Dynamic Load Balancing.	145
4.6.4 Comparison Between Static and Dynamic Load Balancing Algorithms	146
4.6.5 Static versus Dynamic Algorithms.	147
4.6.6 Deterministic versus Probabilistic Algorithms	147
4.6.7 Centralized versus Distributed Algorithms.	148
4.7 Issues in Designing Load balancing algorithm	148
4.7.1 Load Estimation Policy	148
4.7.2 Process Transfer Policy.	149
4.7.3 Location Policies	151
4.7.4 State Information Exchange Policy	152
4.7.5 Priority Assignment Policy	153
4.8 Load sharing approach	153
4.8.1 Load Estimation Policy	153
4.9 Process Management	154
4.9.1 Threads	155
4.10 Virtualization	158
4.11 Clients	161
4.11.1 Thin-Client Network Computing.	161
4.11.2 Client-Side Software for Distribution Transparency	163
4.12 Servers	164


 Dr. G. T. Thampi
 PRINCIPAL
 Thadomal Shahani Engineering College
 Bandra (W), Mumbai - 400 059.



4.12.1	General Design Issues	164
4.12.2	Cluster of Servers	166
4.12.3	Distributed Servers	167
4.13	Code Migration	170
4.13.1	Approaches to Code Migration	170
4.13.2	Models for Code Migration	171
4.13.3	Migration and Local Resources	172
4.13.4	Migration in Heterogeneous Systems	175
	Summary	177
	Review Questions	177
5.	CONSISTENCY, REPLICATION AND FAULT TOLERANCE	179
5.1.	Introduction	180
5.2.	Consistency	180
5.2.1	Consistency Models and their Evolution	181
5.2.1	Data-Centric Consistency Models	183
5.2.2	Client-Centric Consistency Models	191
5.3.	Replication	197
5.3.1	Replica Location	197
5.3.2	Replication Models	198
5.3.3	Replica Consistency	198
5.4.	Fault Tolerance	199
5.4.1	Failure Models	200
5.4.2	Failure Detection	202
5.4.3	Failure Masking	203
5.4.4	Resilience by Process Groups	204
5.4.5	Reliable Communication: Client - Server	212
5.4.6	Reliable Group Communication	215
5.4.7	Atomic Multicast	219
5.4.8	Recovery	223
	Summary	226
	Review Questions	226


 Dr. G. T. Thampi
 PRINCIPAL
 Thadomal Shahani Engineering College
 Bandra (W), Mumbai - 400 050.



6. DISTRIBUTED FILE SYSTEMS AND NAME SERVICES	227
6.1. Introduction to Distributed File systems	228
6.1.1 Features of a Good DFS	228
6.1.2 File Services	229
6.1.3 Naming	230
6.1.4 Naming Schemes	231
6.1.5 File Sharing Semantics	231
6.1.6 Caching	233
6.1.7 File Replication	235
6.2. Case Study on Distributed File Systems	236
6.2.1 NFS	236
6.2.2 Andrew File System	243
6.3. Introduction to Name Services and Domain Name System	247
6.3.1 Name Service	248
6.3.2 The Domain Name System	251
6.4. Directory services	253
6.5. Case study: The Global Name Service	254
6.6. X.500 Directory Service	257
6.7. Google Case Study	259
6.7.1 Components of Google Infrastructure	260
Summary	276
Review Questions	276
LAB	277
INDEX	340

(Handwritten signature)

Dr. G. T. Thamp
PRINCIPAL
 Thadomal Shahani Engineering College
 Bandra (W), Mumbai - 400 050.



Information Management

SEM VII

IT

AU

NoSQL

hadoop



TECHNICAL[™]
PUBLICATIONS

An Up-Thrust for Knowledge

Asha Bharambe
Bhushan Jadhav
Anjali Yeole

WILEY

Information Management

Asha Bharambe

Assistant Professor

Vivekanand Education Society's Institute of Technology
Chembur (E), Mumbai

Bhushan Jadhav

Assistant Professor

Thadomal Shahani Engineering College
Bandra (W), Mumbai

Anjali Yeole

Assistant Professor

Vivekanand Education Society's Institute of Technology
Chembur (E), Mumbai



Dr. G. T. Thampi
PRINCIPAL

Thadomal Shahani Engineering College
Bandra (W), Mumbai - 400 050.



WILEY

*This book is dedicated to my parents Mr. Yadeo Barhate and Mrs. Usha Barhate,
my in-laws Mr. Laxman Bharambe and Mrs. Mangala Bharambe,
my husband Mr. Aniket Bharambe and my son Atharva.*

—Asha Bharambe

*This book is dedicated to my parents Mr. Ashok Jadhav and Mrs. Shanta Jadhav,
my wife Sonali Jadhav and my son Atharva.*

—Bhushan Jadhav

*This book is dedicated to my parents Mr. Devidas Pawar and Mrs. Alka Pawar,
my in-laws Mr. Madhukar Yeole and Late Kusum Yeole,
my husband Mr. Shrikant Yeole and my daughters Shrutika and Sanvee.*

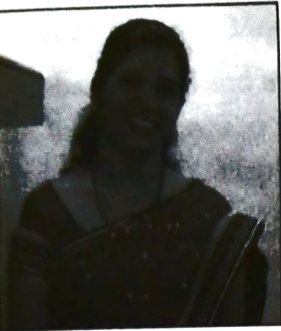
—Anjali Yeole



Dr. G. T. Thampi
PRINCIPAL
Thadomal Shahani Engineering College
Bandra (W), Mumbai - 400 050.



About the Authors



Asha Bharambe is currently working as Assistant Professor at Vivekanand Education Society's Institute of Technology, Chembur (E), Mumbai. She is also pursuing PhD in Information Technology from Mumbai University. She has received her undergraduate degree in Computer Science from SNDT Women's University and postgraduate in Computer Engineering from Mumbai University. Prof. Bharambe has taught both graduate and undergraduate students and has a teaching experience of more than 15 years. She has published many research papers in national and international conferences on

Data Mining, Big Data and Artificial Intelligence. Her areas of interest are Data Mining, Big Data and Natural Language Processing.



Bhushan Jadhav is currently working as Assistant Professor at Thadomal Shahani Engineering College, Bandra (W), Mumbai. He is also pursuing PhD in Computer Science Engineering from Thadomal Shahani Engineering College, Mumbai University. He received his undergraduate and postgraduate education from Mumbai University. Prof. Jadhav has taught both postgraduate and undergraduate students and has teaching experience of more than 8 years. He has published many research papers in national and international conferences on Virtualization and Cloud Computing. His book *Parallel and Distributed*

Systems has received very good response at Mumbai University. His areas of interest are Database Management, Parallel Computing, Distributed Computing, Virtualization and Cloud Computing.




Anjali Yeole is currently working as Assistant Professor at Vivekanand Education Society's Institute of Technology, Chembur (E), Mumbai. She is also pursuing PhD in Computer Science Engineering from Thadomal Shahani Engineering College, Mumbai University. She received her undergraduate in Computer Science from SNDT University and postgraduate education in Computer Engineering from Veermata Jijabai Technological Institute, Mumbai University. Prof. Yeole has taught both postgraduate and undergraduate students and has teaching experience of more than 12 years. She has published

many research papers in national and international conferences on Security, Cloud Computing and Web Applications. Her areas of interest are Parallel Computing, Distributed Computing, Virtualization and Cloud Computing, System Security, Computer Network, and Web Technology.

Contents

Preface	vii
About the Authors	ix
Syllabus	xi
Chapter 1 Database Design and Modelling	1
Learning Objectives	1
1.1 Introduction	1
1.2 Database Design	1
1.3 Database Modelling	2
1.3.1 ER Model	3
1.3.2 ER to Relational Data Model	6
1.3.3 Normalisation	14
1.3.4 Extended Entity Relationship Diagram	16
1.3.5 Case Study: Library Management System	17
1.4 Business Rules	19
1.4.1 Identifying Business Rules	20
1.5 Java Database Connectivity (JDBC)	20
1.5.1 Types of JDBC Drivers	21
1.6 Accessing Database Using JDBC	23
1.6.1 Registering JDBC Driver	23
1.6.2 Creating Database Connection	23
1.6.3 Executing Queries	24
1.6.4 Processing the Results	27
1.6.5 Closing the Database Connection	27
1.7 Stored Procedures	29
Summary	30
Multiple Choice Questions	31
Solved Questions	31


Dr. G. T. Thampi
PRINCIPAL
Thadomal Shahani Engineering College
Bandra (W), Mumbai - 400 050.



Chapter 2 Introduction to Big Data

Learning Objectives

- 2.1 Introduction
 - 2.2 Need for Big Data
 - 2.3 Characteristics of Big Data
 - 2.3.1 *High-Level Hadoop Architecture*
 - 2.3.2 *Characteristics of Hadoop*
 - 2.4 Components and Ecosystem
 - 2.5 Running Hadoop
 - 2.6 HDFS
 - 2.6.1 *Preparing HDFS Writes*
 - 2.6.2 *Reading Data from HDFS*
 - 2.7 MapReduce
 - 2.7.1 *MapReduce Framework*
 - 2.7.2 *MapReduce Flow*
 - 2.7.3 *Execution of a Job*
 - 2.7.4 *MapReduce Program Components*
 - 2.8 YARN
 - 2.8.1 *YARN MapReduce Application Execution Flow*
 - 2.8.2 *Advantages of YARN*
 - 2.9 NoSQL
 - 2.9.1 *Types of NoSQL Databases*
 - 2.10 Hive
 - 2.10.1 *Hive Architecture*
 - 2.10.2 *Hive Applications*
 - 2.10.3 *Difference with RDBMS*
- Summary
- Multiple Choice Questions
- Solved Questions
- Review Questions
- Answers to Multiple Choice Questions



Dr. G. T. Thampi
PRINCIPAL
Thadomal Shahani Engineering College
Bandra (W), Mumbai - 400 050.



Chapter 3 Data Security and Privacy 63

Learning Objectives	63
3.1 Introduction	63
3.2 Program Security	64
3.2.1 Fixing Faults	64
3.2.2 Unexpected Behavior	65
3.2.3 Types of Flaws	65
3.3 Overview: Covert Channels – Programs that Leak Information	77
3.3.1 Storage Channels	79
3.3.2 Timing Channels	79
3.3.3 Shared Resource Matrix	79
3.3.4 Information Flow Method	79
3.3.5 Controls Against Program Threats	80
3.3.6 Good Design	81
3.4 Operating System	81
3.4.1 History of Protection in Operating Systems	82
3.4.2 Protected Objects	82
3.4.3 Security Methods of Operating Systems	83
3.5 Firewalls	92
3.5.1 Design Goals of Firewalls	92
3.5.2 Techniques Used by Firewalls for Access Control and Enforcement of Security Policy	93
3.5.3 Capabilities of a Firewall	93
3.5.4 Limitations of Firewalls	93
3.5.5 Types of Firewalls	93
3.5.6 Example Firewall Configurations	96
3.6 Network Security	98
3.6.1 Introduction to Network	98
3.6.2 TCP/IP Vulnerability	99
3.6.3 Security Measures	104
3.6.4 Different Way of Attacking on Network	105
3.6.5 Protocols for Security	110
3.7 Intrusion Detection System (IDS)	113
3.7.1 Goals of an Ideal IDS	114
3.7.2 Strengths and Weaknesses of IDS	114

3.7.3	<i>Based on IDS Operation</i>	115
3.7.4	<i>Based on Where It is Placed</i>	116
3.7.5	<i>Stealth Mode</i>	118
3.8	Data Privacy Principles	119
3.8.1	<i>Types of Privacy</i>	120
3.8.2	<i>Data Privacy Principles</i>	121
3.9	Data Privacy Laws and Compliance	122
3.9.1	<i>Data Protection Laws</i>	122
3.9.2	<i>Compliance</i>	123
	Summary	124
	Multiple Choice Questions	124
	Solved Questions	126
	Review Questions	128
	Answers to Multiple Choice Questions	130

Chapter 4 Information Governance

131

Learning Objectives

4.1	Introduction	131
4.1.1	<i>Need for Master Data Management</i>	131
4.1.2	<i>Definition of Master Data Management</i>	132
4.1.3	<i>Characteristics and Benefits of Master Data Management</i>	133
4.1.4	<i>Master Data Management Versus Data Warehouse</i>	133
4.2	Stages of Master Data Management Implementation	134
4.3	Master Data Management Architectural Dimensions	135
4.3.1	<i>Design and Deployment Dimension</i>	137
4.3.2	<i>Use Pattern Dimension</i>	138
4.3.3	<i>Information Scope or Data Domain Dimension</i>	139
4.3.4	<i>Master Data Management Reference Architecture</i>	140
4.4	Risk Management	141
4.4.1	<i>Privacy, Regulatory Requirements and Compliance</i>	143
4.4.2	<i>Implications of Data Security and Privacy Regulations on Master Data Management</i>	144
4.5	Data Governance	150
4.5.1	<i>Goals of Data Governance</i>	150
4.6	Data Synchronization	151
4.7	Data Quality Management	152

Summary	154
Multiple Choice Questions	155
Solved Questions	156
Review Questions	166
Answers to Multiple Choice Questions	166
Chapter 5 Information Architecture	167
Learning Objectives	167
5.1 Introduction	167
5.2 Principles of Information Architecture and Framework	167
5.2.1 <i>Definitions of Information Architecture</i>	168
5.2.2 <i>Components of Information Architecture</i>	170
5.3 Organization System	171
5.3.1 <i>Classification of Organization System</i>	171
5.3.2 <i>Organization Schemes</i>	172
5.3.3 <i>Organization Structure</i>	176
5.4 Navigation Systems	180
5.4.1 <i>Types of Navigation Systems</i>	181
5.5 Labeling Systems	185
5.5.1 <i>Sources of Labeling Systems</i>	188
5.6 Searching Systems	189
5.7 Conceptual Design	189
5.8 Granularity of Contents	191
Summary	191
Multiple Choice Questions	193
Solved Questions	194
Review Questions	203
Answers to Multiple Choice Questions	204
Chapter 6 Information Lifecycle Management	205
Learning Objectives	205
6.1 Introduction	205
6.2 Data Retention Policies	205
6.2.1 <i>Purpose</i>	206

6.2.2	<i>Scope</i>	24
6.2.3	<i>Policy Contents</i>	21
6.2.4	<i>Managing the Data Retention Policy</i>	21
6.3	<i>Data Retention in Telecommunication Industry</i>	2
6.3.1	<i>Internet Service Provider (ISP) License</i>	2
6.3.2	<i>Unified Access Service License (UASL)</i>	2
6.3.3	<i>Sample Retention Records</i>	2
6.3.4	<i>Laws Related to Data Retention Policy in India</i>	2
6.4	<i>Confidential and Sensitive Data Handling</i>	2
6.4.1	<i>Handling of Sensitive Data</i>	2
6.4.2	<i>Access Decisions</i>	2
6.4.3	<i>Types of Disclosures</i>	2
6.4.4	<i>Handling Data</i>	2
6.4.5	<i>Law Provision in India Defining Sensitive Data and Its Handling</i>	1
6.5	<i>Lifecycle Management Costs</i>	2
6.6	<i>Archive Data Using Hadoop</i>	2
6.7	<i>Testing and Delivering Big Data Applications for Performance and Functionality</i>	2
6.8	<i>Challenges with Data Administration</i>	2
	<i>Summary</i>	
	<i>Multiple Choice Questions</i>	
	<i>Solved Questions</i>	
	<i>Review Questions</i>	
	<i>Answers to Multiple Choice Questions</i>	

Dr. G. T. Thampi
PRINCIPAL
Thadomal Shahani Engineering College
Bandra (W), Mumbai-400 050.



INTERNET OF EVERYTHING

Includes Labs and Cases



INTERNET

INTERNET OF EVERYTHING

Bhushan Jadhav


Anjali Yeole

Gopal Pardesi


Vaishali Khairnar

Dhananjay Kalbande







Dr. Bhushan Jadhav is an Assistant Professor at Thadomal Shahani Engineering College, Bandra, Mumbai. He has more than 11 years of teaching experience and has completed his Ph.D in area of Cloud Computing & Big Data Analytics. His area of interest and research includes cutting edge technologies such as DevOps, Internet of Things (IoT), Cloud Computing, Big Data Analytics, Python Programming, R Programming and Advance Information Security. He has published more than 8 research papers in National and International journals and conferences He has published six books with various publishers. He is certified trainer for Star Certification's Python Programming, Cloud Computing and DevOps Courses.



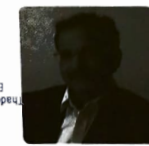
Dr. Anjali Yeole is currently working as Assistant Professor at Vivekanand Education Society's Institute of Technology, Chembur (E), Mumbai. She has received her PhD in Computer Science Engineering from Thadomal Shahani Engineering College under Mumbai University, India. Her research area is "Internet of Things". She received her undergraduate in Computer science from SNDT university and postgraduate education in Computer engineering from VJTI, Mumbai University. She has taught both Masters and undergraduate students and has teaching experience of more than 17 years. She has published many research papers in national and international conferences and journals on IoT, security, cloud computing and web applications. She has authored a book "Information management" Her areas of interest are IoT, Parallel Computing, Distributed Computing, Virtualization and Cloud Computing, System Security, Computer Network and Web Technology.



Dr. Gopal Pardesi is currently working as Associate Professor at Thadomal Shahani Engineering College, Bandra, Mumbai. He has more than 28 years of teaching experience. He has completed his Ph.D in the field of Information and Communication Technologies. His areas of interest are Wireless Networks, Project Management, Microprocessors and Microcontrollers, Embedded Systems and Machine learning.



Dr. Vaishali Khairnar is currently associated as Head of Department Information Technology Department at Terna College of Engineering, Nerul, Navi Mumbai. She has done her Doctoral work from Nirma University, Ahmedabad, Gujrat on Vehicular Ad-Hoc networks. She has more than 19 years of collective experience in industry, teaching as well as research. She has more than 50 papers published in national and international journals and conferences to her credit. Her area of interest includes wireless networks, IOT, Open source tools, Ad-Hoc Sensor networks, Internet Programming, Mobile Application and Development & Storage Network Management and Retrieval. She has authored four books published in India, UK and Singapore. She has been a resource person for several workshops and short term training programmes. She is paper reviewer on different editor bodies at national/ international journals and conferences. She is active member of CSI- Mumbai Chapter and ACM.



Dr. Dhananjay Kalbande is currently a Professor in Computer Engineering and Dean (Industry Relations), Sardar Patel Institute of Technology, Andheri (West), Mumbai, India. He was Head of the Department from April 2012 to Oct 2019. He completed his B.E. in Computer Technology from Nagpur University in 1997 and Master of Engineering in Information Technology in May 2005, from Vivekanand Education Society's Institute of Technology (VESIT), Mumbai University, Mumbai, India. He has obtained a Ph.D in Technology from University of Mumbai, Mumbai in 2011. He has been awarded a Post-Doctorate (PDF) from Tata Institute of Social Sciences (T.I.S.S.) in 2016. He has also been honoured as a Senior Research Fellow (SRF) on the NCW-TISS Project, funded by National Commission for Women, Govt. of India at T.I.S.S., from July 23, 2016 to Oct 10, 2016. He was a Research Fellow on the CliX Project at T.I.S.S., funded by Tata Trust and M.I.T.(U.S.A.) from Feb 20, 2017 to May 19, 2017. He has over 19+ Years of experience in teaching & research. His research interests include Soft computing (Neural Networks, Fuzzy Logic) Computer Network, Human Machine Interaction Decision making and business Intelligence. Mobile application development for social cause. ICT for semi-rural development for social cause. He has authored four books namely Graphical User Interface (Pareen Publications), MIS (Pareen Publications), Human Machine Interaction and Digital Forensic with Wiley India Pvt. Ltd. He has delivered and conducted workshops, seminars, Tutorials and Expert Talks on NS2 Neural Network, VB.Net and ADO.Net, Transfer Learning. Skinzy is his brainchild which has been turned into reality into Healthcare Start-up operated from Mumbai. Skinzy's flagship product "DermaPhoto" is an AI based mobile application that can detect skin diseases. He has patented 6 innovative ideas of research work.

Contents

1. INTRODUCTION	1
1.1. Introduction to IoT and IoE	2
1.2. History of IoE	4
1.2.1. Telemetry (1845 to 1990s)	5
1.2.2. M2M systems (1990s to 2010)	5
1.2.3. Internet of Things (2010 to present)	5
1.2.4. Internet of Everything (IoE) 2014 till present	6
1.2.5. Difference between IoT and M2M	6
1.3. Objects in IoE	8
1.3.1. Sensors	10
1.3.2. Data	13
1.3.3. Actuator	13
1.3.4. People/ Process	13
1.4. Identifiers in the IoE	13
1.4.1. Thing Identifier	14
1.4.2. Communication Identifier	16
1.4.3. User Identifier	17



Dr. G. T. Thampi
PRINCIPAL
Thadomal Shahani Engineering College
Bandra (W), Mumbai - 400 050.



1.5. IoE-enabling Technologies	17
1.5.1. Sensors	17
1.5.2. Wireless Sensor Network	18
1.5.3. Cloud Computing	19
1.5.4. Big Data Analytics	20
1.5.5. Communication Protocols	21
1.5.6. Messaging Protocol	26
1.5.7. Embedded Computing Boards	28
1.5.8. Web services and REST-based web services	30
Summary	30
Important Questions	31
2. RFID TECHNOLOGIES	33
2.1. Introduction	34
2.2. RFID	34
2.3. IOT and RFID Technology	35
2.4. RFID System	37
2.5. Components of RFID System	38
2.5.1. RFID Antenna	38
2.5.2. RFID Reader	40
2.5.3. RFID Transponder	42
2.6. RFID Tag	43
2.7. Principle of RFID	46
2.8. RFID Middleware	47
2.9. Issues	49
Summary	51
Important Questions	52
3. RFID APPLICATIONS	53
3.1. Introduction	54
3.2. Applications of RFID System	54
3.3. Challenges of RFID Technology	60
3.4. Hardware Issues and Protocols	61



Dr. G. T. Thampi
PRINCIPAL
 Thadomal Shahani Engineering College
 • Bandra (W), Mumbai - 400 050.

3.5. RFID Anti-Collision Protocols for Tag Identification	62
3.6. Pure Aloha	64
3.7. Slotted Aloha	65
3.8. Framed Slotted Aloha and Dynamic Framed Slotted Aloha	66
3.9. Tree-Based Protocols	67
3.10. Query Tree (QT) Protocol	67
3.11. Smart Trend Traversal (STT) Protocol	68
3.12. Query Window Tree Protocol	69
3.13. Binary Search (BS) Protocol	70
3.14. Bitwise Arbitration(BAT) Anti-collision Protocols	72
Summary	76
Important Questions	78
4. WIRELESS SENSOR NETWORKS	79
4.1. Introduction	80
4.1.1. Wireless Sensor Networks (WSNs)	80
4.1.2. Advantages of Wireless Sensor Networks	81
4.1.3. Limitations of Wireless Sensor Networks	81
4.1.4. Comparison of Wireless Sensor Networks and Adhoc Networks	82
4.1.5. Wireless Sensor Network Applications	82
4.2. Node	83
4.2.1. Components of a Sensor Node	84
4.2.2. WSN Network Topologies	86
4.3. Architecture and Communication	87
4.3.1. Application Layer	87
4.3.2. Transport Layer	88
4.3.3. Network Layer	88
4.3.4. Data Link Layer	88
4.3.5. Physical Layer	88
4.4. Types of Wireless Sensor Networks	89
4.4.1. Terrestrial WSNs	89
4.4.2. Underground WSNs	89

4.4.3. Underwater WSNs	90
4.4.4. Mobile WSNs	90
4.4.5. Multimedia WSNs	91
4.5. Networking Nodes	91
4.5.1. Medium Access Control Layer	92
4.6. Securing Communication	92
4.7. Standards and Fora	93
4.8. Networking and Internet	94
4.8.1. IP Addressing	94
4.9. Wireless Protocols	96
4.9.1. MQTT	96
4.9.2. CoAP	98
4.10. REST Transferring Data	101
4.10.1. Introduction to RESTful web services	102
4.10.2. RESTful Methods	103
4.10.3. Requirements of REST	104
4.10.4. Advantages of REST	105
4.10.5. Disadvantages of REST	105
Summary	106
Important Questions	108
5. MOBILITY AND SETTING	109
5.1. Introduction	110
5.2. Localization	110
5.2.1. Localization Techniques	115
5.3. Mobility Management	117
5.3.1. Mobile IP	118
5.3.2. Mobility Management at Different Layers	120
5.4. Localization and Handover Management in RFID	124
5.4.1. RFID-enabled Localization	125
5.4.2. RFID-enabled Movement Detection	126
5.4.3. RFID Collision Problem	127

5.5. Technology Consideration	128
5.5.1. Path Loss Model	128
5.5.2. Antenna Radiation Pattern	129
5.5.3. Multiple Tags-to-reader Collisions	130
5.5.4. Multiple Readers-to-tag Collision	133
5.5.5. Reader-to-reader Collision	134
5.5.6. Interface for Specific Material	135
5.6. Performance Evaluation	135
5.6.1. Simulation Setup	135
5.6.2. Performances Result	136
5.7. Identification of IoT	136
5.7.1. IPv6	136
5.7.2. Why does IPv6 suit IoT?	138
5.7.3. Uniform Resource Identifier (URI)	138
Summary	139
Important Questions	140
6. DATA ANALYTICS FOR IOE	141
6.1. Introduction	142
6.2. Apache Hadoop	144
6.2.1. Architecture of Hadoop	147
6.3. MapReduce Programming Model	149
6.3.1. Using Hadoop MapReduce for Batch Data Analysis	152
6.4. Apache Oozie	153
6.4.1. Features of Oozie	156
6.5. Apache Spark	156
6.5.1. Spark Clusters	158
6.6. Apache Storm	159
6.7. Using Apache Storm for Real Time Data Analysis	161
6.8. Case Study on Structural Health Monitoring	163
6.8.1. Different Sensors Used in Healthcare Monitoring System	163
6.8.2. Internet of Things Devices for Healthcare	169
6.8.3. An IoT-Aware platform for Structural Health Monitoring	171

Dr. G. T. Thampi
 PRINCIPAL
 Thadomal Shahani Engineering College
 Bandra (W), Mumbai - 400 050.



6.9. Tools for IoT	
6.9.1. CHEF	
6.9.2. Introduction to Puppet	
6.10. IoT System Network Management	
6.10.1. NETCONFIG	
6.10.2. YANG	
6.10.3. NETCONF- YANG Case Studies	
6.11. IoT Code Generator	
Summary	
Important Questions	
INTRODUCTION TO LAB 'INTERNET OF EVERYTHING'	191
Experiment 1: Basics of Cooja Simulator	194
Experiment 2: Sensor Activation through Button	196
Experiment 3: Simple UDP RPL Broadcast with Sky mote	204
Experiment 4: 6LoWPAN Protocol with Packet Analyzer	210
Experiment 5: Collect View	210
INDEX	214

CLOUD COMPUTING AND SERVICES



Bhushan A. Jadhav
Dr. Deven Shah
Arup Vithal



www.staredusolutions.org



CLOUD COMPUTING AND SERVICES



Dr. G. T. Thampi
PRINCIPAL
Thadomal Shahani Engineering College
Bandra (W), Mumbai - 400 050.



Bhushan A. Jadhav | Dr. Deven Shah | Arup Vithal

Contents

1. INTRODUCTION TO CLOUD COMPUTING	1
➤ Introduction to Cloud Computing	2
→ Definition of cloud Computing	2
→ Characteristics of Cloud Computing	3
➤ Cloud and other similar configurations	4
→ Peer to Peer Architecture	4
→ Client Server Architecture	4
→ Grid Computing	5
➤ Components of Cloud Computing	6
→ Client – The End User	6
→ Cloud Network – The Link	6
→ Cloud Application Programming Interfaces (APIs) – The Gateway	6
→ Large Scale Hardware – Cloud Storage and VM Access	7
➤ Cloud Types	8
→ NIST Model	8
→ Cloud Cube Model	9
➤ Cloud Deployment Models	10
→ Public Cloud	10
→ Private Cloud	11
→ Hybrid Cloud	12
→ Community Cloud	12
→ Comparison between various Cloud Deployment Models	12



- Cloud Computing Services 13
 - "Infrastructure as a Service (IaaS)" 13
 - "Platform as a Service (PaaS)" 14
 - "Software as a Service (SaaS)" 14
- NIST Architecture of Cloud Computing 14
 - Cloud Consumer 14
 - Cloud Provider 15
 - Cloud Auditor 16
 - Cloud Broker 17
 - Cloud Carrier 17
- Advantages of Cloud computing 20
- Cloud Computing Challenges 22

2. VIRTUALIZATION 27


- Introduction 28
- Characteristics of Virtualization 31
- Taxonomy of Virtualization 33
 - Full Virtualization 33
 - Paravirtualization 34
 - Hardware-Assisted Virtualization 35
 - Operating System Virtualization 36
 - Application Server Virtualization 36
 - Application Virtualization 37
 - Network Virtualization 37
 - Storage Virtualization 37
 - Service Virtualization 38
- Understanding the importance of Type I & Type II Hypervisors 38
 - Hosted Structure (Type II Hypervisor) 39
 - Bare-metal Structure (Type I Hypervisor) 40
- Implementation Levels of Virtualization 41
- Resource Virtualization 45
 - CPU Virtualization 45


 Dr. G. T. Thampi
 PRINCIPAL
 Thadomal Shahani Engineering College
 Bandra (W), Mumbai - 400 050.



→ Memory Virtualization	49
→ Device and IO Virtualization	50
➤ Virtualization Vs Cloud Computing	51
➤ Pros and Cons of Virtualization	54
➤ Technology Examples	55
→ KVM Architecture	56
→ Xen Architecture	57
→ VMware	57
→ Hyper-V	59
→ Comparison Table between VMware, Hyper V, KVM and Xen	60
3. CLOUD COMPUTING SERVICES	65
➤ Introduction	66
➤ Exploring Cloud Computing Services	66
➤ SPI Model	67
→ Software as a Service	68
→ Platform as a Service	69
→ Infrastructure as a Service	72
➤ Anything as a Service or Everything as a Service (XaaS)	74
→ Security as a Service (SECaaS)	74
→ Identity Management as a Service (IdMaaS)	75
→ Database as a Service	75
→ Collaboration as a Service	77
→ Compliance as a Service (CaaS)	77
→ Monitoring as a Service	78
→ Communication as a Service	78
→ Disaster Recovery as a Service	79
→ Backup as a Service	79
→ Storage as a Service	80
→ Network as a Service	80
→ Analytics as a Service (AaaS)	81

4. CLOUD IMPLEMENTATION, PROGRAMMING AND MOBILE CLOUD COMPUTING	87
➤ Introduction	88
➤ OpenStack Cloud Architecture	88
→ Components of OpenStack	89
→ Open Stack Architecture	90
→ Features and Benefits of OpenStack	94
→ Modes of Operations of OpenStack	94
➤ Programming Support for Google Apps Engine	95
→ GFS	95
→ Bigtables	98
→ Chubby	101
→ Google APIs	104
➤ Mobile Cloud Computing	104
→ Definition	105
→ Basic Components of Mobile Cloud Computing	106
→ Additional Components of Mobile Cloud Computing	108
→ Architecture of Mobile Cloud Computing	112
→ Benefits of Mobile Cloud Computing	114
→ Challenges of Mobile Cloud Computing	116
5. EXPLORING THE COMPONENTS OF AMAZON WEB SERVICES	123
➤ Amazon Web Services	124
➤ AWS Cloud Computing Platform	126
→ Elastic Compute Cloud (EC2)	126
→ Simple Storage Service (S3)	131
→ Elastic Block Storage (EBS)	134
→ Amazon Virtual Private Cloud (Amazon VPC)	137
→ Elastic Load Balancing (ELB)	140
6. CLOUD BACKUP & SOLUTIONS	147
➤ Cloud Backup Solutions and their Features	148
→ Where is the data stored in the cloud?	148
→ How is data stored?	148


 Dr. G. T. Thampi
 PRINCIPAL
 Thadomal Shahani Engineering College
 Bandra (W), Mumbai - 400 050.



→	Difference between Cloud Sync and Cloud Backup	149
→	Benefits and Risks.	150
↘	Cloud Data Management Interface (CDMI)	151
↘	Cloud Storage Gateways (CSGs).....	153
→	Use of CSG	154
→	Advantages of Using a CSG	156
↘	Comparison between different Cloud Platforms: Amazon Web Services and OpenStack ...	157
→	AWS Services	157
→	OpenStack Services	163
→	Comparing Amazon Web Services and Open Stack	171

LAB

180

INDEX

243


Dr. G. T. Thampi
PRINCIPAL
 Thadomal Shahani Engineering College
 Bandra (W), Mumbai - 400 050.



Books by Author – Shilpa Ingoley



Dr. G. T. Thampi
PRINCIPAL
Thadomal Shahani Engineering College
Bandra (W), Mumbai - 400 050.





Sem
4



Information Technology



Computer Network & Network Design

- Manoj S. Kavedia

- Shilpa Ingoley

☆ With All Latest Solved University Q. Papers

2021
EDITION

Course Code
(ITC402)
(Compulsory Subject)



e-books
(PDF download)



Download App



SCAN TO VISIT



ISBN
978-93-90376-29-2



M4-27A



Price ₹ 365/-

with
MCQ's



FREE
DOWNLOAD

Sample chapter from our Android App

TECH-NEO
PUBLICATIONS

Where Authors Inspire Innovation

A Sachin Shah Venture



Dr. G. T. Thampi
PRINCIPAL
Thadomal Shahani Engineering College
Bandra (W), Mumbai-400 050.



About the Authors



Er. Manoj S. Kavedia

Assistant Professor,
Electronics and Telecommunication Department,
Thadomal Shahani Engineering College, Bandra, Mumbai
(Total 24 years of Teaching Experience)




Shilpa Ingoley

Assistant Professor,
M.E. Computer Engineering (Mumbai University) Computer Department
Thadomal Shahani Engineering College, Bandra, Mumbai
(Total 21 years of Teaching Experience)

Dr. G. T. Thampi
PRINCIPAL
Thadomal Shahani Engineering College
Bandra (W), Mumbai-400 050.



MU Sem **4**  **Information Technology**

Name of Subject	Author
Applied Mathematics-IV	Baphana R M (Adjunct Faculty, COEP, Pune)
Computer Network and Network Design	Manoj S. Kavedia, Shilpa Ingoley
Operating System	Manoj S. Kavedia
Automata Theory	Ashish Budhrani
Computer Organization and Architecture	Velankar Shrikant, Shah Urvashi

For Orders Contact :

Krishna Book Collections

Ground Floor, Krishna Nivvas Building, Behind BEST Nivvas Building, Near Napoo Hall, Chandavarkar Road, Matunga East, Mumbai 400019

E-mail : dharmeshsota05@gmail.com

Mobile No.: Dharmesh Sota - 9820741455, Tulsidas Sota - 9833133921 / 9833082745 / 9833082761


Dr. G. T. Thampi
PRINCIPAL
Thadomal Shahani Engineering College
Bandra (W), Mumbai-400 050.



Books are available on   

University of Mumbai

Computer Network and Network Design

(Course Code : ITC402)

Semester IV - Information Technology

Strictly as per the New Syllabus (REV-2019 'C' Scheme) of
Mumbai University w.e.f. academic year 2020-2021



Er. Manoj S. Kavedia

Assistant Professor

Department of Electronics and Telecommunication

Thadomal Shahani Engineering College (TSEC), Bandra, Mumbai

Shilpa Ingoley

M.E. Computer Engineering (Mumbai University)

Assistant professor,

Computer Department

Thadomal Shahani Engineering College-Bandra

Dr. G. T. Thampi
PRINCIPAL

Thadomal Shahani Engineering College
Bandra (W), Mumbai-400 050.



 **TECH-NEO
PUBLICATIONS**

Where Authors Inspire Innovation

A Sachin Shah Venture


M4-27



Index

◆	Chapter 1 : Introduction to Computer Network	1-1 to 1-53
◆	Chapter 2 : Physical Layer and Data Link Layer	2-1 to 2-81
◆	Chapter 3 : Network Layer.....	3-1 to 3-62
◆	Chapter 4 : Transport Layer and Session Layer.....	4-1 to 4-36
◆	Chapter 5 : Presentation Layer and Application Layer	5-1 to 5-39
◆	Chapter 6 : Network Design Concepts	6-1 to 6-50

□□□


 Dr. G. T. Thampi
 PRINCIPAL
 Thadomal Shahani Engineering College
 Bandra (W), Mumbai-400 059.

MU Sem **5** **Computer Engineering**

New Syllabus
2021-22

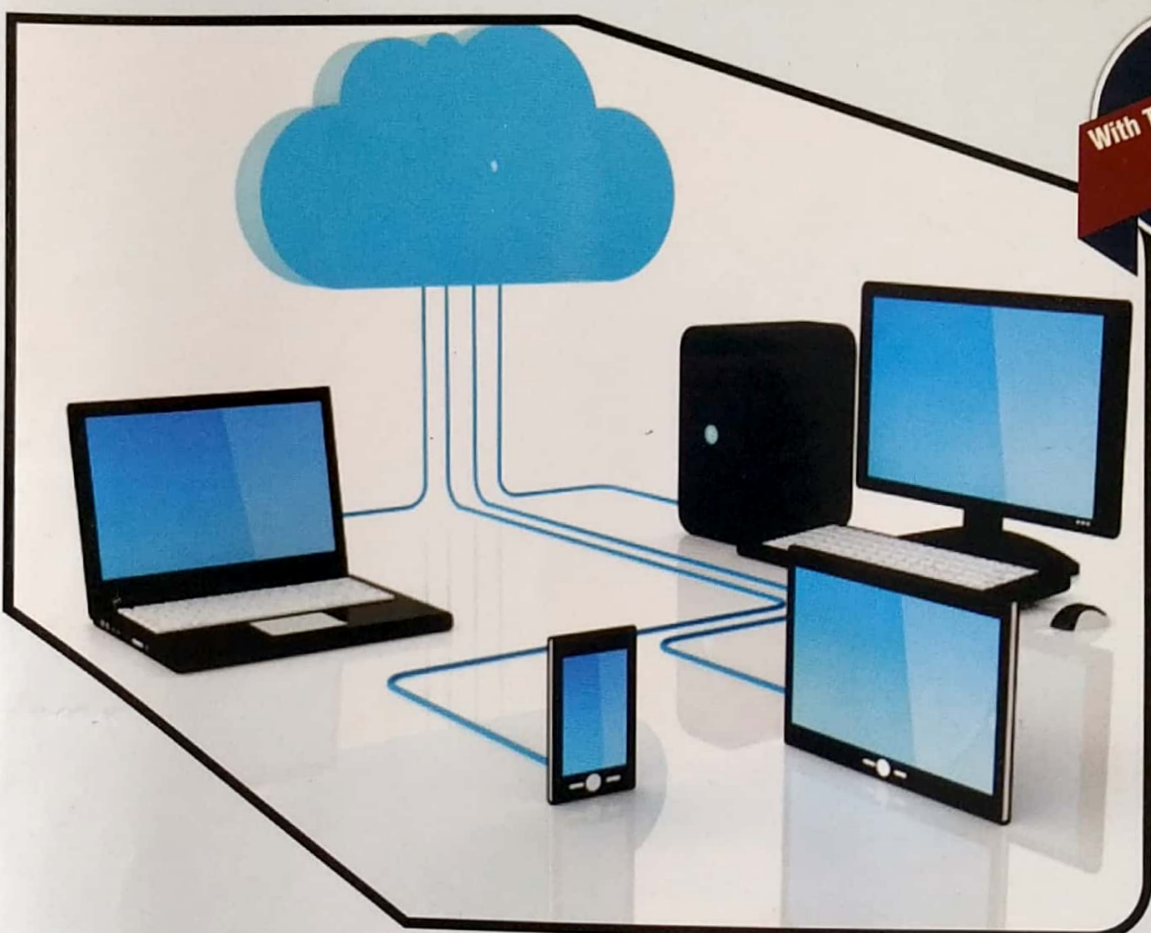
Compulsory Subject (CSC503)

Computer Network

Manoj S. Kavedia

Shilpa Ingoley

(Thadomal Shahani Engineering College, Bandra, Mumbai)



With Typical
MCQ's

TECH-NEO
PUBLICATIONS
Where Authors Inspire Innovation
A Sachin Shah Venture
www.techneobooks.in
info@techneobooks.in

- ☆ With Solved Latest **UNIVERSITY QUESTION PAPERS.**
- ☆ Self Explanatory **Diagrams.**
- ☆ Comparisons in **Tabular Form.**
- ☆ **Multiple Choice Questions.**

M5-55A

Price ₹ 325/-

Dr. G. T. Thampi
PRINCIPAL
Thadomal Shahani Engineering College
Bandra (W), Mumbai-400 050.



About the Authors



Er. Manoj S. Kavedia

Assistant Professor,
Electronics and Telecommunication Department,
Thadomal Shahani Engineering College, Bandra, Mumbai
(Total 24 years of Teaching Experience)



Shilpa Ingoley

Assistant Professor,
M.E. Computer Engineering (Mumbai University) Computer Department
Thadomal Shahani Engineering College, Bandra, Mumbai
(Total 21 years of Teaching Experience)

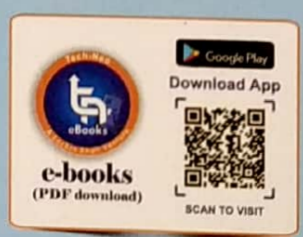

Dr. G. T. Thampi
PRINCIPAL
Thadomal Shahani Engineering College
Bandra (W), Mumbai-400 050.





Sem 5 Computer Engineering

Course Code	Compulsory Subjects
CSC501	Theoretical Computer Science
CSC502	Software Engineering
CSC503	Computer Network
CSC504	Data warehousing and Mining

Course Code	Departmental Elective
CSD05012	Internet Programming
CSD05013	Advance Database Management System



For Orders Contact :
Krishna Book Collections
 Ground Floor, Krishna Niwas Building, Behind BEST Niwas Building, Near Napoo Hall,
 Chandavarkar Road, Matunga East, Mumbai 400019. **E-mail :** dharmeshsota05@gmail.com
Mobile No.: Dharmesh Sota - 98207 41455, Tulsidas Sota - 98331 33921 / 98330 82745 / 98330 82761

Books are available on   

University of Mumbai

Computer Network

(Code : CSC503)

Semester V - Computer Engineering

Strictly as per the Choice Based Credit and Grading System
(Revise 2019) of Mumbai University w.e.f. academic year 2021-2022

Er. Manoj S. Kavedia

Asst. Professor,

Department of Electronics and Telecommunication

Thadomal Shahani Engineering College (TSEC), Bandra, Mumbai

ME Electronics

Pursing PHD in Internet of Things

Shilpa Ingoley

M.E. Computer Engineering (Mumbai University)

Assistant professor,

Computer Department

Thadomal Shahani Engineering College-Bandra

Dr. G. T. Thampi
PRINCIPAL
Thadomal Shahani Engineering College
Bandra (W), Mumbai - 400 050.



 **TECH-NEO**
PUBLICATIONS

Where Authors Inspire Innovation

A Sachin Shah Venture

M5-55



Index

- ◆ Chapter 1 : Introduction to Networking..... 1-1 to 1-56
- ◆ Chapter 2 : Physical Layer..... 2-1 to 2-29
- ◆ Chapter 3 : Data Link Layer 3-1 to 3-46
- ◆ Chapter 4 : Network Layer 4-1 to 4-71
- ◆ Chapter 5 : Transport Layer 5-1 to 5-35
- ◆ Chapter 6 : Application Layer..... 6-1 to 6-26

► **Multiple Choice Questions**



Part of **SPRINGER NATURE**

PROCEEDINGS | JOURNALS | BOOKS

Search

Series: Advances in Intelligent Systems Research

Proceedings of the International Conference on Communication and Signal Processing 2016 (ICCASP 2016)

PROCEEDINGS OF THE INTERNATIONAL CONFERENCE ON COMMUNICATION AND SIGNAL
PROCESSING 2016 (ICCASP 2016)



Human Skin Detection Using RGB, HSV and YCbCr Color Models

Authors

S. Kolkur, D. Kalbande, P. Shimpi, C. Bapat, J. Jatakia

Corresponding Author

S. Kolkur

Available Online December 2016.

DOI

<https://doi.org/10.2991/iccasp-16.2017.51> How to use a DOI?

Keywords

Skin Detection, Color Models, Image Processing, Classifier

Abstract

Human Skin detection deals with the recognition of skin-colored pixels and regions in a given image. Skin color is often used in human skin detection because it is invariant to orientation and size and is fast to process. A new human skin detection algorithm is proposed in this paper. The three main

<https://www.atlantispress.com/proceedings/iccasp-16/25871632>

Human skin detection algorithm is proposed in
parameters for recognizing a skin pixel are RGB (Red, Green, Blue), HSV (Hue, Saturation, Value) and YCbCr (Luminance, Chrominance) color models. The objective of proposed algorithm is to improve the recognition of skin pixels in given images. The algorithm not only considers individual ranges of the three


color parameters but also takes into account combi-national ranges which provide greater accuracy in recognizing the skin area in a given image.

Copyright

© 2017, the Authors. Published by Atlantis Press.

Open Access

This is an open access article distributed under the CC BY-NC license (<http://creativecommons.org/licenses/by-nc/4.0/>).

 Download article (PDF)



Volume Title

Proceedings of the International Conference on Communication and Signal Processing 2016 (ICCASP 2016)

Series

Advances in Intelligent Systems Research

Publication Date

December 2016

ISBN

978-94-6252-305-0

ISSN

1951-6851

DOI

<https://doi.org/10.2991/iccasp-16.2017.51> [How to use a DOI?](#)

<https://www.atlantis-press.com/proceedings/iccasp-16/25871632>



Dr. G. T. Thampi
PRINCIPAL
Thadomal Shahani Engineering College
Bandra (W), Mumbai - 400 050.



Copyright

© 2017, the Authors. Published by Atlantis Press.

Open Access

This is an open access article distributed under the CC BY-NC license (<http://creativecommons.org/licenses/by-nc/4.0/>).

Cite this article

ris

enw

bib

TY - CONF
 AU - S. Kolkur
 AU - D. Kalbande
 AU - P. Shimpi
 AU - C. Bapat
 AU - J. Jatakia
 PY - 2016/12
 DA - 2016/12
 TI - Human Skin Detection Using RGB, HSV and YCbCr Color Models
 BT - Proceedings of the International Conference on Communication and Signal Processing 2016 (ICCASP 2016)
 PB - Atlantis Press
 SP - 324
 EP - 332
 SN - 1951-6851
 UR - <https://doi.org/10.2991/iccasp-16.2017.51>
 DO - <https://doi.org/10.2991/iccasp-16.2017.51>
 ID - Kolkur2016/12
 ER -

Dr. G. T. Thampi
PRINCIPAL
Thadomal Shahani Engineering College
Bandra (W), Mumbai - 400 050.



download .ris COPY TO CLIPBOARD

Dr. G. T. Thampi
PRINCIPAL
Thadomal Shahani Engineering College
Bandra (W), Mumbai - 400 050.






Atlantis Press

Atlantis Press – now part of Springer Nature – is a professional publisher of scientific, technical & medical (STM) proceedings, journals and books. We offer world-class services, fast turnaround times and personalised communication. The proceedings and journals on our platform are Open Access and generate millions of downloads every month.

For more information, please contact us at: contact@atlantis-press.com

- ▶ PROCEEDINGS
- ▶ JOURNALS
- ▶ BOOKS
- ▶ POLICIES
- ▶ MANAGE COOKIES/DO NOT SELL MY INFO
- ▶ ABOUT
- ▶ NEWS
- ▶ CONTACT
- ▶ SEARCH

[Home](#) [Privacy Policy](#) [Terms of use](#)   

Copyright © 2006-2022 Atlantis Press – now part of Springer Nature

ht

<https://www.atlantis-press.com/proceedings/iccasp-16/25871632>



Dr. G. T. Thampi
PRINCIPAL
Thadomal Shahani Engineering College
Bandra (W), Mumbai - 400 050.



Efficacy of a Classical and a Few Modified Machine Learning Algorithms in Forecasting Financial Time Series



Shilpa Amit Verma, G. T. Thampi and Madhuri Rao

Abstract Financial markets and economy forecast are closely related to each other. Forecast of prices of financial assets is therefore of importance for any economy-planning be it global, national or individual. There are various global, local and psychological factors that affect financial markets making its forecasting a non-trivial, complex problem. Numerous machine learning techniques have been applied by various researchers for a last few decades for making forecasts in various fields including the financial one, with varying degree of success. In the present article, time-series data of NIFTY50 of the National Stock Exchange (NSE) of India is considered as a reference data. Forecasting of its prices is done by applying the classical Gradient Descent Method (GDM) and by a few herein proposed modifications of it. The modifications are essentially using variants of the mean square error function of the classical GDM. All the proposed variants, Mean median (MMD) error function, Minkowski (MKW) error function, Logcosh (LCH) error function and Cauchy (CCY) error function, result in significant improvement in all the efficacy parameters of forecasting. Two widely varying time horizons, monthly and daily, have been considered. Significant enhancement in forecasting efficacy is obtained with the application of the Modified GDM methods in all the data sets: training, testing and out-of-sample.

Keywords Gradient descent method · Forecasting · Machine learning · Stock market · NIFTY50 · Time series

S. A. Verma (✉)

Computer Department, Mumbai University Thadomal Shahani Engineering College, Bandra, Mumbai, India

e-mail: shilpaverma65@gmail.com

G. T. Thampi · M. Rao

IT Department, Mumbai University Thadomal Shahani Engineering College, Bandra, Mumbai, India

e-mail: gthampi@yahoo.com

M. Rao

e-mail: my_rao@yahoo.com

© Springer Nature Switzerland AG 2020

N. Dey et al. (eds.), *Internet of Things, Smart Computing and Technology:*

A Roadmap Ahead, Studies in Systems, Decision and Control 266, https://doi.org/10.1007/978-3-030-39047-1_1

https://doi.org/10.1007/978-3-030-39047-1_1

Dr. G. T. Thampi
PRINCIPAL
Thadomal Shahani Engineering College
Bandra (W), Mumbai - 400 050.



1 Introduction

Machine learning is a procedure or an algorithm where machines directly perform the task at hand with minimal human intervention. There are several aspects of the real world problems which require algorithms to ‘learn’ and improve on its performance based on experience which is referred to as ‘training’. The main aim of training in any machine learning algorithm would therefore be to ‘adapt’, given a certain degree of complexity. There are many complex tasks like speech recognition, image processing, transliteration, healthcare, time series forecasting etc. which require state of the art techniques (machine learning algorithms) that can learn from experience and perform the task at hand. In an interesting study, Jain and Bhatnagar have effectively and efficiently employed AI for monitoring the health and security of patient in a hospital environment [1]. In this method a warning system is activated during unwanted scenarios. In another significant study, Kamal et al. have undertaken the study and prediction of coding regions from diseases infected biological data employing the classifiers namely the support vector machine (SVM), principal component analysis (PCA) technique, Fisher’s discriminant analysis (FDA) and compared the results with the neural mapping skyline filtering (NMSF) [2]. The results obtained from NMSF surpassed others emphasizing the importance of this methodology. In a quite detailed study, Dey N. et al. have employed machine learning methods for smart applications like transliteration for the multilingual support in printing of utility bills, data mining, IOT and security and achieved a high degree of accuracy [3]. Dey N. et al. have also employed IOT and big data driven technologies for next generation healthcare and amply demonstrated the efficacy and importance of these methods for future healthcare industry [4].

Financial markets facilitate, the world around, exchange of financial assets like equities, bonds, commodities and currencies etc. The term equity may be used to describe one’s ownership in a company. The three terms equity price, stock price and share price are often used interchangeably and they would carry the same meaning in the present work. Financial markets forecasts are generally required for economics planning. Stock market forecasting therefore attracts worldwide attention of researchers, investors and traders. Although financial markets of developed countries have been quite comprehensively studied, those of developing countries like India have not been considered to that extent. Moreover these markets are generally more volatile and hence more difficult to forecast.

Share price forecasting may be done in two widely differing ways: fundamental analysis and technical analysis. Fundamental analysis of a company is done by evaluating its past economic performance, expected future demand of its products, credibility of its management, government policies and local and global economics scenario. In technical analysis, on the other hand, future prices of stocks are forecasted solely on the basis of the trends/patterns of their past prices. Historically, according to the proponents of Efficient Market Hypothesis (EMH) and Random Walk Hypothesis (RWH) day-to-day share prices of a company are nothing but

random fluctuations around a central value and therefore, in the short run, it is not possible to forecast the markets with more than 50% accuracy. Subsequently however there have been numerous studies providing evidence contrary to EMH and RWH. A large accumulated and growing experience also suggests that these hypotheses may not be strictly applicable as some analysts regularly make forecasts with more than 50% accuracy. Since last many decades, technical analysts have been observing graphs of share prices determining their support and resistance points based on some rather rough empirical rules. Academicians, corporate houses and traders have been making a lot of efforts to forecast stock prices on various time horizons. Stock price movements are nonlinear, complex and at times discontinuous and, in the short run, they are largely driven by sentiments, fear and greed, of numerous investors and traders and by crowd psychology rendering the forecast a difficult task. Any shift in demand and supply of a company's share results in changes of its market price which often deviates from its intrinsic value. There are no formal mathematical models/equations to describe movements of stock prices. Artificial intelligence (AI) based methods are inherently suited to make forecasts in such scenarios. Advantages of AI methods lie in their ability to model complex and nonlinear stock prices without any prior knowledge of the processes generating them [5]. This advantage has recently been accentuated due to availability of large and fast digital computers and substantial progress in research in numerical computational methods. AI methods can be classified roughly into three categories: Artificial Neural Network (ANN), Fuzzy logic (FL) and Genetic algorithm (GA). Owing to their characteristics of being extremely powerful in extracting trends and patterns in unknown environments, ANNs have become preferred tools for prediction of the financial markets. ANNs are frequently referred to as universal approximators as they are capable of approximating any function. Trained ANNs can be considered experts in the domain of their use [6]. However within the ANN domain, choosing the best model for the problem in hand is an important task [7]. The ultimate test for the best choice is naturally the highest forecasting accuracy. A lot of research work has been done and published in open scientific literature on application of AI methods in forecasting financial markets of developed countries. In the present work focus has been on an Indian stock market index NIFTY50 and efforts have been made to enhance forecasting efficacy by deploying a few error functions different from the one used in the classical GDM. Forecasting efficacy parameters are used to measure degree of success in forecasting. Most popular efficacy parameters are briefly described below [8].

2 Efficacy Parameters of Market Forecasting

Let y_{dt} and y_{ft} denote the actual (desired) and forecasted prices of the stock/index at time 't' respectively. The forecast error is then defined by $e_t = (y_{dt} - y_{ft})$. Let the total number of input-output sets, constructed from the stock prices, be denoted by



ABSTRACT :

Multihomed devices are common in today's environment but are underutilized. Uninterrupted application requirements have leap bounds in terms of throughput requirements. Multipath TCP (MPTCP) is a recent and successfully built standard at transport layer, to achieve the above requirement using multipathing. Long-lived flows carry heavy payload and short-lived flows look for quick response. Scheduling algorithm should consider these requirements and accordingly implement varying strategies to fulfill these needs. Long-lived flows need MPTCP, to get maximum throughput. Short-lived flows can perform with TCP or with slow subpath of MPTCP. To distinguish between short- and long-lived flows and distribute their traffic on appropriate subflow of MPTCP, an intelligent packet scheduling algorithm is required. Research is climbing toward building optimum scheduler for MPTCP. Many packet scheduling algorithms are investigated in this paper for proper path selection, increased throughput, energy efficiency, bandwidth aggregation and receiver buffer optimization, by which issues are listed for them to develop better strategy using newer and advanced algorithms.



Lecture Notes in Mechanical Engineering

Damodar Maity
Pradeep G. Siddheshwar
Sunanda Saha *Editors*

Advances in Fluid Mechanics and Solid Mechanics

Proceedings of the 63rd Congress
of ISTAM 2018

 Springer

Dr. G. T. Thampi
PRINCIPAL
Thadomal Shahani Engineering College
Bandra (W), Mumbai-400 050.

Lecture Notes in Mechanical Engineering



Dr. G. T. Thampi
PRINCIPAL
Thadomal Shahani Engineering College
Bandra (W), Mumbai - 400 050.



Lecture Notes in Mechanical Engineering (LNME) publishes the latest developments in Mechanical Engineering - quickly, informally and with high quality. Original research reported in proceedings and post-proceedings represents the core of LNME. Volumes published in LNME embrace all aspects, subfields and new challenges of mechanical engineering. Topics in the series include:

- Engineering Design
- Machinery and Machine Elements
- Mechanical Structures and Stress Analysis
- Automotive Engineering
- Engine Technology
- Aerospace Technology and Astronautics
- Nanotechnology and Microengineering
- Control, Robotics, Mechatronics
- MEMS
- Theoretical and Applied Mechanics
- Dynamical Systems, Control
- Fluid Mechanics
- Engineering Thermodynamics, Heat and Mass Transfer
- Manufacturing
- Precision Engineering, Instrumentation, Measurement
- Materials Engineering
- Tribology and Surface Technology

To submit a proposal or request further information, please contact the Springer Editor in your country:

China: Li Shen at li.shen@springer.com

India: Priya Vyas at priya.vyas@springer.com

Rest of Asia, Australia, New Zealand: Swati Meherishi at swati.meherishi@springer.com

All other countries: Dr. Leontina Di Cecco at Leontina.dicecco@springer.com

To submit a proposal for a monograph, please check our Springer Tracts in Mechanical Engineering at <http://www.springer.com/series/11693> or contact Leontina.dicecco@springer.com

Indexed by SCOPUS. The books of the series are submitted for indexing to Web of Science.

More information about this series at <http://www.springer.com/series/11236>



Damodar Maity · Pradeep G. Siddheshwar ·
Sunanda Saha
Editors

Advances in Fluid Mechanics and Solid Mechanics

Proceedings of the 63rd Congress
of ISTAM 2018

 Springer



Dr. G. T. Thampi
PRINCIPAL
Thadomal Shahani Engineering College
Bandra (W), Mumbai - 400 050.



Editors

Damodar Maity
Indian Institute of Technology Kharagpur
Kharagpur, West Bengal, India

Pradeep G. Siddheshwar
Bangalore University
Bengaluru, Karnataka, India

Sunanda Saha
Dayananda Sagar University
Bengaluru, Karnataka, India



Dr. G. T. Thampi
PRINCIPAL
Thadomal Shahani Engineering College
Bandra (W), Mumbai - 400 050.



ISSN 2195-4356

ISSN 2195-4364 (electronic)

Lecture Notes in Mechanical Engineering

ISBN 978-981-15-0771-7

ISBN 978-981-15-0772-4 (eBook)

<https://doi.org/10.1007/978-981-15-0772-4>

© Springer Nature Singapore Pte Ltd. 2020

This work is subject to copyright. All rights are reserved by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed.

The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.

The publisher, the authors and the editors are safe to assume that the advice and information in this book are believed to be true and accurate at the date of publication. Neither the publisher nor the authors or the editors give a warranty, expressed or implied, with respect to the material contained herein or for any errors or omissions that may have been made. The publisher remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

This Springer imprint is published by the registered company Springer Nature Singapore Pte Ltd. The registered company address is: 152 Beach Road, #21-01/04 Gateway East, Singapore 189721, Singapore



Dr. G. T. Thampi
PRINCIPAL
Thadomal Shahani Engineering College
Bandra (W), Mumbai - 400 050.



Editorial Board

Dr. P. A. Lakshmi Narayana, Associate Professor, Department of Mathematics, Indian Institute of Technology, Hyderabad

Andy Chan, Professor, Dean of Engineering, Nottingham University, Malaysia Campus

Dr. Annamma Abraham, Professor, Department of Mathematics, BMSIT&M, Yelahanka, Bangalore

Dr. Baljeet Singh, Associate Professor, Department of Mathematics, Post Graduate Government College, Sector 11, Chandigarh, India

Dr. M. S. Jagadeesh Kumar, Associate Professor, Department of Mathematics, VIT, Vellore

Dr. Peri Kameswara Kameswaran, Assistant Professor, Department of Mathematics, VIT, Vellore

Dr. Damodar Maity, Professor, Civil Engineering, IIT, Kharagpur

Dr. Pradeep G. Siddheshwar, Professor, Department of Mathematics, Bangalore University

Dr. T Rajashekhar, Assistant Professor, Mathematics, IIT, Kharapur

Dr. Ramesh Sundaram, Chief Scientist and Head Advanced Composites Division, National Aerospace Laboratories, Kodihalli, Old Airport Road, Bangalore

Bandari Shanker, Department of Mathematics, Osmania University

Dr. K. M. Mrityunjaya Swamy, Associate Professor, Mechanical Engineering,
Dayananda Sagar University

Dr. P. Vivekananda Shanmuganathan, Professor, Mechanical Engineering,
Dayananda Sagar University

Dr. Vinayak B. Hemadri, Associate Professor, Mechanical Engineering, Dayananda
Sagar University



Dr. G. T. Thampi
PRINCIPAL
Thadomal Shahani Engineering College
Bandra (W), Mumbai - 400 050.



Contents

Sun Gravity-Assist to Trans-Lunar Injection Orbits	1
Harishkumar Sellamuthu, Subramanian Arumugam and Ram Krishan Sharma	
On the Recurrence Signatures of Flapping Wings Exposed to Gusty Simple Shear Flow	11
Manabendra M. De, J. S. Mathur and S. Vengadesan	
Unsteady Heat Transfer from a Non-isothermal Axisymmetric Body Immersed in Porous Media Saturated by Nanofluid	27
Shobha Bagai and Mridu Sharma	
Nevanlinna Theory for Finding Meromorphic Solutions of Cubic-Quintic Ginzburg–Landau Equation Arising in Nonlinear Dynamics	39
Adaviswamy Tanuja	
Geometry of Variably Inclined Inviscid MHD Flows	47
Anirban Roy and R. Hari Baskar	
Kinematic Analysis of Theo Jansen Mechanism-Based Eight-Leg Robot	75
Keval Bhavsar, Dharmik Gohel, Pranav Darji, Jitendra Modi and Umang Parmar	
Wave Trapping by Trapezoidal Porous Breakwater	83
Santanu Koley	
Heat and Mass Transfer Due to Double-Diffusion Convection in a Square Porous Enclosure Occupied by Casson Fluid	91
Madhu Aneja and Sapna Sharma	
Convergence of Eigenfunction Expansions for Membrane Coupled Gravity Waves	101
Santanu Koley, K. Panduranga and Dipak K. Satpathi	

Numerical Analysis of Variations on Design Modifications of Train and Tunnel Geometries to Reduce Aerodynamic Drag on Train	109
Vaibhav Rastogi and Nityananda Nandi	
Analysis of Exact Solutions of Electromagneto hydrodynamic Flow and Heat Transfer of Non-Newtonian Casson Fluid in Microchannel with Viscous Dissipation and Joule Heating	117
Motahar Reza and Amalendu Rana	
Pre-clinical Analysis of Implanted Ankle Joint Using Finite Element Method	129
Subrata Mondal and Rajesh Ghosh	
Dynamic Problem of Fractional Thermoelasticity in Bounded Cylindrical Domain with Relaxation Time	139
Gaurav Mittal and V. S. Kulkarni	
A Study on Free Vibration Behavior of Microbeam Under Large Static Deflection Using Modified Couple Stress Theory	155
Sujash Bhattacharya and Debabrata Das	
Fatigue Life Estimation of a Box Girder Bridge Using Coupled and Uncoupled Bridge–Vehicle Dynamics	165
Anjaly J. Pillai, Suvendu Parida and Sudip Talukdar	
Size-Dependent Responses of Timoshenko Beam Incorporating the Strain Gradient Theories of Elasticity	175
Sai Sidhardh	
Dynamic Response of Axisymmetric Functionally Graded Viscothermoelastic Hollow Cylinder Due to Heat Sources by Using Series Solution	185
Himani Mittal and Dinesh Kumar Sharma	
Modeling of a Novel Lower Limb Exoskeleton System for Paraplegic Patients	199
Mrinal Gupta, Jyotindra Narayan and S. K. Dwivedy	
Characterization of Banana and Bagasse Fiber-Reinforced Hybrid Epoxy Composites	211
R. Prem Chand, Y. P. Ravitej and J. V. Shiva Mani Kanta	
Crack Growth Simulation in Quasi-brittle Materials Using a Localizing Gradient Damage Model	223
Alok Negi and Sachin Kumar	
Delamination Damage Analyses of Lap Shear Joints Made with Flat Fibre-Reinforced Polymer Composite Laminates Subjected to Transverse Load	233
Sumeet Kumar Pati, A. K. Pradhan and M. K. Pandit	

Dynamic Characteristics of Twisted Composite Panels—A Finite Element Study 245
 K. S. Shivakumar Aradhya and S. Moorthi

Analytical Solution for Two-Dimensional Axisymmetric Thermoelastic Behavior in the Multilayer Composite Hollow Sphere 255
 N. J. Wange, S. P. Pawar and M. N. Gaikwad

Investigation of Torsional Stability and Camber Test on a Meter Gauge Flat Wagon 271
 Apurba Das and Gopal Agarwal

Nonlinear Dynamic Buckling and Failure Study of Laminated Composite Plates Subjected to Axial Impulse Loads 281
 Vasanth Keshav, S. N. Patel and Rajesh Kumar

Adhesion Failure Analysis in Lap Shear Joint Specimen Subjected to Transverse Loading Made of Curved FGM 297
 Pritam Kumar Kundu, Arun Kumar Pradhan and Mihir Kumar Pandit

A Gradient-Damage Model for Cyclic Behavior of Concrete 309
 A. H. Monnamitheen Abdul Gafoor and D. Dinkler

Reductions of Bending Stresses and Wear in an Aerodynamic Involute Spur Gear Profile 319
 Y. P. Ravitej, O. Abhilash and Naveen kumar

Probability of Failure of a Beam Subjected to Randomly Moving Loads 329
 Alben Jose Kezhiyur, S. Talukdar and Anjaly J. Pillai

Numerical Simulation and Wind Tunnel Experiment on Pressure and Velocity Distribution Around the NACA0012 Airfoil for Optimising an Aerodynamic Model 337
 Motahar Reza, Anindita M. Bhattacharyya, Deepak K. Sadangi and Aman Kumar



ADVANCES IN INTELLIGENT AND *SOFT COMPUTING* 166

David C. Wyld
Jan Zizka
Dhinaharan Nagamalai (Eds.)

Advances in Computer Science, Engineering and Applications

Proceedings of the second International
Conference on Computer Science,
Engineering and Applications (ICCSEA 2012),
May 25-27, 2012, New Delhi, India, Volume 1



Dr. G. T. Thampi
PRINCIPAL

Thadomal Shahani Engineering College
Bandra (W), Mumbai - 400 050.



Improved Watermark Extraction from Audio Signals by Scaling of Internal Noise in DCT Domain	235
<i>Rajib Kumar Jha, Badal Soni, Rajlaxmi Choudhan, Kiyoharu Aizawa</i>	
Performance of Adders with Logical Optimization in FPGA	245
<i>R. Uma, P. Dhavachelvan</i>	
Robust Iris Templates for Efficient Person Identification	255
<i>Abhishek Gangwar, Akanksha Joshi, Renu Sharma, Zia Saqib</i>	
Thesaurus Based Web Searching	265
<i>K.V.N. Sunitha, A. Sharada</i>	
Soft Computing, Artificial Intelligence	
An Adaptive Design Pattern for Genetic Algorithm Based Autonomic Computing System	273
<i>B. Naga Srinivas Repuri, Vishnuvardhan Mannava, T. Ramesh</i>	
Cross-Layer Design in Wireless Sensor Networks	283
<i>S. Jagadeesan, V. Parthasarathy</i>	
Semantic Based Category-Keywords List Enrichment for Document Classification	297
<i>Upasana Pandey, S. Chakraverty, Richa Mihani, Ruchika Arya, Sonali Rathee, Richa K. Sharma</i>	
Selection of Fluid Film Journal Bearing: A Fuzzy Approach	311
<i>V.K. Dwivedi, Satish Chand, K.N. Pandey</i>	
Implementation of New Biorthogonal IOFDM	321
<i>A.V. Meenakshi, R. Kayalvizhi, S. Asha</i>	
Application of Real Valued Neuro Genetic Algorithm in Detection of Components Present in Manhole Gas Mixture	333
<i>Varun Kumar Ojha, Paramarta Datta, Hiranmay Saha, Sugato Ghosh</i>	
Concept Adapting Real-Time Data Stream Mining for Health Care Applications	341
<i>Dipti D. Patil, Jyoti G. Mudkanna, Dnyaneshwar Rokade, Vijay M. Wadhai</i>	
Efficient Domain Search for Fractal Image Compression Using Feature Extraction Technique	353
<i>Amol G. Baviskar, S.S. Pawale</i>	



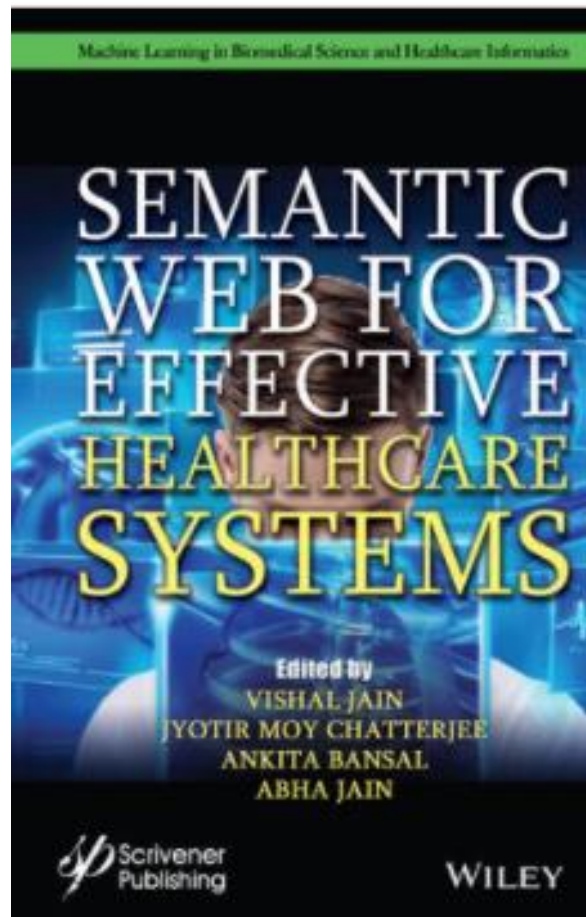
Fast Fingerprint Image Alignment	
<i>Jaspreet Kour, M. Hanmandlu, A.Q. Ansari</i>	
Colour and Texture Feature Based Hybrid Approach for Image Retrieval	
<i>Dipti Jadhav, Gargi Phadke, Satish Devane</i>	
Application of Software Defined Radio for Noise Reduction Using Empirical Mode Decomposition	
<i>Sapan H. Mankad, S.N. Pradhan</i>	
An Approach to Detect Hard Exudates Using Normalized Cut Image Segmentation Technique in Digital Retinal Fundus Image	
<i>Diptoneel Kayal, Sreeparna Banerjee</i>	
Latency Study of Seizure Detection	
<i>Yusuf U. Khan, Omar Farooq, Priyanka Sharma, Nidal Rafiuddin</i>	
Analysis of Equal and Unequal Transition Time Effects on Power Dissipation in Coupled VLSI Interconnects	
<i>Devendra Kumar Sharma, Brajesh Kumar Kaushik, Richa K. Sharma</i>	
Image Analysis of DETECHIP® – A Molecular Sensing Array	
<i>Marcus Lyon, Mark V. Wilson, Kerry A. Rouhier, David J. Symonsbergen, Kiran Bastola, Ishwor Thapa, Andrea E. Holmes, Sharmin M. Sikich, Abby Jackson</i>	
A Gaussian Graphical Model Based Approach for Image Inpainting	15
<i>Krishnakant Verma, Mukesh A. Zaveri</i>	
A Survey on MRI Brain Segmentation	16
<i>M.C. Jobin Christ, R.M.S. Parvathi</i>	
Can Ear and Soft-Biometric Traits Assist in Recognition of Newborn?	17
<i>Shrikant Tiwari, Aruni Singh, Sanjay Kumar Singh</i>	
Multi Segment Histogram Equalization for Brightness Preserving Contrast Enhancement	19
<i>Mohd. Farhan Khan, Ekram Khan, Z.A. Abbasi</i>	
Various Implementations of Advanced Dynamic Signature Verification System	203
<i>Jia Whan Kim</i>	
Performance of Face Recognition Algorithms on Dummy Faces	211
<i>Arund Singh, Shrikant Tiwari, Sanjay Kumar Singh</i>	
Locally Adaptive Regularization for Robust Frame Super Resolution Reconstruction	
<i>S. Chandra Mohan, K. Rajan, P. ...</i>	

Improved Watermark Embedding in DCT Domain to Resist Internal Noise in DCT Domain	
<i>Rajib Kumar Jha, Badal Kumar ...</i>	
Performance of Additive	
<i>R. Uma, P. Dharmachelvan</i>	
Robust Iris Templates	
<i>Abhishek Gangwar, Ak ...</i>	
Thesaurus Based Word	
<i>K.V.N. Saritha, A. Sh ...</i>	
Soft Computing	
An Adaptive Design	
<i>B. Naga Srinivas</i>	
Cross-Layer Design	
<i>S. Jagadeesan</i>	
Semantic Based	
<i>Upasana Patil, Richa K. Sh ...</i>	
Selection	
<i>V.K. Dwivedi</i>	
Implement	
<i>A.V. M ...</i>	
Applic	
<i>Com ...</i>	
<i>Varu ...</i>	
Co	
<i>Ap ...</i>	
<i>D ...</i>	

Contents

Signal, Image Processing and Pattern Recognition	
Automatic FAPs Determination and Expressions Synthesis	1
<i>Narendra Patel, Mukesh A. Zaveri</i>	
Generation of Orthogonal Discrete Frequency Coded Waveform Using Accelerated Particle Swarm Optimization Algorithm for MIMO Radar	13
<i>B. Roja Reddy, M. Uttara Kumari</i>	
Text Independent Speaker Recognition Model Based on Gamma Distribution Using Delta, Shifted Delta Cepstrals	25
<i>K. Suri Babu, Srinivas Yarramalle, Suresh Varma Penumatsa</i>	
Skin Segmentation Based Elastic Bunch Graph Matching for Efficient Multiple Face Recognition	31
<i>Sayantan Sarkar</i>	
A Study of Prosodic Features of Emotional Speech	41
<i>X. Arputha Rathina, K.M. Mehata, M. Ponnaivaikko</i>	
Gender Classification Techniques: A Review	51
<i>Preeti Rai, Pritee Khanna</i>	
Text Dependent Voice Based Biometric Authentication System Using Spectrum Analysis and Image Acquisition	61
<i>Somsubhra Gupta, Soutrik Chatterjee</i>	
Interactive Investigation Support System Design with Data Mining Extension	71
<i>Somsubhra Gupta, Saikat Mazumder, Sourav Mondal</i>	
Pattern Recognition Approaches to Japanese Character Recognition	83
<i>Saumendu Das, Sreeparna Banerjee</i>	





CONTENTS :

TABLE OF CONTENTS

Preface xv

Acknowledgment xix

1 An Ontology-Based Contextual Data Modeling for Process Improvement in Healthcare 1

A. M. Abirami and A. Askarunisa

1.1 Introduction 1

1.1.1 Ontology-Based Information Extraction 3

1.1.2 Ontology-Based Knowledge Representation 4

1.2 Related Work 5

1.3 Motivation 8

1.4 Feature Extraction 9

1.4.1	Vector Space Model	10
1.4.2	Latent Semantic Indexing (LSI)	11
1.4.3	Clustering Techniques	12
1.4.4	Topic Modeling	12
1.5	Ontology Development	17
1.5.1	Ontology-Based Semantic Indexing (OnSI) Model	17
1.5.2	Ontology Development	18
1.5.3	OnSI Model Evaluation	19
1.5.4	Metrics Analysis	23
1.6	Dataset Description	24
1.7	Results and Discussions	25
1.7.1	Discussion 1	29
1.7.2	Discussion 2	29
1.7.3	Discussion 3	30
1.8	Applications	31
1.9	Conclusion	32
1.10	Future Work	33
	References	33

2 Semantic Web for Effective Healthcare Systems: Impact and Challenges 39

Hemendra Shankar Sharma and Ashish Sharma

2.1	Introduction	40
2.2	Overview of the Website in Healthcare	45
2.2.1	What is Website?	45
2.2.2	Types of Website	45
2.2.2.1	Static Website	45
2.2.2.2	Dynamic Website	46
2.2.3	What is Semantic Web?	46



2.2.4	Role of Semantic Web	47
2.2.4.1	Pros and Cons of Semantic Web	49
2.2.4.2	Impact on Patient	51
2.2.4.3	Impact on Practitioner	52
2.2.4.4	Impact on Researchers	52
2.3	Data and Database	53
2.3.1	What is Data?	54
2.3.2	What is Database?	54
2.3.3	Source of Data in the Healthcare System	54
2.3.3.1	Electronic Health Record (EHR)	55
2.3.3.2	Biomedical Image Analysis	56
2.3.3.3	Sensor Data Analysis	57
2.3.3.4	Genomic Data Analysis	57
2.3.3.5	Clinical Text Mining	58
2.3.3.6	Social Media	59
2.3.4	Why Are Databases Important?	60
2.3.5	Challenges With the Database in the Healthcare System	61
2.4	Big Data and Database Security and Protection	61
2.4.1	What is Big Data	61
2.4.2	Five V's of Big Data	62
2.4.2.1	Volume	62
2.4.2.2	Variety	63
2.4.2.3	Velocity	63
2.4.2.4	Veracity	64
2.4.2.5	Value	65
2.4.3	Architectural Framework of Big Data	65
2.4.4	Data Protection Versus Data Security in Healthcare	67



2.4.4.1 Phishing Attacks	67
2.4.4.2 Malware and Ransomware	67
2.4.4.3 Cloud Threats	67
2.4.5 Technology in Use to Secure the Healthcare Data	68
2.4.5.1 Access Control Policy	69
2.4.6 Monitoring and Auditing	69
2.4.7 Standard for Data Protection	70
2.4.7.1 Healthcare Standard in India	70
2.4.7.2 Security Technical Standards	71
2.4.7.3 Administrative Safeguards Standards	71
2.4.7.4 Physical Safeguard Standards	71
References	71

3 Ontology-Based System for Patient Monitoring 75

R. Mervin, Tintu Thomas and A. Jaya

3.1 Introduction	76
3.1.1 Basics of Ontology	77
3.1.2 Need of Ontology in Patient Monitoring	78
3.2 Literature Review	78
3.2.1 Uses of Ontology in Various Domains	78
3.2.2 Ontology in Patient Monitoring System	80
3.3 Architectural Design	80
3.3.1 Phases of Patient Monitoring System	82
3.3.2 Reasoner in Patient Monitoring	87
3.4 Experimental Results	88
3.4.1 SPARQL Results	89
3.4.2 Comparison Between Other Systems	89
3.5 Conclusion and Future Enhancements	90

References 91

4 Semantic Web Solutions for Improvised Search in Healthcare Systems 95

Nidhi Malik, Aditi Sharan and Sadika Verma

4.1 Introduction 95

4.1.1 Key Benefits and Usage of Technology in Healthcare System 96

4.2 Background 97

4.2.1 Significance of Semantics in Healthcare Systems 97

4.2.2 Scope and Benefits of Semantics in Healthcare Systems 98

4.2.3 Issues in Incorporating Semantics 98

4.2.4 Existing Semantic Web Technologies 99

4.3 Searching Techniques in Healthcare Systems 100

4.3.1 Keyword-Based Search 100

4.3.2 Controlled Vocabularies Based Search 101

4.3.3 Improving Searches With Semantic Web Solutions 101

4.3.4 Health Domain-Specific Resources for Semantic Search 102

4.3.4.1 Ontologies 103

4.3.4.2 Libraries 103

4.3.4.3 Search Engines 103

4.4 Emerging Technologies/Resources in Health Sector 108

4.4.1 Elasticsearch 109

4.4.2 BioBERT 109

4.4.3 Knowledge Graphs 110

4.5 Conclusion 110

References 111

5 Actionable Content Discovery for Healthcare 115

Ujwala Bharambe and Anuradha Srinivasaraghavan

5.1 Introduction 116

5.2 Actionable Content 117



Dr. G. T. Thampi
PRINCIPAL
Thadomal Shahani Engineering College
Bandra (W), Mumbai - 400 059.



5.2.1 Actionable Content in Theory	117
5.2.2 Actionable Content in Practice	122
5.3 Health Analytics	124
5.3.1 Artificial Intelligence/Machine Learning-Based Predictive Analytics	125
5.3.2 Semantic Technology for Prescriptive Health Analytics	126
5.4 Ontologies and Actionable Content	127
5.4.1 Ontologies in Healthcare Domain	129
5.5 General Architecture for the Discovery of Actionable Content for Healthcare Domain	130
5.5.1 Ontology-Driven Actionable Content Discovery in Healthcare Domain	131
5.5.2 Case Study for Actionable Content Discovery in Cancer Domain	134
5.6 Conclusion	136
References	136

6 Intelligent Agent System Using Medicine Ontology 139

Tintu Thomas and R. Mervin

6.1 Introduction to Semantic Search	140
6.1.1 What is an Ontology in Terms of Medicine?	140
6.1.2 Needs and Benefits of Ontology in Medical Search	141
6.2 Semantic Search	142
6.2.1 How NLP Works in Semantic Search?	142
6.2.2 Part of Speech Tagging and Chunking	142
6.2.3 Sentence Parsing	143
6.2.4 Discussion About the Various Semantic Search in Medical Databases	144
6.2.5 Discussion About the Retrieval Tools Used in Semantic Search in Medline	145
6.3 Structural Pattern of Semantic Search	146
6.3.1 Architectural Diagram	147
6.3.2 Agent Ontology	148
6.3.3 Rule-Based Approach	149

6.3.4 Reasoners-Based Approach	151
6.4 Implementation of Reasoners	152
6.5 Implementation and Results	153
6.6 Conclusion and Future Prospective	153
References	154

7 Ontology-Based System for Robotic Surgery—A Historical Analysis 159

Ajay Agarwal and Amit Kumar Mishra

7.1 Historical Discourse of Surgical Robots	160
7.2 The Necessity for Surgical Robots	162
7.3 Ontological Evolution of Robotic Surgical Procedures in Various Domains	163
7.4 Inferences Drawn From the Table	164
7.5 Transoral Robotic Surgery	166
7.6 Pancreatoduodenectomy	167
7.7 Robotic Mitral Valve Surgery	168
7.8 Rectal Tumor Surgery	170
7.9 Robotic Lung Cancer Surgery	170
7.10 Robotic Surgery in Gynecology	171
7.11 Robotic Radical Prostatectomy	171
7.12 Conclusion	172
7.13 Future Work	172
References	172

8 IoT-Enabled Effective Healthcare Monitoring System Using Semantic Web 175

Sapna Juneja, Abhinav Juneja, Annu Dhankhar and Vishal Jain

8.1 Introduction	176
8.2 Literature Review	177
8.3 Phases of IoT-Based Healthcare	178
8.4 IoT-Based Healthcare Architecture	179
8.5 IoT-Based Sensors for Health Monitoring	180

8.6 IoT Applications in Healthcare	182
8.7 Semantic Web, Ontology, and Its Usage in Healthcare Sector	183
8.8 Semantic Web-Based IoT Healthcare	183
8.9 Challenges of IoT in Healthcare Industry	185
8.10 Conclusion	186
References	186

9 Precision Medicine in the Context of Ontology 191

Rehab A. Rayan and Imran Zafar

9.1 Introduction	192
9.2 The Rationale Behind Data	195
9.3 Data Standards for Interoperability	197
9.4 The Evolution of Ontology	198
9.5 Ontologies and Classifying Disorders	199
9.6 Phenotypic Ontology of Humans in Rare Disorders	201
9.7 Annotations and Ontology Integration	202
9.8 Precision Annotation and Integration	203
9.9 Ontology in the Contexts of Gene Identification Research	204
9.10 Personalizing Care for Chronic Illness	207
9.11 Roadblocks Toward Precision Medicine	208
9.12 Future Perspectives	209
9.13 Conclusion	209
References	210

10 A Knowledgebase Model Using RDF Knowledge Graph for Clinical Decision Support Systems 215

Ravi Lourdusamy and Xavierlal J. Mattam

10.1 Introduction	216
10.2 Relational Database to Graph Database	217
10.2.1 Relational Database for Knowledge Representation	218

10.2.2 NoSQL Databases	220
10.2.3 Graph Database	223
10.3 RDF	225
10.3.1 RDF Model and Technology	226
10.3.2 Metadata and URI	226
10.3.3 RDF Stores	228
10.4 Knowledgebase Systems and Knowledge Graphs	230
10.4.1 Knowledgebase Systems	230
10.4.2 Knowledge Graphs	232
10.4.3 RDF Knowledge Graphs	233
10.4.4 Information Retrieval Using SPARQL	234
10.5 Knowledge Base for CDSS	235
10.5.1 Curation of Knowledge Base for CDSS	236
10.5.2 Proposed Model for Curation	236
10.5.3 Evaluation Methodology	238
10.6 Discussion for Further Research and Development	239
10.7 Conclusion	239
References	240

11 Medical Data Supervised Learning Ontologies for Accurate Data Analysis 249

B. Tarakeswara Rao, R. S. M. Lakshmi Patibandla, V. Lakshman Narayana and Arepalli Peda Gopi

11.1 Introduction	250
11.2 Ontology of Biomedicine	251
11.2.1 Ontology Resource Open Sharing	254
11.3 Supervised Learning	255
11.4 AQ21 Rule in Machine Learning	256
11.5 Unified Medical Systems	259
11.5.1 Note of Relevance to Bioinformatic Experts	259

11.5.2 Terminological Incorporation Principles 260

11.5.3 Cross-References External 261

11.5.4 UMLS Data Access 262

11.6 Performance Analysis 262

11.7 Conclusion 265

References 265

12 Rare Disease Diagnosis as Information Retrieval Task 269

Jaya Lakkakula, Rutuja Phate, Alfiya Korbu and Sagar Barage

12.1 Introduction 270

12.2 Definition 271

12.3 Characteristics of Rare Diseases (RDs) 272

12.4 Types of Rare Diseases 273

12.4.1 Genetic Causes 274

12.4.2 Non-Genetic Causes 275

12.4.3 Pathogenic Causes (Infectious Agents) 275

12.4.4 Toxic Agents 275

12.4.5 Other Causes 276

12.5 A Brief Classification 276

12.6 Rare Disease Databases and Online Resources 277

12.6.1 European Reference Network: ERN 277

12.6.2 Genetic and Rare Diseases Information Center: GARD 278

12.6.3 International Classification of Diseases, 10th Revision: ICD-10 279

12.6.4 Orphanet-INSERM (Institut National de la Santé et de la Recherche Médicale) 280

12.6.5 Medical Dictionary for Regulatory Activities: MedDRA 280

12.6.6 Medical Subject Headings: MeSH 281

12.6.7 Online Mendelian Inheritance in Man: OMIM 282

12.6.8 Orphanet Rare Disease Ontology: ORDO 282



12.6.9 UMLS: Unified Medical Language System	282
12.6.10 SNOMED-CT: Systematized Nomenclature of Human and Veterinary Medicine— Clinical Terms	283
12.7 Information Retrieval of Rare Diseases Through a Web Search and Other Methods	284
12.7.1 What is Information Retrieval (IR)?	284
12.7.2 Listed Below Are Some of the Methods for Information Retrieval	284
12.7.2.1 Web Search for a Diagnosis	284
12.7.2.2 Cause of Diagnostic Errors in Web-Based Tools	285
12.7.2.3 Nonprofessional Use of Web Tool for Diagnosis	285
12.7.2.4 Performance of Web Search Tools	285
12.7.2.5 Design of Watson	286
12.8 Tips and Tricks for Information Retrieval	287
12.9 Research on Rare Disease Throughout the World	288
12.10 Conclusion	290
References	290

13 Atypical Point of View of Semantic Computing in Healthcare 293

L. Mayuri and K. M. Mehata

13.1 Introduction	294
13.2 Mind the Language	295
13.2.1 Why Words Matter	296
13.2.2 What Words Matter	296
13.2.3 How Words Matter	297
13.3 Semantic Analytics and Cognitive Computing: Recent Trends	297
13.3.1 Semantic Data Analysis	298
13.3.2 Semantic Data Integration	299
13.3.3 Semantic Applications	300
13.4 Semantics-Powered Healthcare SOS Engineering	302
13.5 Conclusion	303

References 304

14 Using Artificial Intelligence to Help COVID-19 Patients 309

Ayush Hans

14.1 Introduction 310

14.2 Method 313

14.3 Results 314

14.4 Discussion 315

14.4.1 What is the Use of AI in Healthcare? 315

14.4.2 How to Use AI for Critical Care Units 315

14.4.2.1 Input Stage 315

14.4.2.2 Process Stage 316

14.4.2.3 Output Stage 317

14.5 Conclusion 320

Acknowledgment 321

References 321

Index 325

