About the Authors

Dr.P.VINAY BHUSHAN is technical wizard in the field of Computer science and Engineering. Has over 11 Years of academic experience in the Teaching as well as research coupled with administrative skills. Presently Working for Mallareddy University, Maisammaguda, Dulapally, Hyderabad, Telangana, India as an Assistant Professor in the stream of Computer Science and Engineering Department. He is awarded PhD in the area of Data Mining from Sri Satya Sai University of Technology and Medical Sciences, Sehore, Madhya Pradesh. As many Patents to his credit and holds life memberships namely IAENG, SWIDC, RSBI. Presented good number of papers at various international conferences and published journals of global repute. His areas of keen interest are Data Mining, Cloud Computing, Network Programming and he is good at OOPS Concepts. He is very well known as student friendly Professor in the entire student Community. He is Certified holder of globally acclaimed courses namely Code Teams, Coursera. To bring to the fore the conceptual talents of the students, so as to make them industry ready.

Dr.NARASIMHA CHARY CH is working as an Associate Professor in the Department of Computer Science & Engineering at Sri Indu College of Engineering & Technology, Facing Main Road, Ibrahimpatnam Mandal, Rangareddy District, Auditorium - Sri Indu Main Block Rd, Sheriguda, Telangana 501510.He was awarded Ph. D from Sri Satya Sai University of Technology & Medical Sciences, Sehore, Pachama, Madhya Pradesh. He has 20+ years of teaching experience, published nearly 20+ papers in international peer-reviewed journals and 10+ Conferences. He guided 60 UG projects. He authored 2 Text Books. He filed 3 Patents, and an Editorial board member. He is a life member of the ISTE, IAENG, TERA, ICSES, SDIWC and CSI-SBC MEMBERSHIP for Hyderabad. He secured infosys campus connect-best faculty award- 2018 bronze medal and infosys campus connect-best faculty award- 2019 silver medal. His research areas include, Data Mining, Big-Data, Cloud Computing, Artificial Intelligence, Machine Learning and Database Management Systems.

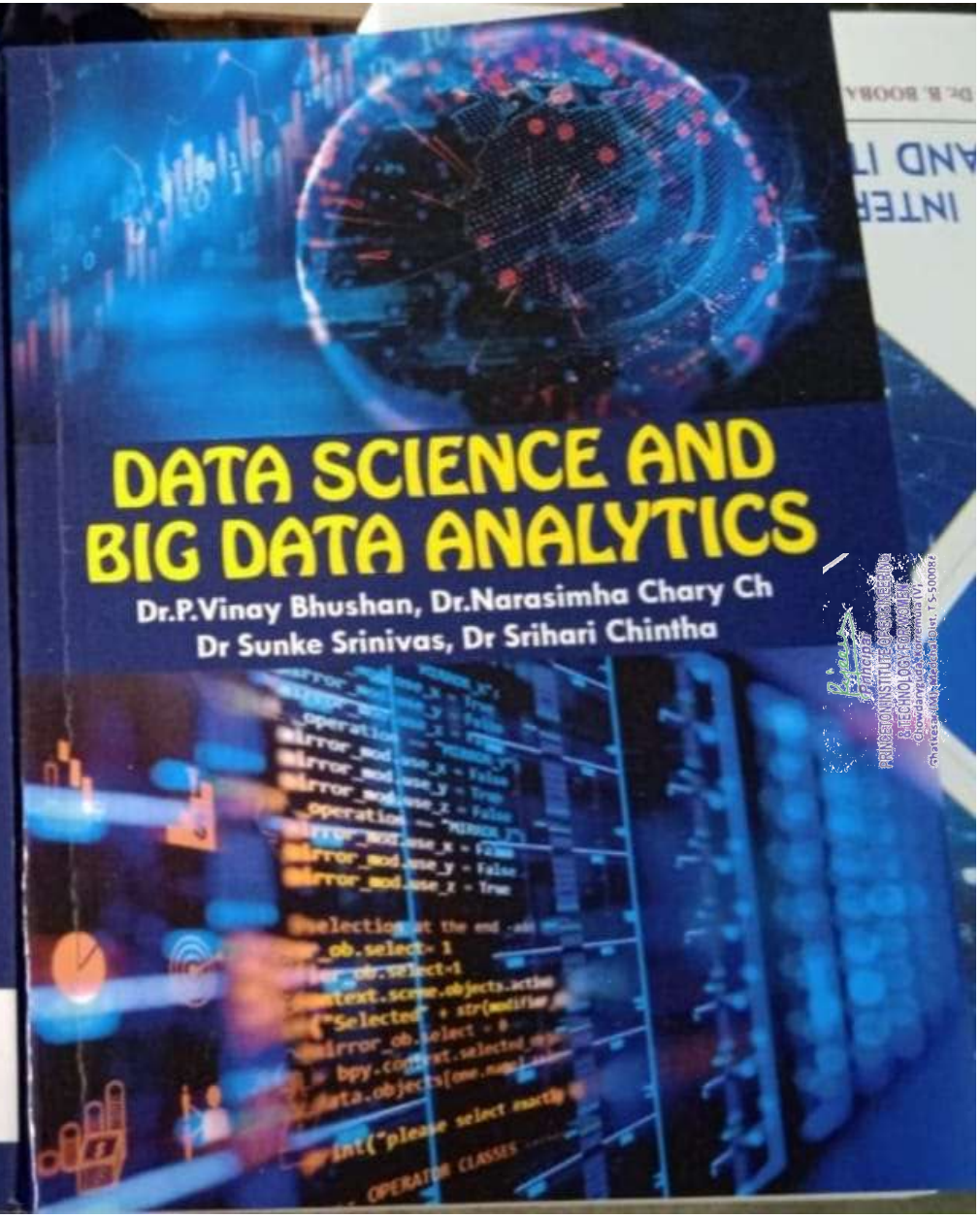
Dr. SUNKE SRINIVAS is a Associate Professor and Head of Data Science Department, Holy Mary Institute of Technology & Science (College of Engineering), Hyderabad, Telangana, India. He received doctorate from Sri Satya Sai University of Technology and Medical Sciences, Madhya Pradesh. He had been in teaching and research for past 20 years. He has published number of research papers in international/ National journals and conferences. He has written a Book Title is "Introduction to Artificial Intelligence". He is a lifetime member of AICTSP, NTA and ISTE. His completed 10 online certification courses from reputed universities and organizations like Coursera. He got 4th SAAP International scientific BEST SENIOR FACULTY OF THE DECADE AWARD and He got award for his TEACHING EXCELLENCE. His research areas include, Data Mining, Big-Data, Cloud Computing, Artificial Intelligence, Machine Learning and Database Management Systems.

Dr SRIHARI CHINTHA is working as an Professor in the Department of Computer Science & Engineering at Princeton Institute of Engineering and Technology for Women, Chowdaryguda Village, Ghatkesar Mandal, Hyderabad - 500088.He was awarded Ph. D from Sunrise University, Alwar, Bagar Rajput, Rajasthan 301028. He has 20+ years of teaching experience, published nearly 20+ papers in international peer-reviewed journals and 10+ Conferences. He guided 60 UG projects. He authored 2 Text Books. He filed 3 Patents, and an Editorial board member. He is a life member of the ISTE, IAENG, TERA, ICSES, SDIWC and CSI-SBC MEMBERSHIP for Hyderabad. His research areas include Wireless Networks, Artificial Intelligence, Machine Learning and Database Management Systems.

DI ACADEMEE

BEDCAN INTERNATIONAL ACADEMIC PUBLISHERS

ISO 9001



DATA SCIENCE AND BIG DATA ANALYTICS

DECCAN INTERNATIONAL ACADEMIC PUBLISHERS
ISO 9001-2015 CERTIFIED
INDIA

Book Title	**DATA SCIENCE AND BIG DATA ANALYTICS**
Authors	**Dr.P.Vinay Bhushan** **Dr.Narasimha Chary Ch** **Dr Sunke Srinivas** **Dr Srihari Chintha**
Book Subject	**DATA SCIENCE AND BIG DATA ANALYTICS**
Book Category	Authors Volume
Copy Right	© Authors
Edition	First Edition, September , 2022
Book Size	B5
Price	Rs.999/-

Published by
DECCAN INTERNATIONAL ACADEMIC PUBLISHERS
India

ISBN Supported by International ISBN Agency,
United House, North Road, London, N7 9DP, UK. Tel. + 44 207 503 6418 &
Raja Ram Mohan Roy National Agency for ISBN
Government of India, Ministry of Human Resource Development,
Department of Higher Education, New Delhi – 110066 (India)

ISBN: 978-93-95191-17-3

PREFACE

This book aims to provide a broad **DATA SCIENCE AND BIG DATA ANALYTICS** for the importance of **DATA SCIENCE AND BIG DATA ANALYTICS** is well known in various engineering fields.

It provides a logical method of explaining various complicated concepts and stepwise methods to explain essential topics. Each chapter is well supported with the necessary illustrations. All the chapters in the book are arranged in a proper sequence that permits each topic to build upon earlier studies.

DATA SCIENCE AND BIG DATA ANALYTICS is a critical research area. The techniques developed in this area so far require to be summarized appropriately. In this book, the fundamental theories of these techniques are introduced.

The brief content of this book is as follows-

CHAPTER 1 INTRODUCTION TO DATA SCIENCE
CHAPTER 2 INTRODUCTION TO BIG DATA
CHAPTER 3 DATA COLLECTION AND DATA PRE-PROCESSING
CHAPTER 4 MODEL EVALUATION GENERALIZATION
CHAPTER 5 WORKING BIG DATA WITH HADOOP
CHAPTER 6 MAP REDUCE
CHAPTER 7 DATA SCIECNE APPLICATIONS AND CASE STUDIES
CHAPTER 8 BIGA DATA ANALYTICES APPILICATIONS

This book is original in style and method. No pains have been spared to make it as compact, perfect, and reliable as possible. Every attempt has been made to make th[illegible] one.

PREFACE

This book aims to provide a broad **DATA SCIENCE AND BIG DATA ANALYTICS** for the importance of **DATA SCIENCE AND BIG DATA ANALYTICS** is well known in various engineering fields.

It provides a logical method of explaining various complicated concepts and stepwise methods to explain essential topics. Each chapter is well supported with the necessary illustrations. All the chapters in the book are arranged in a proper sequence that permits each topic to build upon earlier studies.

DATA SCIENCE AND BIG DATA ANALYTICS is a critical research area. The techniques developed in this area so far require to be summarized appropriately. In this book, the fundamental theories of these techniques are introduced.

The brief content of this book is as follows-

CHAPTER 1 INTRODUCTION TO DATA SCIENCE
CHAPTER 2 INTRODUCTION TO BIG DATA
CHAPTER 3 DATA COLLECTION AND DATA PRE-PROCESSING
CHAPTER 4 MODEL EVALUATION GENERALIZATION
CHAPTER 5 WORKING BIG DATA WITH HADOOP
CHAPTER 6 MAP REDUCE
CHAPTER 7 DATA SCIECNE APPLICATIONS AND CASE STUDIES
CHAPTER 8 BIGA DATA ANALYTICES APPILICATIONS

This book is original in style and method. No pains have been spared to make it as compact, perfect, and reliable as possible. Every attempt has been made to make the book a unique one.

In particular, this book can be handy for practitioners and engineers interested in this area. Hopefully, the chapters presented in this book have just done that.

ACKNOWLEDGMENTS

Take it from me, writing a book takes time, patience, and motivation in equal measures. The challenges can sometimes be overwhelming, and it becomes straightforward to lose focus. However, analytics, patterns, and uncovering the hidden meaning behind data have always attracted me. When one considers the possibilities offered by comprehensive analytics and the inclusion of what may seem to be unrelated databases, the effort involved seems almost inconsequential.

We also have to acknowledge [illegible] many vendors in the Internet of Things arena who inadvertently [illegible] along my journey to expose the value contained [illegible]

Writing takes a great deal of energy and can quickly consume all of

LAMBERT PUBLICATIONS

FUNDAMENTAL OF DATA SCIENCE

Prof. Pinkal Jain
Dr. G. Malleshama
Dr Rajeev Shrivastava
Dr. Santosh Kumar

Lambert Publication's

ISBN: 978-93-91265-72-4

Fundamental of Data Science

By ...

Prof. Pinkal Jain

Dr. G. Malleshama

Dr Rajeev Shrivastava

Dr. Santosh Kumar

Lambert Publication's

First Edition 2022 – Rs. 250 /- (Two Hundred and Fifty Only)

Fundamental of Data Science

By Prof. Pinkal Jain, Dr. G. Malleshama, Dr Rajeev Shrivastava, Dr. Santosh Kumar

ISBN: 978-93-91265-72-4

www.ijarsct.co.in

Fundamental of Data Science

Course Objectives

- To understand the various stages in Data science process.
- To study the applications of Data Science.
- To learn how to setup the environment and implement in Python and R
- To learn to write programs in Python and R for data science projects.
- To know the process of data visualization & data manipulation w.r.to data science.

UNIT - I: Introduction to Data Sciences – The data science process – Roles in data science project – Stages of Data Science Project – Defining the goal – Data collection and Management – Modelling – Model evaluation – Presentation and documentation – Model deployment and Maintenance Applying Data science in Industry – Benefits from Business centric Data Science – Data Analytics and Types – Common Challenges in Analytics – Distinguishing between Business Intelligence and Data Science Using Data Science to Extract meaning from Data – Machine learning – Math Probability and Statistical Modelling – Using clustering to subdivide data – Modeling with instances – Building models that operate on IOT

UNIT- II: Data science tools environment - Python – overview - Setting up Data science toolbox, Essential concepts and tools-Obtaining Data – creating reusable command line tools – scrubbing data – Managing your Data workflow – Drake - Exploring Data – Parallel pipelines – Modeling Data.

UNIT – III: Techniques using Python Tools -Linear Algebra – Statistics – Probability – Hypothesis and Inference – Gradient Descent – working with Data – Machine Learning – k – Nearest Neighbours – Naive Bayes – Simple Linear Regression – Multiple Regression – Logistic Regression

UNIT - IV: Techniques using R Tools - R programming overview - Loading data into R – Modeling methods – choosing and evaluating models – Memorization methods - Linear and logistic Regression – Unsupervised methods – Delivering results – Documentation and Deployment.

UNIT-V: Data Manipulation and visualization – Data manipulation using pandas – visualization using matplotlib.

TEXT BOOKS:

1. J. Janssens, Data science at the command line, First edition. Sebastopol, CA: O'Reilly,2014..
2. J. Grus, Data Science from Scratch: First Principles with Python, 1 edition. Sebastopol, CA: O'Reilly Media,2015.
3. N. Zumel and J. Mount, Practical data science with R. Shelter Island, NY: Manning Publications Co,2014.

REFERENCE BOOKS:

1. L. Pierson and J. Porway, Data science, 2nd edition. Hoboken, NJ: John Wiley and Sons, Inc, 2017.
2. C. O'Neil and R. Schutt, Doing Data Science: Straight Talk from the Frontline, 1 edition. Beijing ; Sebastopol: O'Reilly Media,2013.
3. J. Vander Plas, Python Data Science Handbook: Essential Tools for Working with Data, First edition. Shroff/O'Reilly,2016.
4. S. R. Das, Data Science: Theories, Models, Algorithms, and Analytics. https://srdas.github.io/MLBook/.

Course Outcomes:

Students will be able to

- Demonstrate the basic knowledge of data science process.
- Setup the software environment for python and R Language and apply various techniques to work with data.
- Manipulate and visualize the data using tools like pandas and matplotlib.
- Develop simple data science applications.
- Analyze the various data science related projects.

****

About Authors

Prof. Pinkal Jain is working as an Assistant Professor in the Department of Computer Science and Engineering at Gyan Ganga College of Technology, Jabalpur, Madhya Pradesh, India. He graduated in B.E. in Computer Science & Engineering from LNCT Jabalpur; Madhya Pradesh India .He secured Master of Technology in Computer Technology and Application at Vindhya institute of technology, Madhya Pradesh India. He is currently pursuing his PhD in Computer Science and Engineering. He has been engaged in research and teaching for more than 10 years. He has published and presented more than 10 papers in National and International Journals and conferences. He is having 02 Patents. His main area of interest includes Artificial Intelligence, Cryptography, Data Science, Data mining, Operating System, Software Engineering, Theory of Computation, Compiler Design and Cloud Computing.

Dr. G. Malleshama is working as Professor and Head of the Department in Indur Institute of Engineering and Technology Siddipet, Telanagna (India). He has completed Ph.D. in Embedded Systems Design From Sunrise University, Alwar, (Rajasthan.) (2017). He has completed his M.Tech in Elecronics and communication in from JNTUH, Hyderabad (2013). He has completed B. Tech in Electronics and Communication Engineering from M.V.S.R Hyderabad (2003).He has 16 years of academic experience and has 15 international journal and 10 National and International conference publications. Prof Mallesham's area of interest is in the field of IOT, AI & ML , VLSI testing and Verification ,a front end VLSI Design..He has guided 18 M.Tech scholars and 2 PhD Student in field of Embedded Design and Communication Engineering. Attended various workshops and Seminar. He has 5 Indian patents grants..

Dr Rajeev Shrivastava BE, ME , Ph.D. , is presently working as a Principal in Princeton Institute of Engineering and Technology for women Hyderabad (JNTUH Affiliated). I have completed B.E. (Electronics & Communication) in 2002 from GRKIST Jabalpur. I have completed M.E (Digital Communication) in 2009 from SRIT Jabalpur. I have completed PhD (ECE) in December 2016 from JVWU Jaipur. My total work experience in academic field is nearly 16 years. Also, I have industrial experience of 3 years. I have published 14 Patents, 50 international papers and presented 32 papers in international/national conferences. I have also published two books on JRF International Book in 2017 and 2018. I also got young scientist award in 2017 and International award for teachers with higher potential in 2018

Dr Santosh Kumar, currently working as Associate Professor (Senior Scale) in the School of Computer Science and Engineering, Manipal University Jaipur. He holds rich experience of 19 years in teaching and research domain in prestigious institutes and universities. He has been in various administrative capacities such as Director (Quality & Compliance), Director (Research) Coordinator & Chairman for different Academic Committees at both undergraduate and postgraduate levels. Dr Santosh has driven various accreditation programs at UG & PG Level. He holds special interest in the implementation of education technology. He has worked on define aims, objectives, outline course contents, method of teaching and assessment. He promotes use of education technology that's cover great understanding of learning and differences between individuals. As a teacher, he has enjoyed working in various fields and deriving satisfaction for contributing to society through education technology. He played a significant role to mentor new teachers. His research area includes education technology, predictive analysis, trend detection, regression, classification and clustering algorithms in data mining and text mining. He had implemented various information retrieval models and have key interest on term pre-processing methods, term weighting, term pruning strategy and related areas.

Principal
PRINCETON INSTITUTE OF ENGINEERING
& TECHNOLOGY FOR WOMEN
Chowdaryguda, Korremula (V),
Ghatkesar (M), Medchal Dist, T S-500088

Table of Contents

****

IJARSCT
Impact Factor: 6.252
Scientific Journal Impact Factor
www.sjifactor.com

Lambert
Publication's

Network Theory

Prof. Ravi Mohan
Ms. Garima Tiwari
Dr Rajeev Shrivastava
Mr. Rakesh Patel

Crossref
DOI: 10.48175/568
www.doi.org
www.ijarsct.co.in

Network Theory

By ...

Prof Ravi Mohan

Shri Ram Institute of Technology, Jabalpur, M.P, India

Ms. Garima Tiwari

Jabalpur Engineering College, Jabalpur, M.P. India

Dr Rajeev Shrivastava

Princeton Institute of Engineering and Technology for Women, Hyderabad

Mr. Rakesh Patel

Jabalpur Engineering College, Jabalpur, M.P. India

Lambert Publication's

First Edition 2022 – Rs. 250 /- (Two Hundred and Fifty Only)

Network Theory

By Prof. Ravi Mohan, Ms. Garima Tiwari, Dr. Rajeev Shrivastava, Mr. Rakesh Patel

ISBN: 978-93-91265-07-6

www.ijarsct.co.in