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.

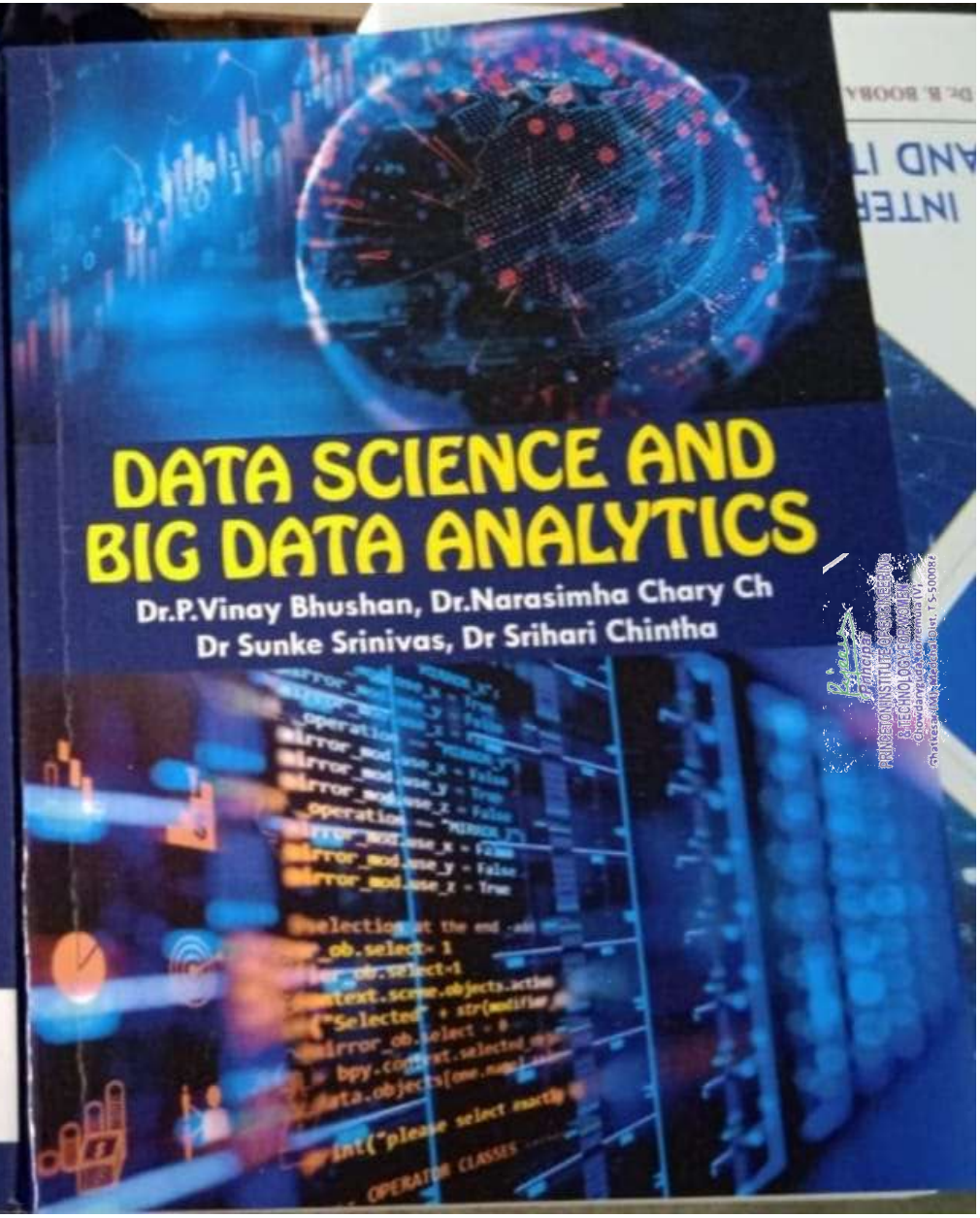
Dr. SUNKE SRINIVAS is a Associate Professor and Head of Data Science Department, Holy Mary Institute of Technology & Science (College of Engineering), Hyderabad, Telangana, India. He received doctorate from Sri Satya Sai University of Technology and Medical Sciences, Madhya Pradesh. He had been in teaching 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