



Mr. T. Rajesh

# BIG DATA ANALYTICS & APPLICATIONS

Archers & Elevators Publishing House  
ISBN:978-81-19385-34-8

# **BIG DATA ANALYTICS AND APPLICATIONS**

Mr.T.Rajesh  
Assistant Professor, Department of CSE,  
GNITS,Hyderabad.

# PREFACE

## Big Data Analytics and Applications

Welcome to "Big Data Analytics and Applications." This textbook aims to provide a comprehensive and accessible introduction to the fascinating field of Big Data Analytics. With the rapid advancements in technology and the increasing integration of data into various domains, it has become essential for students, researchers, and practitioners to understand the foundations and applications.

This book is designed to cater to a wide range of readers, including undergraduate and graduate students, professionals from diverse fields seeking to understand big data, and enthusiasts who are curious about the capabilities and implications of this transformative technology. The primary goal is to present a balanced blend of theoretical concepts, practical techniques, and real-world examples to foster a deeper understanding of data.

Big Data Analytics and Applications is organized into twelve comprehensive chapters that cover a wide spectrum of topics. It begins with an introduction to big data, tracing its historical development and highlighting the key concepts and terminology. The subsequent chapters delve into intelligent agents, problem-solving and search algorithms, knowledge representation and reasoning, machine learning, natural language processing, computer vision, robotics and intelligent systems, big data in healthcare, big data in business applications, big data ethics, and responsible big data. The final chapter explores future perspectives and trends in big data, offering insights into cutting-edge research areas and the potential impact of big data on society.

Throughout this book, we emphasize a practical and hands-on approach, providing examples, case studies, and exercises to reinforce the understanding of big data principles and techniques. It is our aim to strike a balance between theory and practice, allowing readers to grasp the fundamental concepts while gaining practical insights into big data applications across diverse domains.

While creating this textbook, we have been mindful of the importance of avoiding plagiarism. All the content within this book is original, and we have taken great care to ensure that sources are properly cited. We encourage readers to explore the references and further readings provided at the end of each chapter to delve deeper into specific topics and gain a more comprehensive understanding.

We would like to express our gratitude to the researchers, practitioners, and educators who have contributed to the development of the field of big data and made it possible to compile this textbook. Their dedication and innovative contributions have paved the way for the advancements we witness today. We also extend our appreciation to the reviewers who provided valuable feedback and helped shape this book.

We hope that "Big Data Analytics and Applications" will serve as a valuable resource in your journey to explore and understand the captivating realm of big data. Our aspiration is that this book will inspire curiosity, foster critical thinking, and empower readers to contribute to the responsible development and application of AI.

## INDEX

S.No	Unit	Topic	Pg.No.
		<b>INTRODUCTION TO BIG DATA AND ANALYTICS</b>	
1	I	Classification of Digital Data, Structured and Unstructured Data - Introduction to Big Data	1
2	I	Why Big Data Traditional Business Intelligence versus Big Data - Data Warehouse and Hadoop	4
3	I	Environment Big Data Analytics: Classification of Analytics – Challenges - Big Data Analytics importance	5
4	I	Data Science - Data Scientist - Terminologies used in Big Data Environments	10
5	I	Basically, Available Soft State Eventual Consistency - Top Analytics Tools	12
		<b>INTRODUCTION TO TECHNOLOGY LANDSCAPE</b>	
7	II	NoSQL, Comparison of SQL and NoSQL, Hadoop - RDBMS Versus Hadoop - Distributed Computing	15
8	II	Challenges – Hadoop Overview - Hadoop Distributed File System - Processing Data with Hadoop -	20
9	II	Managing Resources and Applications with Hadoop YARN - Interacting with Hadoop Ecosystem	22
		<b>INTRODUCTION TO MONGODB AND MAPREDUCE PROGRAMMING</b>	
10	III	MongoDB: Why MongoDB - Terms used in RDBMS and MongoDB - Data Types - MongoDB Query Language	24
11	III	MapReduce: Mapper – Reducer – Combiner – Partitioner – Searching – Sorting – Compression	36
		<b>INTRODUCTION TO HIVE AND PIG</b>	
12	IV	Hive: Introduction – Architecture - Data Types - File Formats - Hive Query Language Statements	52
13	IV	Partitions – Bucketing – Views - Sub- Query – Joins – Aggregations - Group by and Having - RC File	70
14	IV V	Implementation - Hive User Deserializ Defined Function - Serialization and ation. Pig: Introduction	75
15	IV	Anatomy – Features – Philosophy - Use Case for Pig - Pig Latin Overview - Pig Primitive Data Types	76
16	IV	Running Pig - Execution Modes of Pig - HDFS Commands - Relational Operators - Eval Function	79
17	IV	Complex Data Types - Piggy Bank - User - Defined Functions - Parameter Substitution - Diagnostic	82
18	IV	Operator - Word Count Example using Pig - Pig at Yahoo! - Pig Versus Hive	93

19	V	INTRODUCTION TO DATA ANALYTICS WITH R Machine Learning: Introduction, Supervised Learning, Unsupervised Learning, Machine Learning	96
20	V	Algorithms: Regression Model, Clustering, Collaborative Filtering, Associate Rule Making, Decision Tree, Big Data Analytics with BigR.	97