

Data Science and Predictive Analytics with Python

Dr.M.Mahalakshmi

Principal

R.B.Gothi Jain College for Women,
Redhills, Chennai

Dr.K.Kasturi

Associate Professor

Applied Computing and Emerging Technologies,
School of Computing Sciences,
Vels Institute of Science Technology and Advanced
Studies(VISTAS), Chennai.

Seema Nagaraj

Assistant Professor

Department of MCA

Bangalore Institute of Technology
KR Road, VV Pura
Bengaluru-560004

AN PUBLICATIONS

ISO Certified International Publisher, Regd in MSME Govt. of India

**No: 29, Moorthy Street, Balavinayagar nagar,
Tiruvallur-602001,
Tamilnadu, India**

Title: Data Science and Predictive Analytics with Python

Authors: Dr.M.Mahalakshmi , Dr.K.Kasturi , Seema Nagaraj

ISBN: 978-81-687834-8-5

Published 2026 by AN PUBLICATIONS

© AN PUBLICATIONS

All rights reserved. No part of this publication may be reproduced, stored in a retrieval system, or transmitted, in any form or by any means (electronic, mechanical, photocopying, recording or otherwise), without the prior written permission of the publisher and Author.

The contents of this book is expressed by the authors and they are the responsible for the same.

AN Publications

No:29, Moorthy street, Balavinayagar Nagar,
Tiruvallur-602001,Tamilnadu, India.

Preface

In today's rapidly evolving digital world, data has become one of the most valuable resources for organizations, researchers, industries, and society. The enormous volume, velocity, and variety of data generated from diverse sources have created both significant opportunities and complex challenges. The ability to transform raw data into meaningful insights and use those insights to make informed decisions has therefore become an essential skill across almost every discipline. Data Science and Predictive Analytics provide the methodologies, tools, and techniques required to address these challenges effectively.

The book Data Science and Predictive Analytics with Python has been designed to provide a comprehensive and practical introduction to the concepts, techniques, and applications of data science and predictive analytics using Python. Python has emerged as one of the most widely adopted programming languages for data analysis, machine learning, artificial intelligence, and scientific computing because of its simplicity, flexibility, extensive libraries, and strong community support. This book integrates theoretical foundations with practical implementation to help readers develop both conceptual understanding and problem-solving skills.

The book begins with the fundamental concepts of data science, data analytics, and the data science life cycle. It introduces different types of data, data sources, data collection, data preprocessing, data cleaning, data transformation, and exploratory data analysis. Readers are gradually introduced to Python programming and important libraries such as NumPy, Pandas, Matplotlib, and Seaborn, which form the foundation for effective data analysis and visualization.

A major focus of the book is predictive analytics and machine learning. The concepts of supervised and unsupervised learning are explained systematically, followed by practical discussions of regression, classification, clustering, dimensionality reduction, and model evaluation. Important algorithms and techniques are presented with suitable examples so that readers can understand not only how a model works but also when and why a particular technique should be applied.

The book also emphasizes the complete predictive analytics workflow, including feature engineering, training and testing datasets, performance evaluation, model selection, overfitting and underfitting, and model optimization. Appropriate evaluation measures and visualization techniques are discussed to help readers critically interpret the performance and limitations of predictive models.

We hope that Data Science and Predictive Analytics with Python will serve as a valuable learning resource and provide readers with the knowledge and practical skills required to transform data into actionable intelligence. We also hope that the concepts and techniques presented in this book will inspire students and researchers to explore innovative applications and contribute meaningfully to the rapidly growing field of data science.

We sincerely welcome constructive suggestions and feedback from readers, educators, researchers, and practitioners, which will help us improve future editions of this book.

Authors

Authors

Dr.M.Mahalakshmi, Principal, R B.Gothi Jain College for Women, Redhills,Chennai. An academician with over 23 years of rich experience in higher education. Areas of expertise include Artificial intelligence, Big Data and Java. Apart from this expertise includes Academic leadership, NAAC & Quality assurance. Published 30 research papers in reputed national & international journals, published 5 books, 17 chapters and presented papers at various conferences. Organized more than 15 conferences, 20 Faculty and student development programmes. Passionate about student development, institutional growth and outcome-based education.

Dr.K.Kasturi is currently working as Associate Professor in the Department of Applied Computing and Emerging Technologies ,School of Computing Sciences,Vels Institute of Science Technology and Advanced Studies(VISTAS), Chennai, India. Her experience includes, as a teaching faculty for more than nineteen years. As part of the research work, she has published more than 40 articles in national and international journals from which 18 are scopus indexed. She has presented more than 20 papers in national and international conferences and published 8 patents with one Grant. She has published 10 book chapters in computer science and Application Streams. At present, she is guiding 6 Ph.D research scholars and Produced 3 PhD awardee.She has received the Young Researcher award, Excellence in Teaching award and Best faculty award . She is a active member in ResearchGate Society and her reads crosses more than 9960 reads.Her research area includes Artificial Intelligence, Machine Learning,Deep Learning, Data Mining,Computer Networks, Network Security, Block chain, Internet of Things and Image Processing. My hearty thanks to my family and VISTAS family for supporting me to complete the work.

Prof.Seema Nagaraj presently working as Asst Professor in the Dept of MCA at Bagalore Institute of Technology, Bangalore. Her proficiency in Teaching is spread over the past 18 years. She has published good number of Scopus indexed journals. Her area of interest includes AI, Machine Learning and Deep Learning, Cryptography & Green Computing.