# Applied Analytics through Case Studies Using SAS and R

## Implementing Predictive Models and Machine Learning Techniques

**Deepti Gupta**

Apress®

*Applied Analytics through Case Studies Using SAS and R*

Deepti Gupta
Boston, Massachusetts, USA

ISBN-13 (pbk): 978-1-4842-3524-9 ISBN-13 (electronic): 978-1-4842-3525-6
https://doi.org/10.1007/978-1-4842-3525-6

Library of Congress Control Number: 2018952360



Managing Director, Apress Media LLC: Welmoed Spahr
Acquisitions Editor: Celestin John
Development Editor: James Markham
Coordinating Editor: Divya Modi

Cover designed by eStudioCalamar

Cover image designed by Freepik (www.freepik.com)

Distributed to the book trade worldwide by Springer Science+Business Media New York, 233 Spring Street, 6th Floor, New York, NY 10013. Phone 1-800-SPRINGER, fax (201) 348-4505, e-mail orders-ny@springer-sbm.com, or visit www.springeronline.com. Apress Media, LLC is a California LLC and the sole member (owner) is Springer Science + Business Media Finance Inc (SSBM Finance Inc). SSBM Finance Inc is a **Delaware** corporation.

For information on translations, please e-mail rights@apress.com, or visit http://www.apress.com/rights-permissions.

Apress titles may be purchased in bulk for academic, corporate, or promotional use. eBook versions and licenses are also available for most titles. For more information, reference our Print and eBook Bulk Sales web page at http://www.apress.com/bulk-sales.

Any source code or other supplementary material referenced by the author in this book is available to readers on GitHub via the book's product page, located at www.apress.com/978-1-4842-3524-9. For more detailed information, please visit http://www.apress.com/source-code.

Printed on acid-free paper

*I am dedicating this book to my Family.*

# Table of Contents

# About the Author

**Deepti Gupta** completed her MBA in Finance & PGPM in Operation Research in 2010. She has worked with KPMG and IBM private limited as a Data Scientist and is currently working as a data science freelancer. Deepti has extensive experience in predictive modeling and machine learning and her expertise is in SAS and R. Deepti has developed data science courses and delivered data science trainings and conducted workshops in both corporate and academic institutions. She has written multiple blogs and white papers. Deepti has a passion for mentoring budding data scientists.

# About the Contributor

**Dr. Akshat Gupta** is currently working as a Senior Applications Engineer at MilliporeSigma in Applications Engineering, Global Manufacturing Sciences and Technology (MSAT) group. He authored the health-care case study (Chapter5) of this book. His focal area of research is cell culture clarification and tangential flow filtration. Dr. Gupta has extensive experience in Design of Experiments (DOE) and statistical analysis. He holds a Bachelor of Technology (B.Tech) degree in Chemical Engineering from the Vellore Institute of Technology, and a Master of Science (MS) and Doctor of Philosophy (Ph.D.) in Chemical Engineering from the University of Massachusetts Lowell. He also has graduate certificates in Modeling and Simulation, and Nanotechnology.

# About the Technical Reviewer

**Preeti Pandhu** has a Master of Science degree in Applied (Industrial) Statistics from the University of Pune. She is SAS certified as a base and advanced programmer for SAS 9 as well as a predictive modeler using SAS Enterprise Miner 7. Preeti has more than 18 years of experience in analytics and training.

She started her career as a lecturer in statistics and began her journey into the corporate world with IDeaS (now a SAS company), where she managed a team of business analysts in the optimization and forecasting domain. She joined SAS as a corporate trainer before stepping back into the analytics domain to contribute to a solution-testing team and research/consulting team. She was with SAS for 9 years. Preeti is currently passionately building her analytics training firm, DataScienceLab (`www.datasciencelab.in`).

# Acknowledgments

Book writing is one of the most interesting and challenging attempt one can take up. This book could not have been completed without the encouragement, guidance, and support of my family. I would like to thank Dr. Akshat Gupta, Ved Prakash Garg, Col. Atul Gupta, Dr. Anvita Garg, Ayush Gupta, RS Miyan, Ansi Miyan, Dr. James Chrostowski, and my colleagues and friends for their productive discussions and suggestions. My special thanks to Celestin John who provided great help on everything ranging from technical support to answering my queries. I appreciate the thoughtful and insightful comments from the editor and the reviewers. Thanks to the Apress team, especially to Divya Modi for all the patience, support, and guidance in completing this project.

# Introduction

Analytics is a big buzz and a need for today's industries to solve their business problems. Analytics helps in mining the structured and unstructured data in order to withdraw the effective insights from the data, which will help to make effective business decisions. SAS and R are highly used tools in analytics across the globe by all industries for data mining and building machine learning and predictive models. This book focuses on industrial business problems and a practical analytical approach to solve those problems by implementing predictive models and machine learning techniques using SAS and R analytical languages.

The primary objective of this book is to help statisticians, developers, engineers, and data analysts who are well versed in writing codes; have a basic understanding of data and statistics; and are planning to transition to a data scientist profile. The most challenging part is practical and hands-on knowledge of building predictive models and machine learning algorithms and deploying them in industries to address industrial business problems. This book will benefit the reader in solving the business problems in various industrial domains by sharpening their analytical skills in getting practical exposure to various predictive model and machine learning algorithms in six industrial domains.

## What's in This Book

This book focuses on industrial business problems and practical analytical approaches to solve those problems by implementing predictive models and machine learning techniques using SAS Studio and R analytical languages. **This book contains six industrial case studies of various domains with data and all the codes in SAS Studio and R languages, which would benefit all readers to practice and implement these models in their own business cases.**

In Chapter 1 the general outline about analytics, the role of analytics in various industries, and a few popular data science and analytical tools are discussed. Chapter 2 describes the role of analytics in the banking industry with a detailed explanation of predicting a bank loan default case study in R and SAS. Chapter 3

describes how analytics contribute in the retail industry and offers a detailed explanation of forecasting a case study in R and SAS. Chapter 4 describes how analytics is reshaping the telecommunications industry and gives a detailed explanation of a case study on predicting customer churn in R and SAS. Chapter 5 describes the application of analytics in the healthcare industry and gives a clear explanation of a case study on predicting the probability of benign and malignant breast cancer using R and SAS. Chapter 6 describes the role of analytics in the airline industry and provides a case study on predicting flight arrival delays (minutes) in R and SAS. Chapter 7 describes the application of analytics in the FMCG industry with a detailed explanation of a business case study on customer segmentation based on their purchasing history using R and SAS.

## Who's the Target Audience?

- Data Scientists who would like to implement machine learning techniques with a practical analytical approach toward a particular industrial problem.
- Statistician, Engineers, and Researchers with a great theoretical understanding of data and statistics and would like to enhance their skills by getting practical exposure to data modeling.
- Data analysts who know about data mining but would like to implement predictive models and machine learning techniques.
- Developers who are well versed with coding but would like to transition to a career in data science.

## What You Will Learn

- Introduction to analytics and data understanding.
- How to approach industrial business problems with an analytical approach.
- Practical and hands-on knowledge in building predictive model and machine learning techniques.
- Building the analytical strategies.