

# **Machine Learning and Artificial Intelligence in Marketing and Sales**

## Praise for *Machine Learning and Artificial Intelligence in Marketing and Sales*

The world of business in general and marketing in particular is going through a radical transformation with the application of machine learning and artificial intelligence touching almost every aspect of business. These advanced tools are no longer solely the domain of statisticians and data scientists. Business practitioners find themselves interacting with these methods and, at the same time, data scientist often find themselves sitting at the management table managing groups and discussing their analysis. *Machine Learning and Artificial Intelligence in Marketing and Sales: Essential Reference for Practitioners and Data Scientists* strikes a, difficult to achieve, balance between providing sufficient information on commonly used but complex machine learning and AI tools, and yet keeping the book accessible and applicable to business practitioners with technical orientation. This is a great introduction book for those who wish to know not only about machine learning and AI, but also what it really is, and how to apply it in marketing and sales settings.

Oded Netzer, Professor of Business, Columbia University

This book is a great resource for Data scientists as a reference to anchor your technical understanding, build your intuition of the core machine learning models and at the same time elevate it for application in the real-world context of Marketing and Sales. It would be a good foundational book for students and applied practitioners of non-ML, non-stat backgrounds to gain confidence in your work and to stand behind the choices you make in your model building process. As the authors say, it does a good job at bridging the DS-real world application gap. I also liked the choice of the three models (NN, RF, SVM) to allow for focus and not overload the reader with tons of other material available. These three will get 80–90% of your job done as a modeler. I also liked the Executive Summary sections which appeal to the applied modeler in me and the Technical Detours which piqued a deeper intellectual curiosity and understanding of the details. Because they are positioned as detours, I could still read and use the book without making the technical parts overwhelming.

Vijay Jayanti, Head of Marketing Data Sciences  
at WhatsApp Inc.

In *Machine Learning and Artificial Intelligence in Marketing and Sales*, Syam and Kaul have teamed up to present a timely explanation of important topics in the evolving high-tech world of big data. For readers well-versed in the Support Vector

Machine, artificial neural nets, and deep learning, the book will be immediately useful. For readers new to these topics, the authors' accessible style lowers entry barriers. The book is required reading for managers, analysts, professors, and consultants involved in marketing and sales.

David J. Curry, Professor of Marketing, University of Cincinnati

Syam and Kaul's book is a comprehensive treatise on data science of marketing, a rich and deeply informative dive into the next generation of marketing analytics solutions. The work comprehensively integrates the theoretical concepts of Machine Learning with practical applications of marketing, making it essential for either ML Engineers solving marketing problems or marketing analysts looking to get a rigorous treatment of the nascent science.

Alex Vayner, Data science and AI expert, Partner, PA Consulting

The authors have skillfully tailored the content to a wide audience. I found this book as a solid reference guide for students and a reference for data science practitioners alike. While the book covers the most important Machine Learning topics in lucid detail, it also provides insightful executive summaries, and, most importantly, showcases applications of each model in the practical world of Sales and Marketing. I will wholeheartedly recommend this book to anyone interested in learning Machine Learning and Artificial Intelligence.

Sunish Mittal, Vice President, Data and Analytics, Aramark

This page intentionally left blank

# **Machine Learning and Artificial Intelligence in Marketing and Sales: Essential Reference for Practitioners and Data Scientists**

**BY**

**NILADRI SYAM**

*University of Missouri, USA*

**RAJEEVE KAUL**

*McDonald's Corporation, USA*



United Kingdom – North America – Japan – India – Malaysia – China

Emerald Publishing Limited  
Howard House, Wagon Lane, Bingley BD16 1WA, UK

First edition 2021

Copyright © 2021 by Emerald Publishing Limited.  
All rights of reproduction in any form reserved.

**Reprints and permissions service**

Contact: [permissions@emeraldinsight.com](mailto:permissions@emeraldinsight.com)

No part of this book may be reproduced, stored in a retrieval system, transmitted in any form or by any means electronic, mechanical, photocopying, recording or otherwise without either the prior written permission of the publisher or a licence permitting restricted copying issued in the UK by The Copyright Licensing Agency and in the USA by The Copyright Clearance Center. Any opinions expressed in the chapters are those of the authors. Whilst Emerald makes every effort to ensure the quality and accuracy of its content, Emerald makes no representation implied or otherwise, as to the chapters' suitability and application and disclaims any warranties, express or implied, to their use.

**British Library Cataloguing in Publication Data**

A catalogue record for this book is available from the British Library

ISBN: 978-1-80043-881-1 (Print)

ISBN: 978-1-80043-880-4 (Online)

ISBN: 978-1-80043-882-8 (Epub)



**ISOQAR**

REGISTERED

Certificate Number 1985  
ISO 14001

ISOQAR certified  
Management System,  
awarded to Emerald  
for adherence to  
Environmental  
standard  
ISO 14001:2004.



INVESTOR IN PEOPLE

*This book is dedicated to the memory of Ma and Baba: NS*

*This book is dedicated to the memory of my parents who encouraged me to remember that learning is a lifelong process of staying abreast of change: RK*

This page intentionally left blank



# Table of Contents

List of Figures, Tables and Illustrations	<i>xiii</i>
Foreword	<i>xvii</i>
Preface	<i>xix</i>
Acknowledgments	<i>xxi</i>
<b>Introduction</b>	<b>1</b>
<b>Chapter 1 Introduction and Machine Learning Preliminaries: Training and Performance Assessment</b>	<b>5</b>
1. Training of Machine Learning Models	5
1.1 Regression and Classification Models	6
1.2 Cost Functions and Training of Machine Learning Models	7
1.3 Maximum Likelihood Estimation	9
1.4 Gradient-Based Learning	10
2. Performance Assessment for Regression and Classification Tasks	13
2.1 Performance Assessment for Regression Models	14
2.2 Performance Assessment for Classification	15
Technical Detour 1	23
Technical Detour 2	23
<b>Chapter 2 Neural Networks in Marketing and Sales</b>	<b>25</b>
1. Introduction to Neural Networks	25
1.1 Early Evolution	25
1.2 The Neural Network Model	26

1.3 Cost Functions and Training of Neural Networks Using Backpropagation	38
1.4 Output Nodes	40
2. Feature Importance Measurement and Visualization	42
2.1 Neural Interpretation Diagram (NID)	43
2.2 Profile Method for Sensitivity Analysis	44
2.3 Feature Importance Based on Connection Weights	45
2.4 Randomization Approach for Weight and Input Variable Significance	48
2.5 Feature Importance Based on Partial Derivatives	49
3. Applications of Neural Networks to Sales and Marketing	49
4. Case Studies	54
Case Study 1: Churn Prediction	54
Case Study 2: Rent Value Prediction	57
Technical Detour 1	58
Technical Detour 2	59
Technical Detour 3	60
Technical Detour 4	60
Technical Detour 5	61
Technical Detour 6	62
Linear Activation Function for Continuous Regression Outputs	62
Sigmoid Activations Function for Binary Outputs	63
Softmax Activation Function for Multi-class Outputs	64
<b>Chapter 3 Overfitting and Regularization in Machine Learning Models</b>	<b>65</b>
1. Hyperparameters, Overfitting, Bias-variance Tradeoff, and Cross-validation	65
1.1 Hyperparameters	66
1.2 Overfitting	66
1.3 Bias-variance Tradeoff	68
1.4 Cross-validation	70
2. Regularization and Weight Decay	72
2.1 $L_2$ Regularization	73
2.2 $L_1$ Regularization	74

2.3 $L_1$ and $L_2$ Regularization as Constrained Optimization Problems	75
2.4 Regularization through Input Noise	76
2.5 Regularization through Early Stopping	77
2.6 Regularization through Sparse Representations	77
2.7 Regularization through Bagging and Other Ensemble Methods	78
Technical Detour 1	78
Technical Detour 2	79
Technical Detour 3	80
Technical Detour 4	80
Weight Decay in $L_2$ Regularization	80
Weight Decay in $L_1$ Regularization	81
Technical Detour 5	81
Technical Detour 6	82
Technical Detour 7	83
Technical Detour 8	83
<b>Chapter 4 Support Vector Machines in Marketing and Sales</b>	<b>85</b>
1. Introduction to Support Vector Machines	85
1.1 Early Evolution	85
1.2 Nonlinear Classification Using SVM	86
2. Separating Hyperplanes	88
3. Role of Kernels in Machine Learning	90
3.1 Kernels as Measures of Similarity	91
3.2 Nonlinear Maps and Kernels	94
3.3 Kernel Trick	98
4. Optimal Separating Hyperplane	99
4.1 Margin between Two Classes	99
4.2 Maximal Margin Classification and Optimal Separating Hyperplane	101
5. Support Vector Classifier and SVM	106
6. Applications of SVM in Marketing and Sales	114
7. Case Studies	120
Case Study 1: Consumer Choice Modeling	121
Case Study 2: Rent Value vs Location	125
Technical Detour 1	127
Technical Detour 2	127

Technical Detour 3	128
Technical Detour 4	128
Technical Detour 5	129
Technical Detour 6	130
Technical Detour 7	130
Technical Detour 8	131
Technical Detour 9	133
Technical Detour 10	133
Technical Detour 11	134
Illustration 3	134
Illustration 4	135
Illustration 5	136
<b>Chapter 5 Random Forest, Bagging, and Boosting of Decision Trees</b>	<b>139</b>
1. Early Evolution of Decision Trees: AID, THAID, CHAID	139
2. Classification and Regression Trees (CART)	143
2.1 Regression Trees	147
2.2 Classification Trees	151
3. Decision Trees and Segmentation	155
4. Bootstrapping, Bagging, and Boosting	158
4.1 Bootstrapping	159
4.2 Bagging	161
4.3 Boosting	165
5. Random Forest	169
6. Applications of Random Forests and Decision Trees in Marketing and Sales	171
7. Case Studies	176
Case Study 1: Caravan Insurance	177
Case Study 2: Wine Quality	178
Technical detour 1:	179
Technical detour 2:	181
Technical detour 3:	182
References	183
Index	191

## List of Figures, Tables and Illustrations

Figure 1.1.	Gradient Descent for a Minimization Problem.	11
Figure 1.2.	Small and Large Step Sizes (Learning Rates) in Gradient Descent.	11
Figure 1.3.	Classifying “+” and “-” and 0–1 Loss.	15
Figure 1.4.	Confusion Matrix Showing True Positives and False Positives.	17
Figure 1.5.	Instances Sorted by Decreasing Order of Predicted Class Probabilities and Confusion Matrix Corresponding to a Threshold of 0.8.	18
Figure 1.6.	Instances Sorted by Decreasing Order of Predicted Class Probabilities and Confusion Matrix Corresponding to a Threshold of 0.584.	19
Figure 1.7.	Receiver Operating Characteristics (ROC) Curve.	19
Figure 1.8.	Area Under the Curve (AUC).	20
Figure 1.9.	Cumulative Response Curve.	21
Figure 1.10.	Lift Chart.	22
Figure 1.11.	Gini Coefficient.	22
Figure 2.1.	Linear Relationship between “Rent Value” and “Distance to City Center”.	27
Figure 2.2.	Nonlinear Relationship between “Rent Value” and “Distance to City Center.”	27
Figure 2.3.	NN with 1 Hidden Node, p Input Nodes and 1 Output Node.	29
Figure 2.4.	Input Information at Hidden Node m Is a Weighted Sum of Input Nodes Plus a Bias.	30
Figure 2.5.	Activation Function $f_m$ Acts at the Hidden Node m.	30

Figure 2.6.	NN with 1 Hidden Layer Containing M Hidden Nodes.	31
Figure 2.7.	Information Flow in NN with 1 Hidden Node.	32
Figure 2.8.	Sigmoid Function.	33
Figure 2.9.	An “Inverted U” Shaped Nonlinear Pattern.	34
Figure 2.10.	Simple NN with 2 Hidden Nodes.	34
Figure 2.11.	NN with Two Hidden Units to Model a Nonlinear Relationship.	35
Figure 2.12.	Positively Weighted (Left) and Negatively Weighted (Right) Sigmoids.	35
Figure 2.13.	NN for Multiclass Classification with 1 Hidden Layer.	36
Figure 2.14.	Learning Slowdown with Quadratic Cost for Binary Output.	41
Figure 2.15.	NID for Neural Network for Predicting Choice of PB.	44
Figure 2.16.	Profiles for Sensitivity of Predicted Response to Input Variable.	45
Figure 2.17.	NN with 3 Inputs Nodes, 2 Hidden Nodes and 1 Output Node.	46
Figure 2.18.	Estimated Weights after Training of Network.	46
Figure 2.19.	Contribution of Each Input Neuron to the Output via Each Hidden Neuron.	47
Figure 2.20.	Relative Contribution of Each Input Neuron to Outgoing Signal.	47
Figure 2.21.	Relative Importance of Each Input.	47
Figure 3.1.	Plots of Prediction Error versus Model Complexity.	67
Figure 3.2.	For a Given Training Data Set, a More Complex Model (Right) Is Likely to Overfit.	68
Figure 3.3.	Fivefold Cross-validation.	71
Figure 3.4.	The Estimated Function (Bold Curve) Passes Closer to Data Points, Say Point a, with Smaller Decay Parameter (Left Panel) than with Larger Decay Parameter (Right Panel).	73
Figure 3.5.	Training for Too Long (Too Many Epochs) Can Raise Validation Error.	77
Figure 4.1.	Plot of Two Nonlinearly Separable Classes.	87

Figure 4.2.	Poor Classification of “+” and “-” Classes Using Logistic Regression.	87
Figure 4.3.	Very Good Classification Accuracy Using Support Vector Machine.	88
Figure 4.4.	Linearly Separable Points “+” and “-”	89
Figure 4.5.	The Inner Product Is a Measure of the Angle between Two Vectors.	92
Figure 4.6.	A New Point $x$ Is Assigned to Class. Whose Class Mean Is Closer to It.	93
Figure 4.7.	Two Classes “+” and “-” Are Not Linearly Separable in Input Space.	96
Figure 4.8.	The Transformed Points Are Linearly Separable in Feature Space.	97
Figure 4.9.	Classes Are Separable in Higher Dimensional Feature Space.	98
Figure 4.10.	Separability between Classes Seen from Two Vantage Points: East–West and North–South Axes.	100
Figure 4.11.	The Margin of the “+” Points from Hyperplane (Bold Line).	100
Figure 4.12.	Same Two Classes Separated by a Wider (Left) and Narrower (Right) Margin.	101
Figure 4.13.	Nearest Points from Hyperplane to Both Classes Lie Exactly on Margin.	102
Figure 4.14.	Support Vectors, Like Point $x_a$ , Represents Consumers That Make Hard Decisions.	103
Figure 4.15.	Different Likelihoods of Churn When We Move Away from Separating Hyperplane.	103
Figure 4.16.	Linearly Separable Points in Input Space.	105
Figure 4.17.	Support Vectors from the Two Classes Are Circled.	105
Figure 4.18.	Optimal Separating Hyperplane in Input Space.	106
Figure 4.19.	Classes That Are Not Linearly Separable.	108
Figure 4.20.	Support Vectors from the Two Classes Are Circled.	111
Figure 4.21.	Optimal Separating Hyperplane in Feature Space.	112
Figure 4.22.	The Data for the XOR Problem.	113
Figure 4.23.	Plot of XOR Points.	113
Figure 4.24.	Support Vector Machine Successfully Separates Classes in XOR Problem.	114

Figure 4.25.	Latitude of Acceptance Noncompensatory Choice Rule.	122
Figure 5.1.	First Binary Partition of a Two-Dimensional Feature Space.	144
Figure 5.2.	Second Binary Partition of a Two-Dimensional Feature Space.	144
Figure 5.3.	Third Binary Partition of a Two-Dimensional Feature Space.	145
Figure 5.4.	Final Partitioned Two-Dimensional Feature Space Showing Nonoverlapping Regions.	145
Figure 5.5.	Tree Diagram of Recursive Binary Partitioning.	146
Figure 5.6.	Arbitrary Partition Does Not Allow Easy Interpretation.	146
Figure 5.7.	Hierarchical Clustering of Seven Objects: A Through G.	157
Figure 5.8.	A Dendrogram Is a Tree-Like Representation of Hierarchical Clustering.	158
Figure 5.9.	Clipping the Dendrogram at Different Levels Gives Different Numbers of Clusters.	159
Table 5.1.	Categorical Data for CHAID Analysis.	153
Table 5.2.	Training Data Set for Bootstrapping.	160
Illustration 4.1.	Nonlinear maps transform nonlinearly separable classes to linearly separable classes in feature space.	95
Illustration 4.2.	Nonlinearly separable data in input space is separable in higher dimensional feature space.	97
Illustration 4.3.	Constructing an optimal separating hyperplane using inner products and support vectors.	104
Illustration 4.4.	Constructing optimal separating hyperplane for nonlinearly separable classes using kernels and support vectors.	110
Illustration 4.5.	Identifying support vectors to compute optimal separating hyperplane for nonlinearly separable classes.	113



# Foreword

The book is written in five chapters covering important concepts, principles, and practices on contemporary machine learning topics. Each page is written in an easy-to-read format using clean and lean sentences unencumbered by complex jargon. Each chapter provides solid theoretical background of the methods selected, and further elaborates the how and the why aspects of model selection, model building and model validation; the step-by-step approaches of leveraging specific techniques such as the Neural Network, the decision trees, and the vector machines; and the pros and the cons of certain machine learning (ML) procedures. Moreover, each topic provides many real-world examples that connect the theory with the applied use. Lists and depth of supporting reference materials are also excellent.

We are at an exciting time where the era of big data, machine learning, artificial intelligence, cloud computing and advanced analytics is ushering unprecedented access to and uses of large volumes of data to improve our predictive power to unlock transformational changes that impact many aspects of our lives in the retail, the financial, the manufacturing, the technology, the healthcare, and other industries.

As both big and small firms alike are fine-tuning their pricing, promotion, distribution, customer-retention, risk-management, and go-to-market strategies, data scientists are increasingly expected to know cutting-edge solutions, equip themselves with many facets of ML techniques and solutions. This book undoubtedly provides the foundational background, the tools, and the necessary tips to grasp many of the ML methods currently in use. In addition, as ML is rapidly and dynamically evolving to impact our daily life, the timeliness of this book is undoubtedly very appropriate.

I have worked as a data scientist for diverse organizations and have taught analytics and ML classes in universities. A few pages into this book, I knew that it is a special treat for my appetite, and it really struck a chord. Many of the well-organized core concepts expounded in the book are not only refreshing but also the kind I wish I had long ago. I also dare to describe this book as a versatile tool and a must-have reference material for both beginners and seasoned data scientists alike, business leaders and those who embrace data and analytics-driven decision-making processes. In addition, analytics teachers and their respective students can benefit from the in-depth analysis of the contemporary data science topics and the plethora of examples provided. I commend the authors for a job well done.

***Dawit Mulugeta, PhD (Biostatistics), VP Analytics and Risk Management, Wells Fargo***

*Dawit Mulugeta is an applied data scientist. Currently he holds a vice president of analytics and risk management role with Wells Fargo. In addition, he teaches popular analytics and ML classes in the Department of Management Sciences as well as in the Department of Accounting and Information Systems of the Max M. Fisher College of Business of the Ohio State University in Columbus. Dawit earned a PhD in biostatistics from the University of Wisconsin in Madison, USA.*

# Preface

*Machine Learning and Artificial Intelligence in Marketing and Sales: Essential Reference for Practitioners and Data Scientists* is intended for a variety of audiences:

- Marketing and sales practitioners who want to develop a deeper appreciation for how machine learning models can be applied to problems in marketing and sales.
- Data scientists who are tasked with implementing data-based solutions to business problems in the domain of marketing and sales.
- Software engineers and IT developers who will implement, or assist in the implementation of, and manage the solutions for marketing and sales organizations in their company.
- Students in master's programs in data science and in MBA programs and management consultants who wish to have a deeper appreciation of how machine learning and AI are impacting the world of marketing and sales.

This book represents a fruitful collaboration between an academic and an industry practitioner. Each author made a genuine attempt to understand and incorporate in the book the viewpoints and tastes of the other. This main purpose of this book is to bridge, what we call, the *Domain Specialist – Data Scientist Gap (DS-DS Gap)*. It is aimed at the translators, that is, the boundary-spanners, among two distinct audiences – the marketing practitioners and the data scientists. It is the experience of the authors that often in companies one of the biggest barriers to success is the ability of the technical and sales/marketing business teams to effectively understand and communicate with one another in solving business problems. Marketing practitioners and data scientists wishing to work together to solve marketing and sales issues using the methods of machine learning need to have a shared understanding of how these methods can be applied to marketing and sales. This book treads the fine line between the very technical books on the one hand, and, on the other hand, the purely qualitative books that merely mention the various AI and machine learning applications in marketing and sales.

Any collaboration between an academic and an industry practitioner, with the former emphasizing theory and the latter emphasizing business applications, always has a fair bit of tension – the desirable kind of tension that has hopefully

made the final product better. We have tried earnestly to strike a balance between the theoretical/technical and applications aspects. Having said that, we have decided to err on the side of applications, anchoring our narrative on the connections between the techniques and their applications in a business setting. We have deliberately kept the technical details to a bare minimum in the main body of the text, and have dealt with technical details through the various “Technical detours” that have been collected together at the end of each chapter. This allows readers who do not wish to tackle the technical issues to be able to read the chapters easily without being distracted or overwhelmed by technical details. In addition, to help readers get a quick overview of the concepts involved, we have added “Executive summaries” as and when needed. As far as possible in each chapter we have tried to emphasize the intuitions behind the, sometimes complex, concepts of machine learning methods.

As far as the marketing and sales practitioners are concerned, the book assumes that they are interested in actual implementation of machine learning models, either by themselves or in collaboration with the data scientists in their organizations. For this reason, this book is not just a high-level overview of machine learning applications to marketing and sales. Thus, by its very nature the chapters assume that the reader is willing to handle some technical material. However, we have made every attempt to make the chapters self-contained by providing the background material needed to understand the chapter contents. Each chapter has a section on the existing applications of machine learning and artificial intelligence (AI) to issues in marketing and sales. We have tried to focus on applications that have been archived in the major peer-reviewed research journals in marketing, operations research, machine learning, expert systems, etc. By their nature, journal articles have details of implementation and data sets and interested readers can go through the articles listed in the references at the end of every chapter for further details. Finally, for those wishing to get hands-on experience of actually running analyses of marketing and sales data using machine learning models, each chapter has a couple of detailed “Case studies” at the end.

# Acknowledgments

Niladri

I would like to gratefully acknowledge the support of my wife, Nivedita, without whose patience and encouragement this book would never have materialized. I would also like to thank my teacher, Professor Bibek Debroy, who introduced me to the power of quantitative modeling and testing of business phenomena in the 1980s, much before the term “analytics” had become fashionable. The *Center for Sales and Customer Development (CSCD)* at the University of Missouri provided the support and proper climate which greatly facilitated the writing of this book.

Rajeeve

This book would not be possible without the unyielding support and encouragement of my wife, Shalini, and the patience and understanding of our son, Harsha. Working on a book while staying abreast with challenging executive roles required them to sacrifice time that we could have spent building life memories – for which I am so grateful. I would also like to thank my many professors who encouraged me to learn and experiment with so many quantitative methods across diverse fields from statistics, to marketing, finance, operations research, etc. I further extend my gratitude to the many incredible executives across so many industries who adopted my quantitative solutions to improve decision in areas including pricing, marketing, supply chain, and digital among others, and the companies that allowed me to follow my curiosity to develop and deploy these models.