

# Machine Learning Risk Assessments in Criminal Justice Settings

Richard Berk

# Machine Learning Risk Assessments in Criminal Justice Settings

 Springer

Richard Berk  
Department of Criminology  
University of Pennsylvania  
Philadelphia, PA, USA

ISBN 978-3-030-02271-6      ISBN 978-3-030-02272-3 (eBook)  
<https://doi.org/10.1007/978-3-030-02272-3>

Library of Congress Control Number: 2018959762

© Springer Nature Switzerland AG 2019

This work is subject to copyright. All rights are reserved by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed.

The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.

The publisher, the authors and the editors are safe to assume that the advice and information in this book are believed to be true and accurate at the date of publication. Neither the publisher nor the authors or the editors give a warranty, express or implied, with respect to the material contained herein or for any errors or omissions that may have been made. The publisher remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

This Springer imprint is published by the registered company Springer Nature Switzerland AG  
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

# Preface

This book is an effort to put in one place and in accessible form the most recent work on forecasts of re-offending by individuals already in criminal justice custody. Much of that work is my own and comes from over two decades of close collaborations with a number of criminal justice agencies. After many requests to provide in one place an account of the procedures I have used, I agreed to write this book. What I hope distinguishes the material from treatments that have come before is the use of machine learning statistical procedures trained on very large datasets, an explicit introduction of the relative costs of forecasting errors as the forecasts are constructed, and an emphasis on maximizing forecasting accuracy. Whether the forecasts that result are “good enough,” I leave to the reader.

An earlier version of the manuscript was published as a SpringerBrief. The book was well received, but after 3 years, an update and expansion were needed. Included now are discussions of neural networks, with extensions into deep learning, and of the trade-offs between transparency, accuracy, and fairness. Throughout, difficult issues are more clearly explained, supported by many new references. The current manuscript is about 75% longer than the SpringerBrief.

The audience for the book is still graduate students and researchers in the social sciences and data analysts in criminal justice agencies. Formal mathematics is used only as necessary or in concert with more intuitive explanations. A working knowledge of the generalized linear model is assumed. All of the empirical examples were constructed using the programming language R, in part because most of the key tools are not readily available in the usual social science point-and-click statistical packages. R runs on a wide variety of platforms and is available at no cost. It can be downloaded from [www.r-project.org/](http://www.r-project.org/).

A large number of individuals have helped me as my criminal justice forecasting activities have evolved. On technical matters, I owe a special debt to David Freedman, who put up with my questions over 10 years of collaboration. More recently, I have benefited enormously from working with colleagues in University of Pennsylvania’s Department of Statistics, especially Larry Brown, Andreas Buja, Ed George, Arun Kuchibhotla, Mikhail Traskin, Linda Zhao. Three other wonderful

colleagues are Michael Kerns and Aaron Roth from the University of Pennsylvania's Department of Information and Computer Science and Cary Coglianese from the University of Pennsylvania Law School. On policy matters, my mentor Peter Rossi provided a foundation on which I still draw.

My hands-on education has come from a number of criminal justice officials starting with several talented individuals working for the California Department of Corrections in the 1990s: George Lehman, Penny O'Daniel, and Maureen Tristan. More recently, Ellen Kurtz of Philadelphia's Adult Department of Probation and Parole was an early supporter and adopter of modern forecasting methods from whom I learned a lot. I have continued to learn from her successor, Charles Hoyt. Catherine McVey, chairman of the Pennsylvania Board of Probation and Parole, played a key role in educating me about the parole decision process and the *real politik* in which parole decisions are made. Jim Alibirio, an IT and data analyst working for the parole board, was a true data maven and wonderful colleague as we together developed the required forecasting procedures. Other practitioners to whom I am especially indebted are Mark Bergstrom, Executive Director of the Pennsylvania Sentencing Commission, and Sarah Hart, a former Deputy Prosecutor (and much more) in the Philadelphia district attorney's office. A sincere thanks to one and all.

All of the analyses reported in the book would have been impossible without the data on which they are based. Collecting those data required hard work by a number of individuals from several criminal justice agencies at state and local levels. An important subset of the work was supported by a grant to the State of Pennsylvania from the National Institute of Justice: "Projecting Violent Re-Offending in a Parole Population: Developing A Real-Time Forecasting Procedure to Inform Parole Decision-Making" (# 2009-IJ-CX-0044). My most sincere thanks to NIJ for the funding and to Patrick Clark, who was a knowledgeable and constructive grant monitor with a great sense of humor.

Finally, I am a notoriously bad proofreader and can always use help with the quality of my arguments and prose. Justin Bleich and Anna Kegler carefully read an earlier draft of this manuscript and provided all manner of assistance. However, the responsibility for any remaining typos or unclear prose is surely mine.

Philadelphia, PA, USA

Richard Berk

# Contents

<b>1</b>	<b>Getting Started</b>	1
1.1	Some Introductory Caveats	1
1.2	What Criminal Justice Risk Assessment Is and Is Not	2
1.3	Difficulties Defining Machine Learning	3
1.4	Setting the Stage	6
1.4.1	A Brief Motivating Example	10
<b>2</b>	<b>Some Important Background Material</b>	15
2.1	Policy Considerations	15
2.1.1	Criminal Justice Risk Assessment Goals	15
2.1.2	Decisions to Be Informed by the Forecasts	18
2.1.3	Outcomes to Be Forecasted	19
2.1.4	Real World Constraints	20
2.1.5	Stakeholders	21
2.2	Data Considerations	21
2.2.1	Stability Over Time	22
2.2.2	Training Data, Evaluation Data, and Test Data	24
2.2.3	Representative Data	25
2.2.4	Large Samples	27
2.2.5	Data Continuity	28
2.2.6	Missing Data	28
2.3	Statistical Considerations	30
2.3.1	Actuarial Methods	30
2.3.2	Building in the Relative Costs of Forecasting Errors	32
2.3.3	Effective Forecasting Algorithms	36
<b>3</b>	<b>A Conceptual Introduction to Classification and Forecasting</b>	41
3.1	Populations and Samples	41
3.2	Classification and Forecasting Using Decision Boundaries	41
3.3	Classification by Data Partitions	45
3.4	Forecasting by Data Partitions	47
3.5	Finding Good Data Partitions	48

3.6	Enter Asymmetric Costs .....	49
3.7	Recursive Partitioning Classification Trees .....	51
3.7.1	How Many Terminal Nodes? .....	54
3.7.2	Classification Tree Instability and Adaptive Fitting .....	55
<b>4</b>	<b>A More Formal Treatment of Classification and Forecasting</b> .....	<b>57</b>
4.1	Introduction .....	57
4.2	What Is Being Estimated? .....	58
4.3	Data Generation Formulations .....	58
4.4	Notation .....	60
4.5	From Probabilities to Classification .....	61
4.6	Computing $\hat{f}(\mathbf{G} \mathbf{X})$ in the Real World .....	63
4.6.1	Estimation Bias .....	63
4.6.2	The Traditional Bias-Variance Tradeoff with Extensions .....	67
4.6.3	Addressing Uncertainty .....	69
4.7	A Bit More on the Joint Probability Model .....	71
4.8	Putting It All Together .....	72
<b>5</b>	<b>Tree-Based Forecasting Methods</b> .....	<b>75</b>
5.1	Introduction .....	75
5.2	Once More on Constructing Data Partitions .....	76
5.3	Building the Costs of Classification Errors .....	78
5.4	Confusion Tables .....	79
5.5	Ensembles of Classification Trees: Random Forests .....	82
5.5.1	Transfer Learning .....	86
5.5.2	Variable Importance for Random Forests .....	86
5.5.3	Partial Response Functions from Random Forests .....	89
5.5.4	Forecasting .....	90
5.5.5	Statistical Inference for Random Forests .....	92
5.5.6	Conclusions About Random Forests .....	92
5.6	Tree-Based Alternatives to Random Forests .....	93
5.6.1	Stochastic Gradient Boosting .....	93
5.6.2	Bayesian Additive Regression Trees .....	97
5.7	Why Do Ensembles of Trees Work So Well? .....	99
5.7.1	Neural Networks and Deep Learning .....	101
<b>6</b>	<b>Transparency, Accuracy and Fairness</b> .....	<b>115</b>
6.1	Introduction .....	115
6.2	Transparency .....	116
6.3	Accuracy .....	121
6.4	Fairness .....	123
6.5	Improving Data Quality .....	127
6.6	Communicating with Stakeholders .....	128
6.7	A Policy Suggestion .....	129

- 7 Real Applications** ..... 131
  - 7.1 A Simplified Example ..... 131
  - 7.2 A More Complex Example ..... 139
  - 7.3 Some Fairness Tradeoffs ..... 151
  
- 8 Implementation** ..... 155
  - 8.1 Making It Real ..... 155
  - 8.2 Demonstration Phase ..... 156
  - 8.3 Hardware and Software ..... 157
  - 8.4 Data Preparation for Forecasting ..... 158
  - 8.5 Personnel ..... 159
  - 8.6 Public Relations ..... 160
  
- 9 Some Concluding Observations About Actuarial Justice and More ...** 163
  - 9.1 Current Trends in Actuarial Justice ..... 163
    - 9.1.1 What About Statistical Inference? ..... 165
    - 9.1.2 Unsafe at Any Speed? ..... 168
  - 9.2 Machine Learning and Artificial Intelligence ..... 170
  - 9.3 Dark Structure and New Science ..... 171
  - 9.4 A Closing Sermonette ..... 172
  
- References** ..... 173