

---

# Machine Learning Design Patterns

*Solutions to Common Challenges in Data  
Preparation, Model Building, and MLOps*

*Valliappa Lakshmanan, Sara Robinson,  
and Michael Munn*

Beijing • Boston • Farnham • Sebastopol • Tokyo

**O'REILLY®**

## Machine Learning Design Patterns

by Valliappa Lakshmanan, Sara Robinson, and Michael Munn

Copyright © 2021 Valliappa Lakshmanan, Sara Robinson, and Michael Munn. All rights reserved.

Printed in the United States of America.

Published by O'Reilly Media, Inc., 1005 Gravenstein Highway North, Sebastopol, CA 95472.

O'Reilly books may be purchased for educational, business, or sales promotional use. Online editions are also available for most titles (<http://oreilly.com>). For more information, contact our corporate/institutional sales department: 800-998-9938 or [corporate@oreilly.com](mailto:corporate@oreilly.com).

**Acquisitions Editor:** Rebecca Novack

**Developmental Editor:** Corbin Collins

**Production Editor:** Beth Kelly

**Copyeditor:** Charles Roumeliotis

**Proofreader:** Holly Bauer Forsyth

**Indexer:** nSight, Inc.

**Interior Designer:** David Futato

**Cover Designer:** Karen Montgomery

**Illustrator:** Kate Dullea

October 2020: First Edition

### Revision History for the First Edition

2020-10-15: First Release

See <http://oreilly.com/catalog/errata.csp?isbn=9781098115784> for release details.

The O'Reilly logo is a registered trademark of O'Reilly Media, Inc. *Machine Learning Design Patterns*, the cover image, and related trade dress are trademarks of O'Reilly Media, Inc.

The views expressed in this work are those of the authors, and do not represent the publisher's views. While the publisher and the authors have used good faith efforts to ensure that the information and instructions contained in this work are accurate, the publisher and the authors disclaim all responsibility for errors or omissions, including without limitation responsibility for damages resulting from the use of or reliance on this work. Use of the information and instructions contained in this work is at your own risk. If any code samples or other technology this work contains or describes is subject to open source licenses or the intellectual property rights of others, it is your responsibility to ensure that your use thereof complies with such licenses and/or rights.

978-1-098-11578-4

[LSI]

---

# Table of Contents

<b>Preface.....</b>	<b>ix</b>
<b>1. The Need for Machine Learning Design Patterns.....</b>	<b>1</b>
What Are Design Patterns?	1
How to Use This Book	3
Machine Learning Terminology	3
Models and Frameworks	4
Data and Feature Engineering	6
The Machine Learning Process	7
Data and Model Tooling	8
Roles	9
Common Challenges in Machine Learning	11
Data Quality	11
Reproducibility	13
Data Drift	14
Scale	16
Multiple Objectives	16
Summary	17
<b>2. Data Representation Design Patterns.....</b>	<b>19</b>
Simple Data Representations	22
Numerical Inputs	22
Categorical Inputs	28
Design Pattern 1: Hashed Feature	32
Problem	32
Solution	33
Why It Works	34
Trade-Offs and Alternatives	35

Design Pattern 2: Embeddings	39
Problem	39
Solution	41
Why It Works	46
Trade-Offs and Alternatives	48
Design Pattern 3: Feature Cross	52
Problem	52
Solution	53
Why It Works	57
Trade-Offs and Alternatives	58
Design Pattern 4: Multimodal Input	62
Problem	62
Solution	63
Trade-Offs and Alternatives	65
Summary	77
<b>3. Problem Representation Design Patterns.....</b>	<b>79</b>
Design Pattern 5: Reframing	80
Problem	80
Solution	80
Why It Works	82
Trade-Offs and Alternatives	84
Design Pattern 6: Multilabel	90
Problem	90
Solution	91
Trade-Offs and Alternatives	93
Design Pattern 7: Ensembles	99
Problem	99
Solution	100
Why It Works	104
Trade-Offs and Alternatives	106
Design Pattern 8: Cascade	108
Problem	108
Solution	110
Trade-Offs and Alternatives	114
Design Pattern 9: Neutral Class	117
Problem	117
Solution	118
Why It Works	118
Trade-Offs and Alternatives	120
Design Pattern 10: Rebalancing	122
Problem	122

Solution	123
Trade-Offs and Alternatives	129
Summary	137
<b>4. Model Training Patterns .....</b>	<b>139</b>
Typical Training Loop	139
Stochastic Gradient Descent	139
Keras Training Loop	140
Training Design Patterns	141
Design Pattern 11: Useful Overfitting	141
Problem	141
Solution	142
Why It Works	144
Trade-Offs and Alternatives	145
Design Pattern 12: Checkpoints	149
Problem	150
Solution	150
Why It Works	152
Trade-Offs and Alternatives	154
Design Pattern 13: Transfer Learning	161
Problem	161
Solution	162
Why It Works	169
Trade-Offs and Alternatives	172
Design Pattern 14: Distribution Strategy	175
Problem	175
Solution	175
Why It Works	181
Trade-Offs and Alternatives	183
Design Pattern 15: Hyperparameter Tuning	187
Problem	187
Solution	190
Why It Works	192
Trade-Offs and Alternatives	194
Summary	198
<b>5. Design Patterns for Resilient Serving.....</b>	<b>201</b>
Design Pattern 16: Stateless Serving Function	201
Problem	203
Solution	205
Why It Works	207
Trade-Offs and Alternatives	209

Design Pattern 17: Batch Serving	213
Problem	213
Solution	214
Why It Works	215
Trade-Offs and Alternatives	217
Design Pattern 18: Continued Model Evaluation	220
Problem	220
Solution	221
Why It Works	227
Trade-Offs and Alternatives	227
Design Pattern 19: Two-Phase Predictions	232
Problem	232
Solution	234
Trade-Offs and Alternatives	241
Design Pattern 20: Keyed Predictions	243
Problem	244
Solution	244
Trade-Offs and Alternatives	247
Summary	248
<b>6. Reproducibility Design Patterns.....</b>	<b>249</b>
Design Pattern 21: Transform	250
Problem	250
Solution	251
Trade-Offs and Alternatives	252
Design Pattern 22: Repeatable Splitting	258
Problem	258
Solution	259
Trade-Offs and Alternatives	260
Design Pattern 23: Bridged Schema	266
Problem	266
Solution	266
Trade-Offs and Alternatives	271
Design Pattern 24: Windowed Inference	273
Problem	273
Solution	275
Trade-Offs and Alternatives	277
Design Pattern 25: Workflow Pipeline	282
Problem	282
Solution	284
Why It Works	288
Trade-Offs and Alternatives	289

Design Pattern 26: Feature Store	295
Problem	295
Solution	296
Why It Works	306
Trade-Offs and Alternatives	308
Design Pattern 27: Model Versioning	310
Problem	310
Solution	311
Trade-Offs and Alternatives	315
Summary	317
<b>7. Responsible AI.....</b>	<b>319</b>
Design Pattern 28: Heuristic Benchmark	320
Problem	320
Solution	321
Trade-Offs and Alternatives	324
Design Pattern 29: Explainable Predictions	326
Problem	326
Solution	327
Trade-Offs and Alternatives	339
Design Pattern 30: Fairness Lens	343
Problem	343
Solution	345
Trade-Offs and Alternatives	354
Summary	358
<b>8. Connected Patterns.....</b>	<b>359</b>
Patterns Reference	359
Pattern Interactions	363
Patterns Within ML Projects	366
ML Life Cycle	366
AI Readiness	373
Common Patterns by Use Case and Data Type	377
Natural Language Understanding	377
Computer Vision	378
Predictive Analytics	378
Recommendation Systems	379
Fraud and Anomaly Detection	380
<b>Index.....</b>	<b>383</b>