

Contents

<i>Author's biography</i>	<i>xiii</i>
<i>Foreword</i>	<i>xv</i>
SECTION 1 STORING AND ACCESSING DATA	1
Chapter 1: The healthcare IT landscape	3
1.1 How we got here and the growth of healthcare IT.....	3
1.2 The role of informatics.....	5
1.3 Common architectural aspects of healthcare IT.....	6
1.4 Device and application levels.....	7
1.5 Communication level.....	9
1.6 Process level.....	9
1.7 Common organizational aspects of healthcare IT.....	11
1.8 Physician and nurse informaticists.....	12
1.9 Regulatory aspects of healthcare IT.....	13
1.10 Challenges and opportunities.....	13
1.11 Conclusion.....	14
References.....	15
Chapter 2: Relational databases	17
2.1 A brief history of SQL and relational databases.....	17
2.2 Overview of the relational model.....	18
2.3 Differences between the relational model and SQL.....	20
2.4 Primary and foreign keys.....	21
2.5 ACID and transactions with data.....	24
2.6 Normalization.....	26
2.7 Conclusion.....	28
References.....	29
Chapter 3: SQL	31
3.1 Getting started with SQL.....	31
3.2 Structure of SQL databases.....	31
3.3 Basic SQL: SELECT, FROM, WHERE, and ORDER BY statements.....	33
3.4 Basic SQL: GROUP BY and general aggregate functions.....	36

3.5	Intermediate SQL: Joins.....	38
3.6	Advanced SQL: Window functions.....	40
3.7	SQL concept: Indexes.....	41
3.8	SQL concept: Schemas.....	42
3.9	Advanced SQL: SubQueries.....	42
3.10	Conclusion.....	43
Chapter 4: Example project 1: Querying data with SQL.....		45
4.1	Introduction and project background.....	45
4.2	Viewing tables.....	46
4.3	Querying tables.....	53
4.3.1	Average number of visits per day per location.....	54
4.3.2	Average patient age and patient sex per location.....	56
4.3.3	Average number of patient visits per provider.....	57
4.3.4	Counts of diagnosis codes and average age per clinic location.....	58
4.4	Conclusion.....	59
Chapter 5: Nonrelational databases.....		61
5.1	Early nonrelational models.....	61
5.2	The rise of modern nonrelational models.....	64
5.3	Key-value stores.....	65
5.4	Document stores.....	66
5.5	Column stores.....	68
5.5.1	Traditional column stores.....	68
5.5.2	Wide column stores.....	69
5.6	Graph databases.....	71
5.7	Conclusion.....	72
	Reference.....	73
Chapter 6: M/MUMPS.....		75
6.1	A brief history and context.....	75
6.2	The M language.....	76
6.3	General concepts regarding arrays and MUMPS.....	78
6.4	Arrays and MUMPS.....	81
6.5	MUMPS, globals, and data infrastructure.....	85
6.6	Conclusion.....	87
	References.....	88
SECTION 2 UNDERSTANDING HEALTHCARE DATA.....		89
Chapter 7: How to approach healthcare data questions.....		91
7.1	Introduction.....	91
7.2	Healthcare as a CAS.....	91
7.3	Drivers of fallacy: Chance and bias.....	95

7.4 Missingness.....	96
7.5 Selecting tractable areas for intervention.....	99
7.6 Data and trust.....	100
7.7 Conclusion.....	101
Chapter 8: Clinical and administrative workflows: Encounters, laboratory testing, clinical notes, and billing.....	103
8.1 Introduction.....	103
8.2 Encounters, patients, and episodes of care.....	103
8.3 Laboratory testing, imaging, and medication administration.....	106
8.4 Clinical notes and documentation.....	109
8.5 Billing.....	111
8.6 Conclusion.....	112
Chapter 9: HL-7, clinical documentation architecture, and FHIR.....	113
9.1 Introduction.....	113
9.2 HL7 and HL7v2.....	113
9.3 RIM, HL7v3, and clinical documentation architecture.....	116
9.4 FHIR.....	123
9.5 DICOM.....	125
9.6 Vendor standards.....	125
9.7 Cloud services.....	126
9.8 Conclusion.....	126
References.....	127
Chapter 10: Ontologies, terminology mappings, and code sets.....	129
10.1 Introduction.....	129
10.2 Diagnostic ontologies: ICD and ICD-CM.....	129
10.3 Procedure ontologies: ICD-PCS, CPT, and HCPCS.....	132
10.4 General ontologies: SNOMED, SNOMED-CT.....	135
10.5 Other specific ontologies: LOINC and NDC.....	138
10.6 Summative ontologies: DRG.....	139
10.7 Conclusion.....	141
References.....	141
SECTION 3 ANALYZING DATA.....	143
Chapter 11: A selective introduction to Python and key concepts.....	145
11.1 Python: What and why.....	145
11.2 A note on Python 2 and 3.....	146
11.3 General structure of the language.....	146
11.4 Type system.....	147
11.5 Control flow.....	148
11.6 Functions.....	149

11.7	Objects.....	149
11.8	Basic data structures.....	151
11.8.1	Lists.....	152
11.8.2	Sets.....	154
11.8.3	Tuples.....	155
11.8.4	Dictionaries.....	155
11.8.5	List and dictionary comprehensions.....	156
11.9	Conclusion.....	157
	Reference.....	157
Chapter 12: Packages, interactive computing, and analytical documents.....		159
12.1	Introduction.....	159
12.2	Packages and package management.....	159
12.3	Key packages.....	163
12.4	Jupyter.....	169
12.5	Analytical documents and interactive computing.....	171
12.6	Conclusion.....	174
Chapter 13: Assessing data quality, attributes, and structure.....		175
13.1	Introduction.....	175
13.2	Importing, cleaning, and assessing data.....	175
13.3	Tidying data.....	186
13.4	Handling missing values.....	188
13.5	Conclusion.....	189
Chapter 14: Introduction to machine learning: Regression, classification, and important concepts.....		191
14.1	The aim of machine learning.....	191
14.2	Regression.....	192
14.3	Functions as hypotheses.....	194
14.4	Error and cost.....	198
14.5	Optimization.....	201
14.6	Classification.....	205
14.7	Additional considerations: Normalization, regularization, and generalizability.....	209
14.8	Conclusion.....	210
Chapter 15: Introduction to machine learning: Support vector machines, tree-based models, clustering, and explainability.....		211
15.1	Introduction.....	211
15.2	Support vector machines.....	211
15.3	Decision trees.....	219
15.4	Clustering.....	221
15.5	Model explainability.....	223
15.6	Conclusion.....	225

Chapter 16: Computational phenotyping and clinical natural language processing...	227
16.1 Introduction.....	227
16.2 Manual review and general review considerations.....	227
16.3 Computational phenotyping: General considerations.....	230
16.4 Supervised methods.....	233
16.5 Unsupervised methods.....	235
16.6 Natural language processing.....	236
16.7 Conclusion.....	242
Chapter 17: Example project 2: Assessing and modeling data.....	243
17.1 Introduction and project background.....	243
17.2 Data collection.....	243
17.3 Data assessment and preparation.....	245
17.4 Model creation.....	246
17.4.1 Logistic regression.....	247
17.4.2 Decision tree.....	254
17.4.3 Support vector machine.....	256
17.5 Exporting and persisting models.....	257
17.6 Conclusion.....	259
Chapter 18: Introduction to deep learning and artificial intelligence.....	261
18.1 Introduction.....	261
18.2 What exactly is deep learning.....	261
18.3 Feed forward networks.....	262
18.4 Training and backpropagation.....	269
18.5 Local versus global minima.....	270
18.6 Convolutional networks.....	272
18.7 Adversarial examples and local minima.....	274
18.8 Recurrent networks.....	275
18.9 Autoencoders and generative adversarial networks.....	275
18.10 Conclusion.....	276
Reference.....	276
SECTION 4 DESIGNING DATA APPLICATIONS.....	277
Chapter 19: Analysis best practices.....	279
19.1 Introduction.....	279
19.2 Workflow.....	279
19.3 Documentation.....	284
19.4 Data governance.....	287
19.5 Conclusion.....	289
Chapter 20: Overview of big data tools: Hadoop, Spark, and Kafka.....	291
20.1 Introduction.....	291
20.2 Hadoop.....	292

Contents

20.3 Spark.....	299
20.4 Kafka.....	304
20.5 Conclusion.....	305
Chapter 21: Cloud technologies.....	307
21.1 Introduction.....	307
21.2 Data storage.....	307
21.3 Compute.....	310
21.4 Machine learning and analysis services.....	314
21.5 Conclusion.....	316
Index.....	317