

| 차례 |

지은이 소개	4
기술 감수자 소개	5
윤킨이 소개	6
윤킨이의 말	7
들어가며	19

1부 | 일래스틱 스택으로 머신러닝 시작하기 25

1장 IT를 위한 머신러닝	27
IT의 역사적 도전 과제 극복	28
엄청나게 많은 데이터 처리	29
자동화된 이상 탐지의 출현	30
비지도 ML 대 지도 ML	32
이상 탐지를 위한 비지도 ML 사용하기	33
특이에 관해 정의하기	34
정상 상태 학습하기	36
확률 모델	36
모델 학습하기	37
디트랜드	40
특이성에 대한 점수화	41
시간 요소	43
데이터 프레임 분석에 지도 ML 적용하기	43
지도 학습 과정	44
요약	45

2장	활성화와 운영화	47
	기술 요구 사항	48
	일래스틱 ML 기능 활성화	48
	자체 관리형 클러스터에서 ML 활성화	48
	클라우드에서 ML 활성화 - 일래스틱서치 서비스	51
	운영화의 이해	59
	ML 노드	59
	작업	61
	시계열 분석에서 데이터 버킷팅	62
	일래스틱 ML에 데이터 공급	63
	제공하는 인덱스	65
	.ml-config	66
	.ml-state-*	66
	.ml-notification-*	66
	.ml-annoataions-*	67
	.ml-stats-*	67
	.ml-anomalies-*	67
	이상 탐지 오케스트레이션	67
	이상 탐지 모델 스냅샷	68
	요약	69

2부 | 시계열 분석 - 이상 탐지와 예측 71

3장	이상 탐지	73
	기술 요구 사항	74
	일래스틱 ML 작업 유형	74
	탐지기 해부	77
	함수	77
	필드	78
	partition 필드	78
	by 필드	79

over 필드	79
공식(formula)	79
이벤트 비율의 변화 탐지	80
카운트 함수 탐색	81
다른 카운트 함수	94
논제로 카운트	94
디스팅트 카운트	96
메트릭 값에서 변화 탐지	97
메트릭 함수	98
min, max, mean, median과 metric	98
varp	98
sum, not-null sum	99
고급 탐지기 함수의 이해	100
레어(rare)	100
프리퀀시 레어(frequency rare)	101
정보 내용(information content)	102
지오그래픽	102
시간	103
범주형 피처로 분석 분할	103
분할 필드 설정	104
partition과 by_field를 사용한 분할의 차이점	106
이중 분할에 한계가 있을까?	107
시간 분석과 모집단 분석의 이해	108
비정형 메시지 범주화 분석	112
범주화에 훌륭한 후보가 되는 메시지 유형	113
범주화에 사용되는 프로세스	114
범주 분석	115
범주화 작업 예제	115
범주화 사용을 피해야 하는 경우	120
API를 통한 일래스틱 ML 관리	121
요약	123

4장	예측	125
	기술 요구 사항	126
	예언과 대비되는 예측	126
	예측 사용 사례	127
	작업의 예측 이론	128
	단일 시계열 예측	131
	예측 결과 검토	146
	다중 시계열 예측	152
	요약	156

5장	결과 해석	157
	기술 요구 사항	158
	일래스틱 ML 결과 인덱스 보기	158
	이상 징후 점수	164
	버킷 수준 스코어링	165
	정규화	167
	인플루언서 수준 점수	167
	인플루언서	170
	레코드 수준 점수	172
	결과 인덱스 스키마의 세부 정보	174
	버킷 결과	175
	레코드 결과	178
	인플루언서 결과	183
	다중 버킷 이상 징후	185
	다중 버킷 이상 징후 예제	186
	다중 버킷 스코어링	187
	예측 결과	188
	예측 결과 쿼리	190
	결과 API	193
	결과 API 엔드포인트	193
	전체 버킷 조회 API	194

범주 조회 API	196
사용자 정의 대시보드와 캔버스 워크패드	198
대시보드 “임베디블”	198
TSVB에서 이상 징후 추적	200
캔버스 워크패드 사용자 정의	203
요약	206

6장 ML 분석에 기반한 얼러팅 **207**

기술 요구 사항	208
얼러팅 개념 이해	208
모든 이상 징후가 얼러팅일 필요는 없다	208
실시간 얼러팅에는 타이밍이 중요하다	209
ML UI에서 얼러트 작성	213
샘플 이상 탐지 작업 정의	214
샘플 작업에 대한 얼러트 생성	220
실시간 이례적인 행위 시뮬레이션	226
얼러트 수신과 검토	229
와치(watch)로 얼러트 만들기	231
레거시 기본 ML 와치의 구조 이해	231
trigger 섹션	232
input 섹션	233
condition 섹션	237
action 섹션	238
사용자 정의 와치는 몇 가지 고유한 기능을 제공할 수 있다	240
연결된 입력과 스크립트 내의 조건	240
연결된 입력 간에 정보 전달	240
요약	242

7장 AIOps와 근본 원인 분석 **245**

기술 요구 사항	246
-----------------------	------------

AIOps 용어의 이해	246
KPI의 중요성과 한계 이해	248
KPI를 넘어서	252
더 나은 분석을 위한 데이터 조직화	254
이상 탐지 데이터피드에 대한 사용자 정의 쿼리	255
수집 시 데이터 강화	258
컨텍스트 정보 활용	260
분석 분할	260
통계적 인플루언서	261
RCA를 위해 모든 것을 통합	262
가동 중단 배경	262
상관관계와 공유된 인플루언서	264
요약	271
8장 다른 일래스틱 스택 앱에서 이상 탐지	273
기술 요구 사항	274
일래스틱 APM의 이상 탐지	274
APM에 대한 이상 탐지 활성화	274
APM UI에서 이상 탐지 작업 결과 조회	280
데이터 인식을 통한 ML 작업 생성	282
로그 앱의 이상 탐지	284
로그 카테고리	285
로그 이상 징후	287
메트릭 앱의 이상 탐지	288
업타임 앱의 이상 탐지	292
일래스틱 시큐리티 앱의 이상 탐지	296
사전 구축된 이상 탐지 작업	296
탐지 얼럿으로서의 이상 탐지 작업	299
요약	301

9장 데이터 프레임 분석 소개

305

기술 요구 사항	306
변환하는 방법 학습	307
왜 변환이 유용한가?	307
변환 작업의 내부 구조	308
전자 상거래 주문을 분석하기 위해 변환 사용	309
더 고급 수준의 피벗과 집계 구성 탐색	314
배치 변환과 연속 변환의 차이점 발견	317
연속 변환을 사용해 소셜 미디어 피드 분석	318
고급 변환 구성에 페인리스 사용	323
페인리스 소개	323
변수, 연산자, 제어 흐름	325
함수	332
파이썬과 일래스틱서치로 작업하기	337
파이썬 일래스틱서치 클라이언트에 대해 간략하게 둘러보기	338
일런드의 개발 목적 이해	340
일런드와 함께하는 첫걸음	341
요약	345
더 읽어보기	346

10장 아웃라이어 탐지

349

기술 요구 사항	350
아웃라이어 탐지의 내부 작동 이해	351
아웃라이어 탐지에 사용하는 4가지 기술 이해	352
거리 기반 기술	352
밀도 기반 기술	353
피처 영향력 이해	355
각 점에 대한 피처 영향력은 어떻게 계산하는가?	356

아웃라이어 탐지는 이상 탐지와 어떻게 다른가?	357
확률 모델 기반 대 인스턴스 기반	357
점수화	358
데이터 특성	358
온라인 대 배치(batch)	358
실제 아웃라이어 탐지 적용	359
Evaluate API로 아웃라이어 탐지 평가	365
아웃라이어 탐지를 위한 하이퍼파라미터 조정	372
요약	376
11장 분류 분석	379
기술 요구 사항	380
분류: 데이터에서 훈련된 모델로	381
데이터에서 분류 모델 학습	382
피처 엔지니어링	385
모델 평가	386
분류의 첫 걸음	387
분류의 내부 구조: 그래디언트 부스트 의사결정 트리	396
의사결정 트리 소개	397
그래디언트 부스트 의사결정 트리	398
하이퍼파라미터	399
결과 해석	402
분류 확률	404
분류 점수	404
피처 중요도	405
요약	407
더 알아보기	408
12장 회귀	409
기술 요구 사항	410

회귀 분석을 사용해 주택 가격 예측	410
회귀를 위한 의사결정 트리 사용	419
요약	422
더 읽어보기	423

13장 추론 425

기술 요구 사항	426
훈련된 모델 API 및 파이썬을 사용해 훈련된 머신러닝 모델을 검사하고 가져오며 내보내기	426
훈련된 모델 API 살펴보기	426
훈련된 모델 API와 파이썬을 사용해 훈련된 모델 내보내기와 가져오기	429
추론 프로세서와 인제스트 파이프라인 이해하기	434
인제스트 파이프라인에서 누락되거나 손상된 데이터 처리	446
예측에 대한 더 많은 통찰력을 얻기 위한 추론 프로세서 구성 옵션 사용하기	447
일련드를 사용해 외부 모델을 일래스틱서치로 가져오기	450
일련드에서 지원하는 외부 모델에 대해 알아보기	450
scikit-learn의 DecisionTreeClassifier로 훈련하고 일련드를 사용해 일래스틱서치로 가져오기	451
요약	457

부록 이상 탐지 팁 459

부록	459
기술 요구 사항	460
분할 작업 대 비분할 작업의 인플루언서 이해하기	460
단축함수를 유리하게 사용하기	466
기간 무시하기	469
예정된 (알려진) 시간 원도 무시하기	470
캘린더 이벤트 생성	470
원하는 타임프레임을 무시하기 위해 데이터피드 중지 및 시작	471
예기치 못한 시간 원도를 사후에 무시하기	472
작업의 복제와 과거 데이터의 재실행	472
작업을 이전 모델 스냅샷으로 되돌리기	472

사용자 정의 규칙과 필터 유리하게 사용하기	475
사용자 정의 규칙 만들기	475
“하향식” 얼러팅 철학에 대한 사용자 지정 규칙의 장점	477
이상 탐지 작업 처리량에 관한 고려 사항	478
사용 사례의 과도한 엔지니어링 방지하기	479
런타임 필드에서 이상 탐지 사용하기	480
요약	484
찾아보기	485



에이콘출판의 기틀을 마련하신 故 정완재 선생님 (1935-2004)