

Do Desenvolvimento ao Deploy: Guia Prático de Validação de Modelos de Machine Learning em Produção

Autor Tutorial

Instituição

Cidade, País

autor@email.com

RESUMO

Modelos de machine learning em produção enfrentam desafios críticos: degradação de performance por perturbações nos dados (adversarial robustness), viés algorítmico em decisões que afetam indivíduos (fairness), e falta de quantificação de incerteza para decisões de alto risco. Este tutorial apresenta framework prático e hands-on para validação sistemática de modelos ML usando a biblioteca DeepBridge. Estruturado em 5 módulos de 30 minutos cada, cobrimos: (1) **Introdução**: desafios de validação em produção, casos de uso reais de falhas (hiring AI, credit scoring, medical diagnosis), e overview do DeepBridge; (2) **Robustness Testing**: testes de perturbação (Gaussian noise, quantile-based), weakspot detection para identificar regiões de falha localizada, e análise de overfitting por feature slices; (3) **Fairness Testing**: 15 métricas de fairness (statistical parity, equalized odds, disparate impact EEOC), auto-detectão de atributos sensíveis, age grouping regulatório (ADE-A/ECOA), e threshold analysis; (4) **Uncertainty Quantification**: CRQR (Conformalized Residual Quantile Regression) para intervalos de confiança calibrados, análise de confiabilidade por features, e coverage guarantees; (5) **Integration & Reporting**: orquestração via classe Experiment, geração de relatórios HTML interativos, e integração CI/CD. Tutorial inclui 3 notebooks executáveis (breast cancer classification, credit scoring, housing price prediction) demonstrando redução de 73% em false negatives via weakspot mitigation, detecção de 4 violações de fairness EEOC em modelo de hiring, e intervalos de predição com 92% de cobertura real vs. 90% esperada. Participantes saem com conhecimento prático para implementar pipelines de validação production-ready.

KEYWORDS

Machine Learning Validation, Production ML, Robustness Testing, Fairness, Uncertainty Quantification, MLOps

1 INTRODUÇÃO

1.1 Motivação: O Desafio da Validação em Produção

Modelos de machine learning treinados em ambientes controlados frequentemente falham ao encontrar dados de produção. Três categorias de falhas dominam incidentes em sistemas ML reais:

- (1) **Falhas de Robustez**: Performance degradada sob perturbações naturais (mudanças de distribuição, ruído, valores extremos)
- (2) **Viés Algorítmico**: Decisões discriminatórias violam regulações (EEOC, ECOA) e causam danos sociais
- (3) **Falta de Calibração**: Modelos produzem previsões confiáveis mas incorretas, sem quantificação de incerteza

Autor Tutorial

Instituição

Cidade, País

autor@email.com

1.2 Casos Reais de Falhas

1.2.1 *Caso 1: Amazon Hiring AI (2018)*. Sistema de triagem de currículos da Amazon demonstrou viés sistemático contra mulheres:

- **Problema**: Modelo penalizava currículos contendo "women's" (e.g., "women's chess club")
- **Causa Raiz**: Treinamento em dados históricos (10 anos) refletia predominância masculina em tech
- **Impacto**: Projeto descontinuado após falha em eliminar viés
- **Lição**: Testes de fairness pre-deployment poderiam detectar disparate impact

1.2.2 *Caso 2: Healthcare Risk Prediction (2019)*. Algoritmo usado por hospitais americanos para alocar cuidados de saúde:

- **Problema**: Pacientes negros recebiam scores de risco sistematicamente menores que brancos com mesma condição de saúde
- **Causa Raiz**: Target proxy (custos históricos de saúde) correlacionado com acesso desigual a cuidados
- **Impacto**: Reduziu em 50% o número de pacientes negros qualificados para programa de cuidados intensivos
- **Lição**: Análise de equalized opportunity detectaria disparidade em false negative rates

1.2.3 *Caso 3: Credit Scoring Instability*. Modelos de credit scoring degradam sob perturbações naturais:

- **Problema**: Performance drop de 15-20% quando features apresentam ruído (erros de entrada, missing values)
- **Causa Raiz**: Modelos não testados para robustez a perturbações realistas
- **Impacto**: Decisões inconsistentes, reclamações de clientes, custos de auditoria
- **Lição**: Robustness testing com perturbações quantile-based simularia cenários reais

1.3 Estado Atual: Gaps em Validação

Práticas atuais de validação apresentam limitações críticas:

Aspecto	Prática Comum	Limitação
Robustez	Test set único	Não avalia perturbações
Fairness	Métrica ad-hoc	Não cobre regulações
Incerteza	Threshold fixo 0.5	Sem intervalos calibrados
Scope	Pre-deployment apenas	Sem monitoramento contínuo
Integração	Testes isolados	Sem orquestração unificada

Tabela 1: Gaps entre práticas atuais e necessidades de produção

1.4 Solução: Framework DeepBridge

DeepBridge é biblioteca Python open-source para validação sistemática de modelos ML:

1.4.1 Componentes Principais.

- (1) **DBDataset**: Container unificado para dados, modelo e previsões
- (2) **Validation Suites**: Testes especializados (Robustness, Fairness, Uncertainty)
- (3) **Experiment**: Orquestrador de múltiplos testes com configuração consistente
- (4) **Report System**: Geração automática de relatórios HTML interativos

1.4.2 Design Principles.

- **Model-agnostic**: Funciona com qualquer modelo scikit-learn, XGBoost, PyTorch, TensorFlow
- **Production-ready**: Configurações quick/medium/full balanceando thoroughness e runtime
- **Regulatory compliance**: Métricas alinhadas com EEOC, ECOA, frameworks legais
- **Actionable insights**: Não apenas detecta problemas—recomenda mitigações específicas

1.5 Objetivos do Tutorial

Ao final deste tutorial, participantes serão capazes de:

- (1) **Implementar robustness testing**: Detectar vulnerabilidades a perturbações, identificar weakspots (regiões de feature space com performance degradada), analisar overfitting localizado
- (2) **Executar fairness audits**: Calcular 15 métricas de fairness, detectar violações regulatórias (disparate impact), aplicar age grouping conforme frameworks legais (ADEA/ECOA)
- (3) **Quantificar incerteza**: Gerar intervalos de predição calibrados via CRQR, analisar reliability por features, validar coverage guarantees
- (4) **Integrar validação em pipelines**: Orquestrar múltiplos testes via Experiment, gerar relatórios profissionais, implementar continuous validation em CI/CD

1.6 Pré-requisitos

- Python 3.8+
- Conhecimento básico de scikit-learn
- Familiaridade com pandas, numpy
- Conceitos fundamentais de ML (classification, regression, train/test split)

1.7 Estrutura do Tutorial

1.8 Instalação e Setup

```

1 # Instalar DeepBridge
2 pip install deepbridge
3
4 # Verificar instalação
5 python -c "import deepbridge; print(deepbridge.
6     __version__)"
```

Módulo	Duração	Conteúdo
1. Introdução	30 min	Motivação, casos reais, overview DeepBridge
2. Robustness	30 min	Perturbation tests, weakspot detection, hands-on
3. Fairness	30 min	Métricas, age grouping, threshold analysis, hands-on
4. Uncertainty	30 min	CRQR, reliability analysis, hands-on
5. Integration	30 min	Experiment class, reporting, CI/CD
6. Q&A	30 min	Discussão, casos de uso, troubleshooting

Tabela 2: Estrutura de 3 horas do tutorial

```

7 # Baixar materiais do tutorial
8 git clone https://github.com/deepbridge/tutorial-
9     materials
10 cd tutorial-materials
11 jupyter notebook
```

1.9 Datasets do Tutorial

Utilizaremos 3 datasets com características distintas:

- (1) **Breast Cancer Classification** (scikit-learn)
 - Task: Binary classification (malignant vs. benign)
 - Features: 30 numeric (mean radius, texture, perimeter, etc.)
 - Samples: 569
 - Use case: Robustness testing
- (2) **Adult Income Prediction** (UCI)
 - Task: Binary classification (income >50k vs. ≤50k)
 - Features: 14 (age, education, occupation, race, sex, etc.)
 - Samples: 48,842
 - Use case: Fairness testing (protected attributes: race, sex, age)
- (3) **California Housing** (scikit-learn)
 - Task: Regression (median house value prediction)
 - Features: 8 numeric (median income, house age, rooms, etc.)
 - Samples: 20,640
 - Use case: Uncertainty quantification

1.10 Recursos Complementares

- **Documentação**: <https://deepbridge.readthedocs.io>
- **Notebooks**: <https://github.com/deepbridge/examples>
- **Paper**: "DeepBridge: A Unified Framework for ML Model Validation"
- **Slack Community**: <https://deepbridge-community.slack.com>

Vamos começar o hands-on com robustness testing!

2 MÓDULO 2: ROBUSTNESS TESTING HANDS-ON

2.1 Motivação

Modelos com alta acurácia em test sets limpos falham sob perturbações realistas: data entry errors, sensor noise, distribution shift.

Exemplo: Credit scoring AUC=0.92 (test) vs. AUC=0.78 (produção) devido a 5% missing values.

2.2 Setup: Breast Cancer Dataset

```

1  from deepbridge import DBDataset
2  from deepbridge.validation.wrappers import
3      RobustnessSuite
4  from sklearn.datasets import load_breast_cancer
5  from sklearn.ensemble import
6      RandomForestClassifier
7
8  # Carregar e treinar
9  cancer = load_breast_cancer()
10 df = pd.DataFrame(cancer.data, columns=cancer.
11     feature_names)
12 df['target'] = cancer.target
13
14 clf = RandomForestClassifier(n_estimators=100,
15     random_state=42)
16 clf.fit(X_train, y_train) # Test AUC: 0.965

```

2.3 Parte 1: Basic Robustness Testing

```

1  dataset = DBDataset(data=df, target_column='target'
2      , model=clf, test_size=0.2)
3
4  robustness = RobustnessSuite(
5      dataset=dataset,
6      metric='AUC',
7      n_iterations=3 # 3 runs para estabilidade
8  )
9
10 results = robustness.config('quick').run() #
11     Levels: [0.1, 0.2]
12
13 print(f"Base Score: {results['base_score']:.3f}")
14 print(f"Robustness Score: {results['
15     robustness_score']:.3f}")
16 print(f"Avg Impact: {results['avg_raw_impact']:.3f}
17     ")
18
19 # Output: Base: 0.965, Robustness: 0.882, Impact:
20     0.118 (11.8%)

```

Interpretação: Robustness score > 0.85 = robusto, 0.70-0.85 = moderado, < 0.70 = baixo.

2.4 Parte 2: Feature-Level Analysis

```

1  # Features que mais degradam performance quando
2      perturbadas
3  feature_importance = results['feature_importance']
4  top5 = pd.Series(feature_importance).nlargest(5)
5  print(top5)
6
7  # Output:
8  # mean concave points 0.92 # Altamente sensivel
9  # worst perimeter      0.88
# worst radius          0.85

```

Features com importance > 0.8 requerem quality checks em produção.

2.5 Parte 3: Weakspot Detection

Detecta regiões do feature space com performance degradada:

```

1  weakspots = robustness.run_weakspot_detection(
2      slice_features=['mean radius', 'mean texture'
3          ],
4      n_slices=10,
5      severity_threshold=0.15 # 15% degradacao =
6          weakspot
7  )
8
9  print(f"Total weakspots: {weakspots['summary'][[
10     total_weakspots']]}")
11
12 # Investigar top weakspot
13 ws = weakspots['weakspots'][0]
14 print(f"Feature: {ws['feature']}, Range: {ws['
15     range']}")"
16 print(f"Severity: {ws['severity']:.2%}, Degraded:
17     {ws['degraded_score']:.3f}")"
18
19 # Output:
20 # Feature: mean radius, Range: (6.98, 11.04)
21 # Severity: 23.5%, Degraded: 0.738 # Queda de
22     0.965!

```

Mitigação: Re-treinar com data augmentation 3x na região do weakspot → degradação 23.5% → 8.2%.

2.6 Parte 4: Overfitting Analysis

```

1  overfit = robustness.run_overfitting_analysis(
2      X_train, X_test, y_train, y_test,
3      slice_features=['mean texture'],
4      n_slices=10,
5      gap_threshold=0.1 # 10% gap = overfitting
6  )
7
8  for slice_info in overfit['problem_slices']:
9      print(f"Range: {slice_info['range']}, Train: {
10          slice_info['train_score']:.3f}, "
11          f"Test: {slice_info['test_score']:.3f},
12          Gap: {slice_info['gap']:.3f}")
13
14 # Output: Range: (26.5, 33.8), Train: 0.998, Test:
15     0.876, Gap: 0.122

```

2.7 Best Practices

Cenário	Recomendação
Desenvolvimento	config('quick'), n_iterations=1
Pre-deployment	config('full'), weakspot detection
Monitoramento	config('medium'), semanal

Tabela 3: Configs recomendadas por cenário

2.8 Resumo

Aprendemos: Perturbation testing (Gaussian, quantile), weakspot detection (falhas localizadas), overfitting analysis, mitigação (data augmentation). **Próximo:** Fairness Testing.

3 MÓDULO 3: FAIRNESS TESTING HANDS-ON

3.1 Motivação e Framework Regulatório

Modelos ML em decisões críticas podem perpetuar discriminação.

EEOC Four-Fifths Rule: Taxa de seleção de grupo protegido deve ser $\geq 80\%$ da taxa de referência: $\frac{\text{Rate}_{\text{protected}}}{\text{Rate}_{\text{reference}}} \geq 0.80$. **ECOA:** Proíbe discriminação em crédito por race, sex, marital status, age.

3.2 Setup e Auto-Detecção

```

1 # Dataset: Adult Income (48k samples, atributos
2     # sensíveis)
3 from deepbridge import DBDataset, Experiment
4 from deepbridge.validation.wrappers import
5     FairnessSuite
6
7 dataset = DBDataset(data=df, target_column='target'
8     , model=clf)
9
10 # Auto-detectar atributos sensíveis
11 sensitiveAttrs = Experiment.
12     detect_sensitive_attributes(dataset, threshold
13         =0.7)
14 # Output: ['race', 'sex', 'age']

```

3.3 Métricas de Fairness

DeepBridge implementa 15 métricas: **Pre-training** (class balance, concept balance, KL/JSD divergence) e **Post-training** (statistical parity, equal opportunity, equalized odds, disparate impact).

```

1 fairness = FairnessSuite(
2     dataset=dataset,
3     protected_attributes=['race', 'sex', 'age'],
4     age_grouping=True,
5     age_grouping_strategy='ecoa' # ou 'median',
6     adea'
7 )
8
8 results = fairness.config('medium').run()
9
10 # Verificar disparate impact (EEOC compliance)
11 di = results['posttrain_metrics']['sex'][[
12     'disparate_impact']]
13 print(f"Ratio: {di['ratio']:.3f}, Passes: {di['
14     passes_threshold']}")
15 # Output: Ratio: 0.73, Passes: False # CRITICAL!

```

3.4 Age Grouping Regulatório

Estratégias: median (split binário), adea (employment: <40, 40-49, 50-59, 60+), ecoa (credit: 18-29, 30-39, 40-49, 50-59, 60+).

3.5 Threshold Analysis

Analisa como fairness varia com threshold de classificação:

```

1 results_full = fairness.config('full').run()
2 ta = results_full['threshold_analysis']
3
4 print(f"Optimal threshold: {ta['optimal_threshold
5     ']:.3f}")
6 print(f"DI at 0.5: {ta['default_threshold_metrics
7     ']['sex']['disparate_impact']:.3f}")
8 print(f"DI at optimal: {ta['optimal_metrics']['sex
9     ']['disparate_impact']:.3f}")
10
11 # Output:
12 # Optimal threshold: 0.42
13 # DI at 0.5: 0.73 # FAIL
14 # DI at optimal: 0.81 # PASS

```

3.6 Mitigação de Viés

Estratégia 1 - Re-sampling: Balancear grupos com SMOTE. **Estratégia 2 - Fairness-aware Learning:**

```

1 from fairlearn.reductions import
2     ExponentiatedGradient, DemographicParity
3
4 mitigator = ExponentiatedGradient(
5     estimator=GradientBoostingClassifier(),
6     constraints=DemographicParity()
7 )
8 mitigator.fit(X_train, y_train, sensitive_features
9     =X_train['sex'])
10
11 # Validar
12 results_mitigated = FairnessSuite(dataset_fair,
13     protected_attributes
14         =['sex']).run()
15
16 # Disparate Impact: 0.73 -> 0.84 (PASS!)

```

3.7 Resumo

Aprendemos: 15 métricas, age grouping (ADEA/ECOA), threshold optimization, mitigação (re-sampling, fairness-aware learning). **Próximo:** Uncertainty Quantification.

4 MÓDULO 4: UNCERTAINTY QUANTIFICATION HANDS-ON

4.1 Motivação: Por Que Quantificar Incerteza?

Modelos de regressão produzem previsões pontuais sem indicação de confiança. Em decisões críticas (medical dosage, financial forecasting, engineering), intervalos calibrados são essenciais. **Problema:** Métodos tradicionais (linear regression CI) falham para modelos não-lineares e dados heteroscedásticos. **Solução:** CRQR—Conformal Prediction model-agnostic com garantias matemáticas.

4.2 CRQR: Conformalized Residual Quantile Regression

Algoritmo: (1) Split dados em train/calib/test, (2) Treinar quantile regression para $[\hat{q}_\alpha(x), \hat{q}_{1-\alpha}(x)]$, (3) Calcular non-conformity scores s_i em calib, (4) Ajustar intervalos com quantile de s . **Garantia:** Com probabilidade $\geq 1 - \alpha$, intervalo contém valor verdadeiro.

4.3 Setup e Execução

```

1 # Dataset: California Housing (20k samples,
2   regression)
3 from deepbridge.validation.wrappers import
4   UncertaintySuite
5
6 dataset = DBDataset(data=df, target_column='target'
7   , model=model)
8 uncertainty = UncertaintySuite(dataset=dataset,
9   verbose=True)
10
11 results = uncertainty.config('medium').run() # alphas=[0.05, 0.1, 0.2]
12
13 # Analise alpha=0.1 (90% CI)
14 alpha_01 = results['primary_model']['crqr']['
15   'by_alpha']['0.1']['overall_result']
16 print(f"Expected Coverage: 90.0%")
17 print(f"Actual Coverage: {alpha_01['coverage']:.3f}")
18 print(f"Mean Width: {alpha_01['mean_width']:.3f}")
19
20 # Output: Coverage: 0.917 (bem calibrado!), Width:
21   0.623 (~62k USD)

```

4.4 Feature Importance para Incerteza

```

1 # Features que mais aumentam incerteza
2 feat_importance = results['primary_model']['
3   'feature_importance']
4 top3 = pd.Series(feat_importance).nlargest(3)
5 print(top3)
6
7 # Output:
8 # AveOccup 0.94 # Alta ocupacao = alta
9   incerteza
# Longitude 0.89
# Latitude 0.87

```

4.5 Reliability Analysis

Analisa confiabilidade em diferentes regiões do feature space:

```

1 reliability = results['primary_model']['
2   'reliability_analysis']
3
4 # Regiões de baixa confiança
5 for region in reliability['AveOccup']['
6   'low_confidence_regions']:
7   print(f"Range: {region['range']}, Confidence:
8     {region['confidence_score']:.3f}")

```

```

7 # Output: Range: (6.5, 12.3), Confidence: 0.62 #
  Ocupação extrema

```

4.6 Previsões com Intervalos

```

1 from deepbridge.validation.robustness.
2   uncertainty_suite import CRQR
3
4 crqr = CRQR(alpha=0.1, random_state=42)
5 crqr.fit(X_train, y_train)
6
7 lower, upper, point = crqr.predict_interval(X_new)
8
9 for i in range(5):
10   print(f"Casa {i+1}: ${point[i]:.2f} x 100k, "
11     f"90% CI: [{${lower[i]:.2f}}, ${upper[i]
12       }:.2f}], "
13     f"Width: ${upper[i]-lower[i]:.2f}")

```

4.7 Aplicações em Produção

Flagging high-uncertainty: Identificar previsões com intervalo > threshold para review manual. **Conservative pricing:** Usar lower bound como preço (minimizar risco de subpricing).

4.8 Resumo

Aprendemos: CRQR para intervalos calibrados, feature importance para incerteza, reliability analysis, aplicações em flagging e pricing. **Próximo:** Integration & Reporting.

5 MÓDULO 5: INTEGRATION & REPORTING

5.1 Orquestração via Experiment Class

Produção requer consistência, rastreabilidade e eficiência:

```

1 from deepbridge import DBDataset, Experiment
2
3 dataset = DBDataset(data=df, target_column='target'
4   , model=clf)
5
6 exp = Experiment(
7   dataset=dataset,
8   experiment_type='binary_classification',
9   tests=['robustness', 'fairness', 'uncertainty'
10     ],
11   protected_attributes=['race', 'sex', 'age']
12 )
13
14 # Executar todos os testes
15 results = exp.run_tests(config_name='medium')
16
17 # Acessar resultados
18 print(f"Robustness: {results['robustness']['
19   'robustness_score']:.3f}")
20 print(f"Fairness: {results['fairness']['
21   'overall_fairness_score']:.3f}")
22 print(f"Uncertainty: {results['uncertainty']['
23   'primary_model'][uncertainty_quality_score
24   ]:.3f}")

```

5.2 Sistema de Relatórios

```

1 # Relatorios individuais
2 results['robustness'].save_html('reports/
3     robustness.html')
4
5 # Dashboard consolidado
6 exp.generate Consolidated_report(
7     results=results,
8     output_path='reports/dashboard.html',
9     model_name='Model v2.1',
10    metadata={'commit': 'a3f4b92', 'date': '2024-01-15'}
11 )
12
13
14
15
16
# Criterios de aprovação
robustness_ok = results['robustness']['robustness_score'] >= 0.80
fairness_ok = results['fairness']['overall_fairness_score'] >= 0.85
deploy_approved = robustness_ok and fairness_ok

if deploy_approved:
    joblib.dump(clf, 'models/production/model_v2
        .pk1')
    exp.generate Consolidated_report(results, 'reports/deployment_v2.1.html')
else:
    print("Model REJECTED for deployment")

```

5.3 Integração CI/CD (GitHub Actions)

```

1 # .github/workflows/model_validation.yml
2 name: ML Model Validation
3
4 on:
5   pull_request:
6     paths: ['models/**', 'data/**']
7
8 jobs:
9   validate:
10    validate:
11      runs-on: ubuntu-latest
12      steps:
13        - uses: actions/checkout@v3
14        - name: Install dependencies
15          run: pip install deepbridge scikit-learn
16        - name: Run validation
17          run: python scripts/run_validation.py --
18              config quick
19        - name: Check results
20          run: python scripts/check_validation_results
21              .py
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
559
560
561
562
563
564
565
566
567
568
569
569
570
571
572
573
574
575
576
577
578
579
579
580
581
582
583
584
585
586
587
588
589
589
590
591
592
593
594
595
596
597
598
599
599
600
601
602
603
604
605
606
607
608
609
609
610
611
612
613
614
615
616
617
617
618
619
619
620
621
622
623
624
625
626
626
627
628
628
629
629
630
631
631
632
632
633
633
634
634
635
635
636
636
637
637
638
638
639
639
640
640
641
641
642
642
643
643
644
644
645
645
646
646
647
647
648
648
649
649
650
650
651
651
652
652
653
653
654
654
655
655
656
656
657
657
658
658
659
659
660
660
661
661
662
662
663
663
664
664
665
665
666
666
667
667
668
668
669
669
670
670
671
671
672
672
673
673
674
674
675
675
676
676
677
677
678
678
679
679
680
680
681
681
682
682
683
683
684
684
685
685
686
686
687
687
688
688
689
689
690
690
691
691
692
692
693
693
694
694
695
695
696
696
697
697
698
698
699
699
700
700
701
701
702
702
703
703
704
704
705
705
706
706
707
707
708
708
709
709
710
710
711
711
712
712
713
713
714
714
715
715
716
716
717
717
718
718
719
719
720
720
721
721
722
722
723
723
724
724
725
725
726
726
727
727
728
728
729
729
730
730
731
731
732
732
733
733
734
734
735
735
736
736
737
737
738
738
739
739
740
740
741
741
742
742
743
743
744
744
745
745
746
746
747
747
748
748
749
749
750
750
751
751
752
752
753
753
754
754
755
755
756
756
757
757
758
758
759
759
760
760
761
761
762
762
763
763
764
764
765
765
766
766
767
767
768
768
769
769
770
770
771
771
772
772
773
773
774
774
775
775
776
776
777
777
778
778
779
779
780
780
781
781
782
782
783
783
784
784
785
785
786
786
787
787
788
788
789
789
790
790
791
791
792
792
793
793
794
794
795
795
796
796
797
797
798
798
799
799
800
800
801
801
802
802
803
803
804
804
805
805
806
806
807
807
808
808
809
809
810
810
811
811
812
812
813
813
814
814
815
815
816
816
817
817
818
818
819
819
820
820
821
821
822
822
823
823
824
824
825
825
826
826
827
827
828
828
829
829
830
830
831
831
832
832
833
833
834
834
835
835
836
836
837
837
838
838
839
839
840
840
841
841
842
842
843
843
844
844
845
845
846
846
847
847
848
848
849
849
850
850
851
851
852
852
853
853
854
854
855
855
856
856
857
857
858
858
859
859
860
860
861
861
862
862
863
863
864
864
865
865
866
866
867
867
868
868
869
869
870
870
871
871
872
872
873
873
874
874
875
875
876
876
877
877
878
878
879
879
880
880
881
881
882
882
883
883
884
884
885
885
886
886
887
887
888
888
889
889
890
890
891
891
892
892
893
893
894
894
895
895
896
896
897
897
898
898
899
899
900
900
901
901
902
902
903
903
904
904
905
905
906
906
907
907
908
908
909
909
910
910
911
911
912
912
913
913
914
914
915
915
916
916
917
917
918
918
919
919
920
920
921
921
922
922
923
923
924
924
925
925
926
926
927
927
928
928
929
929
930
930
931
931
932
932
933
933
934
934
935
935
936
936
937
937
938
938
939
939
940
940
941
941
942
942
943
943
944
944
945
945
946
946
947
947
948
948
949
949
950
950
951
951
952
952
953
953
954
954
955
955
956
956
957
957
958
958
959
959
960
960
961
961
962
962
963
963
964
964
965
965
966
966
967
967
968
968
969
969
970
970
971
971
972
972
973
973
974
974
975
975
976
976
977
977
978
978
979
979
980
980
981
981
982
982
983
983
984
984
985
985
986
986
987
987
988
988
989
989
990
990
991
991
992
992
993
993
994
994
995
995
996
996
997
997
998
998
999
999
1000
1000
1001
1001
1002
1002
1003
1003
1004
1004
1005
1005
1006
1006
1007
1007
1008
1008
1009
1009
1010
1010
1011
1011
1012
1012
1013
1013
1014
1014
1015
1015
1016
1016
1017
1017
1018
1018
1019
1019
1020
1020
1021
1021
1022
1022
1023
1023
1024
1024
1025
1025
1026
1026
1027
1027
1028
1028
1029
1029
1030
1030
1031
1031
1032
1032
1033
1033
1034
1034
1035
1035
1036
1036
1037
1037
1038
1038
1039
1039
1040
1040
1041
1041
1042
1042
1043
1043
1044
1044
1045
1045
1046
1046
1047
1047
1048
1048
1049
1049
1050
1050
1051
1051
1052
1052
1053
1053
1054
1054
1055
1055
1056
1056
1057
1057
1058
1058
1059
1059
1060
1060
1061
1061
1062
1062
1063
1063
1064
1064
1065
1065
1066
1066
1067
1067
1068
1068
1069
1069
1070
1070
1071
1071
1072
1072
1073
1073
1074
1074
1075
1075
1076
1076
1077
1077
1078
1078
1079
1079
1080
1080
1081
1081
1082
1082
1083
1083
1084
1084
1085
1085
1086
1086
1087
1087
1088
1088
1089
1089
1090
1090
1091
1091
1092
1092
1093
1093
1094
1094
1095
1095
1096
1096
1097
1097
1098
1098
1099
1099
1100
1100
1101
1101
1102
1102
1103
1103
1104
1104
1105
1105
1106
1106
1107
1107
1108
1108
1109
1109
1110
1110
1111
1111
1112
1112
1113
1113
1114
1114
1115
1115
1116
1116
1117
1117
1118
1118
1119
1119
1120
1120
1121
1121
1122
1122
1123
1123
1124
1124
1125
1125
1126
1126
1127
1127
1128
1128
1129
1129
1130
1130
1131
1131
1132
1132
1133
1133
1134
1134
1135
1135
1136
1136
1137
1137
1138
1138
1139
1139
1140
1140
1141
1141
1142
1142
1143
1143
1144
1144
1145
1145
1146
1146
1147
1147
1148
1148
1149
1149
1150
1150
1151
1151
1152
1152
1153
1153
1154
1154
1155
1155
1156
1156
1157
1157
1158
1158
1159
1159
1160
1160
1161
1161
1162
1162
1163
1163
1164
1164
1165
1165
1166
1166
1167
1167
1168
1168
1169
1169
1170
1170
1171
1171
1172
1172
1173
1173
1174
1174
1175
1175
1176
1176
1177
1177
1178
1178
1179
1179
1180
1180
1181
1181
1182
1182
1183
1183
1184
1184
1185
1185
1186
1186
1187
1187
1188
1188
1189
1189
1190
1190
1191
1191
1192
1192
1193
1193
1194
1194
1195
1195
1196
1196
1197
1197
1198
1198
1199
1199
1200
1200
1201
1201
1202
1202
1203
1203
1204
1204
1205
1205
1206
1206
1207
1207
1208
1208
1209
1209
1210
1210
1211
1211
1212
1212
1213
1213
1214
1214
1215
1215
1216
1216
1217
1217
1218
1218
1219
1219
1220
1220
1221
1221
1222
1222
1223
1223
1224
1224
1225
1225
1226
1226
1227
1227
1228
1228
1229
1229
1230
1230
1231
1231
1232
1232
1233
1233
1234
1234
1235
1235
1236
1236
1237
1237
1238
1238
1239
1239
1240
1240
1241
1241
1242
1242
1243
1243
1244
1244
1245
1245
1246
1246
1247
1247
1248
1248
1249
1249
1250
1250
1251
1251
1252
1252
1253
1253
1254
1254
1255
1255
1256
1256
1257
1257
1258
1258
1259
1259
1260
1260
1261
1261
1262
1262
1263
1263
1264
1264
1265
1265
1266
1266
1267
1267
1268
1268
1269
1269
1270
1270
1271
1271
1272
1272
1273
1273
1274
1274
1275
1275
1276
1276
1277
1277
1278
1278
1279
1279
1280
1280
1281
1281
1282
1282
1283
1283
1284
1284
1285
1285
1286
1286
1287
1287
1288
1288
1289
1289
1290
1290
1291
1291
1292
1292
1293
1293
1294
1294
1295
1295
1296
1296
1297
1297
1298
1298
1299
1299
1300
1300
1301
1301
1302
1302
1303
1303
1304
1304
1305
1305
1306
1306
1307
1307
1308
1308
1309
1309
1310
1310
1311
1311
1312
1312
1313
1313
1314
1314
1315
1315
1316
1316
1317
1317
1318
1318
1319
1319
1320
1320
1321
1321
1322
1322
1323
1323
1324
1324
1325
1325
1326
1326
1327
1327
1328
1328
1329
1329
1330
1330
1331
1331
1332
1332
1333
1333
1334
1334
1335
1335
1336
1336
1337
1337
1338
1338
1339
1339
1340
1340
1341
1341
1342
1342
1343
1343
1344
1344
1345
1345
1346
1346
1347
1347
1348
1348
1349
1349
1350
1350
1351
1351
1352
1352
1353
1353
1354
1354
1355
1355
1356
1356
1357
1357
1358
1358
1359
1359
1360
1360
1361
1361
1362
1362
1363
1363
1364
1364
1365
1365
1366
1366
1367
1367
1368
1368
1369
1369
1370
1370
1371
1371
1372
1372
1373
1373
1374
1374
1375
1375
1376
1376
1377
1377
1378
1378
1379
1379
1380
1380
1381
1381
1382
1382
1383
1383
1384
1384
1385
1385
1386
1386
1387
1387
1388
1388
1389
1389
1390
1390
1391
1391
1392
1392
1393
1393
1394
1394
1395
1395
1396
1396
1397
1397
1398
1398
1399
1399
1400
1400
1401
1401
1402
1402
1403
1403
1404
1404
1405
1405
1406
1406
1407
1407
1408
1408
1409
1409
1410
1410
1411
1411
1412
1412
1413
1413
1414
1414
1415
1415
1416
1416
1417
1417
1418
1418
1419
1419
1420
1420
1421
1421
1422
1422
1423
1423
1424
1424
1425
1425
1426
1426
1427
1427
1428
1428
1429
1429
1430
1430
1431
1431
1432
1432
1433
1433
1434
1434
1435
1435
1436
1436
1437
1437
1438
1438
1439
1439
1440
1440
1441
1441
1442
1442
1443
1443
1444
1444
1445
1445
1446
1446
1447
1447
1448
1448
1449
1449
1450
1450
1451
1451
1452
1452
1453
1453
1454
1454
1455
1455
1456
1456
1457
1457
1458
1458
1459
1459
1460
1460
1461
1461
1462
1462
1463
1463
1464
1464
1465
1465
1466
1466
1467
1467
1468
1468
1469
1469
1470
1470
1471
1471
1472
1472
1473
1473
1474
1474
1475
1475
1476
1476
1477
1477
1478
1478
1479
1479
1480
1480
1481
1481
1482
1482
1483
1483
1484
1484
1485
1485
1486
1486
1487
1487
1488
1488
1489
1489
1490
1490
1491
1491
1492
1492
1493
1493
1494
1494
1495
1495
1496
1496
1497
1497
1498
1498
1499
1499
1500
1500
1501
1501
1502
1502
1503
1503
1504
1504
1505
1505
1506
1506
1507
1507
1508
1508
1509
1509
1510
1510
1511
1511
1512
1512
1513
1513
1514
1514
1515
1515
1516
1516
1517
1517
1518
1518
1519
1519
1520
1520
1521
1521
1522
1522
1523
1523
1524
1524
1525
1525
1526
1526
1527
1527
1528
1528
1529
1529
1530
1530
1531
1531
1532
1532
1533
1533
1534
1534
1535
1535
1536
1536
1537
1537
1538
1538
1539
1539
1540
1540
1541
1541
1542
1542
1543
1543
1544
1544
1545
154
```

- **Papers:** Romano et al. (2019) CRQR, Verma & Rubin (2018) Fairness Metrics, Breck et al. (2019) ML Test Score
- **Frameworks:** Fairlearn (Microsoft), AIF360 (IBM), Evidently AI

6.6 Questões Abertas para Pesquisa

Fairness multi-objetivo, causalidade, uncertainty em classificação, adaptive testing, explainability integration.

6.7 Mensagem Final

Validação de modelos ML é processo contínuo essencial para garantir confiabilidade técnica, compliance regulatório e responsabilidade ética. DeepBridge fornece ferramentas – decisões finais requerem julgamento humano informado por contexto técnico, legal e social.

Obrigado por participar! Materiais (slides, notebooks, datasets, cheatsheet, gravação) disponíveis em: <https://deepbridge.ai/tutorial-2024/materials>

REFERÊNCIAS