

| | |
|---------------|---|
| ●序文 | 3 |
| ●本書について | 4 |

■ CHAPTER 01

PyCaretとは

| | |
|-------------------------|----|
| □01 モジュールとパッケージ | 18 |
| ▶モジュールとは | 18 |
| ▶パッケージとは | 19 |
| ▶モジュールとパッケージの違い | 20 |
| □02 ライブラリの概要 | 21 |
| ▶ライブラリとは | 21 |
| ▶ライブラリのインストール | 22 |
| □03 PyCaretの概要 | 23 |
| ▶PyCaretとは | 23 |
| ▶PyCaretのインストール方法 | 23 |
| □04 本章のまとめ | 26 |

■ CHAPTER 02

環境構築

| | |
|---------------------------------------|----|
| □05 Google Colaboratoryの概要 | 28 |
| ▶Google Colaboratoryとは | 28 |
| ▶Google Colaboratoryのメリット・デメリット | 28 |
| □06 Google Colaboratoryの使い方 | 30 |
| ▶起動方法 | 30 |
| ▶実行方法 | 31 |
| ▶テキストの活用 | 31 |
| ▶シェルコマンド | 32 |
| ▶GPUの選択 | 33 |
| ▶ファイルの読み書き方法 | 34 |
| □07 本章のまとめ | 38 |

■ CHAPTER 03

機械学習の流れ

| | |
|-------------------------|----|
| □08 機械学習プロジェクトの流れ | 40 |
| □09 目的の定義 | 41 |
| □10 データの収集 | 42 |
| □11 機械学習の概要 | 43 |
| ▶機械学習の種類 | 46 |
| ▶教師あり学習 | 46 |
| ▶教師なし学習 | 46 |
| ▶強化学習 | 47 |
| □12 データの前処理 | 48 |
| □13 モデル学習 | 52 |
| □14 モデル評価 | 54 |
| □15 デプロイ | 57 |
| □16 本章のまとめ | 58 |

■ CHAPTER 04

教師あり学習 回帰分析

| | |
|-----------------------------------|----|
| □17 回帰分析の概要 | 60 |
| ▶質的変数と量的変数 | 60 |
| ▶連続型と離散型 | 60 |
| ▶回帰分析 | 60 |
| □18 PyCaretに含まれる回帰モデル | 61 |
| □19 Linear Regression(線形回帰) | 62 |
| ▶線形回帰分析のイメージ | 62 |
| ▶用語の整理 | 64 |
| ▶パラメータの求め方 | 65 |
| ▶残差二乗和の計算 | 68 |
| ▶重回帰分析 | 70 |
| ▶実践 | 73 |
| ▶多項式の例 | 74 |

| | | |
|------|-----------------------------|-----|
| □ 20 | Lasso回帰・Ridge回帰・Elastic net | 75 |
| | ▶ 過学習 | 75 |
| | ▶ 正則化 | 81 |
| | ▶ ラッソ回帰 | 82 |
| | ▶ リッジ回帰 | 83 |
| | ▶ L1正則化とL2正則化 | 83 |
| | ▶ ラッソ回帰とリッジ回帰の性質の違い | 84 |
| | ▶ Elastic Net | 89 |
| | ▶ scikit-learnによる実践 | 90 |
| □ 21 | MLP Regressor(多層パーセプトロン回帰) | 92 |
| | ▶ 取り扱う問題 | 92 |
| | ▶ 入力層・中間層・出力層 | 93 |
| | ▶ 活性化関数 | 95 |
| | ▶ 重みの最適化と損失関数 | 97 |
| | ▶ 中間層を追加したニューラルネットワーク | 98 |
| | ▶ scikit-learnによる実装 | 100 |
| □ 22 | その他の回帰手法 | 101 |
| □ 23 | PyCaretによる実践 | 102 |
| □ 24 | 本章のまとめ | 105 |

■ CHAPTER 05

教師あり学習 分類

| | | |
|------|----------------------|-----|
| □ 25 | 分類とは | 108 |
| □ 26 | PyCaretによる分類モデル | 109 |
| □ 27 | ロジスティック回帰 | 110 |
| | ▶ 線形回帰で解いてみる | 111 |
| | ▶ ロジスティック回帰のアイデア | 112 |
| | ▶ ロジスティック回帰の実践 | 118 |
| □ 28 | Naive Bayes(ナイーブベイズ) | 119 |
| | ▶ ベイズの定理 | 119 |
| | ▶ ベイズの定理の分類問題への応用 | 121 |
| | ▶ ナイーブな仮定 | 122 |
| | ▶ ナイーブベイズを計算してみる | 123 |
| | ▶ ゼロ頻度問題 | 124 |
| | ▶ ナイーブベイズの実践 | 125 |

| | | |
|------|------------------------------|-----|
| □ 29 | KNN(K近傍法) | 126 |
| | ▶ 大学の可否を予測する例 | 126 |
| | ▶ 回帰問題への応用 | 127 |
| | ▶ K近傍法の実践 | 128 |
| □ 30 | 決定木 | 129 |
| | ▶ 決定木のイメージ | 129 |
| | ▶ 不純度 | 131 |
| | ▶ 回帰問題への応用 | 135 |
| | ▶ 決定木の実践 | 136 |
| □ 31 | Random Forest | 137 |
| | ▶ バギング | 137 |
| | ▶ ランダムフォレストの実践 | 138 |
| | ▶ Extra Trees Classifier | 139 |
| □ 32 | ブースティング | 140 |
| | ▶ AdaBoost | 140 |
| | ▶ 勾配ブースティング | 141 |
| | ▶ 勾配ブースティングの実践 | 143 |
| □ 33 | Linear Discriminant Analysis | 144 |
| □ 34 | PyCaretによる分類の実践 | 146 |
| □ 35 | 本章のまとめ | 149 |

■ CHAPTER 06

教師なし学習 クラスタリング

| | | |
|------|--------------------------------------|-----|
| □ 36 | クラスタリングとは | 152 |
| | ▶ クラスタリングのイメージ | 152 |
| □ 37 | PyCaretに含まれるクラスタリングモデル | 154 |
| □ 38 | K平均法(K-Means) | 155 |
| | ▶ K-Meansのアルゴリズム | 155 |
| | ▶ K-Meansの実践 | 158 |
| □ 39 | 階層的クラスタリング(Agglomerative Clustering) | 159 |
| | ▶ 階層的クラスタリングのアルゴリズム | 159 |
| | ▶ 階層的クラスタリングの実践 | 164 |

| | | |
|-------|---|-----|
| □ 4.0 | Density-Based Spatial Clustering of Applications with Noise(DBSCAN) | 165 |
| ▶ | DBSCANのアルゴリズム | 165 |
| ▶ | DBSCANの実践 | 172 |
| ▶ | OPTICS Clustering | 173 |
| □ 4.1 | PyCaretで扱うことのできるその他のクラスタリング手法 | 174 |
| ▶ | Mean shift Clustering | 174 |
| ▶ | BIRCH Clustering | 174 |
| ▶ | Spectral Clustering | 175 |
| ▶ | Affinity Propagation | 176 |
| ▶ | K-Modes Clustering | 176 |
| □ 4.2 | PyCaretによるクラスタリング実践 | 177 |
| □ 4.3 | 本章のまとめ | 180 |

■ CHAPTER 07

データの前処理

| | | |
|-------|---------------|-----|
| □ 4.4 | PyCaretでの前処理 | 182 |
| □ 4.5 | データの準備 | 183 |
| ▶ | エンコーディング | 183 |
| ▶ | データ型の指定 | 189 |
| ▶ | 欠損値処理 | 190 |
| ▶ | 不均衡データの処理 | 193 |
| □ 4.6 | スケーリング | 195 |
| ▶ | 特徴量のスケーリング | 195 |
| ▶ | 目的変数の変換 | 197 |
| □ 4.7 | 特徴量エンジニアリング | 198 |
| ▶ | 多項式特徴量 | 198 |
| ▶ | Group特徴量 | 199 |
| ▶ | 数値特徴量のビン分割 | 199 |
| ▶ | 変数の結合 | 200 |
| □ 4.8 | 特徴量選択 | 202 |
| ▶ | 重要度に基づいた特徴量選択 | 202 |
| ▶ | 多重共線性 | 203 |
| ▶ | 主成分分析 | 204 |
| ▶ | 分散の低い特徴量の削除 | 205 |
| □ 4.9 | 本章のまとめ | 206 |

■ CHAPTER 08

モデルチューニング

| | | |
|-------|--------------------|-----|
| □ 5.0 | 交差検証(クロスバリデーション) | 208 |
| □ 5.1 | モデル学習 | 210 |
| ▶ | compare_models() | 210 |
| ▶ | create_model | 213 |
| □ 5.2 | ハイパーパラメータチューニング | 217 |
| ▶ | tune_model | 217 |
| ▶ | 探索範囲の指定 | 220 |
| ▶ | オリジナルモデルとの比較 | 220 |
| □ 5.3 | アンサンブル学習 | 222 |
| ▶ | ensemble_models | 222 |
| ▶ | ブースティング | 223 |
| ▶ | blend_models | 224 |
| ▶ | stack_models | 226 |
| □ 5.4 | optimize_threshold | 230 |
| □ 5.5 | calibrate_model | 231 |
| □ 5.6 | 本章のまとめ | 233 |

■ CHAPTER 09

評価指標

| | | |
|-------|------------------------|-----|
| □ 5.7 | 回帰における評価指標 | 236 |
| ▶ | RMSE(MSE) | 237 |
| ▶ | MAE | 238 |
| ▶ | RMSLE | 239 |
| ▶ | MAPE | 240 |
| ▶ | R^2 | 241 |
| ▶ | R^2 (決定係数)とそれ以外の評価指標 | 242 |
| ▶ | RMSEとMAEの比較 | 243 |
| ▶ | RMSEとRMSLEの比較 | 246 |
| ▶ | RMSEとMAPEの比較 | 247 |
| ▶ | 回帰における評価指標についてまとめ | 248 |

| | | |
|-----|--|-----|
| □58 | 分類における評価指標 | 249 |
| | ▶混同行列 | 249 |
| | ▶ACC(Accuracy) | 251 |
| | ▶Recall | 252 |
| | ▶Precision | 252 |
| | ▶RecallとPrecisionの関係 | 253 |
| | ▶F1 | 255 |
| | ▶不均衡データに対する対応 | 256 |
| | ▶Kappa | 257 |
| | ▶MCC(Matthews Correlation Coefficient) | 258 |
| | ▶偽陽性率と真陽性率 | 259 |
| | ▶ROC曲線 | 260 |
| | ▶AUC | 263 |
| | ▶分類モデルにおける評価指標についてまとめ | 265 |
| □59 | 評価指標を追加したい場合 | 266 |
| □60 | plot_model | 268 |
| | ▶回帰モデル | 268 |
| | ▶分類モデル | 275 |
| | ▶クラスタリングモデル | 280 |
| | ▶グラフの保存 | 282 |
| □61 | evaluate_model | 283 |
| □62 | dashboard | 284 |
| □63 | 本章のまとめ | 285 |

■ CHAPTER 10

デプロイ

| | | |
|-----|----------------|-----|
| □64 | デプロイに関する関数一覧 | 288 |
| □65 | finalize_model | 289 |
| □66 | predict_model | 291 |
| | ▶テストデータを使用 | 291 |
| | ▶未知のデータで予測 | 291 |
| | ▶クラスの予測確率 | 292 |
| | ▶確率閾値の設定 | 293 |
| □67 | check_drift | 294 |
| □68 | save_model | 296 |

| | | |
|-----|-----------------|-----|
| □69 | deploy_model | 297 |
| | ▶AWS | 297 |
| | ▶GCP | 298 |
| | ▶Azure | 299 |
| □70 | load_model | 300 |
| □71 | save_experiment | 302 |
| □72 | load_experiment | 303 |
| □73 | convert_model | 304 |
| □74 | create_api | 306 |
| □75 | create_docker | 308 |
| □76 | create_app | 311 |
| □77 | 本章のまとめ | 312 |

■ APPENDIX

Pythonの基礎や補足解説

| | | |
|-----|------------------|-----|
| □78 | Pythonの基礎 | 314 |
| | ▶算術演算子 | 314 |
| | ▶比較演算子 | 315 |
| | ▶データ型 | 315 |
| | ▶リスト | 316 |
| | ▶タプル | 317 |
| | ▶辞書 | 318 |
| | ▶条件文 | 319 |
| | ▶関数の定義 | 320 |
| | ▶ループ処理 | 320 |
| | ▶リスト内包表記 | 321 |
| □79 | 主要なライブラリの使用方法 | 322 |
| | ▶Numpyライブラリの使い方 | 322 |
| | ▶Pandasの使い方 | 326 |
| | ▶Pythonの可視化ライブラリ | 332 |
| □80 | 分析手法の補足 | 339 |
| | ▶最小二乗法 | 339 |
| | ▶シグモイド関数の導出 | 340 |
| | ▶ロジスティック回帰の損失関数 | 341 |
| | ▶主成分分析(PCA) | 341 |

| | | |
|-------|----------------------|-----|
| □ B 1 | PyCaretの解説補足 | 351 |
| | ▶ 異常検知モデル | 351 |
| | ▶ 「get_data」関数 | 353 |
| ● | 索引 | 355 |
| ● | 参考文献 | 359 |