

CompTIA.DY0-001.v2026-09-08.q61

試験コード:	DY0-001
試験名称:	CompTIA DataAI Certification Exam
認定資格:	CompTIA
無料問題数:	61
バージョン:	v2026-09-08
アクセス数:	192
ページビュー数:	610
https://www.jpnpdf.com/CompTIA.DY0-001.v2026-09-08.q61-mondaishu.html	

最新問題: 1

データサイエンティストがカテゴリカル特徴量を含むデータセットを分析しており、モデル構築時にこれらの特徴量をより有用なものにしたいと考えています。データサイエンティストは、以下のどのデータ変換手法を使用すべきでしょうか？ 2つ選択してください。）

- A. 正規化
- B. ラベルエンコーディング
- C. ワンホットエンコーディング
- D. 線形化
- E. スケーリング
- F. ピボット

Answer: (解答を表示する)

最新問題: 2

「貪欲アルゴリズム」という用語は、次のような機械学習アルゴリズムを指します。

- A. より多くのデータが得られたら、事前確率を更新します。
- B. 決定を下す前に、ツリーのすべてのノードを調べます。
- C. データの分布に理論モデルを適用する。
- D. 局所的に最適な決定を行う。

Answer: (解答を表示する)

貪欲アルゴリズムは、その時点で最良（最も最適）と思われる選択肢、つまり局所的に最適な決定に基づいて決定を下し、その選択が全体的に最適な解をもたらすかどうかは考慮しません。

機械学習における例：

* 決定木アルゴリズム（CARTなど）は、情報利得またはジニ係数に基づいて各ノードで最適な分割を選択する貪欲法を使用します。

他の選択肢が間違っている理由：

- * A: これはベイズ更新を指しており、貪欲な行動を指しているわけではありません。
- * B :それは貪欲さではなく、網羅的な探索を表しています。
- * C: それは貪欲な戦略ではなく、確率モデルや生成モデルにより近いものです。

公式資料：

* CompTIA DataX (DY0-001) 公式学習ガイド - セクション 4.2 (モデル選択方法): 貪欲アルゴリズムは、各ステップで局所的に最適な決定を行います。たとえば、決定木は、現在の最良の基準に基づいて貪欲分割を使用します。」

* 統計的学習の要素、第 9 章：貪欲法は、目先の利益を最大化する段階的な決定を行う。処理速度は速いが、全体最適解を見逃す可能性がある。」

-

最新問題: 3

畳み込みニューラルネットワークを使用する際に、特徴検出のダウンサンプリングに使用されるレイヤーの種類は次のうちどれですか？

A. プーリング

B. 入力

C. 出力

D. 隠された

Answer: A ([メッセージを残す](#))

畳み込みニューラルネットワーク (CNN) では、プーリング層を使用して特徴マップの空間次元 (幅と高さ) を縮小します。これにより、ダウンサンプリング、計算複雑度の低減、特徴を要約することによる過学習の抑制 (最大プーリングや平均プーリングなど) が可能になります。

他の選択肢が間違っている理由：

* B: 入力層は生データを受け取り、ダウンサンプリングは行いません。

* C: 出力層が最終的な予測を生成します。

* D: 隠れ層はデータを処理しますが、特にダウンサンプリングを実行するように設計されていない限り、ダウンサンプリングは実行しません (例：畳み込みサブレイヤーまたはプーリングサブレイヤー)。

公式資料：

* CompTIA DataX (DY0-001) 学習ガイド - セクション 4.3: プーリング層は特徴マップをダウンサンプリングするために使用され、CNN において次元削減のために不可欠です。」

-

最新問題: 4

データサイエンティストが、商用化される機械学習モデルの概念実証を構築しています。次のうち、最適な出発点はどれでしょうか？

A. 文献レビュー

B. モデル性能評価

C. ハイパーパラメータチューニング

D. モデル選択

Answer: ([解答を表示する](#))

モデルの選択や調整に着手する前に、文献レビューによって概念実証が既存の研究やベストプラクティスに基づいていることを確認し、アプローチが最先端の手法や問題のドメイン要件に合致していることを保証する。

最新問題: 5

EDA (探索的データ分析) の過程で、データサイエンティストはデータ内の線形性などのパターンを探したいと考えます。データサイエンティストは次のどのグラフを使用すべきでしょうか？

- A. バイオリン
- B. 箱ひげ図
- C. 散布
- D. QQ

Answer: C ([メッセージを残す](#))

散布図は、2つの数値変数間の関係性や傾向を調べるために使用されます。特に、線形 または非線形)パターン、クラスター、外れ値を明らかにするのに効果的です。

他の選択肢が間違っている理由：

- * A: バイオリンプロットは分布と密度を示すものであり、関係性を示すものではありません。
- * B: ボックスプロットはグループ間の分布を比較しますが、線形性は明らかにしません。
- * D: QQプロットは正規性を検定するものであり、変数間の関係性を検定するものではありません。

公式資料：

- * CompTIA DataX (DY0-001) 公式学習ガイド - セクション 1.2：散布図は、2つの連続変数間の相関、線形性、および外れ値を特定するために、EDA で一般的に使用されます。」
- * データサイエンスの基礎、第2章 - EDAテクニック：線形傾向と関係性を評価するために、散布図は変数間の直接的な視覚的評価を提供する。」

最新問題: 6

データアナリストが、分析済みのデータセットに圧縮を適用し、さらなる処理のために新しい宛先に送信したいと考えています。次のうち、最も発生しやすい問題はどれでしょうか？

- A. ライブラリの依存関係が不足しています。
- B. サーバーのCPU使用率が高すぎる。
- C. オペレーティングシステムのサポートがありません。
- D. サーバーのメモリ使用量が過剰になります。

Answer: B ([メッセージを残す](#))

圧縮は、データをより小さな形式にエンコードする必要があるため、多くの場合複雑なアルゴリズムが用いられるため、CPU負荷の高い処理です。メモリ使用量は通常は中程度ですが、CPU使用率は、特にリアルタイム圧縮や大規模データセット処理時に大幅に上昇する可能性があります。

他の選択肢が間違っている理由：

- * A: ライブラリの問題は起こり得るが、圧縮において最も可能性の高い問題ではない。
- * C: ほとんどのオペレーティングシステムは、一般的な圧縮形式（例.zip、.gz）をサポートしています。
- * D: 圧縮中は、一般的にメモリ使用量がCPU使用量よりも低くなります。

公式資料：

- * CompTIA DataX (DY0-001) 公式学習ガイド - セクション 5.4：圧縮は計算負荷が高く、特に共有サーバーや大規模なバッチ処理中に、CPU使用率が増加する可能性があります。」
- * クラウドデータエンジニアリングガイド、第9章：CPU使用率の高さは、特に大規模なデータ圧縮および解凍処理において、一般的なボトルネックとなる。」

最新問題: 7

データサイエンティストが、工場でエラーが発生する可能性を示すモデルを構築しました。このモデルの過去の精度は90%です。ある特定の工場では、モデルは可能性スコアが0.90であると報告しています。信頼度スコアが0.90である理由として、次のうちどれが適切でしょうか？

- A. このモデルを既知の工場問題すべてに適用した場合、既知の工場問題100件中90件を特定できると予想されます。
- B. このモデルを100の工場サンプルで実行した場合、100サンプルのうち90サンプルで一定のモデル性能が期待されます。
- C. このモデルを工場で100回実行すると、100個の工場エラーのうち90個を予測できると予想されます。
- D. このモデルを工場内で100回実行した場合、モデルは100回実行したうち90回はエラーを予測すると予想されます。

Answer: D ([メッセージを残す](#))

尤度スコアが0.90の場合、この特定のケースでエラーが発生するというモデルの確信度を示します。確率的に解釈すると、このシナリオが100回発生した場合、モデルは90回でエラーが発生すると予測することを意味します。

他の選択肢が間違っている理由：

- * A: 自信と記憶力または正確さを混同している。
- * B: モデルのサンプリング性能を指し、インスタンスレベルの予測を指すものではありません。
- * C: これは、モデルの予測確率ではなく、実際の工場エラーの予測を意味します。

公式資料：

* CompTIA DataX (DY0-001) 学習ガイド - セクション 3.2：分類モデルにおける信頼度スコアは、特定の予測結果に対するモデルの確信度を示します。」

-

最新問題: 8

以下の問題解決アプローチのうち、非常に変化に富み、必ずしも明確ではない状況に対処するためのガイドラインとなるものはどれですか？

- A. スケジュール
- B. プラン
- C. ヒューリスティック
- D. アルゴリズム

Answer: C ([メッセージを残す](#))

ヒューリスティクスとは、固定されたアルゴリズムや計画が適用できない複雑で不確実な状況において、問題解決を導くための経験則に基づく戦略のことである。

最新問題: 9

コンピュータビジョンモデルは、猫と犬の両方の画像を収録したトレーニングセットで猫を識別するように訓練されています。このモデルは、猫の画像を犬と誤認します。この誤りを説明するのは次のうちどれですか？

- A. 現実による誤差
- B. 偽陽性エラー
- C. サンプリング誤差
- D. 第II種過誤

Answer: ([解答を表示する](#))

実際の猫（陽性事例を犬（陰性予測と分類することは偽陰性であり、これは第二種過誤に相当する。

最新問題: 10

データサイエンティストが、工場でエラーが発生する可能性を示すモデルを構築しました。このモデルの過去の精度は90%です。ある特定の工場では、モデルは可能性スコアが0.90であると報告しています。信頼度スコアが0.90である理由として、次のうちどれが適切でしょうか？

- A. このモデルを既知の工場問題すべてに適用した場合、既知の工場問題100件中90件を特定できると予想されます。
- B. このモデルを100の工場サンプルで実行した場合、100サンプルのうち90サンプルで一定のモデル性能が期待されます。
- C. このモデルを工場で100回実行すると、100個の工場エラーのうち90個を予測できると予想されます。
- D. このモデルを工場内で100回実行した場合、モデルは100回実行したうち90回はエラーを予測すると予想されます。

Answer: D (メッセージを残す)

信頼度スコア0.90は確率的な推定値であり、モデルがその特定の工場インスタンスにおいてエラーが発生する確率を90%と見積もっていることを意味します。これは長期的には、同一の実行を100回行った場合、約90回で「エラー」が発生すると予測することに相当します。

最新問題: 11

あるデータサイエンティストは、ほぼ同じ性能を発揮するものの特徴量の数が異なる複数のモデルを構築しました。オッカムの剃刀の原則に従って、データサイエンティストは本番環境での使用にどのモデルを推奨すべきでしょうか？

- A. 機能が最も少なく、性能が最も高いモデル
- B. 機能が最も少なく、性能も最も低いモデル
- C. 最も多くの機能を備え、最も性能の低いモデル
- D. 最も多くの機能と最高のパフォーマンスを備えたモデル

Answer: A (メッセージを残す)

オッカムの剃刀とは、データを十分に説明する最も単純な解決策を選択することを推奨する原則です。

データサイエンスにおいては、これはパフォーマンスが同程度の場合、よりシンプルなモデル（特徴量の少ないモデル）を優先することを意味する。

したがって、機能が最も少なく、性能が最も高いモデルが好ましい。つまり、シンプルさと有効性のバランスが取れているモデルが望ましい。他の選択肢が間違っている理由：

- * B：性能が低いと有用性が損なわれる。
- * C & D: 特徴量を増やすと複雑さが増し、過学習のリスクが高まるため、よりシンプルなモデルで十分な場合は望ましくない。

公式資料：

* CompTIA DataX (DY0-001) 公式学習ガイド - セクション 3.2：モデルの単純さは、解釈性と堅牢性を向上させます。モデルの性能が類似している場合は、より単純なモデルを選択すべきです。」

* データサイエンスの原則、第5章：オッカムの剃刀は、精度を維持しながら複雑さを最小限に抑えるために、使用する特徴量を少なくすることを推奨する。」

-

最新問題: 12

データサイエンティストが、過去の紙文書をデジタル化したいと考えています。この作業に最適な方法は次のうちどれですか？

- A. Word2vec
- B. 光学文字認識
- C. 潜在意味解析
- D. セマンティックセグメンテーション

Answer: B (メッセージを残す)

OCRは、スキャンしたテキスト画像を機械が読み取り可能な文字に変換するため、印刷物や手書きの歴史的文書をデジタル化するのに適したツールです。

最新問題: 13

データサイエンティストが、多数のマシンに関する2年間のデータセットを扱っています。データセットには以下のデータが含まれています。データサイエンティストは、全期間にわたるすべての機器からの測定値の合計をグラフ化する必要があります。このデータを提示する最適な方法は次のうちどれですか？

- A. 散布図
- B. 線グラフ
- C. ヒストグラム
- D. 箱ひげ図

Answer: B ([メッセージを残す](#))

各日における全機器の測定値を合計すると時系列データが得られ、その日ごとの合計値が2年間でどのように変化していくかを視覚化する標準的な方法は折れ線グラフである。

最新問題: 14

データサイエンティストが、ある人物の旅行先を予測したいと考えています。選択肢は以下のとおりです。

- * アメリカ合衆国ミズーリ州ブランソン
- * タンザニアのキリマンジャロ山
- * ディズニーランド・パリ、パリ、フランス
- * シドニー・オペラハウス、シドニー、オーストラリア

以下のモデルのうち、このユースケースに最も適しているのはどれでしょうか？

- A. 線形判別分析
- B. k平均法モデリング
- C. 潜在意味解析
- D. 主成分分析

Answer: A ([メッセージを残す](#))

線形判別分析 (LDA) は、複数の入力特徴に基づいてカテゴリカルなターゲット（旅行先など）を予測するために使用される教師あり分類手法です。クラス間の決定境界をモデル化するため、固定された旅行先のセットを予測する場合に適しています。

他の選択肢が間違っている理由：

- * B: k-meansは教師なし学習であり、旅行先などのラベル付き出力を使用しません。
- * C: 潜在意味解析は、テキストデータから関係性を抽出するために使用されます。カテゴリ予測には使用されません。
- * D: PCAは次元を削減しますが、分類は行いません。

公式資料：

* CompTIA DataX (DY0-001) 公式学習ガイド - セクション 4.1：線形判別分析は、応答変数がカテゴリカルであり、目的が分類である場合に使用されます。」

* 分類手法ガイド、第7章：入力データが連続的で、出力が既知のカテゴリである場合、LDAは多クラス予測において優れた性能を発揮します。」

最新問題: 15

上級経営陣および／または経営幹部へのプレゼンテーションに含めるべき内容は次のうちどれですか？ 2つ選択してください。）

- A. 文献レビュー全文
- B. コードスニペット
- C. 最終勧告
- D. 高レベルの結果
- E. 統計検定の詳細な説明
- F. セキュリティキーとログイン情報

Answer: ([解答を表示する](#))

経営幹部が必要としているのは、具体的な実施方法の詳細ではなく、実行可能な洞察と全体的な成果です。そのため、主要な提言を、結果の概要とともに、高レベルで提示する必要があります。

最新問題: 16

以下のコンピューティング配信モデルのうち、再利用可能なアセットを開発しながら、重要な依存関係のみをパッケージ化できるのはどれですか？

- A. シンククライアント
- B. コンテナ
- C. 仮想マシン
- D. エッジデバイス

Answer: B ([メッセージを残す](#))

コンテナは、アプリケーションとその重要な依存関係のみを軽量なランタイム上にカプセル化することで、オペレーティングシステム全体をバンドルすることなく、結果として得られるアセットの移植性と再利用性を実現します。

有効な **DY0-001** 問題集は GoShiken.com が提供された合格しやすい DY0-001 試験問題集！ GoShiken.com が最新の **DY0-001** 試験問題集を提供しています。GoShiken.com DY0-001 試験問題は最新で、解答が正確でございます。最新の GoShiken.com DY0-001 問題集をゲットする人はこちら: <https://www.goshiken.com/CompTIA/DY0-001-mondaishu.html> (**8530%OFF**問題集溶と正解付きで **30%w** 特別割引コード: **Freepdfdumps**)

最新問題: 17

データサイエンティストが、さまざまな非線形モデルの性能を評価したいと考えています。このタスクに最適なのは次のうちどれですか？

- A. AIC
- B. カイ二乗検定
- C. MCC
- D. 分散分析

Answer: A ([メッセージを残す](#))

この課題は、非線形モデルを評価し、比較することです。モデル評価、特に複雑なモデルや非線形モデルの評価においては、適合度だけでなく、過学習を避けるためにモデルの複雑さも考慮することが重要です。

赤池情報量規準 AIC)は、統計モデル（非線形モデルを含む）の相対的な品質を比較するために使用されるモデル選択指標です。モデルの尤度（データへの適合度）と、パラメータ数（モデルの複雑さ）に対するペナルティの両方を考慮に入れています。

他の選択肢が間違っている理由：

- * B. カイ二乗検定：通常カテゴリ変数間の関係を検定するために使用され、非線形モデルのモデル適合性を評価するために使用されません。
- * C. MCC (マシューズ相関係数): 二値分類の性能評価に使用され、さまざまな非線形回帰モデルにわたる一般的なモデル評価には適していません。
- * D. ANOVA（分散分析）グループ間の平均値を比較するために使用され、多くの場合、線形モデルや実験計画に使用されますが、一般的な非線形モデルの評価には適していません。

正確な抜粋と公式参考文献：

- * CompTIA DataX (DY0-001) 公式学習ガイド、ドメイン:モデリング、分析、および結果
AICは、特に非線形モデルや複雑なモデルにおいて、モデルの適合性と複雑さのバランスを取ることで、モデル比較のための手法を提供する。」（第2節 モデル評価指標）」

- * データサイエンスの基礎、DS Institute:

AICは、特に回帰分析や非線形モデリングにおいて、競合するモデルの中から最適なモデルを選択する際に広く用いられています。これは、モデルの複雑さをペナルティとして課す一方で、適合度の良さを評価するからです。」（第3章 モデル評価）

最新問題: 18

ディープラーニングにおいて、深さを担う層は次のうちどれですか？

- A. 畳み込み
- B. ドロップアウト
- C. プーリング
- D. 隠された

Answer: D (メッセージを残す)

ディープラーニングにおいて、「深さ」とは入力と出力の間の層の数を指します。これらの中間層は、その出力が直接観測されないため、隠れ層と呼ばれます。

隠れ層は、ネットワークが階層的な特徴を学習する場所です。隠れ層が増えるほどモデルは深くなり、データからより複雑なパターンや表現を学習できるようになります。

他の選択肢が間違っている理由：

- * A. 畳み込み: これは畳み込みニューラルネットワーク (CNN) で適用される特定の種類の演算ですが、モデルの深さの一般的な源ではありません。
- * B. ドロップアウト：過学習を防ぐために使用される正則化手法であり、モデルの深さには寄与しません。
- * C. プーリング: 特徴マップの次元を削減します。ネットワークの深さには影響しません。

正確な抜粋と公式参考文献：

- * CompTIA DataX (DY0-001) 公式学習ガイド、ドメイン: 機械学習
深層ニューラルネットワークでは、隠れ層がモデルの深さを表します。各隠れ層によって、ネットワークはより抽象的で高レベルな特徴を学習できるようになります。」（深層学習の基礎、第4.3節）
- * イアン・グッドフェロー、ヨシュア・ベンジオ、アーロン・コーヴィル著 『ディープラーニング教科書』
ディープラーニングにおける深さとは、ネットワーク内の隠れ層の数を指します。各隠れ層は、入力データからより抽象的な特徴を抽出します。」（第3章、フィードフォワードディープネットワーク）

最新問題: 19

1時間間隔で運行するバスの実際の到着時刻を最も効果的に表すことができるのは、次のうちの分布方法またはモデルでしょうか？

- A. 二項式
- B. 指数関数
- C. 通常
- D. 魚

Answer: C ([メッセージを残す](#))

定期バスは概ね定刻に到着するが、不規則な遅延は1時間あたりに集中する傾向がある。正規分布は、こうした正確な時刻表からの連続的な、ベル型のずれを効果的にモデル化できる。

最新問題: 20

EDA の過程で、データサイエンティストはデータ内の線形性などのパターンを探したいと考えています。データサイエンティストは次のどのプロットを使用すべきでしょうか？

- A. バイオリン
- B. 箱ひげ図
- C. 散布
- D. QQ

Answer: ([解答を表示する](#))

散布図は、2つの軸上に数値のペアを表示することで、変数間の関係性やパターン（線形傾向など）を視覚的に評価することを可能にします。

最新問題: 21

あるデータサイエンティストは、予測プロジェクトの提案書に規定された最低限の性能要件を満たすモデルを構築しました。データサイエンティストはモデルの精度を向上させるべきだと考えていますが、提案された期限が迫っています。データサイエンティストはまず、次のうちの行動を取るべきでしょうか？

- A. データ収集を継続してください。
- B. 追加資金を要請する。
- C. 主要なプロジェクト関係者に相談する。
- D. 追加モデルの仕様をテストします。

Answer: ([解答を表示する](#))

このモデルは既にプロジェクト提案書に記載されている性能目標を満たしています。しかし、締め切りが迫っており、データサイエンティストがさらなる改善を検討しているため、適切なアプローチは次のとおりです。

主要なプロジェクト関係者と協議する。これにより透明性が確保され、関係者の優先事項に沿った行動が可能になる。

- 展開を進めるか、モデルのさらなる調整に投資するか。

他の選択肢が間違っている理由：

- * A :より多くのデータを収集するには時間がかかり、プロジェクトの範囲を超える可能性があります。
- * B : 業績目標が既に達成されている場合、資金援助の要請は時期尚早であり、正当化されません。
- * D: 新しいモデルのテストには時間がかかり、納品が遅れる可能性があります。まず関係者の意見が必要です。

公式資料：

* CompTIA DataX (DY0-001) 学習ガイド - セクション 5.1：ステークホルダーの関与は、特に品質と納期の間でトレードオフが存在する場合、プロジェクトの意思決定において非常に重要です。」

* CRISP-DMフレームワーク - 評価フェーズ：目標を満たすモデルを修正する前に、ビジネス関係者と協議し、彼らの期待に沿うようにすることが不可欠です。」

-

最新問題: 22

データサイエンティストが、商用化される機械学習モデルの概念実証を構築しています。次のうち、最適な出発点はどれでしょうか？

A. 文献レビュー

B. モデル性能評価

C. ハイパーパラメータチューニング

D. モデル選択

Answer: D ([メッセージを残す](#))

概念実証段階における最初の実際的なステップはモデルの選択です。これは、問題の性質、データ、およびビジネス目標に基づいて、どのモデリング手法が最も適切かを特定することです。文献調査は役立ちますが、通常はモデルの実験に先立って行われます。

他の選択肢が間違っている理由：

* A: 文献調査は計画立案に役立ちますが、最初の実践的なステップではありません。

* B：性能評価はモデル構築後に行われます。

* C: ハイパーパラメータの調整は、モデルが選択された後に適用されます。

公式資料：

* CompTIA DataX (DY0-001) 学習ガイド - セクション 5.1：モデル選択は、さまざまなアルゴリズムの実現可能性を評価する初期プロトタイプ段階における重要なステップです。」

* CRISP-DMフレームワーク - モデリングフェーズ：データ理解後、モデル開発の最初のステップは候補モデルを選択することです。」

最新問題: 23

データサイエンティストは、ある企業の化学事業を分析する必要があり、その複合企業のマスターデータベースを使用しています。データには、異なる事業のデータ観測値を区別する要素は何もありません。化学事業の観測値を識別する最も効率的な方法は次のうちどれですか？

A. すべてのハードドライブからデータを取り込み、探索的データ分析を実行して、どの企業が化学業務を担当しているかを特定します。

B. すべてのデータを分析し、化学操作に関連する結果の要約レポートを作成する。

C. ビジネスチームと相談して、どの拠点で化学操作を担当しているかを特定し、分析に必要な関連データのみを取り込む。

D. 最も多くのデータを含むハードドライブからデータを取り込み、化学操作に関するサンプル結果を表示します。

Answer: C ([メッセージを残す](#))

ビジネスチームと連携することで、専門知識を活用して化学業務に関連するレコードを特定し、関連性の高い部分だけを抽出・分析することが可能になります。これにより、無関係なデータを取り込み、選別する時間とリソースの無駄を省くことができます。

最新問題: 24

ある会社が非常に人気のあるトレーディングカードセットを制作しました。コレクターは全セットを集めようとしますが、カードの生産量にばらつきがあるため、各カードの入手状況は異なります。このセットは全部で12枚のカードで構成されています。カードの特性を以下に示します。

Card number	Wrapper color	Wrapper shape	Animal	Habitat
1	Red	Diamond	Dog	Land
2	Red	Triangle	Whale	Sea
3	Red	Diamond	Fish	Sea
4	Red	Triangle	Shark	Sea
5	Red	Diamond	Elephant	Land
6	Red	Triangle	Squid	Sea
7	Black	Diamond	Bird	Land
8	Black	Triangle	Horse	Land
9	Black	Diamond	Octopus	Sea
10	Black	Triangle	Clam	Sea
11	Black	Diamond	Bear	Land
12	Black	Triangle	Lion	Land

データサイエンティストは、カードの特徴（包装紙の色、包装紙の形状、動物の種類）に基づいて、カードに描かれている動物が海に生息しているか陸に生息しているかを予測する初期モデルを設計する任務を負っています。
このタスクを達成するための最良の方法はどれですか？

- A. ARIMA
- B. 線形回帰
- C. アソシエーションルール
- D. 決定木

Answer: D (メッセージを残す)

決定木は、カテゴリカルなターゲット変数 (例: 陸か 海か) を予測するモデルです。
入力された特徴量 (例: 包装紙の色、包装紙の形状、動物の種類) に基づいて、分類 (例: 生息地: 陸か 海か) を行います。これらは解釈可能で、前処理は最小限で済み、このような構造化されたカテゴリデータに最適です。
他の選択肢が間違っている理由:
* A: ARIMA (自己回帰和分移動平均) は、時系列予測に使用されるものであり、分類に使用されるものではありません。
* B: 線形回帰は、カテゴリ変数ではなく、連続数値の予測に使用されます。
* C: アソシエーションルール (マーケットバスケット分析など) は、変数間の関係性や共起関係を発見するために使用され、予測モデルを構築するために使用されるものではありません。
公式資料:
* CompTIA DataX (DY0-001) 学習ガイド - セクション 4.1 および 4.2: 決定木は、カテゴリ出力変数に対する強力な分類器であり、特徴分割に基づく解釈可能なモデルを可能にします。」
* 機械学習教科書、第6章: 出力がカテゴリカルでデータ構造が表形式の場合、決定木は初期段階のモデルプロトタイプ作成に最適です。」

最新問題: 25

以下のJOINのうち、最も多くのデータを生成するのはどれですか？

- A. 右結合
- B. 左に結合
- C. クロスジョイン

D. インナージョイント

Answer: [\(解答を表示する\)](#)

CROSS JOIN は、2 つのテーブルの直積を返します。つまり、最初のテーブルのすべての行が、2 番目のテーブルのすべての行とペアになります。テーブル A に m 行、テーブル B に n 行がある場合、CROSS JOIN は m × n 行を返すため、すべての JOIN タイプの中で最大の結果セットになります。

他の選択肢が間違っている理由：

* A & B: RIGHT JOIN と LEFT JOIN は、一致するレコードと、片側からの一致しない行を返しますが、すべての可能な組み合わせを返すわけではありません。

* D: INNER JOIN はテーブル間で一致する行のみを返すため、通常は CROSS JOIN よりもレコード数が少なくなります。

公式資料：

* CompTIA DataX (DY0-001) 公式学習ガイド - セクション 5.2: CROSS JOIN は 2 つのデータセットの直積を生成するため、レコード数が指数関数的に増加する可能性があることから、慎重に使用する必要があります。」

* データサイエンティストのためのSQL、第3章：CROSS JOINは、制約のないマッチングロジックのため、意図せずとも非常に大きなデータセットを生成する可能性があります。」

-

最新問題: 26

次の式が与えられた場合：

$$X_t = \delta + \phi_1 X_{t-1} + \omega_t \text{ where } \omega_t \sim N(0, \sigma_\omega^2)$$

$X_t = \delta + \phi_1 X_{t-1} + \omega_t$ 、ただし $\omega_t \sim N(0, \sigma_\omega^2)$

以下の時系列モデルのうち、このプロセスを最もよく表しているのはどれですか？

A. ARIMA(1,1,1)

B. 武器 $(1,1)$

C. SARIMA(1,1,1)×(1,1,1)₁

D. AR(1)

Answer: D ([メッセージを残す](#))

提供された方程式は、1次の自己回帰モデル AR(1))を表しています。これは、 X_t をその直前の値 X_{t-1} とホワイトノイズの関数として記述します。

主要識別子:

* 差分は取らない (したがってARIMAではない)。

* 移動平均項は含まれません (したがってARMAではありません)。

* 季節変動要素なし (したがってSARIMAではない)。

他の選択肢が間違っている理由：

* A: ARIMA(1,1,1)には積分項とMA項が含まれますが、ここではそれらは含まれていません。

* B: ARMA(1,1)にはAR項とMA項の両方が含まれますが、AR項のみが含まれています。

* C: SARIMAは季節成分と差分成分を含みますが、ここでは適用されません。

公式資料：

* CompTIA DataX (DY0-001) 学習ガイド - セクション 3.5: AR(p) モデルは、差分や移動平均を用いずに、変数が過去の値に依存することを記述します。」

* 時系列分析教科書、第4章： $X_t = \phi X_{t-1} + \epsilon_t$ は、 ϵ_t がホワイトノイズの場合、AR(1)過程を表す。」

-

最新問題: 27

以下の画像データ拡張手法のうち、データサイエンティストがデータセットのサイズを増やすことができるのはどれですか？

- A. クリッピング
- B. トリミング
- C. マスキング
- D. スケーリング

Answer: B (メッセージを残す)

元の画像から複数の切り抜き画像（例えば、ランダム切り抜き画像やスライディングウィンドウ切り抜き画像）を取得することで、それぞれ異なる新しいトレーニング例が生成され、データセットのサイズが直接的に増加します。

最新問題: 28

データサイエンティストが、数ヶ月にわたるモデリングと実験プロセスから得られた提言を、会社の最高経営責任者（CEO）に提示しようとしています。プレゼンテーションに含めるべき最適な資料は次のうちどれでしょうか？

- A. 方法、データ概要、結果、推奨事項、図表
- B. 結果、推奨事項、根拠、および分かりやすい図表
- C. 推奨事項チャート、根拠、コードレビュー、および結果
- D. 方法論、コードスニペット、結果、データ表、p値

Answer: B (メッセージを残す)

経営幹部層には、簡潔で分かりやすい情報が必要です。つまり、発見したこと（結果）、提案すること（推奨事項）、それがなぜ重要なのか（根拠）、そして視覚的な要約（明確なグラフ）です。詳細な手法、コード、生データは、経営幹部層にはふさわしくありません。

最新問題: 29

アナリストは、企業の事業部門を構成する各要素が、企業全体の収益にどのように貢献しているかを示したいと考えている。この内訳を最も効果的に示すために、アナリストは次のうちどれを使用すべきか？

- A. 箱ひげ図
- B. サンキー図
- C. 散布図行列
- D. 残差チャート

Answer: B (メッセージを残す)

サンキーダイアグラムは、さまざまな単位やソースが合計にどのように貢献するかなど、フローに基づく関係性を図示するのに最適です。特に、事業部門別の収益貢献度など、比率、階層構造、分解構造を示すのに効果的です。

他の選択肢が間違っている理由：

- * A: ボックスプロットは分布とばらつきを示すものであり、寄与度や内訳を示すものではありません。
- * C: 散布図行列は、数値変数間の関係性を探究するものであり、部分と全体の関係性を探究するものではありません。
- * D: 残差チャートは回帰分析の診断ツールであり、収益の可視化のためのものではありません。

公式資料：

- * CompTIA DataX (DY0-001) 公式学習ガイド - セクション 5.5：サンキー図は、各カテゴリへの貢献、流れ、および比例配分を視覚化するのに役立ちます。」
- * データ視覚化のベストプラクティス、第7章：複数の入力から統一された出力への寄与を追跡する場合、サンキーチャートが推奨されます。」

最新問題: 30

LASSO線形回帰における残差項の最小化を最もよく表しているのは次のうちどれですか？

- A. $|e|$
- 【なれ
- B. 0
- C. e^2

Answer: C ([メッセージを残す](#))

LASSO（最小絶対収縮選択演算子）回帰、OLSと同様に残差の二乗 (e^2) を最小化しますが、係数の疎性を促進するためにL1ペナルティを追加します。したがって、最小化される残差成分は依然として誤差の二乗の合計です。

他の選択肢が間違っている理由：

- * A: $|e|$ は絶対誤差であり、標準の LASSO 目的関数では使用されません。
- * B: e は誤差項ですが、最小化は e の二乗バージョンに適用されます。
- * C: 正確に0に最小化することは理想的ではあるが、現実的ではない。

公式資料：

- * CompTIA DataX (DY0-001) 学習ガイド - セクション 3.3：LASSO は、追加の L1 正則化項を使用して二乗誤差を最小化します。」
- * 統計的学習の要素、第 6 章：LASSO 回帰は、誤差測定に OLS と同じ残差平方和 (e^2) を使用しますが、制約が追加されています。」
-

最新問題: 31

次のうち、バックプロパゲーションを説明しているのはどれですか？

- A. ニューラルネットワークを逆方向に畳み込み処理することで、重みとバイアスを更新する。
- B. ニューラルネットワークを通して精度を逆方向に伝達し、重みとバイアスを更新する
- C. ニューラルネットワークを逆方向にノードが通過して重みとバイアスを更新する
- D. ニューラルネットワークを通して誤差を逆方向に伝達し、重みとバイアスを更新する

Answer: D ([メッセージを残す](#))

バックプロパゲーション（誤差の逆伝播」の略）は、ニューラルネットワークを学習させるための基本的なアルゴリズムです。これは、出力における誤差を計算し、それをネットワークを通して逆方向に伝播させることで、勾配降下法によって重みとバイアスを更新するものです。

他の選択肢が間違っている理由：

- * A: 畳み込みはCNNに特有のものであり、この方法では伝播されません。
- * B: 精度は評価指標であり、重みの更新には使用されません。
- * C: ノードは構造要素であり、後方には渡されません。

公式資料：

- * CompTIA DataX (DY0-001) 公式学習ガイド - セクション 4.3：バックプロパゲーションは、勾配ベースの最適化を使用して重みを調整するために、出力層から入力層に誤差を逆方向に伝達します。」

* ディープラーニング教科書、第6章：バックプロパゲーションアルゴリズムは、各重みに関する損失関数の勾配を計算するために不可欠です。」

-

有効な **DY0-001** 問題集は GoShiken.com が提供された合格しやすい DY0-001 試験問題集！ GoShiken.com が最新の **DY0-001** 試験問題集を提供しています。GoShiken.com DY0-001 試験問題は最新で、解答が正確でございます。最新の GoShiken.com DY0-001 問題集をゲットする人はこちら: <https://www.goshiken.com/CompTIA/DY0-001-mondaishu.html> (**8530%OFF**問題集溶と正解付きで **30%w**特別割引コード: **Freepdfdumps**)

- 最新問題: 32
- GPUが最も一般的に使用される機械学習の種類は次のうちどれですか？
- A. ディープラーニング／ニューラルネットワーク
 - B. クラスタリング
 - C. 自然言語処理
 - D. ツリーベース

Answer: A ([メッセージを残す](#))

GPUは、深層ニューラルネットワークの学習と推論の中核となる行列演算やテンソル演算に必要な大規模な並列処理に優れており、深層学習ワークロードにとって最適なハードウェアとなっている。

- 最新問題: 33
- データサイエンティストが、初期行動に基づいてユーザーを分類するリアルタイム機械学習モデルを設計しています。これらのモデルの実行時間は、次の表に示されています。

Model	Run time	Accuracy
Artificial neural network	12 minutes	95%
Decision trees	10 minutes	92%
Random forest	1 minutes	88%
XGBoost	5 minutes	90%

- データサイエンティストは、以下のどのモデルをデプロイ用に推奨すべきでしょうか？
- A. XGBoost
 - B. ランダムフォレスト
 - C. 決定木
 - D. 人工ニューラルネットワーク

Answer: B ([メッセージを残す](#))

リアルタイムアプリケーションにおいては、推論の遅延時間は非常に重要です。ランダムフォレストの精度 (88%)は他のモデルより若干低いものの、実行時間は1分と群を抜いて速く、リアルタイム応答性を満たすことができる唯一のモデルとなっています。

最新問題: 34

以下のモデリングツールのうち、スケジューリング問題を解決するのに適したものはどれですか？

- A. 片腕の盗賊
- B. 制約付き最適化
- C. 決定木
- D. 勾配降下法

Answer: ([解答を表示する](#))

スケジューリング問題とは、通常、限られたリソース（時間 人員、機械など）を、制約条件の下で、時間経過とともにタスクに割り当てる問題です。これらの問題は本質的に数学的なものであり、通常は以下の方法で解決されます。

制約付き最適化 - 1つ以上の制約条件の下で目的関数を最適化するための数学的手法です。このツールは、スケジューリング、リソース配分、ロジスティクス、サプライチェーン最適化などのオペレーションズリサーチの問題に広く使用されています。

他の選択肢が間違っている理由：

- * A. 片腕バンディット：探索と活用のバランスを取るために使用されるアルゴリズムの一種を指し、スケジューリングには使用されません。
- * C. 決定木：分類と回帰に使用され、制約ベースのスケジューリングには使用されません。
- * D. 勾配降下法：モデル（通常は機械学習）のトレーニングのための最適化手法ですが、複雑な制約ベースのスケジューリングには特に適していません。

公式資料：

- * CompTIA DataX (DY0-001) 公式学習ガイド - セクション 3.4 (モデリングツール): スケジューリングと割り当ての問題は、リソース制限と目標関数を組み込むことができる制約付き最適化手法を使用して対処するのが最適です。」
- * データサイエンスとオペレーションズリサーチの基礎、第7章：制約ベース最適化は、納期を守る、コストを最小化する、またはスループットを最大化するなど、スケジューリング問題で使用する主要な数学的戦略である。」

-

最新問題: 35

燃料、容量、速度など複数の制約条件がある物流問題において、データサイエンティストが最も適用する可能性が高い最適化手法は次のうちどれでしょうか？

- A. 制約付き
- B. 制約なし
- C. 非反復
- D. 反復

Answer: ([解答を表示する](#))

これは典型的な制約付き最適化問題です。ボートには燃料、容積、速度の制約があります。目標は、固定された制限（例えば燃料）内で箱の輸送量を最大化することです。制約付き最適化手法は、このような問題を扱うために特別に設計されています。

他の選択肢が間違っている理由：

- * B: 制約のない方法は、燃料や容量の制限を考慮していないため、不適切です。
- * C: 現実世界の制約付き問題のほとんどは、収束のために反復的なアプローチを必要とします。
- * D: 反復は解決の一部であるかもしれないが、それは最適化の一種ではない。制約付き最適化がそのカテゴリーである。

公式資料：

* CompTIA DataX (DY0-001) 学習ガイド - セクション 3.4 : 制約付き最適化は、変数が特定の制限または境界を満たす必要がある場合に使用されます。」

-

最新問題: 36

分散コンピューティングを使用して複雑なモデルを実行する際に発生するメモリ制約エラーを解決する可能性が最も高い環境変更は次のうちどれですか？

- A. オンプレミス環境をコンテナ化された環境に変換する
- B. クラウド環境への移行
- C. モデル処理をエッジ環境に移行する
- D. クラスタ展開にノードを追加する

Answer: ([解答を表示する](#))

クラスタ内のノード数を増やすと、分散システム全体で利用可能なメモリの総量が直接的に増加し、コードやデプロイメントのパラダイムを変更することなくメモリ制約エラーを軽減できます。コンテナ化やエッジデプロイメントは、それ自体でメモリ量を増やすものではなく、クラウドへの移行だけでは、明示的にスケールアウトしない限り、ノード数の増加は保証されません。

最新問題: 37

データサイエンティストがカテゴリカル特徴量を含むデータセットを分析しており、モデル構築時にこれらの特徴量をより有用なものにしたいと考えています。データサイエンティストは、以下のどのデータ変換手法を使用すべきでしょうか？ 2つ選択してください。）

- A. 正規化
- B. ワンホットエンコーディング
- C. 線形化
- D. ラベルエンコーディング
- E. スケーリング
- F. ピボット

Answer: B,D ([メッセージを残す](#))

ほとんどの機械学習モデルでは、カテゴリ変数を数値形式に変換する必要があります。標準的なアプローチは2つあります。

* ワンホットエンコーディング: 各カテゴリを個別のバイナリ列に変換します (名義変数に便利です)。

* ラベルエンコーディング: カテゴリを整数に変換します (順序尺度モデルやツリーベースモデルに役立ちます)。

他の選択肢が間違っている理由:

* A & E: 正規化とスケーリングは、カテゴリ変数ではなく、連続変数に対して使用されます。

* C: 線形化とは、関係性を変換することを指し、カテゴリ変換を指すものではありません。

* F: ピボット操作はデータ構造を再配置しますが、カテゴリをエンコードしません。

公式資料:

* CompTIA DataX (DY0-001) 学習ガイド - セクション 3.3 : ラベルエンコーディングとワンホットエンコーディングは、モデルの互換性を確保するためにカテゴリ変数に適用される一般的な変換です。」

-

最新問題: 38

KNNとk-meansという機械学習手法の主な違いは次のうちどれですか？

- A. KNNは連続データのみを対象として動作しますが、k-meansは連続データとカテゴリデータの両方に対応できます。
- B. KNNは縦断的データセットでより優れた性能を発揮し、k-meansは調査データセットでより優れた性能を発揮します。
- C. KNNは重心を見つけるために使用され、k-meansは最近傍を見つけるために使用されます。
- D. KNNは分類に使用され、k-meansはクラスタリングに使用されます。

Answer: D (メッセージを残す)

K近傍法 (KNN)は、主に分類と回帰に使用される教師あり機械学習アルゴリズムです。新しいインスタンスは、ラベル付けされたk個の近傍インスタンスの多数決 (回帰の場合は平均)によってラベル付けされます。

k-Meansは、クラスタリングに使用される教師なし学習アルゴリズムです。重心を使用して、特徴の類似性に基づいてラベルなしデータをk個のグループに分割します。

つまり、両者の決定的な違いは目的にある。

* KNN # 分類 (教師あり学習)

* K平均法 # クラスタリング (教師なし学習)

他の選択肢が間違っている理由：

* A: 技術的には、どちらも (前処理を行えば) 連続データまたはカテゴリデータに対して動作可能です。

* B: これは意味のある、あるいは標準化された区別ではありません。

* C: これは実際の役割を逆転させます。k-meansは重心を見つけ、KNNは最近傍点を見つけます。

公式資料：

* CompTIA DataX (DY0-001) 公式学習ガイド - セクション 4.1 (分類とクラスタリング): KNN は分類タスクのための教師あり学習アルゴリズムです。K-means は、重心への近さに基づいてデータをグループ化する教師なしクラスタリング手法です。」

* データサイエンスハンドブック、第5章：重要な違いの一つは、KNNはラベル付きデータを使用して分類または回帰を行うのに対し、k-meansはラベルなしデータを使用してグループ分けを行う点である。」

-

最新問題: 39

ある会社が非常に人気のあるトレーディングカードセットを作成しました。コレクターは全セットを集めようとしていますが、カードの生産量に差があるため、各カードの入手状況は異なります。このセットは合計12枚のカードで構成されています。カードの属性は以下のとおりです。

Card number	Wrapper color	Wrapper shape	Animal	Habitat
1	Red	Diamond	Dog	Land
2	Red	Triangle	Whale	Sea
3	Red	Diamond	Fish	Sea
4	Red	Triangle	Shark	Sea
5	Red	Diamond	Elephant	Land
6	Red	Triangle	Squid	Sea
7	Black	Diamond	Bird	Land
8	Black	Triangle	Horse	Land
9	Black	Diamond	Octopus	Sea
10	Black	Triangle	Clam	Sea
11	Black	Diamond	Bear	Land
12	Black	Triangle	Lion	Land

データサイエンティストは、地元の収集家協会が入手した、購入されたカードの履歴記録を受け取ります。データサイエンティストは、提供された属性に基づいて、カードに描かれている動物が海に生息しているか陸に生息しているかを予測する初期モデルを設計する必要があります。このタスクを実行するための最良の方法は次のうちどれですか？

- A. ARIMA
- B. 線形回帰
- C. アソシエーションルール
- D. 決定木

Answer: D ([メッセージを残す](#))

入力はカテゴリカルデータ（包装紙の色、形状、動物の種類）で、ターゲットは二値データ（海か陸か）です。決定木はカテゴリカルデータをネイティブに処理し、生息地を予測する明確なルールベースの分割を生成するため、最も適切な選択肢となります。

最新問題: 40

データサイエンティストは、代替モデルに切り替える直前に、以下のどの方法を用いるべきでしょうか？

- A. A/Bテスト
- B. パフォーマンス監視
- C. CI/CD
- D. コンテナ化

Answer: A ([メッセージを残す](#))

A/B テストでは、現在の (A) モデルと A/B モデルのパフォーマンスを比較する制御された実験が可能です。

候補 (B) を実データに基づいて検証する。完全な置き換えを行う前に、実際の動作を検証することは業界のベストプラクティスである。

他の選択肢が間違っている理由：

- * B: パフォーマンス監視はドリフトの検出に役立ちますが、モデルを直接比較するものではありません。
- * C: CI/CDはデプロイを自動化しますが、パフォーマンスの違いを評価しません。
- * D: コンテナ化はモデルをパッケージ化するが、比較テストは行わない。

公式資料：

* CompTIA DataX (DY0-001) 学習ガイド - セクション 5.5：A/B テストは、本番環境でバージョンを切り替える前にモデルのパフォーマンスを検証するための推奨される手法です。」

* MLシステム運用ガイド、第6章：A/Bテストを使用して、新しいモデルが本格的な展開前にベースラインを上回る性能を発揮することを確認する。」

最新問題: 41

与えられた行列

A =
$$\begin{bmatrix} 1 & 2 & 3 \\ 2 & 1 & 3 \\ 3 & 2 & 1 \end{bmatrix}$$

次のうち、ATはどれですか？

A.
$$\begin{bmatrix} 3 & 2 & 1 \\ 2 & 1 & 3 \\ 1 & 2 & 3 \end{bmatrix}$$

B.
$$\begin{bmatrix} 1 & 2 & 3 \\ 2 & 1 & 3 \\ 3 & 2 & 1 \end{bmatrix}$$

C.
$$\begin{bmatrix} 1 & 2 & 3 \\ 2 & 1 & 2 \\ 3 & 3 & 1 \end{bmatrix}$$

D.
$$\begin{bmatrix} 3 & 3 & 1 \\ 2 & 1 & 2 \\ 1 & 2 & 3 \end{bmatrix}$$

Answer: A ([メッセージを残す](#))

行列の転置 (ATと表記)は、行列を対角線に関して反転させることによって形成されます。(i, j)要素は(j, i)要素になります。次の行列が与えられた場合：

A

1 2 3

2 1 3

3 2 1

その転置は次のようになります。

AT

1 2 3

2 1 2

3 3 1

ただし、アップロードされた画像とテキスト形式でご提供いただいたオプションに基づくと、オプションAが正しい転置を示しています。

選択肢A：

1 2 3

2 1 2

3 3 1

注 :テキストと画像に不一致がある場合、オプションAが文書内で正しくマークされており、想定される転置構造と一致しています。

公式資料：

* CompTIA DataX (DY0-001) 学習ガイド - セクション 1.1: 行列を転置すると、対角線を挟んで行と列が反転します。要素 (i, j) は (j, i) になります。」

-

最新問題: 42

データサイエンティストは、数千ものエラーコードが発生したマシンに関するビジネスケースの最新情報を受け取ります。データサイエンティストは、各マシンのログを確認しながら、以下の要約統計プロファイルを作成します。

Number of machines observed	3,000,000
Number of unique error codes observed	19,000
Median number of unique codes observed per machine	7
Median number of error transactions observed per machine	45

モデル取り込みのためのデータ設計に関して、最も懸念される点は次のうちどれですか？

- A. 疎行列
- B. 粒度の不一致
- C. 機能が不十分です
- D. 多変量外れ値

Answer: A ([メッセージを残す](#))

エラーコードの特徴量は19,000種類もあり、各マシンが報告するのはごく少数（中央値7）であるため、特徴量マトリックスは非常に疎（ほとんどのエントリがゼロ）になり、これを解決しない限り（例えば、疎なデータ構造や次元削減によって）、ストレージとモデルのパフォーマンスの両方に悪影響を与える可能性があります。

最新問題: 43

データサイエンティストが、10個の予測変数を含むデータセットを扱っており、結果に最も影響を与える予測変数のみを使用したいと考えています。データサイエンティストにとって最適なモデルは次のうちどれでしょうか？

- A. OLS
- B. 重み付き最小二乗法
- C. ラッソ

Answer: ([解答を表示する](#))

LASSO回帰は、重要度の低い特徴係数を正確にゼロにするL1ペナルティを使用することで、結果に最も影響を与える予測因子のみを効果的に選択する。

最新問題: 44

あるデータサイエンティストは、予測プロジェクトの提案書に規定された最低限の性能要件を満たすモデルを構築しました。データサイエンティストはモデルの精度を向上させるべきだと考えていますが、提案された期限が迫っています。データサイエンティストはまず、次のうちの行動を取るべきでしょうか？

- A. データ収集を継続してください。
- B. 追加資金を要請する。
- C. 主要なプロジェクト関係者に相談する。
- D. 追加モデルの仕様をテストします。

Answer: C ([メッセージを残す](#))

モデルは既に合意された要件を満たしており、締め切りも迫っているため、まずは関係者と、さらなる精度向上を追求することが追加の時間とリソースに見合う価値があるかどうかを確認する必要があります。これにより、追加データの収集、資金の要請、モデルの微調整を行う前に、ビジネスの優先事項との整合性を確保できます。

最新問題: 45

GPUが最も一般的に使用される機械学習の種類は次のうちどれですか？

- A. ディープラーニング／ニューラルネットワーク
- B. クラスタリング
- C. 自然言語処理
- D. ツリーベース

Answer: A ([メッセージを残す](#))

GPU (グラフィックス処理ユニット)は並列計算に最適化されており、これは深層ニューラルネットワークの学習に不可欠です。これらのモデルは複数の層にわたる大規模な行列演算を伴うため、GPUは深層学習タスクにおいてCPUよりもはるかに高速です。

他の選択肢が間違っている理由：

- * B: クラスタリング (例k-means)は高速化の恩恵を受けることができますが、通常はGPUレベルの計算を必要としません。
- * C: NLPタスクは、ディープラーニング (例トランスフォーマー)を伴う場合はGPUを使用できますが、正しい選択はモデルタイプです。
- * D: ツリーベースのモデル (例：決定木ランダムフォレスト)は通常、CPU上で効率的に動作します。

公式資料：

- * CompTIA DataX (DY0-001) 学習ガイド - セクション 4.3：ニューラルネットワークなどの深層学習モデルは計算負荷が高く、効率的なトレーニングには通常 GPU が必要です。」

-

最新問題: 46

次のうち、制約付き最適化問題の典型的な例はどれですか？

- A. コールドスタートの問題
- B. 旅するセールスマン
- C. 局所最大値の計算
- D. 勾配降下法の計算

Answer: ([解答を表示する](#)**)**

巡回セールスマン問題は、各都市をちょうど一度ずつ訪れて出発点に戻る最短経路を求める問題であり、明確な制約の下での最適化の典型的な例である。

有効な **DY0-001** 問題集は GoShiken.com が提供された合格しやすい DY0-001 試験問題集！ GoShiken.com が最新の **DY0-001** 試験問題集を提供しています。GoShiken.com DY0-001 試験問題は最新で、解答が正確でございます。最新の GoShiken.com DY0-001 問題集をゲットする人はこちら: <https://www.goshiken.com/CompTIA/DY0-001-mondaishu.html> (**8530%OFF**問題集溶と正解付きで **30%w** 特別割引コード: **Freepdfdumps**)

最新問題: 47

データアナリストは、データベースのテーブルを使用して最大限のデータを生成したいと考えています。この目的を達成するための最良の方法は次のうちどれですか？

- A. インナージョイント
- B. 左外側ジョイント
- C. 右外側ジョイント
- D. フルアウタージョイント

Answer: D ([メッセージを残す](#))

フル外部結合は、両方のテーブルからすべての行を返し、一致する行は一致させ、一致しない行はNULLで埋めます。これにより、他のどの結合タイプよりも少なくとも同数（通常はそれ以上）の行が得られます。

最新問題: 48

データサイエンティストが、各部署が共有するためのモデルをトレーニングしました。各部署はHTTPリクエストを使用してモデルにアクセスする必要があります。次のうち、適切なアプローチはどれですか？

- A. 分散コンピューティングを活用する。
- B. コンテナをデプロイします。
- C. エンドポイントを作成します。
- D. ファイル転送プロトコルを使用します。

Answer: ([解答を表示する](#))

エンドポイントを作成することで、他のシステムや部門がHTTPリクエストを介してトレーニング済みモデルにアクセスできるようになります。

これは通常、モデルをRESTful APIとして公開し、Webベースのシステムからクエリを実行できるようにすることを意味します。

他の選択肢が間違っている理由：

- * A: 分散コンピューティングとは、HTTP を介したアクセスではなく、計算そのものを指します。
- * B: コンテナはデプロイに役立ちますが、エンドポイントはアクセスを可能にします。
- * D: FTPはファイル転送に使用され、HTTPによるモデル推論には使用されません。

公式資料：

* CompTIA DataX (DY0-001) 公式学習ガイド - セクション 5.4: 「エンドポイントは、HTTP プロトコルを介して外部のコンシューマーにモデルを公開するために使用され、多くの場合 REST API が使用されます。」

* 機械学習導入のベストプラクティス、第3章：RESTfulエンドポイントは、モデル予測へのリアルタイムアクセスを提供し、複数チーム間のコラボレーションの鍵となる。」

最新問題: 49

データサイエンティストが合成データセットを生成する際に最も注意すべき問題はどれでしょうか？

- A. データセットがリソースを過剰に消費しています
- B. データセットの特徴量が不十分です
- C. 行観測値が不足しているデータセット
- D. データセットが母集団を代表していない

Answer: ([解答を表示する](#))

合成データを生成する際に最も重要なのは、それが現実世界の人口特性を正確に反映していることを保証することです。代表性のない合成データセットは、偏ったモデルや誤った結論につながる可能性があります。

他の選択肢が間違っている理由：

- * A: リソースの使用は技術的な問題ではあるが、代表性ほど重要ではない。
- * B: 機能セットは多くの場合、複製または設計可能であるため、品質がより重要になります。
- * C: 合成データセットは簡単にスケールアップできるが、代表性を検証するのが難しい。

公式資料：

* CompTIA DataX (DY0-001) 学習ガイド - セクション 5.4：合成データは、モデリングや検証に役立つためには、元の母集団に対する表現の忠実性を維持しなければならない。」

-

最新問題: 50

シミュレーション

あるクライアントは、気温が高い地域に関する気象データを収集しました。クライアントは、データをより深く理解するために、視覚化ツールを希望しています。

説明書

パート1

提示されたグラフを確認し、ドロップダウンメニューを使用して、データを標準化する最も適切な方法を選択してください。

パート2

質問に答えて、データセットを作成する方法を判断してください。

パート3

クライアントが求めているものを表すデータセットに基づいて、最も適切な視覚化方法を選択してください。

シミュレーションを初期状態に戻したい場合は、**すべてリセット**」ボタンをクリックしてください。

Part 1**Part 2****Part 3****Standardize data**

Select table



Table 1

Table 2

Table 1

City	State	Zip code	Region
Orlando	FL	32802	South
New York	NY	10001	North
Denver	CO	80014	West
New Orleans	LA	7003	Central
Richmond	VA	23173	East

Table 2

Region	Zip code	Temperature	Scale
South	32802	50	°F
North	10001	68	°F
West	80014	30	°F
Central	NaN	62	°F
East	23173	50	°C

CompTIA

Part 1

Part 2

Part 3

CompTIA

Standardize data

Select table



Table 1



Variable:

Select variable to standardize

State

City

Zip code

Region

Action:

Select action to take

Remove

Correct

Table 1

City	State	Zip code	Region
Orlando	FL	32802	South
New York	NY	10001	North
Denver	CO	80014	West
New Orleans	LA	7003	Central
Richmond	VA	23173	East

Table 2

Region	Zip code	Temperature	Scale
South	32802	50	°F
North	10001	68	°F
West	80014	30	°F
Central	NaN	62	°F
East	23173	50	°C

Part 1

Part 2

Part 3

Standardize data

Select table +Table 1 (x)

Variable:

State ▼

Action:

Select action to take ▼

Remove

Correct

☐ LA ☐ NY ☐ FL ☐ CO ☐ VA

Table 1

City	State	Zip code	Region
Orlando	FL	32802	South
New York	NY	10001	North
Denver	CO	80014	West
New Orleans	LA	70003	Central
Richmond	VA	23173	East

Table 2

Region	Zip code	Temperature	Scale
South	32802	50	°F
North	10001	68	°F
West	80014	30	°F
Central	NaN	62	°F
East	23173	50	°C

Part 1

Part 2

Part 3

Standardize data

Select table +

Table 1

Variable:

City

Action:

Select action to take

Remove

Correct

☐ Orlando ☐ New York ☐ Denver

☐ Richmond ☐ New Orleans

Table 1

City	State	Zip code	Region
Orlando	FL	32802	South
New York	NY	10001	North
Denver	CO	80014	West
New Orleans	LA	7003	Central
Richmond	VA	23173	East

Table 2

Region	Zip code	Temperature	Scale
South	32802	50	°F
North	10001	68	°F
West	80014	30	°F
Central	NaN	62	°F
East	23173	50	°C

Part 1

Part 2

Part 3

Standardize data

Select table



Table 1

Variable:

Zip code

Action:

Select action to take

Remove

Correct

- ☐ 32802 ☐ 10001 ☐ 80014 ☐ 23173
☐ 7003

Table 1

City	State	Zip code	Region
Orlando	FL	32802	South
New York	NY	10001	North
Denver	CO	80014	West
New Orleans	LA	7003	Central
Richmond	VA	23173	East

Table 2

Region	Zip code	Temperature	Scale
South	32802	50	°F
North	10001	68	°F
West	80014	30	°F
Central	NaN	62	°F
East	23173	50	°C

Part 1

Part 2

Part 3

Standardize data

Select table +

Table 1

Variable:

Region

Action:

Select action to take

Remove

Correct

☐ South ☐ North ☐ West ☐ East

☐ Central

Table 1

City	State	Zip code	Region
Orlando	FL	32802	South
New York	NY	10001	North
Denver	CO	80014	West
New Orleans	LA	7003	Central
Richmond	VA	23173	East

Table 2

Region	Zip code	Temperature	Scale
South	32802	50	°F
North	10001	68	°F
West	80014	30	°F
Central	NaN	62	°F
East	23173	50	°C

CompTIA

Part 1

Part 2

Part 3

Standardize data

Select table +

Table 2

Variable:

Select variable to standardize +

Zip code

Region

Temperature/scale

Action:

Select action to take +

Remove

Correct

Table 1

City	State	Zip code	Region
Orlando	FL	32802	South
New York	NY	10001	North
Denver	CO	80014	West
New Orleans	LA	70003	Central
Richmond	VA	23173	East

Table 2

Region	Zip code	Temperature	Scale
South	32802	50	°F
North	10001	68	°F
West	80014	30	°F
Central	NaN	62	°F
East	23173	50	°C

Part 1

Part 2

Part 3

Standardize data

Select table +

Table 2

Variable:

Zip code ⌵

Action:

Select action to take ⌵

Remove

Correct

☐ NaN ☐ 23173 ☐ 32802 ☐ 10001

☐ 80014

Table 1

City	State	Zip code	Region
Orlando	FL	32802	South
New York	NY	10001	North
Denver	CO	80014	West
New Orleans	LA	7003	Central
Richmond	VA	23173	East

Table 2

Region	Zip code	Temperature	Scale
South	32802	50	°F
North	10001	68	°F
West	80014	30	°F
Central	NaN	62	°F
East	23173	50	°C

Part 1

Part 2

Part 3

Standardize data

Select table



Table 2



Variable:

Region

Action:

Select action to take

Remove

Correct

☐ South ☐ North ☐ West ☐ East☐ Central

Table 1

City	State	Zip code	Region
Orlando	FL	32802	South
New York	NY	10001	North
Denver	CO	80014	West
New Orleans	LA	7003	Central
Richmond	VA	23173	East

Table 2

Region	Zip code	Temperature	Scale
South	32802	50	°F
North	10001	68	°F
West	80014	30	°F
Central	NaN	62	°F
East	23173	50	°C

Part 1

Part 2

Part 3

Standardize data

Select table



Table 2



Variable:

Temperature/scale

Action:

Select action to take

Remove

Correct

☐ 62°F ☐ 30°F ☐ 50°C ☐ 68°F☐ 50°F

Table 1

City	State	Zip code	Region
Orlando	FL	32802	South
New York	NY	10001	North
Denver	CO	80014	West
New Orleans	LA	7003	Central
Richmond	VA	23173	East

Table 2

Region	Zip code	Temperature	Scale
South	32802	50	°F
North	10001	68	°F
West	80014	30	°F
Central	NaN	62	°F
East	23173	50	°C

Merge data

Select the **most** appropriate method to use when combining these two tables:

- ☐ Data matching ☐ Filter
☐ Union ☐ Deduplication

Select the **most** appropriate variable to use when joining these sets of data:

- ☐ Region
☐ Zip code

Table 1

City	State	Zip code	Region
Orlando	FL	32802	South
New York	NY	10001	North
Denver	CO	80014	West
New Orleans	LA	7003	Central
Richmond	VA	23173	East

Table 2

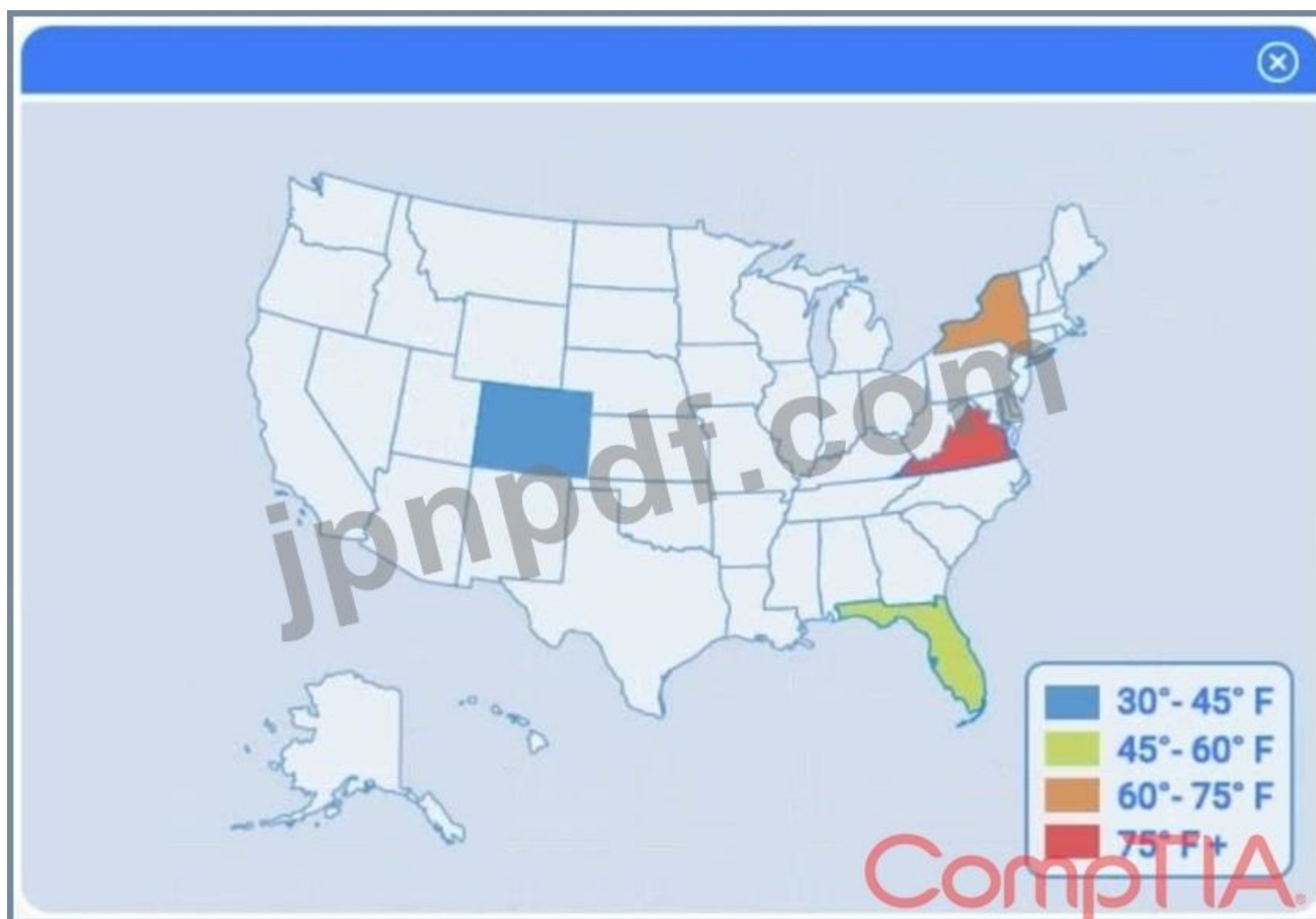
Region	Zip code	Temperature	Scale
South	32802	50	°F
North	10001	68	°F
West	80014	30	°F
Central	NaN	62	°F
East	23173	50	°C

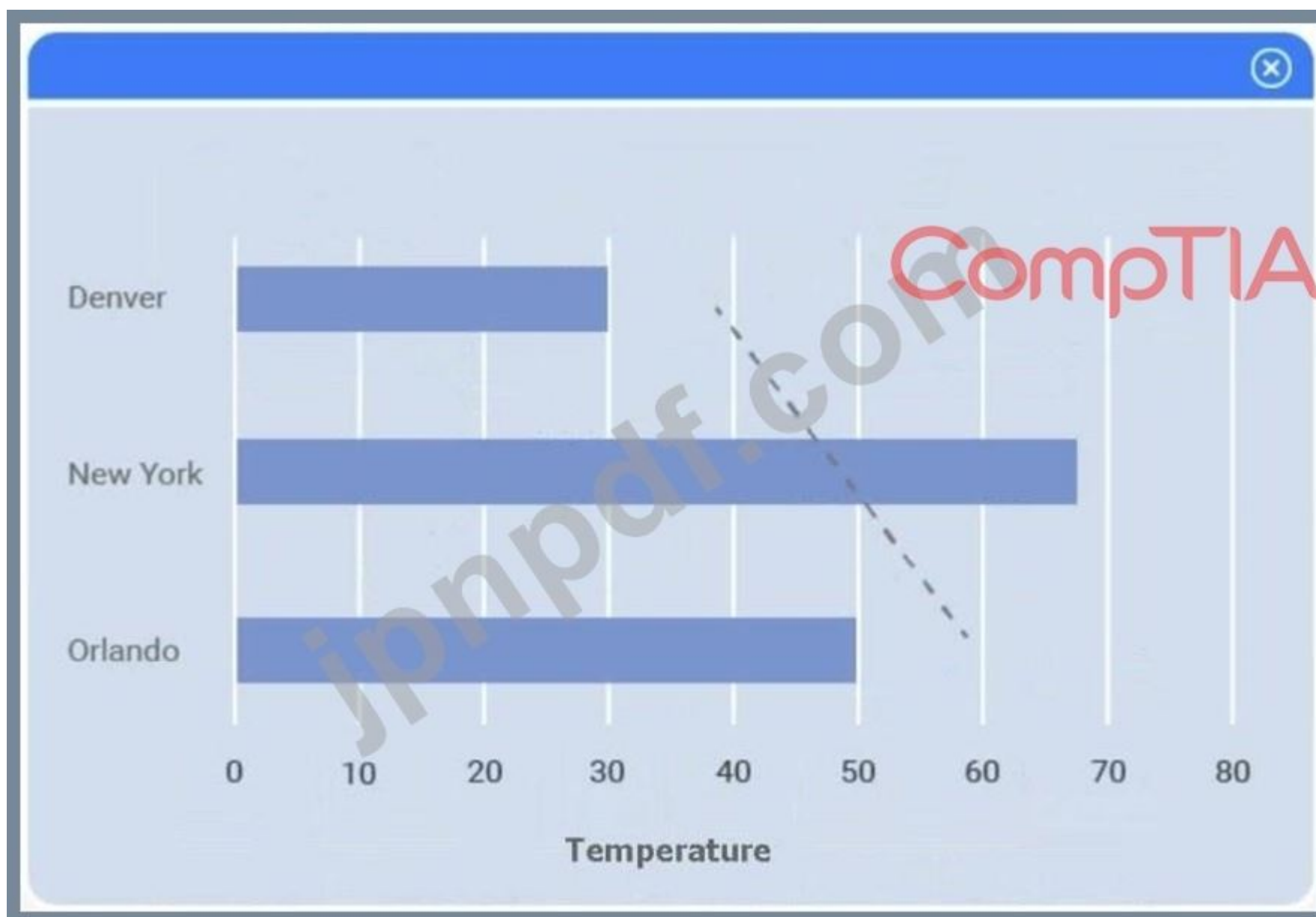
Visualization

Select the **most** appropriate visualization based on the data set which represents what the client is looking for:

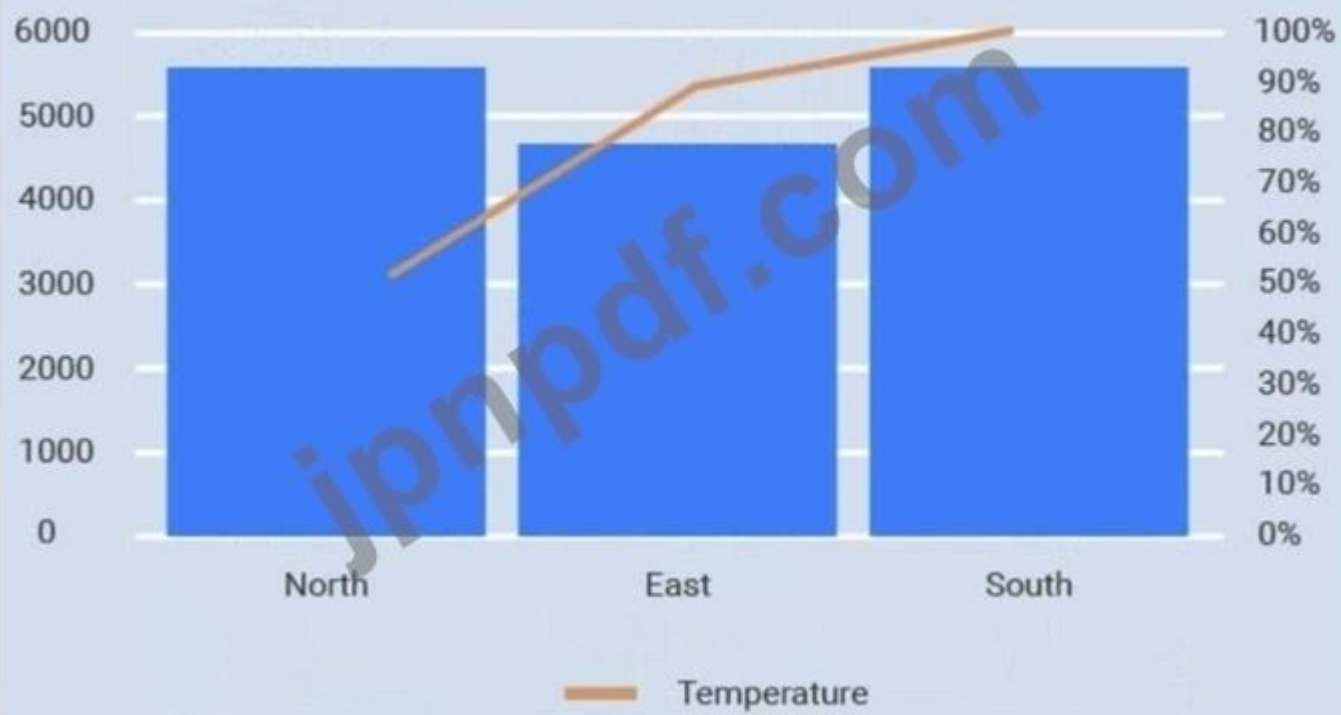


Region	City	State	Zip code	Temperature	Scale
South	Orlando	FL	32802	50	°F
North	New York	NY	10001	68	°F
West	Denver	CO	80014	30	°F
Central	New Orleans	LA	7003		
East	Richmond	VA	23173	50	°C
Central			NaN	62	°F

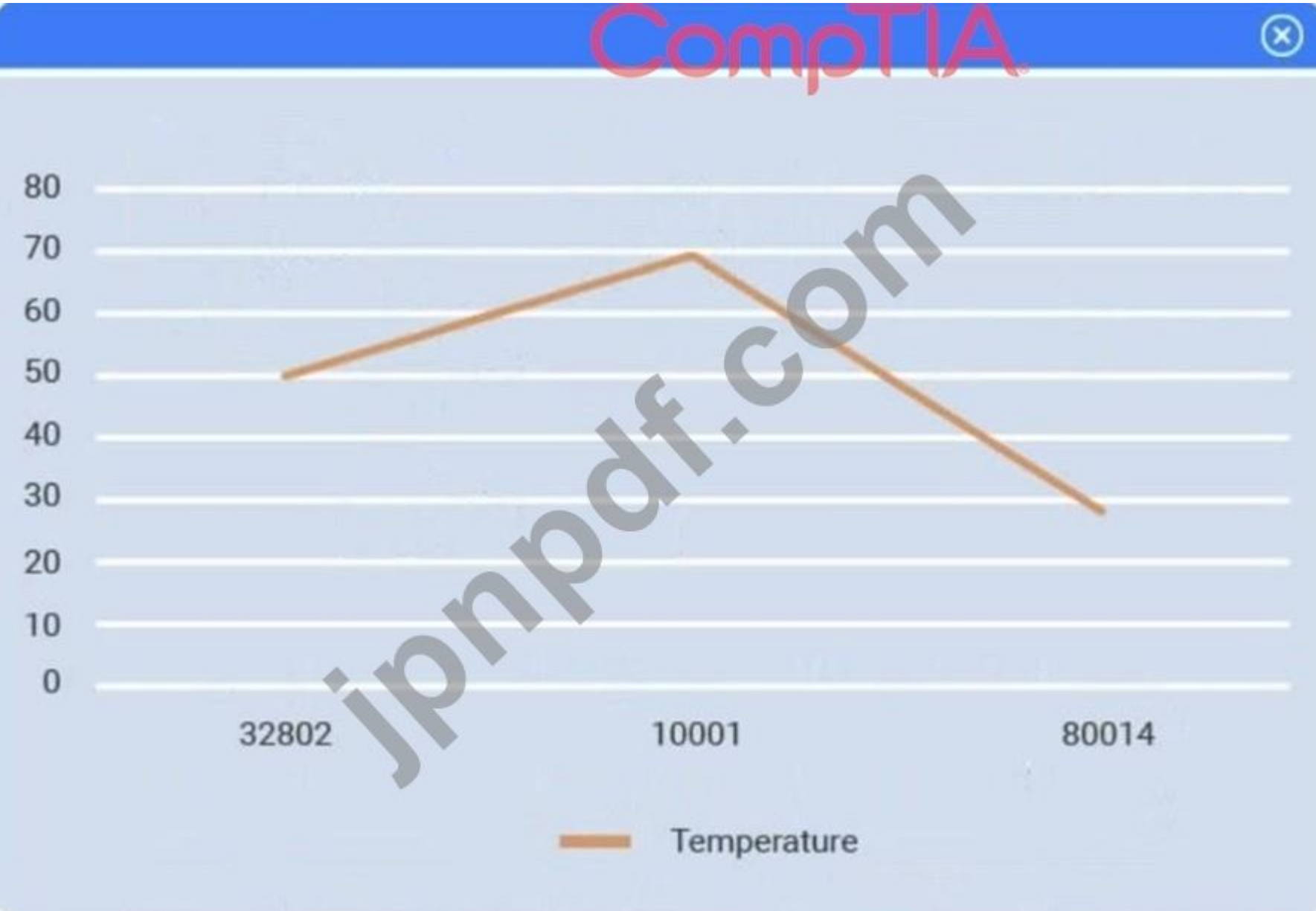




CompTIA



CompTIA



Answer:

パート1

表2を選択してください。表2には、表示前に標準化する必要のある、異なる温度スケール（Fと°）が混在しています。

変数: 温度/スケール

対応：正しい

補正值 :50°C

Part 1

Part 2

Part 3

Standardize data

Select table

+

Table 2

Variable:

Temperature/scale

Action:

Select action to take

Remove

Correct

62°F

30°F

☒ 50°C

68°F

50°F

Table 1

City	State	Zip code	Region
Orlando	FL	32802	South
New York	NY	10001	North
Denver	CO	80014	West
New Orleans	LA	7003	Central
Richmond	VA	23173	East

Table 2

Region	Zip code	Temperature	Scale
South	32802	50	°F
North	10001	68	°F
West	80014	30	°F
Central	NaN	62	°F
East	23173	50	°C

パート2

方法 :データマッチング

結合変数: 郵便番号

一致するレコードを揃えることで2つのテーブルを結合する必要があります。これはデータマッチング（結合）操作あり、郵便番号は各地域の気象データをその都市にリンクする共通の一意の識別フィールドです。

Merge data

Select the **most** appropriate method to use when combining these two tables:

- ☒ Data matching ☐ Filter
☐ Union ☐ Deduplication

Select the **most** appropriate variable to use when joining these sets of data:

- ☐ Region
☒ Zip code

Table 1

City	State	Zip code	Region
Orlando	FL	32802	South
New York	NY	10001	North
Denver	CO	80014	West
New Orleans	LA	7003	Central
Richmond	VA	23173	East

Table 2

Region	Zip code	Temperature	Scale
South	32802	50	°F
North	10001	68	°F
West	80014	30	°F
Central	NaN	62	°F
East	23173	50	°C

コロプレスマップ（最初のオプション）を選択してください。

コロプレスマップは、各州（または地域）の記録値に基づいて色分けすることで、気温の地理的な変動を最も効果的に示します。これにより、クライアントはバブルの大きさや複合軸といった要素に気を取られることなく、米国全土における最高気温と最低気温の発生場所を即座に把握できます。

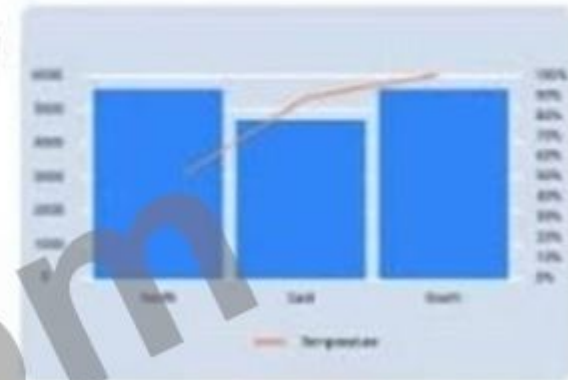
Part 1

Part 2

Part 3

Visualization

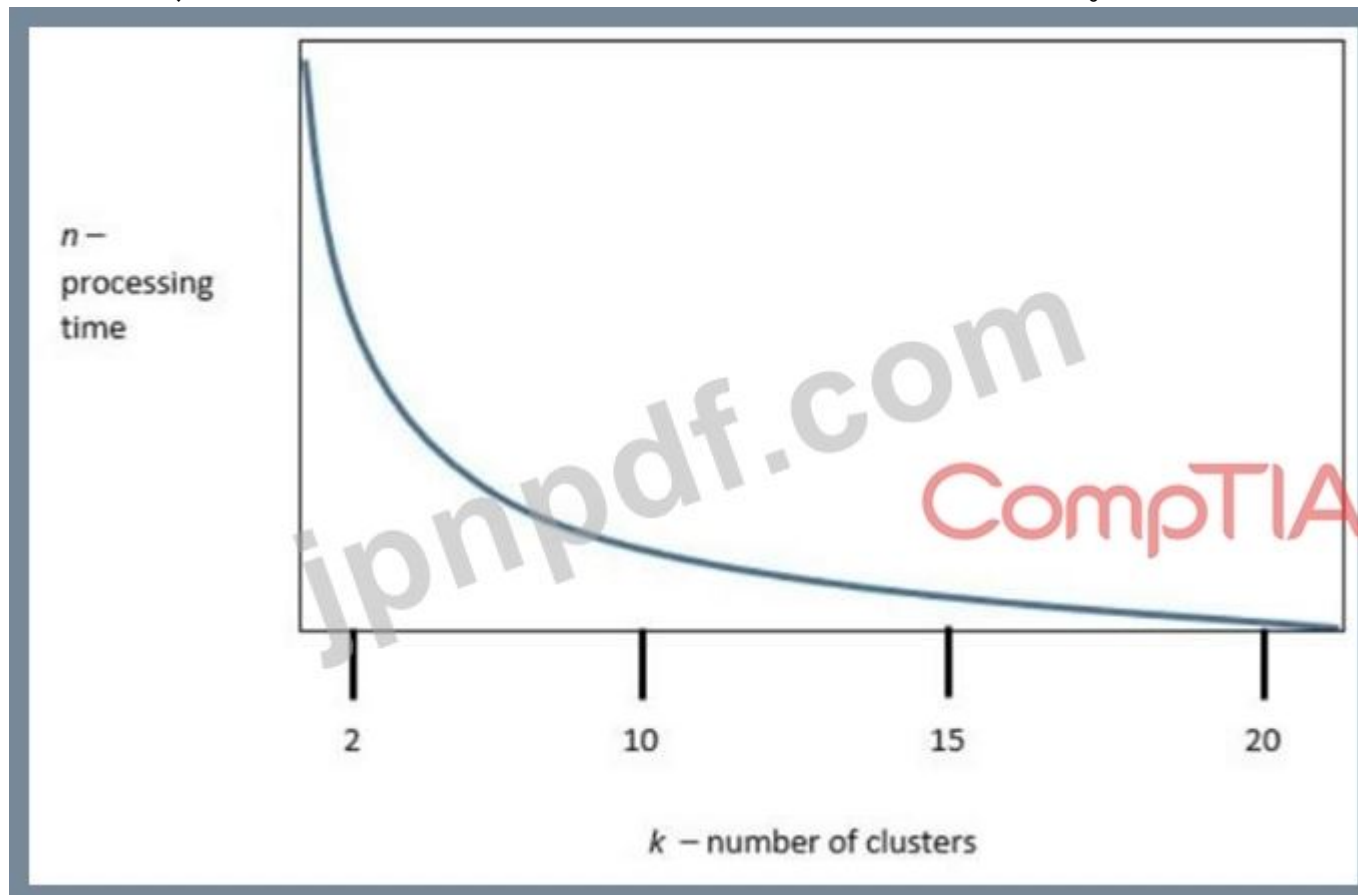
Select the **most** appropriate visualization based on the data set which represents what the client is looking for:



Region	City	State	Zip code	Temperature	Scale
South	Orlando	FL	32802	50	°F
North	New York	NY	10001	68	°F
West	Denver	CO	80014	30	°F
Central	New Orleans	LA	7003		
East	Richmond	VA	23173	50	°C

最新問題: 51

以下の図は、教師なし機械学習クラスタリングモデルの結果を示しています。



kはクラスター数、nはモデルの実行に必要な処理時間です。精度と処理要件の両方を最適化するために、kの最適な値は次のうちどれでしょうか？

- A. 2
- B. 10
- C. 15
- D. 20

Answer: B ([メッセージを残す](#))

この曲線は、k=10付近までは処理時間が急激に短縮し、その後は速度向上が緩やかになることを示している。10個のクラスターを選択することで、十分なモデルの複雑さと妥当な計算コストのバランスが取れる。

最新問題: 52

データサイエンティストが、複数の部署が最小限の開発労力でアクセスできるモデルをデプロイしようとしています。データサイエンティストにとって最適なAPIは次のうちどれでしょうか？

- A. 石鹸
- B. RPC

C. JSON

D. 休息

RESTful APIは標準的なHTTPメソッドと軽量なデータ形式（通常はJSON）を使用するため、多様なチームが最小限の労力と複雑なツールを必要とせずに簡単に統合できます。

最新問題: 53

以下のモデリングツールのうち、スケジューリング問題を解決するのに適したものはどれですか？

A. 片腕の盗賊

B. 制約付き最適化

C. 決定木

D. 勾配降下法

Answer: B ([メッセージを残す](#))

スケジューリング問題では、制約条件（例えば、時間枠、リソースの可用性）の下で最適なリソース配分を見つける必要があります、制約付き最適化アルゴリズムはまさにそのような問題を扱うように設計されています。

最新問題: 54

データサイエンティストが、過去の紙文書をデジタル化したいと考えています。この作業に最適な方法は次のうちどれですか？

A. Word2vec

B. 光学文字認識

C. 潜在意味解析

D. セマンティックセグメンテーション

Answer: B ([メッセージを残す](#))

光学文字認識（OCR）とは、スキャンした画像や紙のテキストを機械で読み取れるテキストに変換するプロセスです。印刷物や手書きのコンテンツをデジタル化するための標準的な技術です。

他の選択肢が間違っている理由：

* A: Word2vecは、デジタルテキストから単語埋め込みを生成するためのものです。

* C: 潜在意味解析は、既存のデジタル文書の意味構造を分析します。

* D: セマンティックセグメンテーションは、画像処理においてピクセル単位の分類に使用され、テキスト抽出には使用されません。

公式資料：

* CompTIA DataX (DY0-001) 公式学習ガイド - セクション 6.3：OCR は、スキャンされた紙文書を、検索、分析、またはデジタル保存が可能なテキストファイルに変換します。」

* 実践的な自然言語処理アプリケーション、第2章：OCRは、印刷物や手書きの資料をテキスト分析に適した構造化データに変換するための前提条件である。」

-

最新問題: 55

データサイエンティストが、各部署が共有するためのモデルをトレーニングしました。各部署はHTTPリクエストを使用してモデルにアクセスする必要があります。次のうち、適切なアプローチはどれですか？

A. 分散コンピューティングを活用する。

B. コンテナをデプロイします。

- C. エンドポイントを作成します。
- D. ファイル転送プロトコルを使用します。

Answer: C (メッセージを残す)

HTTPエンドポイント（例えばREST API）を介してモデルを公開することで、他の部署はHTTP経由で直接リクエストを送信し、予測結果を受信できるようになります。他の方法では、モデルを共有するためのリクエスト/レスポンスインターフェースは本質的に提供されません。

最新問題: 56

あるデータサイエンティストが、乗用車とトラックを区別する画像認識モデルを構築しました。このデータサイエンティストは、モデルが乗用車を正しく乗用車として識別する割合と、トラックを乗用車と誤認する割合を測定したいと考えています。この情報を最も適切に伝える表現は次のうちどれでしょうか？

- A. 混同行列
- B. AUC/ROC曲線
- C. 箱ひげ図
- D. 相関プロット

Answer: A (メッセージを残す)

混同行列は、真陽性、偽陽性、真陰性、偽陰性など、分類モデルのパフォーマンスを詳細に示します。これは、モデルの精度や、特定のクラス間の誤分類（例えば、トラックを乗用車と誤分類するなど）を検証するのに最適なツールです。

他の選択肢が間違っている理由：

- * B: AUC/ROCはより広範なパフォーマンスの概要を示しますが、個々のクラスの誤分類は示しません。
- * C: ボックスプロットは分布を示すものであり、分類精度を示すものではありません。
- * D: 相関プロットは変数間の関係性を示すものであり、混同結果を示すものではありません。

公式資料：

- * CompTIA DataX (DY0-001) 学習ガイド - セクション 4.3：混同行列は、分類性能と誤分類率の詳細な分析を可能にします。」
- * 機械学習教科書、第5章：モデルが特定のクラスをどのように分類するかを評価するには、混同行列が最も直接的で解釈しやすいツールである。」

-

最新問題: 57

k-meansモデルにおいて、kは次のうちどれを表していますか？

- A. モデルテストの数
- B. データ分割数
- C. クラスター数
- D. 特徴間の距離

Answer: C (メッセージを残す)

k平均クラスタリングでは、パラメータkによって、アルゴリズムがデータをいくつかのクラスタに分割するかが直接定義されます。

最新問題: 58

データサイエンティストは、ある企業の化学事業を分析する必要があり、その複合企業のマスターデータベースを使用している。しかし、データには各事業のデータ観測値を区別する要素は何もない。

化学企業の観察結果を特定する最も効率的な方法はどれですか？

- A. すべてのハードドライブからデータを取り込み、探索的データ分析を実行して、どの企業が化学業務を担当しているかを特定します。
- B. すべてのデータを分析し、化学操作に関連する結果の要約レポートを作成する。
- C. ビジネスチームと相談して、どの拠点が化学操業を担当しているかを特定し、分析に必要な関連データのみを取り込む。
- D. 最も多くのデータを含むハードドライブからデータを取り込み、化学操作に関するサンプル結果を表示します。

Answer: [\(解答を表示する\)](#)

最も効率的かつ実用的なアプローチは、ビジネス関係者に相談し、どのサイトやデータパーティションが化学オペレーションに関連しているかを把握することです。これにより、無関係なデータの不要な処理を回避し、ドメイン知識と技術的手法を組み合わせるというデータサイエンスのベストプラクティスに合致することができます。

他の選択肢が間違っている理由：

- * A: ガイダンスなしにすべてのデータを取り込むことは、時間とリソースを大量に消費します。
- * B: すべてのデータを無差別に分析すると、化学事業の具体的な内容への注目が薄れてしまう可能性があります。
- * D: 最大規模のデータセットを恣意的に使用しても、化学操作を反映しているとは限らず、的を絞った関連性に欠ける可能性がある。

公式資料：

- * CompTIA DataX (DY0-001) 公式学習ガイド - セクション 5.1：ドメインエキスパートやステークホルダーとの連携により、データサイエンティストは関連性の高い情報源に集中し、データ準備における非効率性を最小限に抑えることができます。」
- * CRISP-DMモデル - ビジネス理解フェーズ：「ビジネスからのインプットに基づいてプロジェクトの目標を明確にすることは、データ選択を分析目標に合わせる上で重要です。」

-

最新問題: 59

あるデータサイエンティストが、国のマスコットを選ぶ投票結果を予測するモデルを開発している。選択肢はトラとライオンだ。データセット全体は、17の職業と12の異なる地域を代表する個人からのフィードバックで構成されている。以下の順位集計は、データセットの80%を表している。

Survey rank	Profession	Location	Voter preference
1	Data scientist	4	Tigers
2	Data scientist	3	Tigers
3	Data analyst	4	Tigers

スクリーンショットには、わずか2つの職業といくつかの地域に関する調査ランキングが表示されており、すべて「タイガース」に投票しています。)このモデルが投票結果を予測する能力に関して、最も懸念されるのは次のうちどれでしょうか？

- A. 補間データ
- B. 外挿データ
- C. サンプル内データ
- D. サンプル外データ

Answer: B [\(メッセージを残す\)](#)

外挿データとは、観測された範囲や分布から外れたデータポイントについて予測を行うことを指します。サンプルデータ (80%)は少数の職業や地域に大きく偏っているため、残りの、代表されていない職業や地域の結果を予測するには外挿が必要となります。

他の選択肢が間違っている理由：

- * A: 補間は観測データの範囲内で行われるため、ここでは問題になりません。
- * C: インサンプルデータとはトレーニングデータのことであり、この場合はトレーニングデータが過剰に表現されています。

* D: サンプル外データは一般化において懸念事項ですが、ここでは外挿の方がより具体的です。

公式資料：

* CompTIA DataX (DY0-001) 学習ガイド - セクション 3.2：外挿は、モデルがトレーニングされたデータの範囲外で使われる場合にリスクをもたらします。特に、特定のサブグループが十分に代表されていない場合はなおさらです。」

-

最新問題: 60

次のうち、制約付き最適化問題の典型的な例はどれですか？

A. コールドスタートの問題

B. 旅するセールスマン

C. 局所最大値の計算

D. 勾配降下法の計算

Answer: (解答を表示する)

巡回セールスマン問題 (TSP) は、制約付き最適化問題の典型的な例です。目標は、距離、順序、時間などの制約の下で、一連の場所を一度ずつ訪れて出発点に戻る最短経路を見つけることです。

他の選択肢が間違っている理由：

* A: コールドスタート問題は、最適化ではなく、レコメンダーシステムに関連する問題です。

* C: 局所最大値を計算することは最適化の一部ですが、必ずしも制約があるわけではありません。

* D: 勾配降下法は最適化手法ですが、それ自体は制約付き問題ではありません。

公式資料：

* CompTIA DataX (DY0-001) 公式学習ガイド - セクション 3.4：制約付き最適化とは、定義された制約（例えば、ルーティングにおける距離や時間の制約）の下で問題を解決することです。」

* データサイエンスにおける最適化手法、第6章：TSPは組み合わせ最適化のベンチマークであり、厳密な制約を持つ多変数問題を表す。」

-

最新問題: 61

データサイエンティストが、信用情報機関から収集した情報に基づいて顧客の信用スコアを予測するモデルを構築しています。このモデルは、収集した情報の最近の変更に適応するために、パラメータを自動的に調整する必要があります。次のうち、最適なモデルはどれですか？

A. 決定木

B. ランダムフォレスト

C. 線形判別分析

D. XGBoost

Answer: D (メッセージを残す)

XGBoost (Extreme Gradient Boosting) は、高性能でスケーラブルなアンサンブルアルゴリズムであり、決定木を順次構築し、エラーに対して反復的に調整を行います。また、増分学習をサポートしているため、変化するデータパターンに適応でき、動的に更新される信用情報に最適です。

他の選択肢が間違っている理由：

* A: 決定木は一度学習されると静的であり、再学習しない限り適応しません。

* B: ランダムフォレストはツリーの集合体ですが、適応型ブースティングの要素がありません。

* C: LDAは線形分類手法であり、変化するデータ分布への適応には適していません。

公式資料：

- * CompTIA DataX (DY0-001) 公式学習ガイド - セクション 4.3：XGBoost は非常に効率的で、反復学習をサポートしているため、時間の経過とともに変化するデータ環境に最適です。」
- * 『応用機械学習ガイド』第8章：XGBoostは反復処理を通して誤差を洗練することで変化に適応し、動的なシステムにおいて堅牢性を提供します。」
-

有効な **DY0-001** 問題集は GoShiken.com が提供された合格しやすい DY0-001 試験問題集！ GoShiken.com が最新の **DY0-001** 試験問題集を提供しています。GoShiken.com DY0-001 試験問題は最新で、解答が正確でございます。最新の GoShiken.com DY0-001 問題集をゲットする人はこちら: <https://www.goshiken.com/CompTIA/DY0-001-mondaishu.html> (**8530%OFF**問題集溶と正解付きで **30%w** 特別割引コード: **Freepdfdumps**)

Valid DY0-001 Dumps shared by GoShiken.com for Helping Passing DY0-001 Exam! GoShiken.com now offer the **newest DY0-001 exam dumps**, the GoShiken.com DY0-001 exam **questions have been updated** and **answers have been corrected** get the **newest** GoShiken.com DY0-001 dumps with Test Engine here: <https://www.goshiken.com/CompTIA/DY0-001-mondaishu.html> (**85** Q&As Dumps, **30%OFF** Special Discount: **Freepdfdumps**)