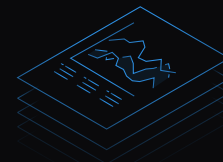




DataRobot Paxata

AI を活用したデータ準備

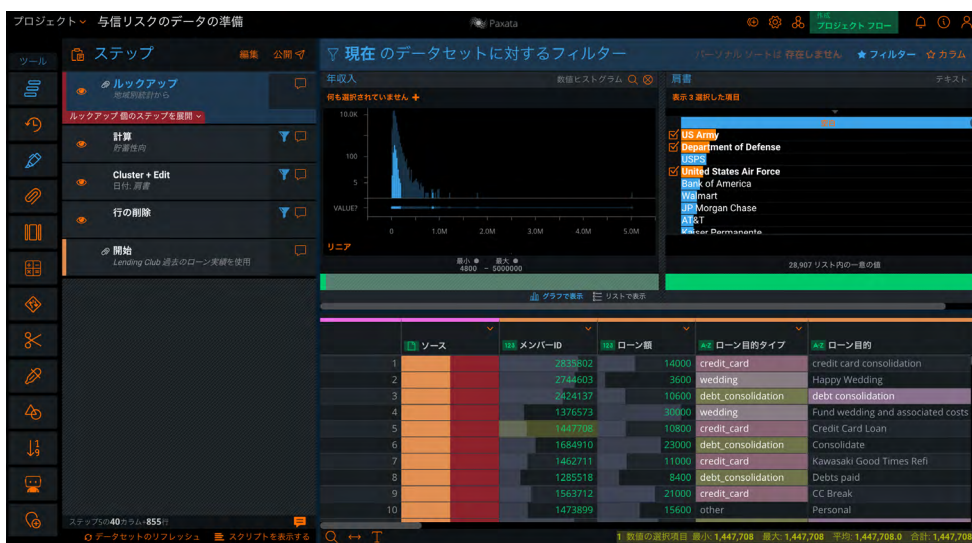


概要

AI は比類のない経済的価値をもたらす可能性を秘めています。ところが、多くの組織では、AI への投資から継続的に利益を生み出すことに苦労しています。組織では、膨大な量と種類のデータを収集していますが、機械学習プロジェクトを実行して高度な分析を行えるデータサイエンティストが少ないのが現状です。そのため、未加工データを有益な情報に変換するより効率的な方法を見つけ出す必要があります。実際、専門家は作業時間の最大 80% を、ビジネスの問題解決に必要なインサイトの獲得ではなく、データの検索、クリーンアップ、作成に費やしています。

1つのプラットフォームでデータを管理し、エンドツーエンドの価値を提供

DataRobot Paxata のセルフサービスのデータ準備を DataRobot の AI プラットフォームとシームレスに統合すれば、シチズンデータサイエンティストがデータをすばやく探索、クリーンアップ、結合、作成できるようになります。これにより、機械学習モデルをトレーニングおよびデプロイしたり本番環境のデータパイプラインを構築し、AI によるイノベーションを促進します。データサイエンスチームは、データソース、データセット、レシピを共同で作成、再利用、共有する際に、エンタープライズクラスのガバナンスとセキュリティを確保して、組織のポリシーを確実に遵守できるようになります。



視覚的な操作でデータを準備

あらゆる人が使用できるセルフサービスのデータ準備

DataRobot Paxata は、シチズンデータサイエンティストとアナリストのために開発されたアプリケーションです。技術やコーディングに関する豊富な経験がなくても、機械学習プロジェクトに必要なデータの準備やパイプラインのデプロイを行えます。

- **ポイントアンドクリックで直感的に操作できる UI:** 表示が一般的に使用されているスプレッドシートに似ており、すべてのデータセットを視覚的かつインタラクティブに操作できます。データのパターンを簡単に見つけ出し、視覚的な情報を利用してデータの品質を評価することが可能です。
- **データのクレンジング、情報付加、整形ができる強力な機能:** 直感的なポイントアンドクリック操作で、未加工データをすばやく変換して、データの不整合を解消し、欠損値を埋め、特徴量を改善します。また、コーディングすることなくデータを集約、重複排除し、ピボットテーブルで使用できるようになります。
- **再利用可能なレシピや共同作業により、複数のユーザーでプロジェクト開発:** データサイエンスチームのメンバーと連携しながら、複数のプロジェクト間でデータセットの開発、データソースの共有、データ準備プロセスの再利用を行えます。

「さまざまなデータソースを関連付け、データから価値を引き出すことは、世界的な銀行にとって重要な戦略の一部です。DataRobot Paxata のおかげで、私たちは未加工データを有用な情報に変換し、これまでよりも迅速に意思決定を下せるようになりました。データに基づいて意思決定を行う企業として自社の能力を高め、あらゆる役職の従業員が指先 1 つでデータを利用できるようにするには、このような分野への投資が重要です」

— Michael Gorriz 氏
グループ担当最高情報責任者



AI による支援で作業時間を短縮

あらかじめ組み込まれた AI とアルゴリズムによって、データ準備プロセス全体を通してユーザーを常にサポートし、インテリジェントな結論を導き出して複雑なデータ構造を簡素化します。また、DataRobot の AI による支援では、カテゴリ変数や特徴量の標準化と改善に役立つガイダンスや推奨事項を提供します。

- **インテリジェントな変数の標準化**：あらかじめ組み込まれた自然言語処理ベースのアルゴリズムを使用して、同じような発音またはつづりが似ている単語や語句を検出し、最もよく使われる値に標準化できるクラスター（名寄せ）を提案します。
- **インテリジェントな結合検出**：単一列または複数列の正確な結合条件やあいまいな結合条件を自動検出し、データセットを組み合わせると特徴量を改善します。
- **インテリジェントなデータ取り込み**：データの取り込み時にファイルタイプを自動で判別し、学習データセット作成のために、複雑な構造を適切な表形式に簡素化します。

AI が複数列の結合を提案

エンタープライズグレードのデータ準備

DataRobot Paxata は、業界で初めてとなる最も包括的な AI プラットフォームとの緊密な統合を実現します。これにより、経験の浅いユーザーがあらゆる規模のデータセットを利用して、データサイエンスチームのメンバーやデータ準備の関係者と連携し、エンドツーエンドのガバナンスを実施できるようになります。そのため、データの透明性を安全に管理および追跡することが可能です。

- **ビッグデータに適した Apache Spark™ を活用**：任意に抽出した少量のデータサンプルだけでなく、すべてのデータボリュームで変数と特徴量をインタラクティブに表示、探索、作成、標準化できます。
- **本番環境のデータパイプラインでインテリジェントな自動化を実現**：スケジュールに従ってデータ準備プロジェクトを一定間隔で自動的に開始します。さらに、今後のデータ準備に関する手順を自動的に検出して組み込み、完全なデータフローを作成します。
- **完全なデータリネージ**：データ準備プロセスを自動で記録して、データリネージの追跡やプロジェクト手順の再利用を実現します。さらに、すべてのデータ準備プロセスの説明可能性を高めることができます。

「このプラットフォームを構築した目的は 1 つです。それは、研究者や腫瘍専門医がわずかな時間でデータを再構築できるようにし、患者のデータを研究データベースにマッピングするのに費やしていた膨大な時間を治療計画の作成に使えるようにすることです」

— Dave Parkhill 氏
Precision Profile 社 CEO

お問い合わせ

DataRobot Japan
www.datarobot.com/jp
info-jp@datarobot.com

© 2020 DataRobot, Inc. All rights reserved.
DataRobot および DataRobot のロゴは DataRobot, Inc. の商標です。他のすべてのマークはそれぞれの所有者の商標または登録商標です。