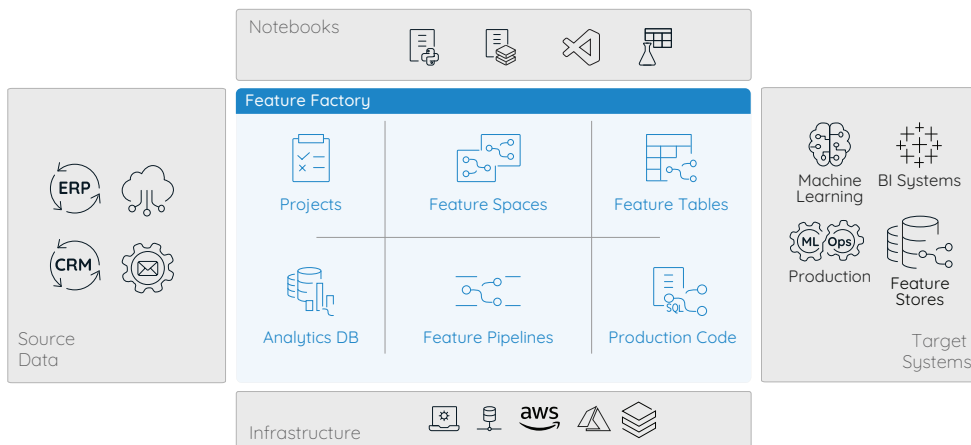


# dotData Feature Factory

データ中心の特微量自動設計が  
データサイエンスを革新する

## 企業のデータを変革

dotData Feature Factoryは、企業のデータサイエンスチームがキュレーションされた特微量を開発するために、データ加工に関するノウハウを再利用可能なアセットとして蓄積する仕組みを提供し、企業におけるデータの扱い方を根本的に変えるものです。データ中心に網羅的に自動構築される特微量空間から、アルゴリズムによってデータに隠れたパターン(特微量)を発見し、特微量発見のスピードと効率、再利用性と再現性、専門家間の連携、品質と透明性を向上します。dotData Feature Factoryは、企業データにおけるサイロ(縦割り構造)を破壊し、データから最大の価値を引き出します。



### 特微量探索の第一歩を踏み出す

本番環境で利用可能な特微量パイプラインを、分析の初日に生成できます。そして、お客様の課題解決に関連するテーブルやカラムから特微量を発見し、それらをすぐにチームと共有して評価できます。これらを数行のPythonコードで始めることができます。

### データ中心の特微量自動設計

従来、特微量の発見と設計は、職人芸とも言える、経験と勤による手作業が中心でした。dotData Feature Factoryは、そのような特微量設計をデータ中心のアプローチへと発展させます。特微量空間をプログラマ的に定義することで、手作業では不可能な圧倒的に広い範囲の特微量仮説を自動生成し、ユーザーのデータや業務に関する知識をデータから補完します。

### 特微量設計を再利用可能なアセット化

Notebookは実験の実行だけでなく、特微量抽出のプロセスを再利用可能なアセット化したり、特微量パイプラインを生成したりできるようになりました。dotData Feature Factoryは、特微量設計のプロセスを蓄積し、特微量抽出のノウハウを再利用可能なアセットに変換します。アセットはプロジェクトに保存され、データサイエンティストだけでなくチームで共有され、属人化されやすい特微量空間や特微量パイプラインといった特微量設計のノウハウをチーム間で連携し開発することができます。

### 本番品質の特微量パイプラインの生成

dotData Feature Factoryは、繰り返し可能な実験を運用できるアセットに変換できる価値を提供します。蓄積されたアセットは、特微量ストア、機械学習モデル、ダッシュボードへ入力するための運用可能な特微量パイプラインとして直接展開できます。

“dotData Feature Factory は、私たちのチームに驚異的な柔軟性とパワーをもたらしました。我々のビジネスにとって真に価値のある特微量を発見し、社内データを最大限に活用して透明性の高い機械学習モデルを構築できるようになりました。”

Exeter Finance LLC

データサイエンス・シニア・  
バイスプレジデント

Karthik Chandrasekhar氏

“dotDataは、多数のデータセットを横断して、新しく透明性の高い特微量を自動的に生成する独自の機能を有しています”

FORRESTER®

dotData社は、AWSから機械学習コンピテンシーパートナーとして認定されている世界でわずか9社のうちの1社です。



## 特長

dotData Feature Factoryは、特徴量を生成、修正、展開するためのプラットフォームです。本プラットフォームは、Pythonインターフェースにより、お使いのIDEやNotebookにインストールして使用できます。ノートパソコン等にインストールしてローカルで実行し、Databricks、Microsoft Azure ML、Amazon EMRなど、Spark対応の環境を含むクラスタやクラウドにスケールできます。

### マルチソース、マルチテーブルの特徴量設計

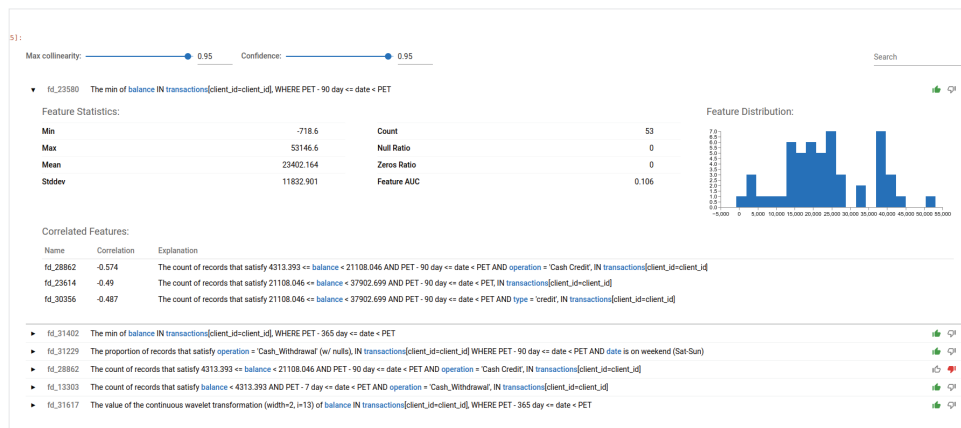
様々なデータソースに接続して、特徴量空間を拡張し、複数表からの特徴量の抽出を可能に。数値、カテゴリ、時系列、テキスト、空間データから、数百万もの特徴量仮説を自動的に生成

### データ加工の自動化

時間がかかり、ミスの発生しやすいデータ加工を、分析データベースに手順として記述し、自動のデータクレンジングによってデータ準備を自動化

### 時系列特徴量

ラグ、周期性、イベントなど、多次元の時間特徴量を、データリークを自動的に防止しながら探索



### ドメイン知識の補完

発見された特徴量を編集し、複数の実験から得られた特徴量を組み合わせて、ドメイン知識と融合

### 特徴量設計を再利用可能なアセット化

データ加工と特徴量変換の全てのステップを分析データベースへ記録し、特徴量に関する「ノウハウ」を再利用可能なアセットとして蓄積

### 企業のビッグデータに対応

dotData Feature Factoryは数十のテーブル、数千のカラム、数十億のレコードを持つ、企業のビッグデータに対応

### インサイトと説明可能性

特徴量の説明文やリーダーボードによって提供される定性的・定量的な情報が特徴量の透明性と説明性を保証し、ビジネスの洞察を導き出す

### 本番品質での特徴量パイプライン

複数のソーステーブルから特徴量を生成するためのパイプラインを、本番品質で自動生成

### Notebookへの統合

dotData Feature Factoryは、どのような環境でもインストールして実行できるPythonライブラリです。ノートパソコンでお使いのPython NotebookやIDEから、またはDatabricks、Azure ML、AWS上で使用可能

“dotDataによって到底発想し得ないような斬新な切り口の特徴量が多く抽出され、タイヤ開発のイノベーションに繋がっています。”

横浜ゴム株式会社  
エグゼクティブフェロー  
研究先行開発本部  
AI研究室 研究室長博士(工学)  
小石 正隆氏



## 問い合わせ先

合同会社 dotData Japan  
〒105-0021  
東京都港区東新橋1-5-2  
汐留シティセンター5階  
<https://jp.dotdata.com>  
[contact-j@dotdata.com](mailto:contact-j@dotdata.com)

© 2023 dotData, Inc. 無断複写・転載を禁じます。dotDataおよびdotDataのロゴは、米国dotData, Inc.または関連会社の米国およびその他各国における登録商標です。本文中の会社名、商品名およびロゴは、各社の商標または登録商標です。