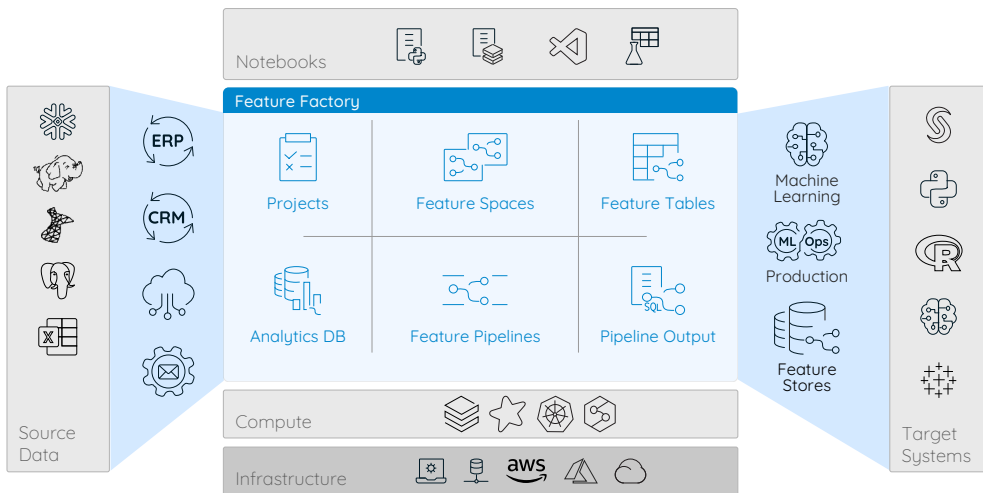.d dotData

dotData Feature Factory

Data-centric & programmatic feature discovery for Data Scientists and Data Engineers

A PARADIGM SHIFT FOR ENTERPRISE DATA

Feature Factory is a fundamental shift in how enterprise data science teams develop curated features and accumulate data know-how as reusable assets. Exploring exhaustive feature spaces and the ability to discover features through a data-centric, programmatic approach leads to enhanced collaboration, better efficiency, increased model quality, greater reusability, reproducibility, scalability, and transparency. Break down silos and capitalize on the wealth of information at your disposal.

JUMP-START FEATURE DISCOVERY

Build reproducible feature pipelines on day one. Identify the tables and columns that tell a story in relationship to your goal, described as features, and immediately share with others for evaluation and use. To get started, all you need to do is describe your data and let computers surface relevant features. All in a few lines of python code.

DATA-CENTRIC & PROGRAMMATIC

Feature engineering is an interactive process that is equal parts exploration and discovery, on repeat. Feature Factory is a python library built on top of a AI powered feature discovery framework that seamlessly integrates into your tool chain, allowing you to programmatically build, modify, and deploy features interactively with confidence all from the comfort of your favorite notebook or IDE.

REUSABLE DISCOVERY ASSETS

Notebooks are for experiments, not for creating reusable feature discovery assets or feature pipelines. Feature Factory captures the art of feature engineering in reusable feature discovery assets and the features themselves. Assets are stored in a project that are no longer trapped in a Data Scientists notebook. Recreate feature experiments across notebooks and IDEs, collaboratively develop feature spaces and feature pipelines, without the need to share notebooks or python code directly.

DEPLOY-READY FEATURE PIPELINES

True value comes from repeatable experimentation turned into operational assets. Feature pipelines assets can be directly deployed as operational feature pipelines to feed feature stores, machine learning models, or dashboards. Or generate SQL code to integrate into custom production applications with full transparency into the feature generation process.

> "You can't just throw more bodies at your data. When done manually, feature discovery is just a repetitive, trial-and-error process that takes a ton of time.
>
> dotData solves a problem I've been trying to solve for 20+ years."

Karthik Chandrasekhar
SVP, Decision Science
Exeter Finance

READY TO GET STARTED?

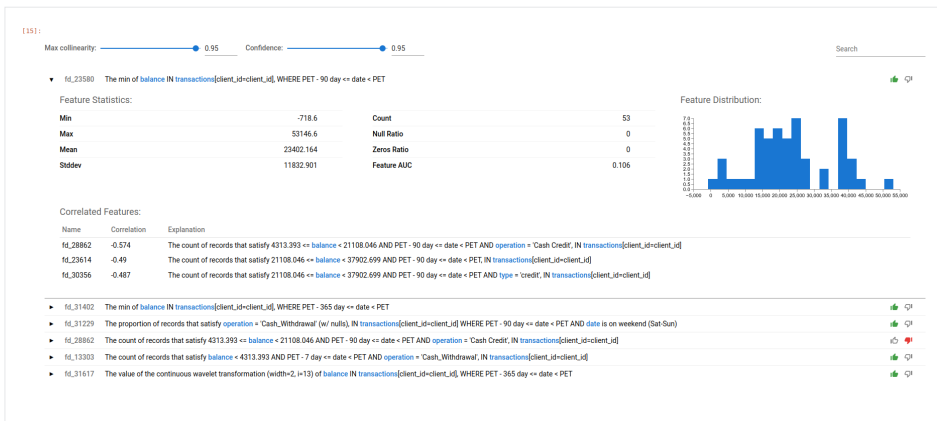**TRY FEATURE FACTORY**

.d

PRODUCT FEATURES

Feature Factory is a platform for building, modifying, and deploying features. A python interface to the Feature Factory platform enables installation and use in your favorite notebook or IDE. Install on your laptop and run locally, and scale to a cluster or cloud, including your choice of Spark-enabled compute environment, such as Databricks, Microsoft Azure ML, and Amazon EMR.

Multi-Source, Multi-Table Feature Engineering

Quickly connect to diverse data sources to enrich feature discovery and unlock iterative feature additions from new sources. Automatically generate millions of feature hypotheses from numeric, categorical, time-series, text, and geospatial data.

Automated Data Wrangling

Avoid time-consuming and error-prone data wrangling, dotData will automatically cleanse, align, and prepare data for feature discovery.

Time-Series Features, Without Leakage

Automatically generate and validate multidimensional time-based features, including holidays, lags or delays, periodicity, seasonality, and more.

Capture Expertise

Edit discovered features and combine features from multiple experiments to leverage domain expertise and knowledge gained from experimentation.

Build Reusable Feature Discovery Assets

Record every data and feature transformation step into your Analytic Database and turn your "know-how" into reusable assets for your organization.

Automated Feature Discovery at Enterprise Scale

Generate and score millions of features from complex tables, relationships, and billions of rows.

Insights and Explainable AI

Produce actionable business insights and explainable features from natural language descriptions and blueprinting of discovered features from source variables.

Production-Ready Feature Pipeline Generation

Automatically generate feature pipelines with full SQL code from source tables, ready for production use.

Integrated into Your Notebook

Feature Factory is a python library that can be installed and run anywhere you need to work. Use it from your favorite python IDE or Notebook on your laptop or remotely on Databricks, Azure ML, and AWS.

Contact Us

dotData, Inc.
1740 S. El Camino Real
200
San Mateo, CA 94402, USA
https://dotdata.com
contact@dotdata.com