ML & Data Science with SAP Data Intelligence
From isolated PoC to Intelligent Enterprise Application

Elisa Masetti, SAP Presales Data Scientist
February 18, 2020
Challenge of Enterprise Data Science

“Business managers surveyed by Gartner identified **time to value** as one of their biggest challenges [...] it takes an average of **52 business days** for their team to build a predictive model, and **longer** to deploy into production.”

*Survey Analysis: Traditional Approaches Dominate Data and Analytics Initiatives, Gartner 2018*

“Over 60% of models developed to operationalize them were **never actually operationalized**.”

*Survey Analysis: Data Science Team, Gartner 2018*
WHY #1: Data Jam
WHY #1: Data Jam
WHY #2: Data Science lives in a “test tube”

Along came Operations...
WHY #2: Data Science lives in a “test tube”

Did anyone check data access rights?

How to automate this? How to monitor? How to maintain this?

Does our architect know about this?

How to transform/embed this in our application technology?

Is this an enterprise architecture?

How to make this process safe, stable and functional?

Always the same discussion. It’s so slow. Costs money. Doesn’t scale. And I’m half way fired or in jail if this is not compliant!

Along came Operations...

I’m not the one to productize this

Here is my python code!
SAP Data Intelligence:

Turn Machine Learning into an Enterprise Process
Data Science Workflow Aided by SAP Data Intelligence

- Streaming Data
- Structured Data
- Unstructured Data

**Connect**
- Personalized Interaction
- Visualize and Respond
- Automate Business Processes

**Learn**
- Data Hub Orchestration
- Agile Data Preparation
- Machine Learning
- Deep Learning
- Deploy Models
- Monitor Performance
- Refine and Enhance
- Replace and Retire

**Scale**
- Consume
The situation
The idea

- **Sensors**
  - Pressure
  - Temperature

- **IR cameras**
  - IR image features

- **Raw material supplier**
- **ERP Data**

- **Raw material**

The process involves monitoring raw material using sensors and IR cameras to detect pressure and temperature. The detected features are sent to the ERP data system, which then decides whether the material is acceptable or not. If acceptable, the material is confirmed; if not, it is rejected.
The unsuccessful prototype

Manufacturing plant
- Raw Material
- Molding Press
- Pressure Sensor
- 1R Camera

Enterprise Data Center
- ERP Data
- SAP HANA

Cloud Data Lake
- Offline Image Processing
- Event Streaming

Not ready for production
- Smart Quality App
The right tool to make it productive - SAP Data Intelligence

On-premise and consumable SAP Cloud Platform service for:

- Centralized Metadata Governance
- Data and Process Pipelining
- Distributed and Heterogeneous Data Processing
- Model Lifecycle Management
Connection Management

Connect to everything leveraging the access control defined in the source system

- Prebuilt connectors for SAP systems, DB2, MySQL, Oracle, S3, HDFS, OpenData and many more
- Possibility to extend to even more data sources with SAP Open Cloud Connectors and SDK
Metadata Explorer

View, understand and organize your data across the IT landscape

- Data Discovery
- Data Profiling
- Data Preparation
- Data Quality Rules
- Metadata Cataloging
Metadata Explorer - Data Discovery

Explore data assets across the connected landscape and use the preview functionality to get a first understanding of your data.
Retrieve more in depth information about your data in order to clarify the structure and understand anomalies:

- Growth trend
- Values distribution across the dataset
- Percentage of null, blank and zero across the columns
Metadata Explorer - Data Preparation

End-to-end self-service data preparation for business users

- Access a dataset or upload a flat file to prepare the data based on a sample dataset
- Filter, replace, combine, convert and more in just a few clicks without any technical scripting skills
- Save the results in one of your connected systems
Understand the quality of your data to help you in data preparation and modeling

- Define quality rules through pre-defined or custom formats
- Define thresholds and rulebooks
- Run rules against datasets
- Build data quality dashboards
Define, govern and manage your metadata assets across enterprise systems

- Automatically index metadata based on content type and enabled text based search
- Use labels at dataset and column level to classify your data assets and use labels as filters in your data catalog
ML Scenario Manager

One single interface to see all AI models across the organization and their versioning in a unified workspace.
Modeler & Monitoring

- Rapidly develop complex data workflows as chains of operators, replacing coding and scripting with an intuitive graphical user interface
- Schedule, monitor and observe the executions of your workflows
Modeler - Pipelines

Extensible
- Standard operators
- Custom/Partner operators with the possibility to use as base Python, R, JavaScript, Node.js, Go, ABAP & more

Scalable
- Distributed
- Containerized

Re-Usable
Modeler - Orchestration

The heavy lifting / logic execution happens either:

- Internally, in the operator container
- Externally, in the connected system
Model Creation

Keep total flexibility for data scientists

- JupyterLab integration to ML Scenario Manager
- Use your preferred language to deal with data exploration and model creation and all the latest open-source libraries
- Leverage all HANA predictive libraries through the configurable operator, in an R or Python-like syntax or by direct scripting
- Consume SAP Leonardo functional services using the pre-defined operator
ML Scenario Manager & Modeler
Model Lifecycle Management

- Build pipelines for training and inference starting from scratch or using templates, and link them to the related ML scenario
- Retrieve model metrics and related artifacts from pipelines executions
- Deploy your models as RESTful APIs to make them available in enterprise applications
- Manage and monitor all of this centrally through the ML Scenario Manager
The new as-is: Predictive quality workflow in manufacturing plant
Early issue detection due to solution performance
- Reduce waste and rework due to higher accuracy of quality checks
What is AutoML?

- Having a **configurable toolbox** for every step in the **ML workflow**
- A **smart way of deciding** which tools to use

**Score:** 92%

Out-of-the-Box deployment via SAP Data Intelligence
Thank you

Elisa Masetti
elisa.masetti@sap.com