



DS + AI Accelerators

Lunch & Learn | August 3, 2023

Bryan Powell & Florencia Silveira

Data Scientists - Data Science Professional Services at Domo, Inc.

Davis Busteed

Data Scientist - RXA

Agenda

- Intro to Domo's DS + AI Accelerators
 - GitHub Repository
- Data Profiler
 - Background
- Data Anonymizer
 - Background & Demo
- Model Insights
 - Background & Demo

Domo's DS + AI Accelerators

- Accelerators
 - Tools to simplify, streamline, automate, & integrate different parts of the data management process.
 - Accelerators automate repetitive work to free Domo users' time to focus on insights and decision making.
 - Requires access to & use of Jupyter Workspaces
- Available at [domo.developer.com](https://developer.domo.com) and the Domo DS GitHub.
 - <https://developer.domo.com/portal/pg9t2vdbix5un-data-science-accelerators>
 - https://github.com/domoinc/domo-data-science-resources/tree/main/data_science_accelerators

Accelerator: Data Profiler

- Creates a profile of the input dataset on a Domo dashboard (HUGE time saver!)
- Conducts an automated exploratory data analysis, giving a first glance at the distribution and relationships between variables.
- Several cards are created and can be used or discarded at the user's discretion.



- Demo of the tool available at

<https://developer.domo.com/portal/pg9t2vdbix5un-data-science-accelerators#data-profiler>

Var1	Age	DistanceFromHome	MonthlyIncome	NumCompaniesWorked	PerformanceRating	RelationshipSatisfaction	StockOptionLevel	TotalWorkingYears
Correlation	Correlation	Correlation	Correlation	Correlation	Correlation	Correlation	Correlation	Correlation
Age	1.00	0.04	0.41	0.30	0.02	0.05	0.07	0.63
DistanceFromHome	0.04	1.00	-0.00	-0.02	0.02	0.00	0.08	0.01
MonthlyIncome	0.41	0.04	1.00	0.15	0.04	0.01	0.02	0.04
NumCompaniesWorked	0.30	-0.00	0.15	1.00	-0.02	0.05	0.03	-0.09
PerformanceRating	0.02	0.02	0.04	-0.02	1.00	-0.01	0.00	0.06
RelationshipSatisfaction	0.05	0.00	0.01	0.05	-0.01	1.00	-0.04	-0.01
StockOptionLevel	0.07	0.08	0.02	0.03	0.00	-0.04	1.00	0.02
TotalWorkingYears	0.63	0.04	0.69	0.26	0.06	0.02	0.02	1.00
WorkLifeBalance	0.01	-0.02	0.04	-0.00	-0.01	0.00	0.00	-0.01
YearsAtCompany	0.26	0.02	0.47	-0.11	0.04	0.00	0.02	0.04
YearsInCurrentRole	0.19	0.02	0.38	-0.09	0.04	-0.01	0.05	0.04
YearsSinceLastPromotion	0.19	0.02	0.33	-0.04	0.06	0.01	0.01	0.06

Accelerator: Data Anonymizer

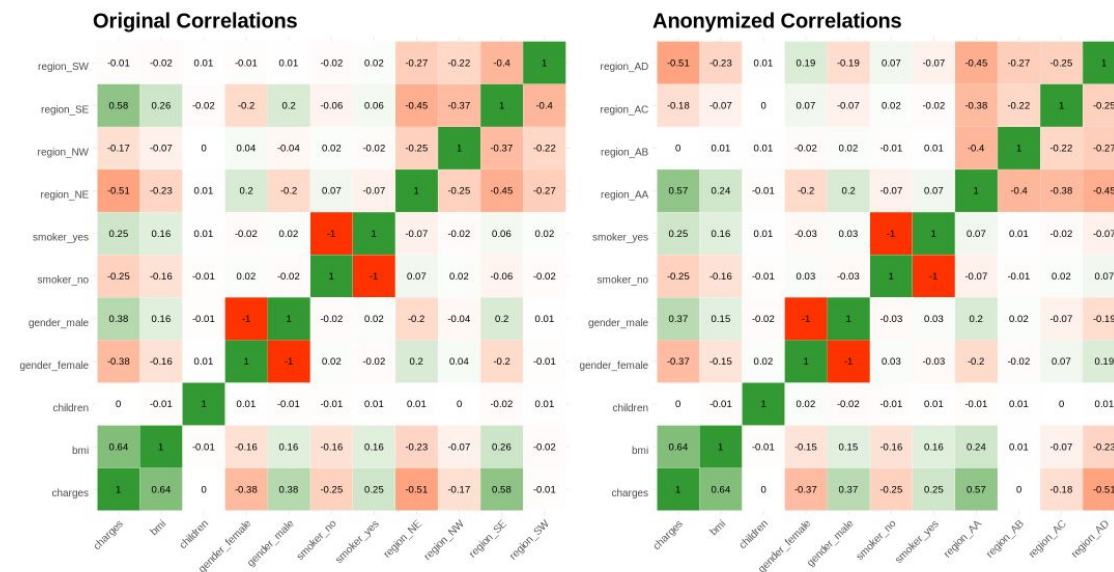
- Use case: You have sensitive data on an individual level that you would like to analyze on an aggregate level or use to train a model for prediction on future individuals.
- The output anonymized dataset maintains the aggregate characteristics of the original dataset without revealing information about individuals within the dataset.

A data.frame: 6 × 7

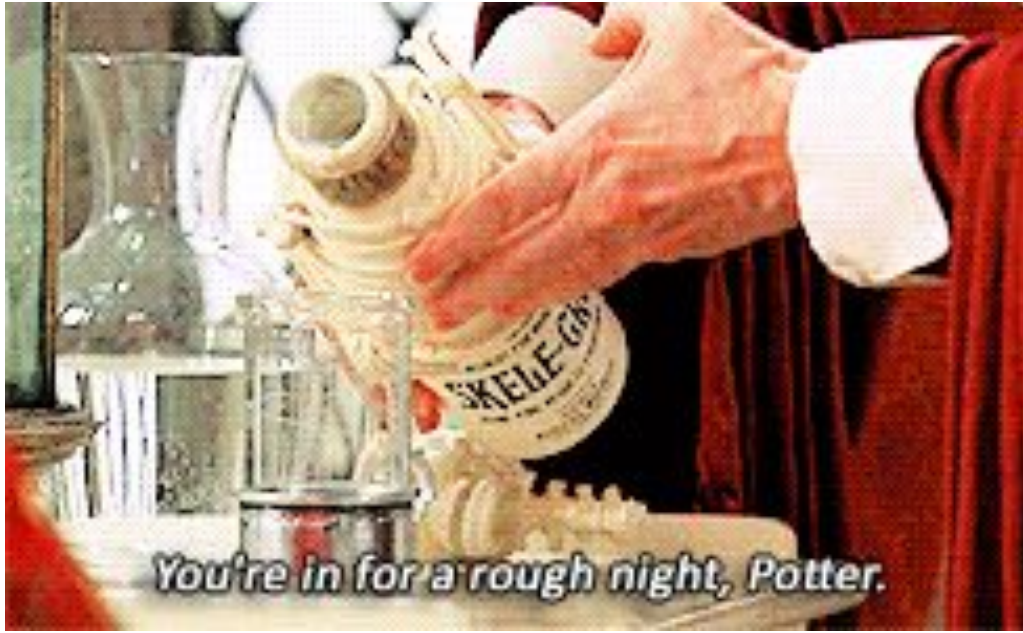
	charges	bmi	gender	smoker	region	children	last_visit
	<int>	<int>	<chr>	<chr>	<chr>	<int>	<chr>
1	33467	37	male	yes	SE	1	2017-07-23
2	5023	20	female	yes	NW	3	2020-03-30
3	6713	24	male	no	NE	4	2018-01-31
4	10056	21	male	no	SE	4	2021-06-02
5	18994	35	male	yes	SE	5	2017-03-05
6	13201	30	male	yes	SE	3	2019-09-13

A data.frame: 6 × 7

	charges	bmi	gender	smoker	region	children	last_visit
	<int>	<int>	<chr>	<chr>	<chr>	<int>	<chr>
1	17949	20	male	yes	AD	1	2017-12-05
2	20944	34	male	yes	AB	3	2018-04-01
3	24344	32	female	yes	AA	2	2020-01-05
4	12891	32	male	yes	AD	1	2020-05-10
5	34377	38	male	yes	AA	2	2017-11-11
6	17072	29	female	yes	AC	3	2017-06-17



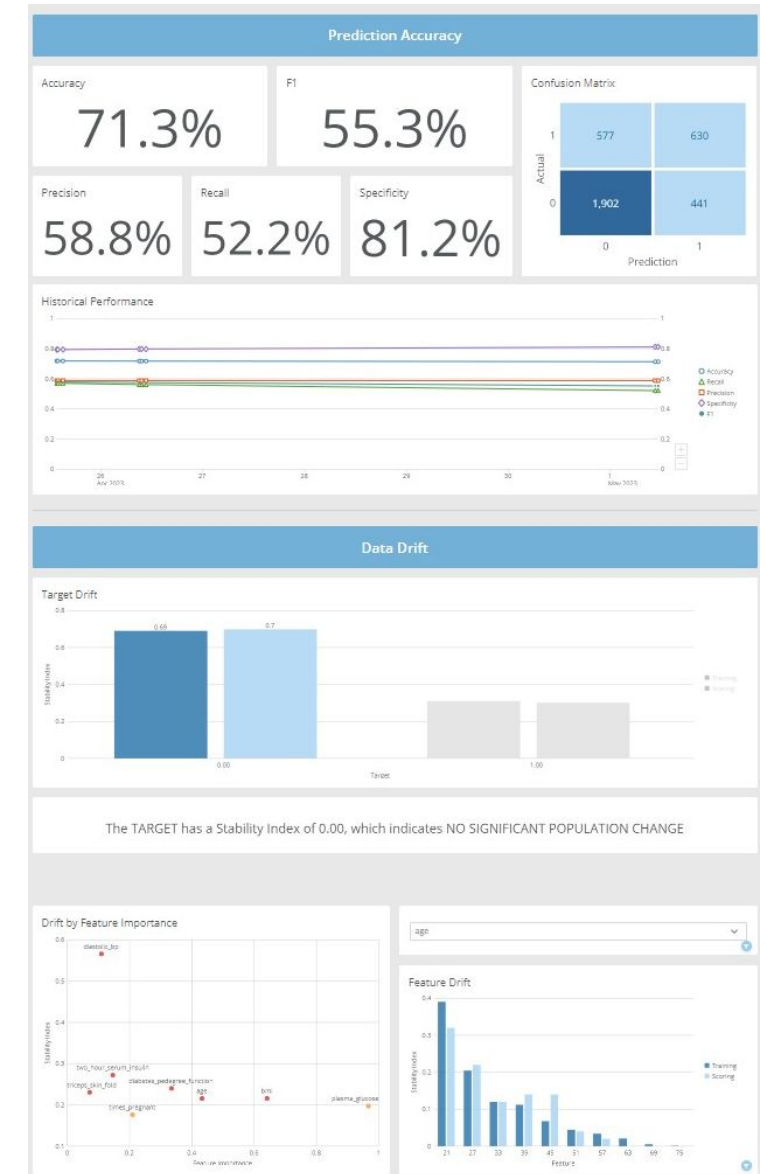
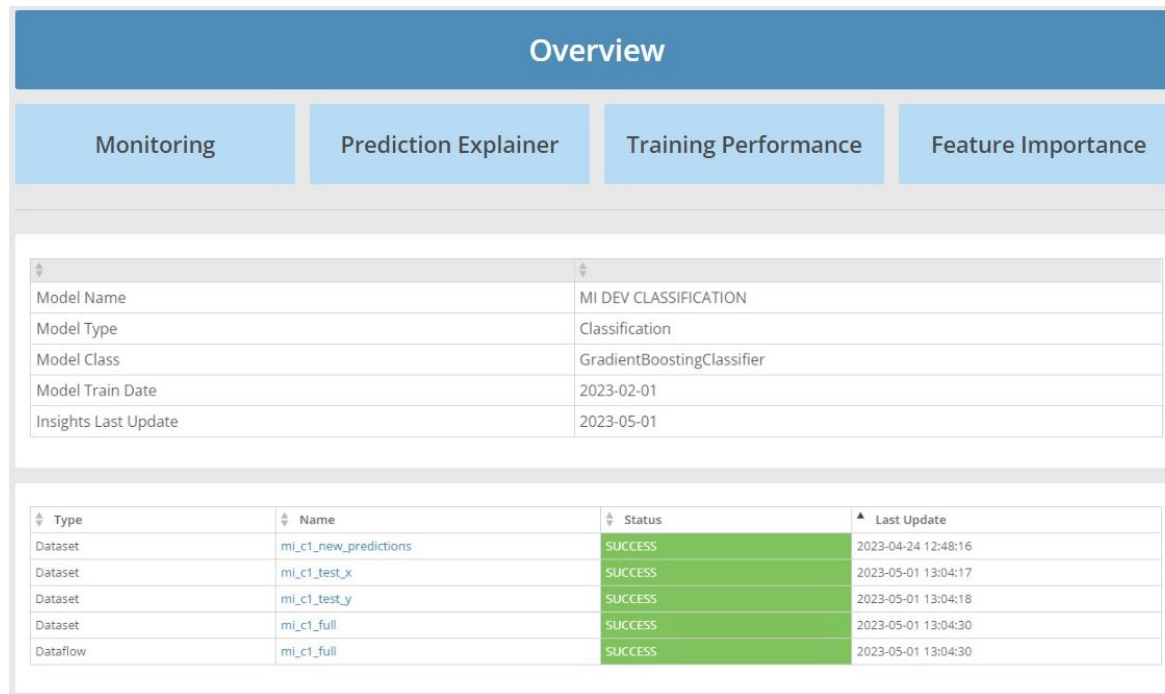
Data Anonymizer



- Demo Use Case
 - You, a Muggle data scientist, have been tasked with creating a model that predicts whether a Hogwarts student is at risk for diabetes.
 - You have training data available from past students, but to protect their privacy you must anonymize this data before training your model.

Accelerator: Model Insights

- Creates an entire dashboard that evaluates and monitors the performance of input data science models.
- The tool details training performance, live performance, model drift monitoring, model feature details, and individual model predictions.
- Compatible with Regression, Binary Classification and Multi-Classification Models.



Accelerator: Model Insights

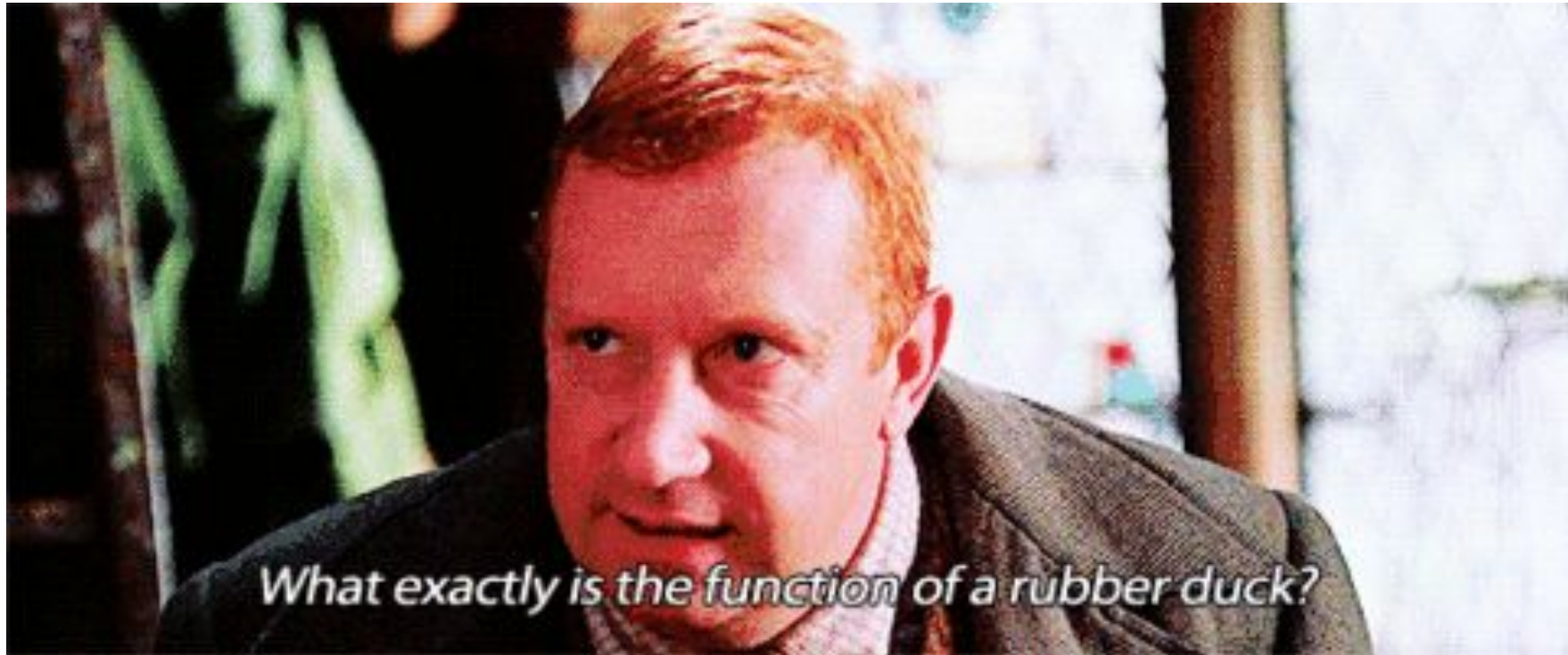
- Inputs:
 - Pickled (saved) model
 - Training data
 - Live data
 - Model predictions on live data
- Outputs:
 - ~30 cards organized neatly into 5 dashboards detailing:
 - Model pipeline health
 - Model drift and performance on live data
 - Training performance
 - Model feature details
 - Individual prediction details

Model Insights Demo



- 10 years ago, Divination Professor Sybill Trelawney foresaw that the Sorting Hat would be cursed, and unable to place students in their respective houses.

Model Insights Demo



- In preparation, Arthur Weasley (Muggle Studies expert and data science dabbler) suggested that incoming students take a subject aptitude exam prior to being sorted into their Hogwarts House.

Model Insights Demo

- Thanks to Mr. Weasley's fascination with all things Muggle, a training set with 10 years of incoming students' aptitude exams and the Sorting Hat's house decisions was compiled before the foretold curse fell.
- A model was built with this data and was used to sort students into their Hogwarts House last year.
- You, a Muggle data scientist, have been tasked with monitoring the performance of this model and deciding whether the model should be retrained.