

The Future of Data Quality: Harnessing Machine Learning to Enhance Investment Operations

Founded in 1997, Meradia provides strategic advisory and implementation services to the global investment management industry.

## PERSPECTIVE. PASSION. IMPACT.

## Investment Operations: Confirming Data Quality at Scale is Vital but Onerous





Simple/dumb rules go out of date quickly.



TIMELINESS

#### Let us take a step back and talk about ducks and radiologists!







From an early age we are trained to look at patterns and breaks in patterns





#### In Investment Operations We Often Look For Breaks In The Patterns

Month	Item	Beginning	Ending
Jan-2021	C0996154	14,545	14,350
Jan-2021	C4086728	83,828	92,565
Jan-2021	C3092123	91,741	105,700
Feb-2021	C5110777	5,025	5,869
Feb-2021	C7572301	16,480	17,441
Feb-2021	C4086728	73,102	92,565
Feb-2021	C3092123	84,722	96,875
Feb-2021	C8664805	67,747	67,717
Mar-2021	C0996154	20,365	27
Mar-2021	C5110777	16,437	19,078
Mar-2021	C8226992	28,612	30,348
Mar-2021	C4086728	85,827	92,565
Mar-2021	C3092123	75,398	88,675

Even in the raw data, egregious items are usually easy to spot.



#### We Use Rules To Identify Differences We Care About

Month	Item	Beginning	Ending	Diff.
Jan-2021	C0996154	14,545	14,350	-1.3%
Jan-2021	C4086728	83,828	92,565	10.4%
Jan-2021	C3092123	91,741	105,700	15.2%
Feb-2021	C5110777	5,025	5,869	16.8%
Feb-2021	C7572301	16,480	17,441	5.8%
Feb-2021	C4086728	73,102	92,565	26.6%
Feb-2021	C3092123	84,722	96,875	14.3%
Feb-2021	C8664805	67,747	67,717	0.0%
Mar-2021	C0996154	20,365	27	-99.9%
Mar-2021	C5110777	16,437	19,078	16.1%
Mar-2021	C8226992	28,612	30,348	6.1%
Mar-2021	C4086728	85,827	92,565	7.9%
Mar-2021	C3092123	75,398	88,675	17.6%

"Warn me whenever the absolute difference is greater than 25%"

Sometimes though, static rules are worse than no rules, generating **too many false positives.** 



In the 'traditional' approach we define rules and react to the results.



generate false positives.

The feedback loop to improve or edit the rules can have delays, friction, or is simply de prioritized.

Over time our data quality process decays and becomes inefficient

We do not operate in a static environment – external and internal factors drive changes and our process decays over time. We asked ourselves this – **can we build a better 'mousetrap'**? One that learns over time from the data and from the interactions with the team?



## Move From 'Rules Acting On Data' To 'Rules Derived From Data'



Our Aim is to Mimic Human Intuition Using Machine Learning Techniques



## There was a Machine Learning World Before ChatGPT and the Focus on AI



1: Quinlan, J. R. C4.5: Programs for Machine Learning. Morgan Kaufmann Publishers, 1993



ML and Gen AI are Different. Former Has Interesting Aspects to Offer. Conventional ML algorithms employed for Operational Efficiency Delivers Value.

	Front Office	Middle & Back Office
Machine Learning	Used since early 1980s. Algorithmic trading and security misprice identification are	Unstructured data processing coupled with workflow integration produces business benefits.
	dominant use cases. Simulation and back testing.	Potential to exploit conventional ML techniques in data quality.
Generative AI	Understand market sentiment. Summarize company and industry fundamentals. Investment banking securities issuance.	Portfolio commentary is dominant. Market narrative use cases are on the rise. Quality checks by scrubbing market data and employing synthesis techniques.

We employ a combination of **classical machine learning algorithms**, **statistical tests** and **data management techniques** to develop a novel model.



What does this mean in practice? A 10-month journey to develop a working prototype in Python that can generate useful results against test data sets.



Today: Able to unearth heuristics and deliver actionable results.



## 10 months in less than 5 minutes...how did we get to Review and Affirm?



We identify the data attributes (features) that we want the model to consider.

The algorithm can compare features, or combinations of features to fixed values.



The algorithm is trained against a known historical data set <u>alongside</u> the expected results.



#### Getting to Real Drivers Behind Data Quality..

#### Exception Status has systemic & manual updates

	Portfolio	Gross Return	Gross Return	Portfolio -	Standard		Market	Total	Net Return	Gross Return -	Yesterday -	Cash	
Date	Туре	Portfolio	Benchmark	Benchmark	Deviation	Flows	Value	Flows/MV %	Portfolio	Net Return	Today's Return	Return	Status
14-Apr	DERV	-8	-8	0	0.804673847	19	195	9.743589744	-8.017075781	0.017075781	2	0	Pass
14-Apr	EQ	0	-4	4	0.804673847	-4	51	7.843137255	-0.003214059	0.003214059	1	0	Pass
14-Apr	FI	-2	0	2	0.734846923	18	190	9.473684211	-2.006817417	0.006817417	4	0	Pass
14-Apr	DERV	3	-1	4	1.878163997	12	130	9.230769231	2.979725066	0.020274934	0	0	Pass
14-Apr	EQ	5	5	0	0.804673847	4	40	10	4.974886781	0.025113219	1	0	Pass
14-Apr	FI	-5	-2	3	0.5	5	62	8.064516129	-5.044132738	0.044132738	1	0	Pass
14-Apr	DERV	-2	-2	0	0.804673847	-2	20	10	-2.150585467	0.150585467	2	0	Pass
15-Apr	EQ	0	0	0	1.878163997	4	43	9.302325581	-0.078760132	0.078760132	1	0	Pass
15-Apr	FI	0	-4	4	0.5	-7	82	8.536585366	-0.096124814	0.096124814	1	0	Pass
15-Apr	DERV	-7	-10	3	0.836660027	4	44	9.090909091	-7.170847257	0.170847257	1	0	Pass
15-Apr	EQ	-2	2	4	0.804673847	-4	47	8.510638298	-2.002720413	0.002720413	2	0	Pass
15-Apr	FI	1	-3	4	1.417744688	-2	29	6.896551724	0.911979737	0.088020263	3	0	Pass
15-Apr	DERV	4	7	3	1.933907961	1	17	5.882352941	3.81281407	0.18718593	1	0	Pass
15-Apr	EQ	3	0	3	0.804673847	-13	135	9.62962963	2.905125802	0.094874198	2	0	Pass
15-Apr	FI	-4	-1	3	0.5	-19	204	9.31372549	-4.042400134	0.042400134	1	0	Pass
15-Apr	DERV	-1	2	3	0.836660027	5	58	8.620689655	-1.076713489	0.076713489	1	0	Pass
16-Apr	EQ	3	1	2	1.946792233	5	61	8.196721311	2.932123711	0.067876289	3	0	Pass
16-Apr	FI	-1	-5	4	0.5	4	42	9.523809524	-1.069231249	0.069231249	1	0	Pass
16-Apr	DERV	-6	-10	4	0.804673847	19	194	9.793814433	-6.0628372	0.0628372	1	0	Pass
16-Apr	EQ	-1	-4	3	0.836660027	12	134	8.955223881	-1.058909756	0.058909756	1	0	Pass
16-Apr	FI	-3	-2	1	1.503329638	9	99	9.090909091	-3.038449857	0.038449857	4	0	Pass
16-Apr	DERV	9	5	4	1.961504525	10	112	8.928571429	8.979127647	0.020872353	5	0	Pass
16-Apr	EQ	6	7	1	0.836660027	14	142	9.85915493	5.919362111	0.080637889	3	0	Fail
16-Apr	FI	-1	3	4	0.5	-14	150	9.333333333	-1.032485347	0.032485347	3	-0.01752	Fail
16-Apr	DERV	-6	7	13	0.804673847	16	168	9.523809524	-6.072425645	0.072425645	5	0	Fail
17-Apr	EQ	3	-2	5	1.946792233	-17	174	9.770114943	2.905105405	0.094894595	0	0	Fail
17-Apr	FI	1	0	1	0.5	14	139	10.07194245	0.912279552	0.087720448	2	0	Fail
17-Apr	DERV			0	0.804673847	-2	31	6.451612903	-0.065274427	0.065274427	6	0	Fail



ML Model

#### **Identify Strength of Existing Rules\***

Standard deviation of Portfolio returns (52%)

Total Flows/MV > 10(27%)

Portfolio Return – Benchmark Return > 8 (18%)

Gross Return - Net Return < 0 (13%)

**Eliminate Redundant Rules** 

Portfolio return > 10

Gross Return = NULL

#### **Recommend Additional Rules**

Portfolio Return – Benchmark Return > 5

Scenario Dependent Monthly/Quarterly Rules

\* Sum greater than 100 as an exception can fail due to multiple rules



# **Pandemic, Inflation Spike Periods**

How do fixed rules stand up?



What does mean for you? Moving **from a fixed rules engine** to an engine that will **learn from the interactions with the team**.

The algorithm generates **multiple rule variants**.

Each rule variant is accompanied by a **measure of its validity**.



Operations can pick new rules to replace, or supplement, the old during every iteration, **training and improving the machine every day**.



Machine Learning + Human Intelligence Solves Scaling Challenges Augmented Learning Enables Continuous Improvement

Instead of decaying over time, our **data quality process can improve every day.** 



Fewer false positives, less wasted time.



# Questions?

I will be around for the remainder of the conference and can tell you what the 10 differences were between the two pictures!





#### Meradia

119 North High Street West Chester, PA 19380 Phone 610-738-7787 info@meradia.com www.meradia.com

## PERSPECTIVE. PASSION. IMPACT.

©2025 Meradia. All rights reserved.