LEARNING TO RANK

A Machine Learning Approach to Rank Assets

- What is Learning to Rank (LTR)?
- How to use Learning to Rank for trading strategies
- Results
- Next Steps

Context

Lot of Asset Classes

Bonds
Interest rates
Cryptpo Currencies
Currencies
Equity
Commodities

Extract

Apply ML

Capture complex patterns

Data Intensive Problem → Machine Learning

Lot of Datas

Price & Volume
Fundamental
Alternative
....

1.000 stocks x daily data x 20 years of history

→ **5.040.000 lines**

Roots of Learning to Rank

Documents	Characteristics
Rock USA Céline Dion The Rolling Stones Red Hot Chili Peppers	• Lyrics • Nationality • Number of hits • …
Dogs Groomer Cat Groomer Quentin's Groomer Top Dog's	• Website Title • Location • Customer reviews • …

Ranking of Documents

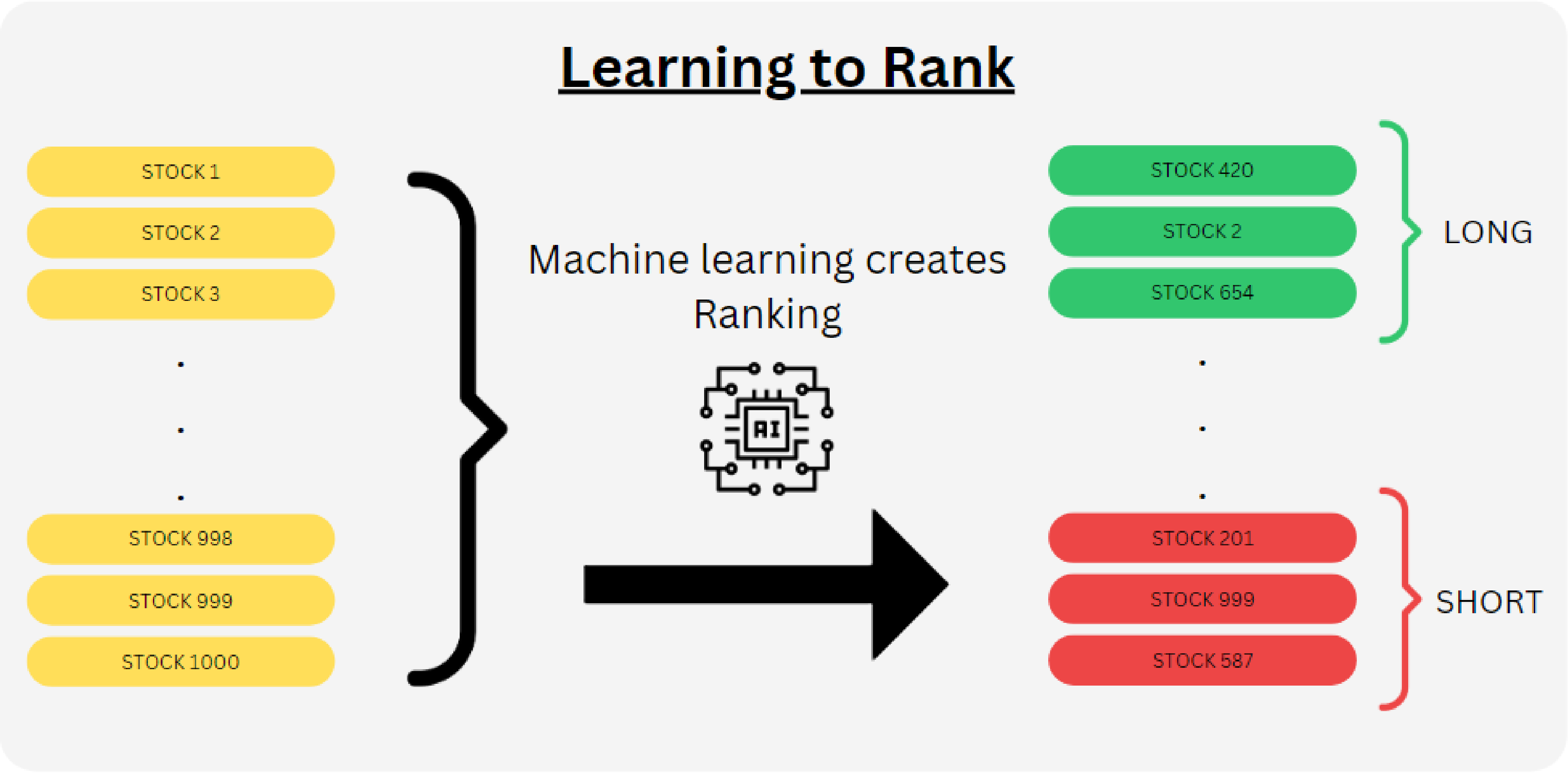
Rock USA

- Red Hot Chili Peppers
- The Rolling Stones
- Céline Dion

Dogs Groomer

- Top Dog's Groomer
- Quentin's Groomer
- Cat Groomer

Learning to Rank in Finance



Trading Strategy Steps

Data Collection and Processing → **Train Predictive Models** → **Assets Selection and Portfolio Construction**

Data Collection and Processing

- Collect and clean/check data
- Create input features
- Data transformations
- Data exploration

Train Predictive Models

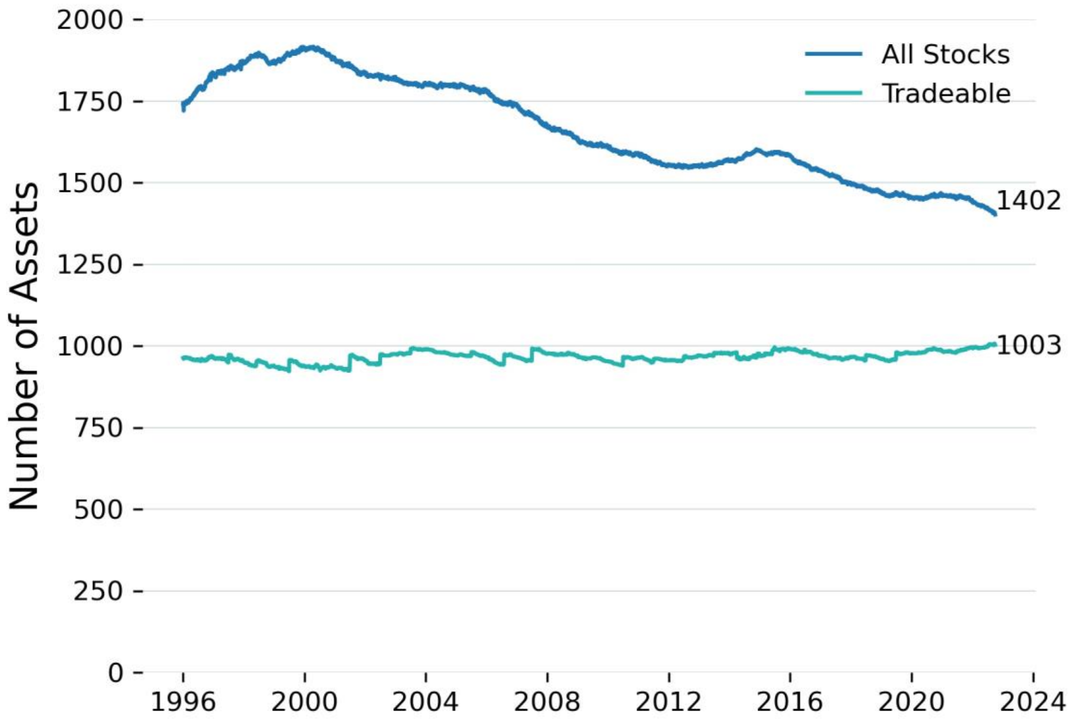
- Train predictive models to get predictions

Assets Selection and Portfolio Construction

- Construct portfolios using the predictions
- Position sizing

Data Description

- Universe: US Equities **from 01/01/1995 – 04/10/2022**
- Daily pricing data for **3250 stocks:**
  - Open, High, Low, Close, Volume, Market Cap
- **Price and volume derived features** over several periods
- **Over 12.250.000 rows** and over **100 columns**
- **Tradeable:** personal criteria



Data Transformation and Preprocessing

Raw Data

Create Predictors

Examples:

- Return
- Sharpe ratio
- Volume
- Technical indicator

Create Targets

Examples:

- Forward Return
- Forward Sharpe ratio (SR)
- …

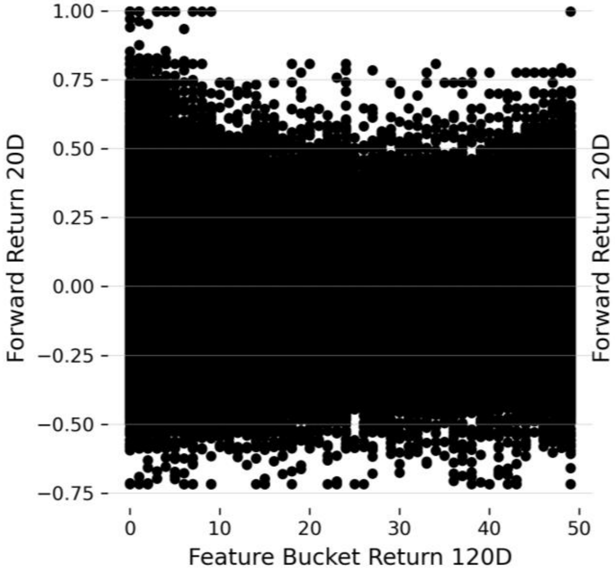
Data Preprocessing

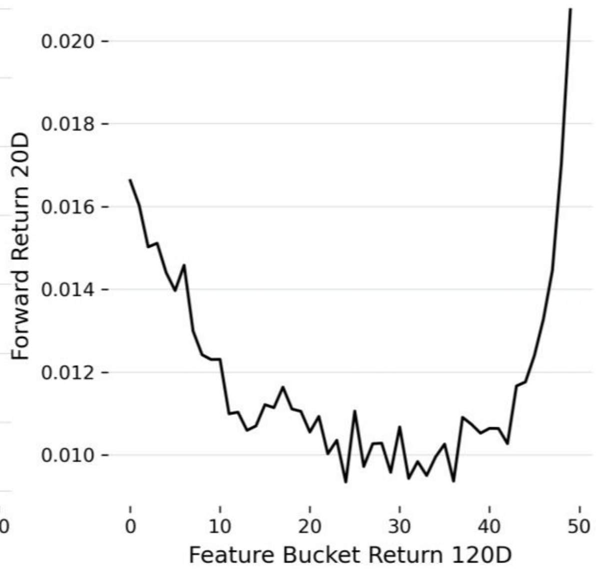
- Cross-sectional ranking
- Handling outliers
- Handling missing values

- Features computed on several horizons
- Missing values: cross sectional imputation with the mean
- Normalization: cross-sectional ranking
  - Automatically handles outliers

Data Exploration - Nonlinearities and Interactions

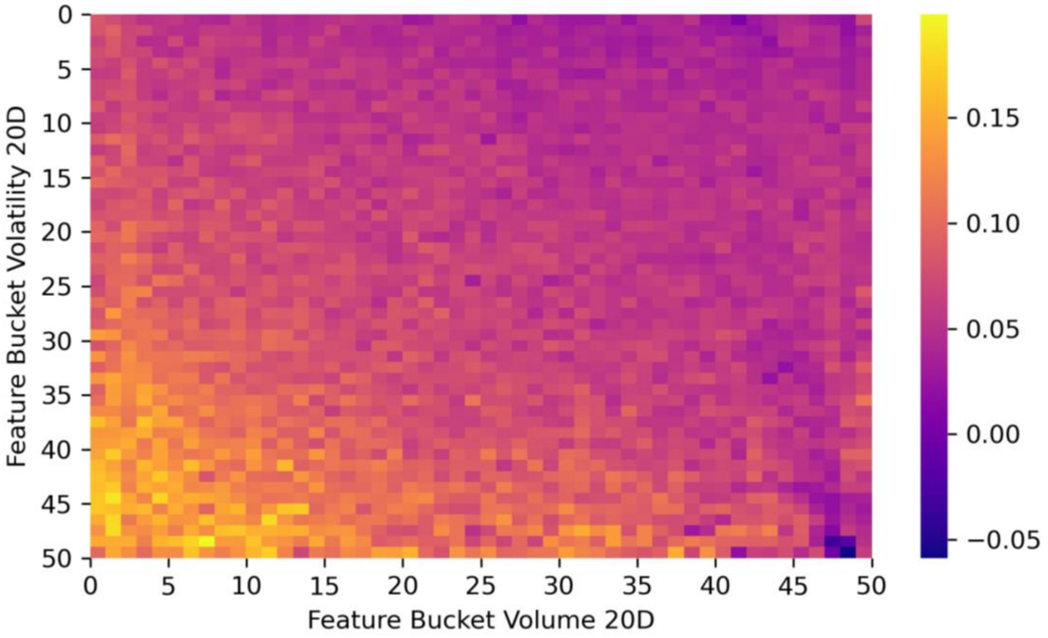
Nonlinearity





Stocks are ranked in N buckets regarding the past 120D return and plot shows the value of the average 20D forward return (right) for each of them

Interactions



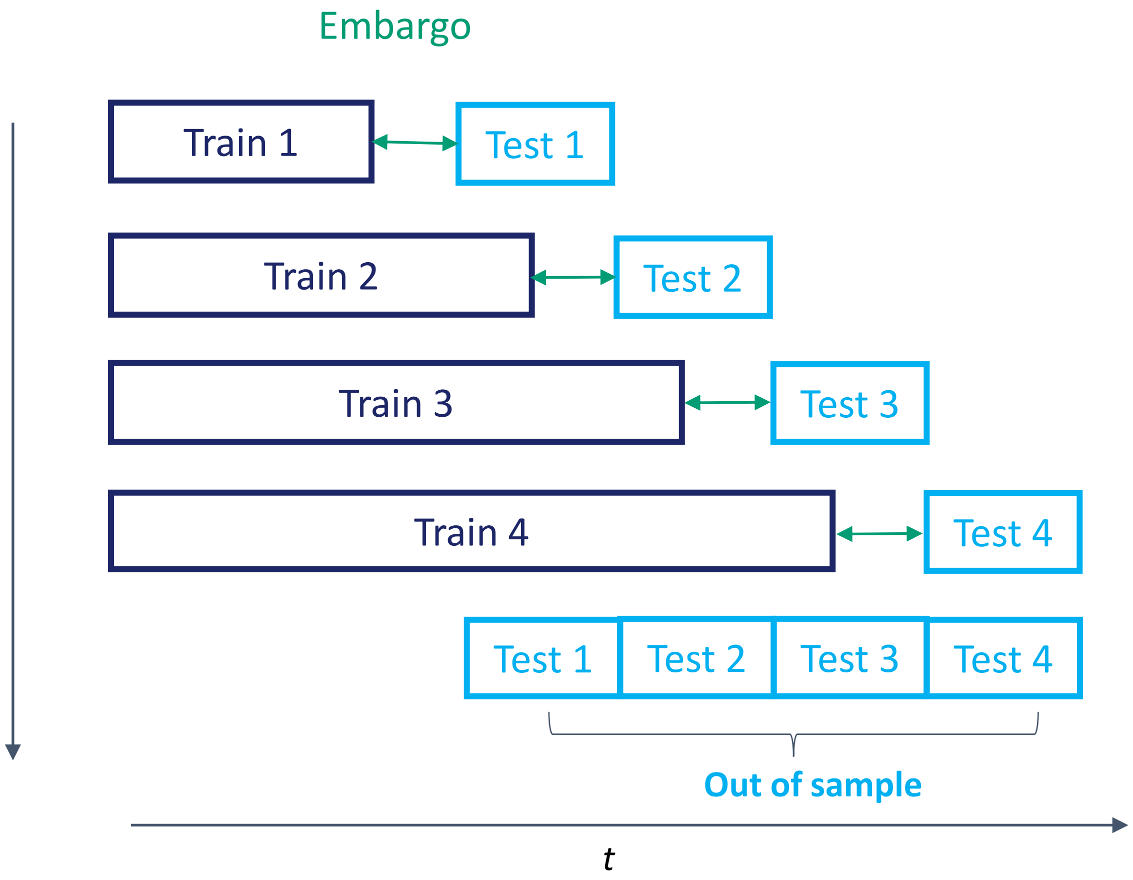
Stocks are ranked in N buckets regarding the past 20D volume (x-axis) and their past 20D volatility (y-axis) and plot shows (depending on the color) the value of the average 20D forward return for each of them

Time-Series Walk Forward Backtesting

Backtest Parameters

- **Minimum Training Period**: 1 Year
- **Test Period:** 1 Year
- **Embargo:** horizon of the target
  - Ex: Forward *SR* 10 D → Embargo ≥ 10 D

Expanding Version



Back to machine learning ranking results
(if you come from slide 12)

Portfolios Creation - Idea

Out of sample

Test 1	Test 2	Test 3	Test 4

$\underbrace{\hphantom{Test 1 Test 2 Test 3 Test 4}}$

Predicted Data for 1 Rebalancing period

Date	Asset ID	Prediction SR 10D	Portfolio
D_1	Asset 1	4/6	2
D_1	Asset 2	6/6	3 (LONG)
D_1	Asset 3	1/6	1 (SHORT)
D_1	Asset 4	2/6	1 (SHORT)
D_1	Asset 5	3/6	2
D_1	Asset 6	5/6	3 (LONG)

- Idea is to backtest every portfolios as a long one, and see that all P&Ls are also continuously well ranked

Backtesting Setup

Models Presented

- **Benchmark:**
  - Russell 1000
  - One-Factor momentum strategy based on the return over several horizons
- **Learning to Rank Models (LTR)**
  - Multiple Linear Regression (MLR)
  - Regression Trees (Reg Trees)
  - No Hyper parameter tuning

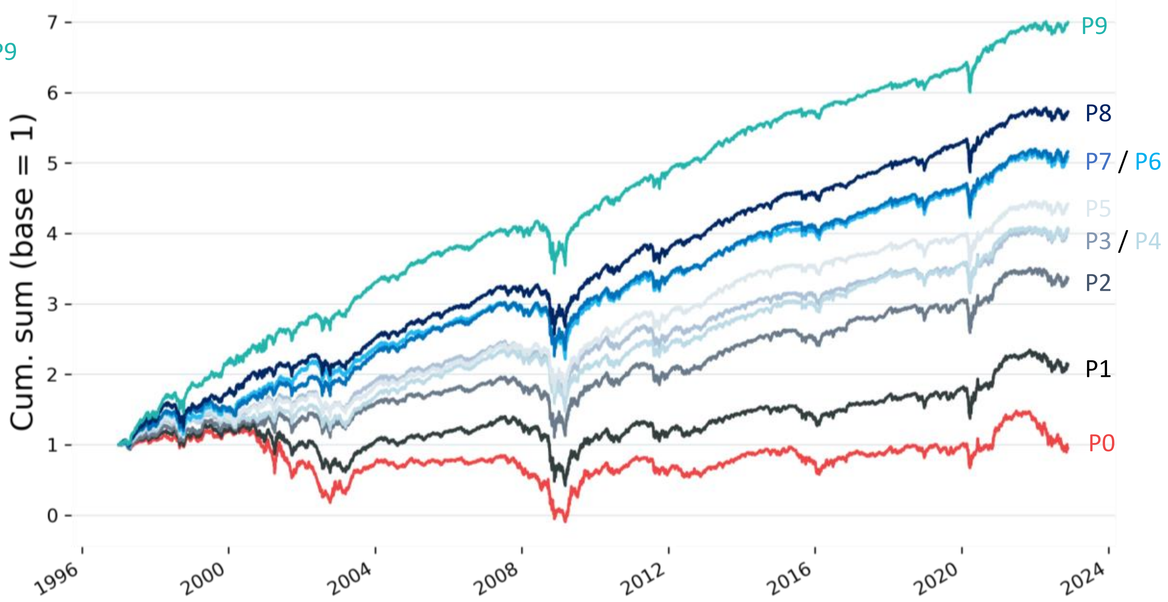
Configurations

- **Universe:** US Equities
- **Prediction:** *relative* 10 day forward Sharp Ratio (*SR)*
- **Expanding walk forward:**
  - 1996 - 2022
- **Default Number of Portfolio:** 10
  - ~ 100 stocks per portfolio
- **Rebalancing Period and Allocation process:**
  - Russell 1000: bi-monthly and equal weighted
  - One-Factor momentum: monthly and equal weighted
  - LTR models: bi-monthly and rank-vol weighting
  - Max weight per stock 0.025

Traditional vs Supervised Portfolios

Momentum Strategy Portfolios Ranking

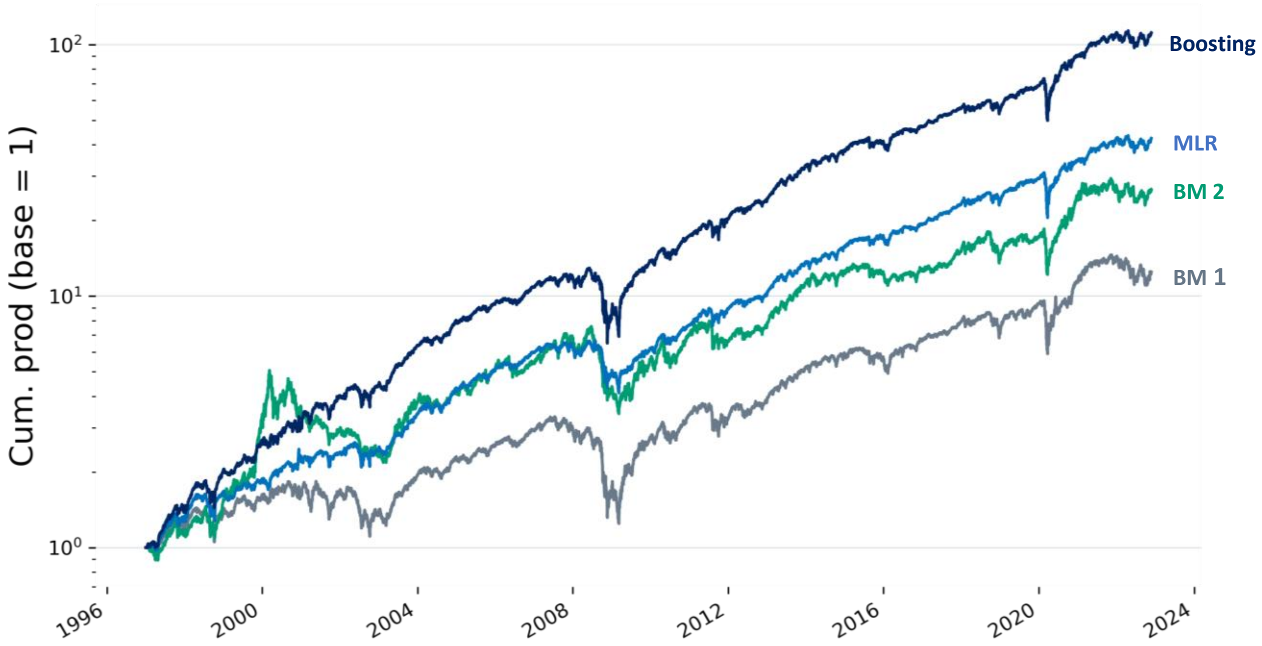
Machine Learning Portfolios Ranking



Reminder portfolios construction

LTR Outperforms Benchmarks

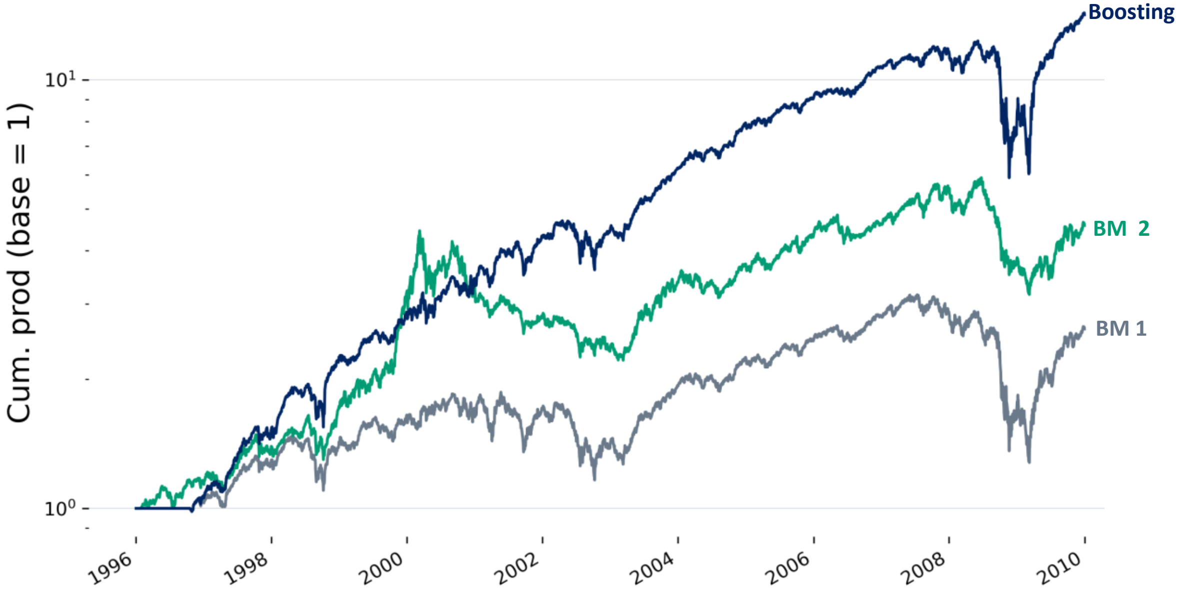
Models Performances without Fees



Strategy	BM 1	BM 2	MLR	Boosting
Return (%)	10.76	13.96	15.54	**19.97**
Volatility (%)	20.8	23.22	15.62	**16.98**
Sharpe Ratio	0.52	0.60	0.99	**1.17**
Max DD (%)	62.43	57.02	37.96	**49.41**
Max TUW (Days)	865	1400	404	**345**

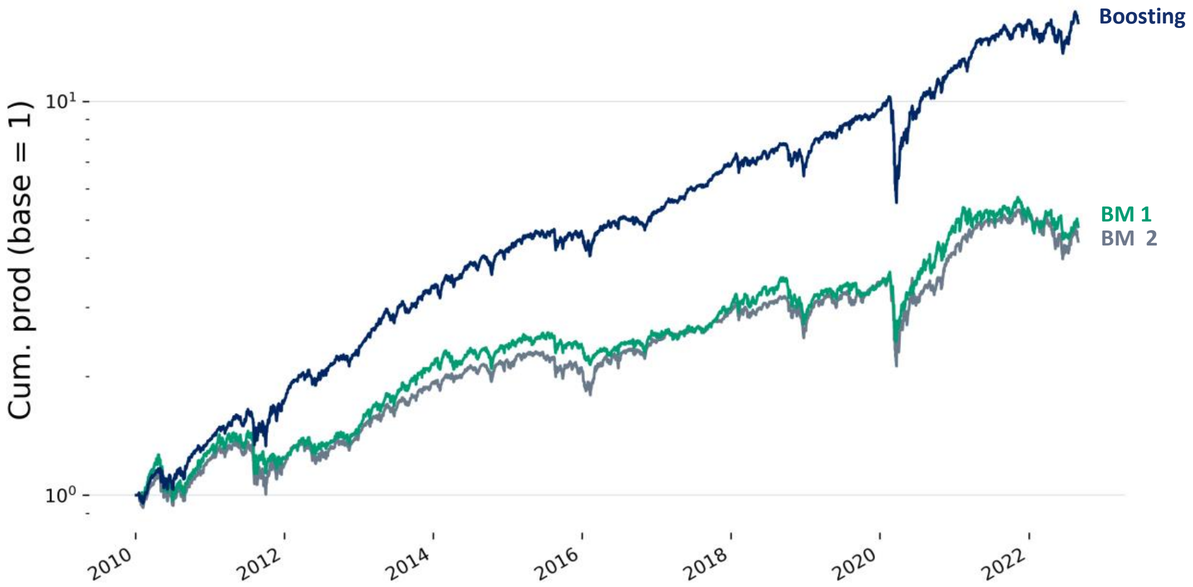
Consistent Performance over Time

Performance 1995 - 2009



Strategy	Return (%)	Volatility (%)	SR	Max DD (%)	Max TUW (Days)
Boosting	**21,29**	**20.54**	**1.04**	**53.12**	**292**
BM 1	7,55	21,83	0,35	63,01	633
BM 2	13,93	25,45	0.55	59.92	1479

Performance 2009 - 2022



Strategy	Return (%)	Volatility (%)	SR	Max DD (%)	Max TUW (Days)
Boosting	**17.18**	**14.14**	**1.21**	**34.87**	**153**
BM 1	12,10	19,37	0.62	39.42	297
BM 2	13,45	20.41	0.72	34.06	364

Easily Survives Transaction fees

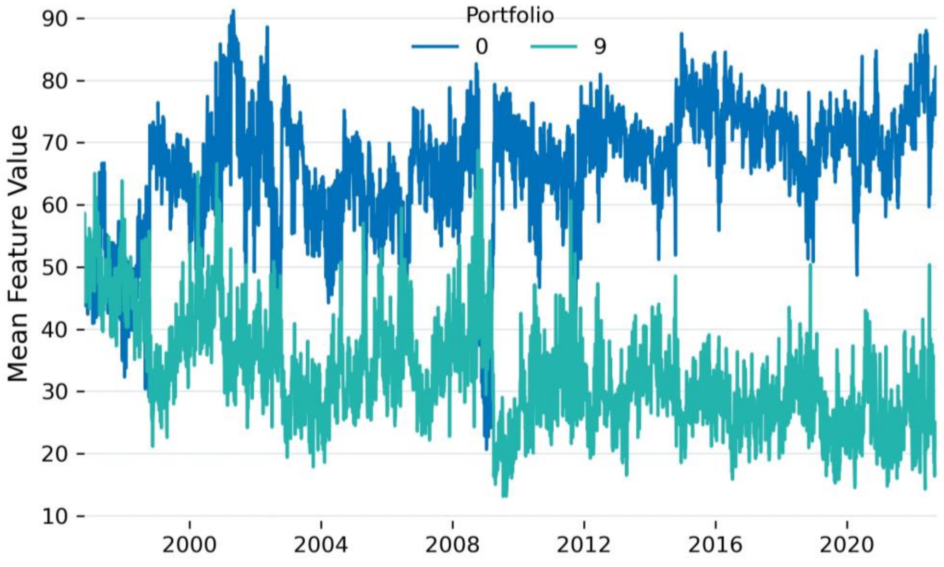
Performance with 5BP's Fees (F)

Strategy (Fees)	BM 1	BM 2	BM 2 (F)	Reg Trees	Reg Trees (F)
Return (%)	10.21	13.47	12.80	**19.97**	18.39
Volatility (%)	21.11	23.46	23.46	**16.98**	16.98
Sharpe Ratio	0.48	0.57	0.55	**1.17**	1.08
Max DD (%)	62.43	57.02	57.79	**49.41**	50.12
Max TUW (Days)	865	1400	1466	**345**	365

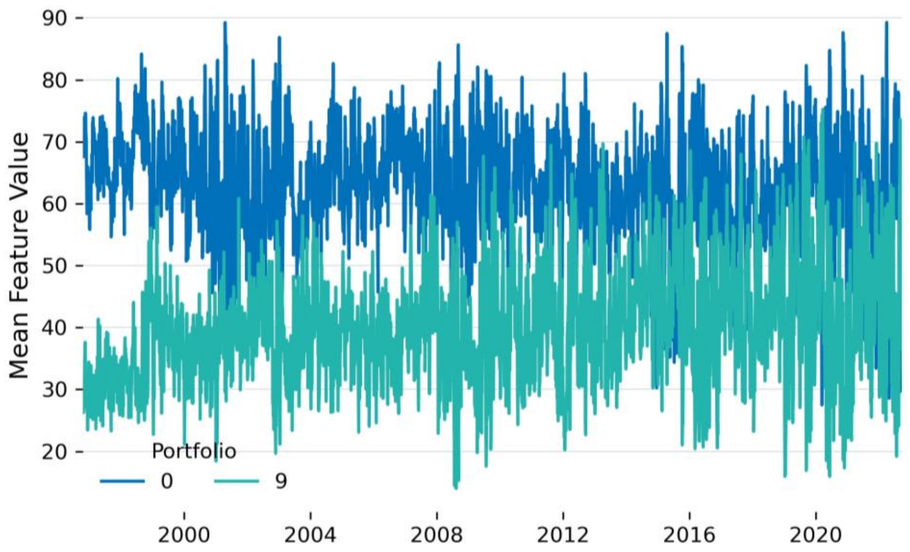
Model Interpretation – A first impression

Represent the mean of the past 10D volatility (left) and return (right) at time t, of the best (9) and worst (0) portfolio

Feature Volatility 10 Days



Feature Return 10 Days



Key Takeaways

Strong consistent performance after transaction costs

- Signal is consistent over time
- More return
- Lower risk in terms of volatility and drawdown
- **Returns: 19,97% p.a. / Sharp Ration: 1,17**

Multifunctional and Generalizable

- Applicable on other targets (returns, volatility…)
- Applicable to other universes and asset classes (European, Asian… market)
- Intuitive to incorporate new predictive features or add more assets

Next Steps

- Start paper trading
- Work on a Long-short portfolio strategy
- Test on other universes
  - e.g. create a universe composed of US, EU and Asian market and target a trading universe composed of 3000 to 5000 stocks each rebalancing date
- Addition of other data sources to involve new interactions and non-linearity
  - e.g. add fundamental data
- Combine multiple models
  - e.g. model trained on several periods, model trained on several target
- Model understanding
  - SHAP values and other approaches