Application in Finance II Topological Data Analysis

Machine-Learning-based Market Crash Early Indicator and How TDA Can Boost the Performance at High Cut-offs

Content

1. The Problem 1.0
2. The Results 1.0
3. The Data
4. The Model
5. The Problem 2.0
6. Recap on TDA Intro
7. The Results 2.0
8. The Results in details

The Problem 1.0 - Motivation

- Market moves in cycles: spikes followed by crashes, vice versa
- There are handful of “crash” indicators summarised by practitioners from experience:
  - Yield Spreads: 2-10 year, CDS
  - Commodity and Rates: Gold and Treasury Bonds as flight-to-safety
  - Volatility spikes
  - Sectorial Rotations
- In the AI/ML era, leveraging the expanded amount of data and computing power, Financial Machine Learning is gathering momentum to augment our insights in understanding market crashes

The Problem 1.0

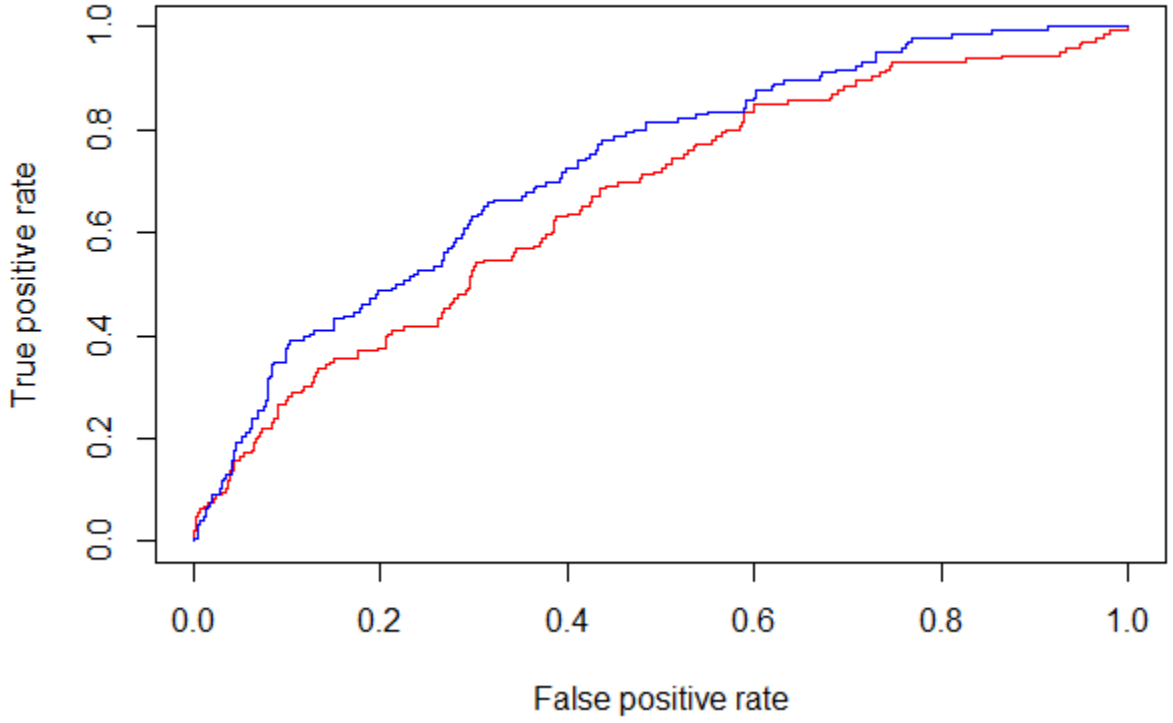
- Based on the public data, can we develop an **ML-powered market crash indicator** to quantify the risk of market crashes in a given forecasting window (e.g. 2w), in terms of probability?

- Something has been done:

1. "Predicting stock market crashes" – Roman Moser, Jan 2019
2. "Forecasting stock market crisis events using deep and statistical machine learning techniques" - Sotirios P. Chatzis et al. June 2018

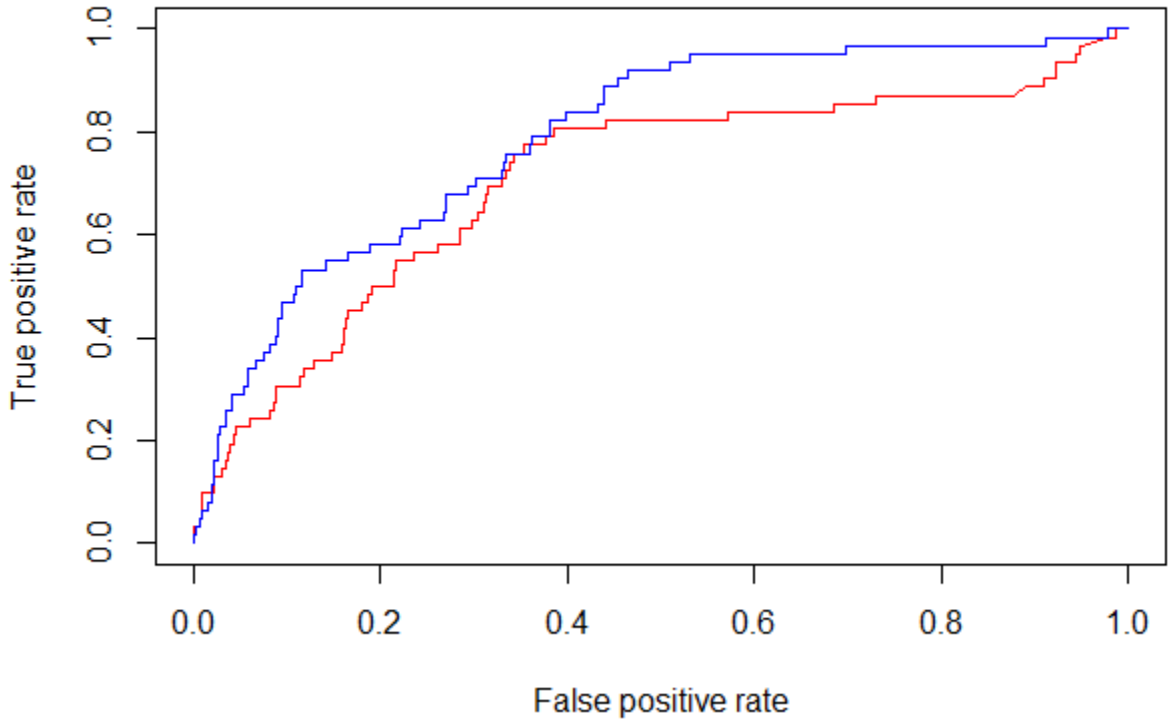
The Results 1.0

AUC Score (on Hang Seng Index down > 5% in 2w) = 72.8%



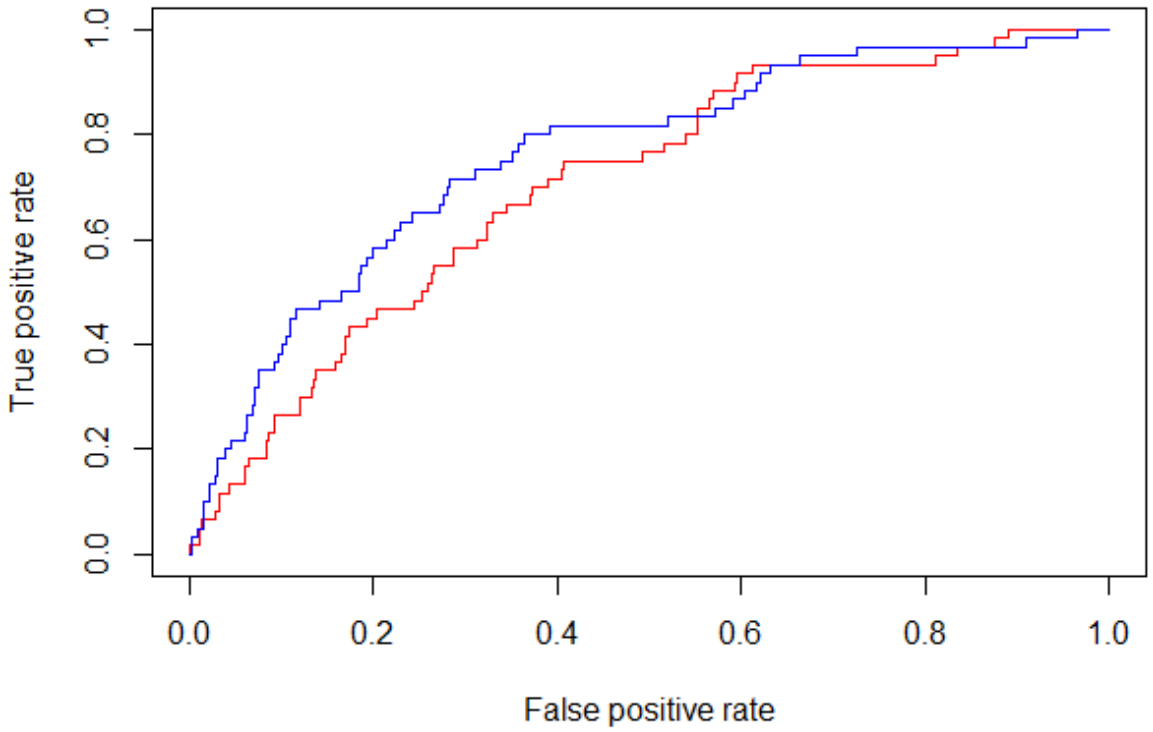
The Results 1.0

AUC Score (on S&P 500 down > 5% in 2w) = 79.0%



The Results 1.0

AUC Score (on FSTE 100 down > 5% in 2w) = 76.1%



The Results 1.0

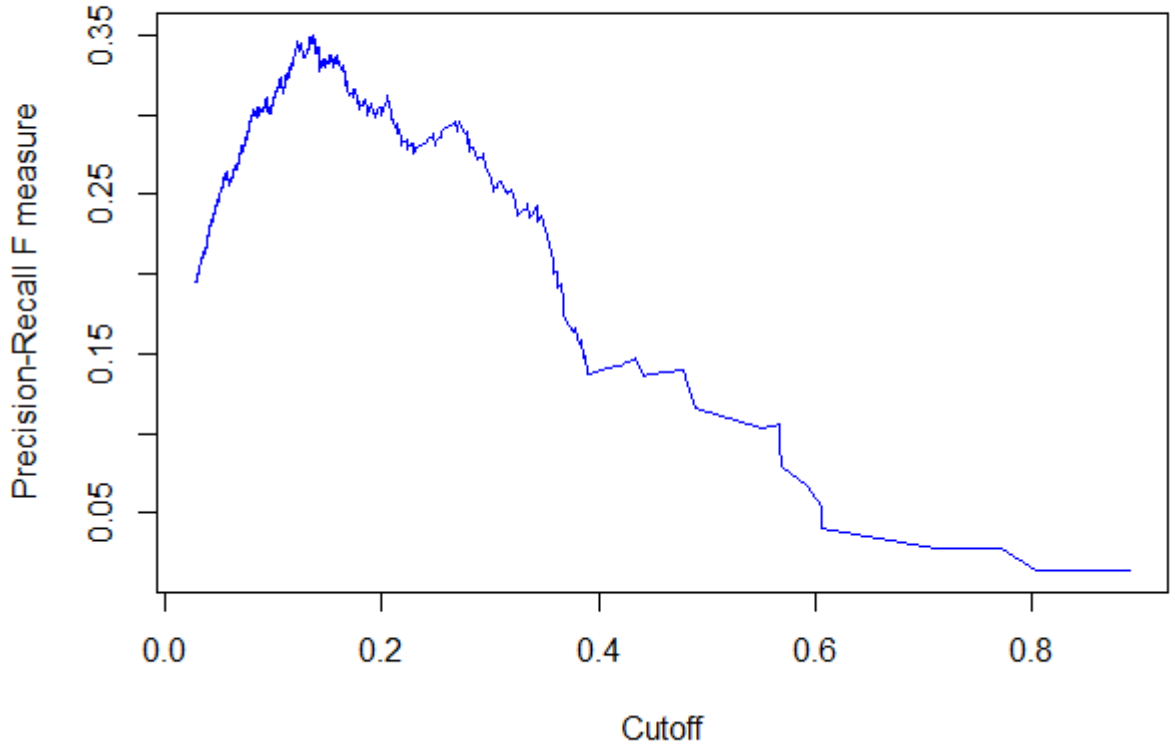
- Comparable to the results in the “Chatzis et al” paper (2nd reference paper above):

Glob20	Logit	CART	RF	SVM	NN	XGBOOST	MXNET
AUROC	0.630	0.654	0.739	0.708	0.677	0.743	0.783

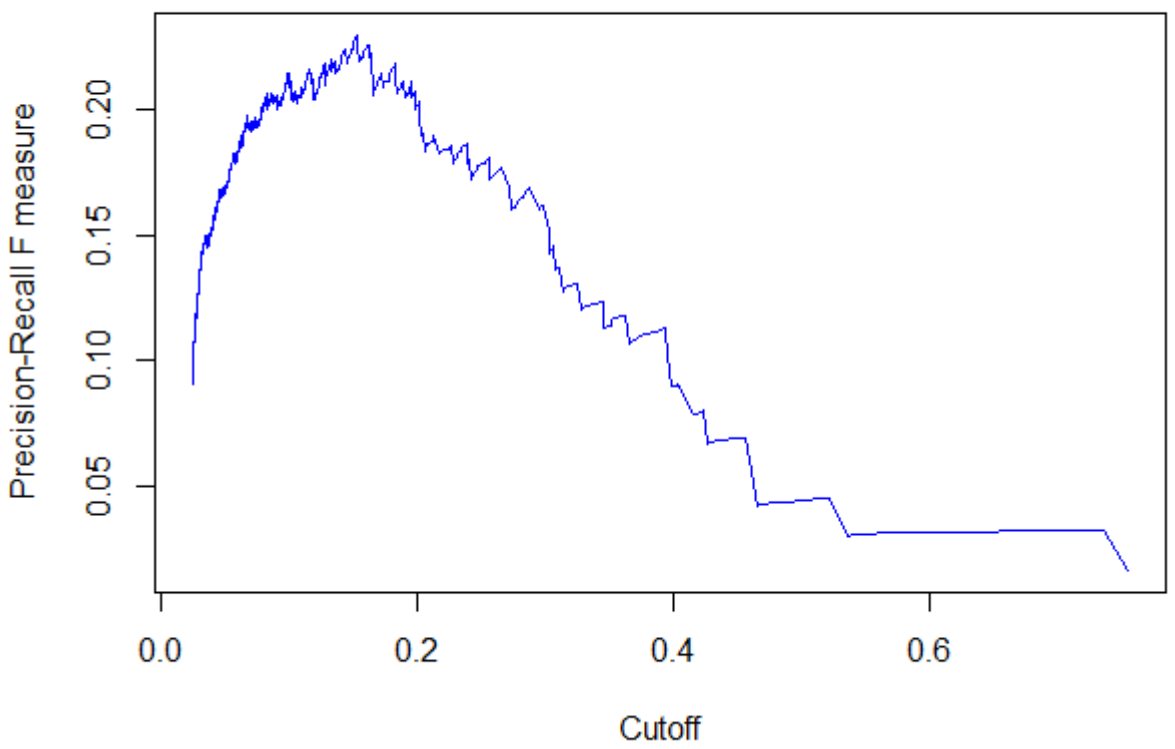
- However, “Financial Machine Learning” and “Market Crashes” are also biased to **false positives** at the high cut-offs

The Results 1.0

F1 vs. Cutoff (on HSI down > 5% in 2w)



F1 vs. Cutoff (on S&P 500 down > 5% in 2w)



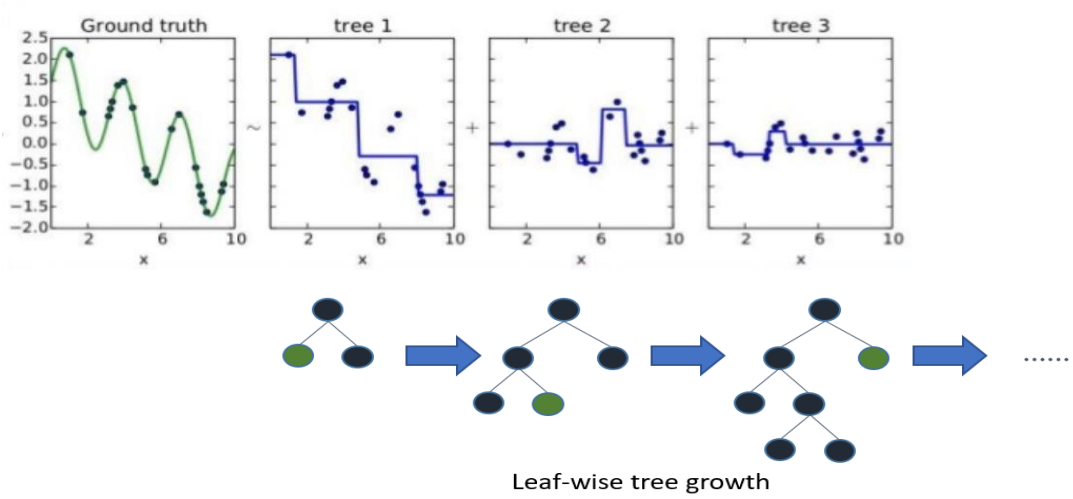
The Data

- 35 years of weekly price (covering major crashes since 1983)
- Asset types covering: Equity, FI, FX, Commodities (& Fama French Factors)
- Parsimonious approach to Feature Engineering (→ c. 10 features/dimensions):
    1. Return/Change/Momentum features
    2. Spread features
    3. Volatility features
- Meta parameters:
    1. forecast window = 2w
    2. TH = 5% (i.e. Y=1 for chosen index drop > 5%, Y=0 otherwise)
    3. training batch window = 26w
    4. rolling step size = 1w

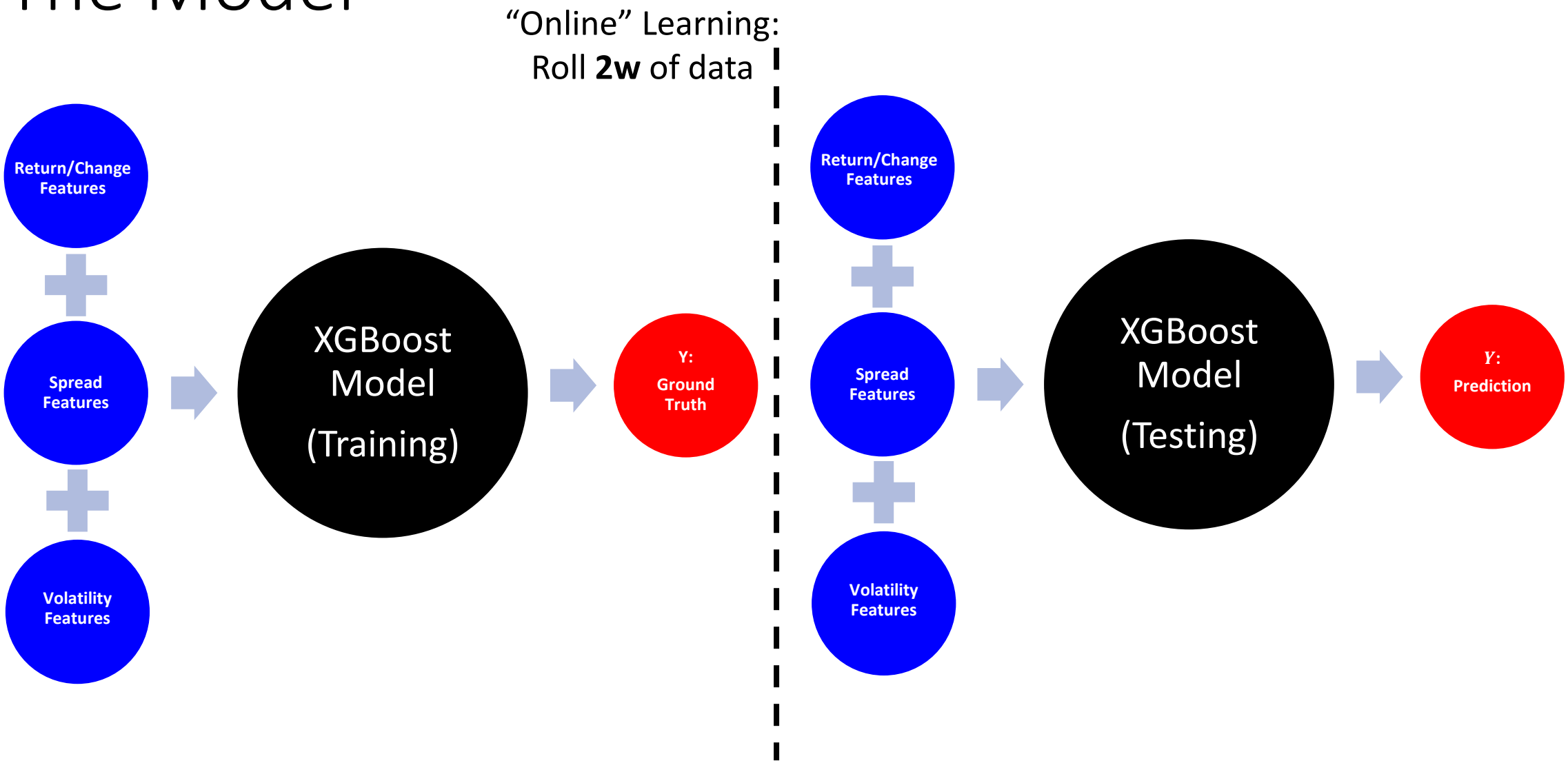
→ c. **3000** data points in total

The Model

- Best Results vs. SVM vs. CNN vs. Regression
- Tree-based
  - Transparent and whitebox as regression vs. deep neural networks
  - Free from normalisation/standardisation of the features vs. regression
- Kagglers' love (used by 17 out of 29 winners)



The Model


"Online" Learning:
Roll **2w** of data
Return/Change Features
Spread Features
Volatility Features
XGBoost Model (Training)
Y: Ground Truth
Return/Change Features
Spread Features
Volatility Features
XGBoost Model (Testing)
Y: Prediction

The Problem 2.0

- How do we improve the model performance at high cutoffs?

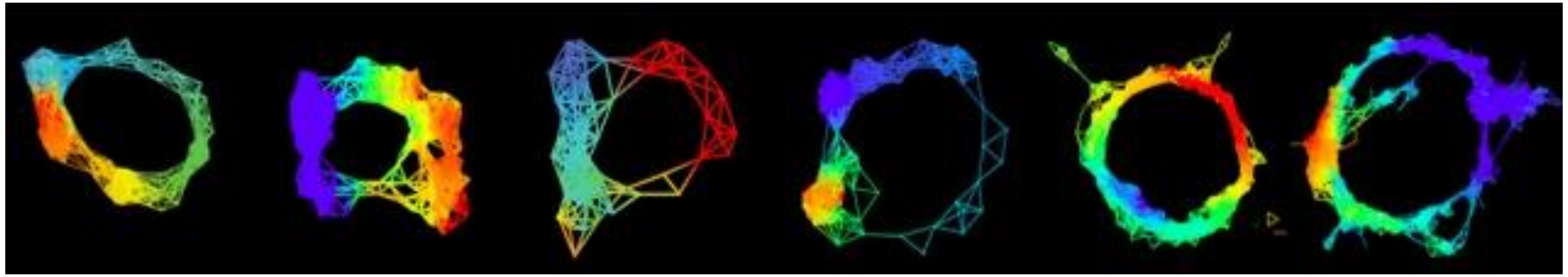
- Beyond all the conventional features/market factors, are we missing anything actually matters?

Recap on Topological Data Analysis

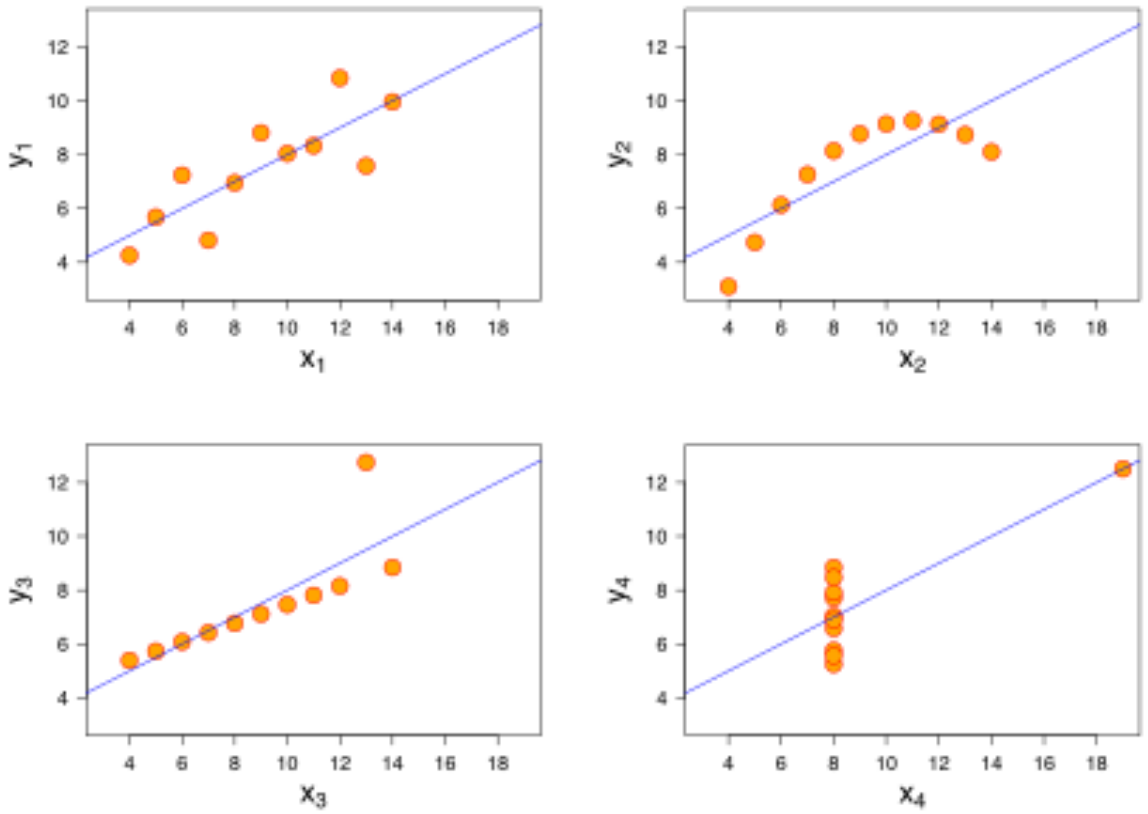
Topologists perhaps are best defined as the people with “disability” to tell doughnuts from mugs or vice versa:

Recap on Topological Data Analysis

"*Data has shapes, shape matters*." - Ayasdi

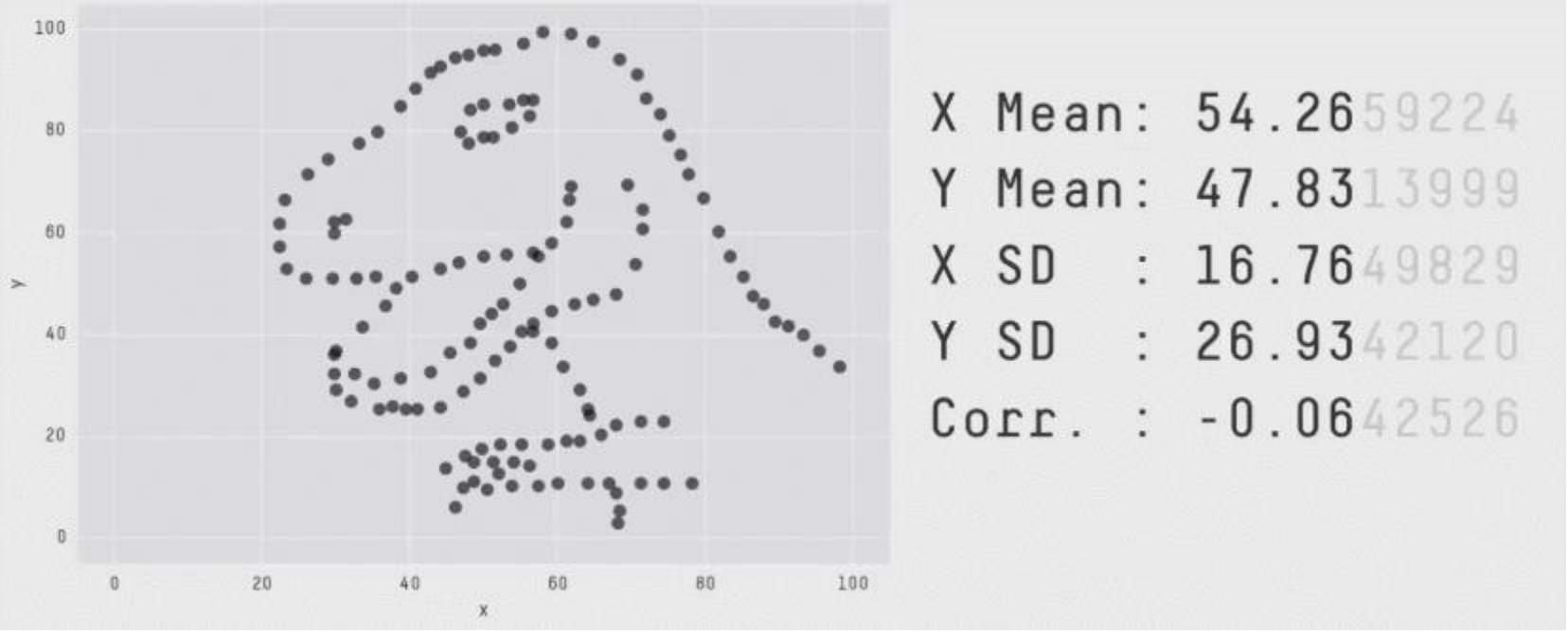


Recap on Topological Data Analysis



Property	Value
μ(X) =	9
σ^2(X) =	11
μ(Y) =	7.5
σ^2(Y) =	4.125
Corr(X, Y) =	0.816
Linear Reg	y = 3 + 0.5 x

Recap on Topological Data Analysis

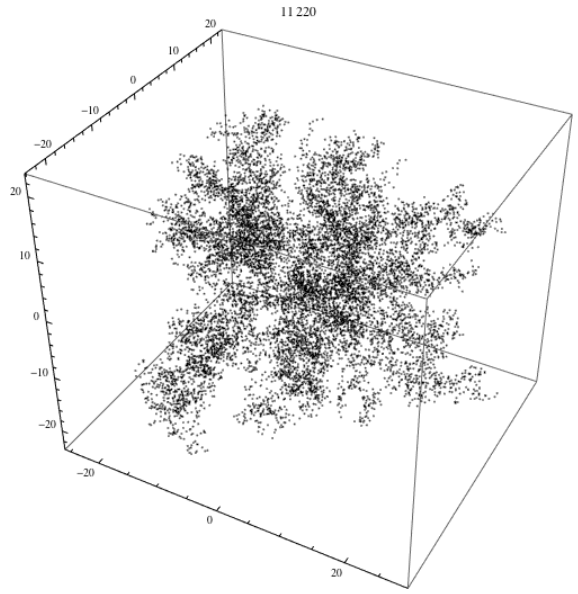


How does TDA work?

Data → Point Cloud → Complex

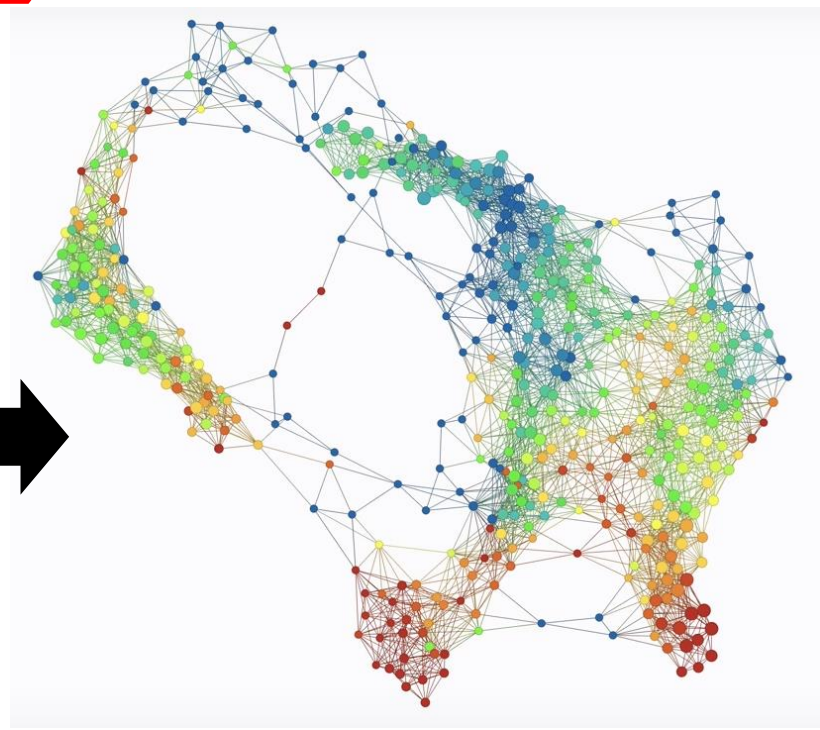
1

2



Vectorisation/
Feature Engineering
→ Space Embedding

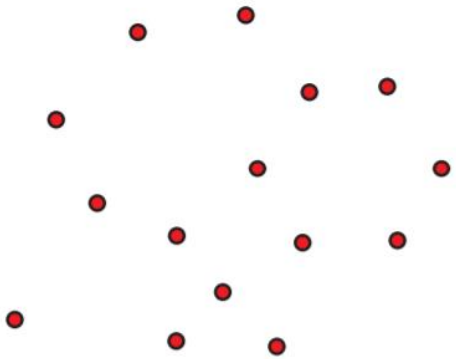
3

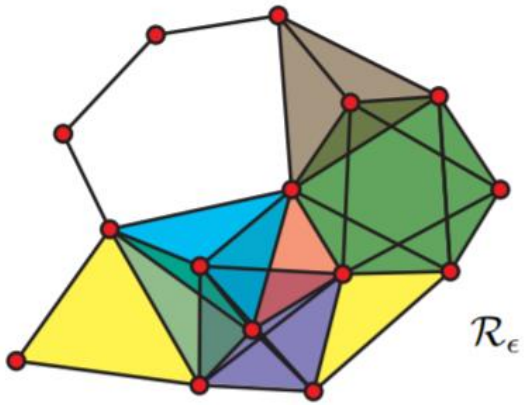


Metric → Structure → Complex

How does TDA work?

Point Cloud → ε → Complex

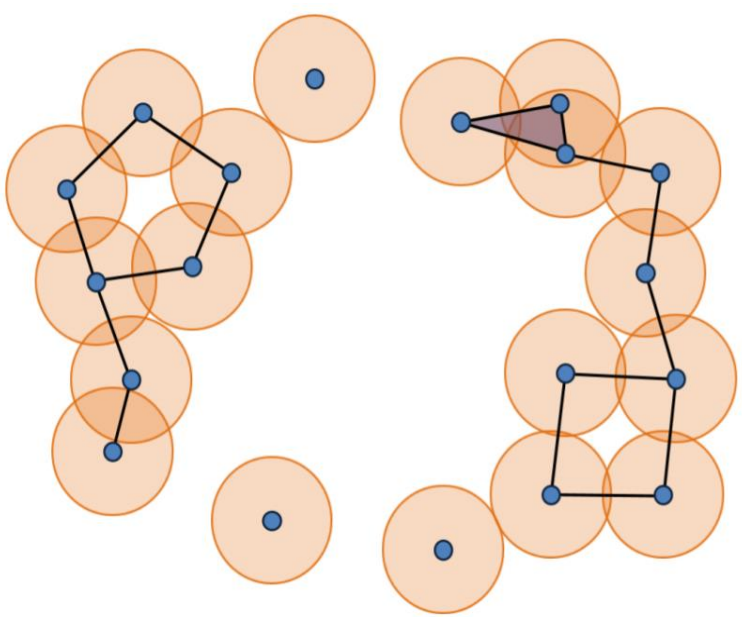


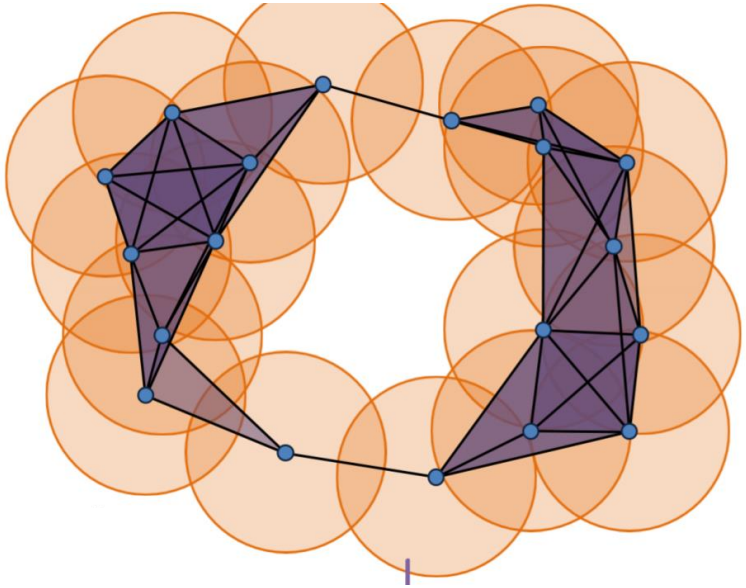


Question: which “ε” ?

How does TDA work?

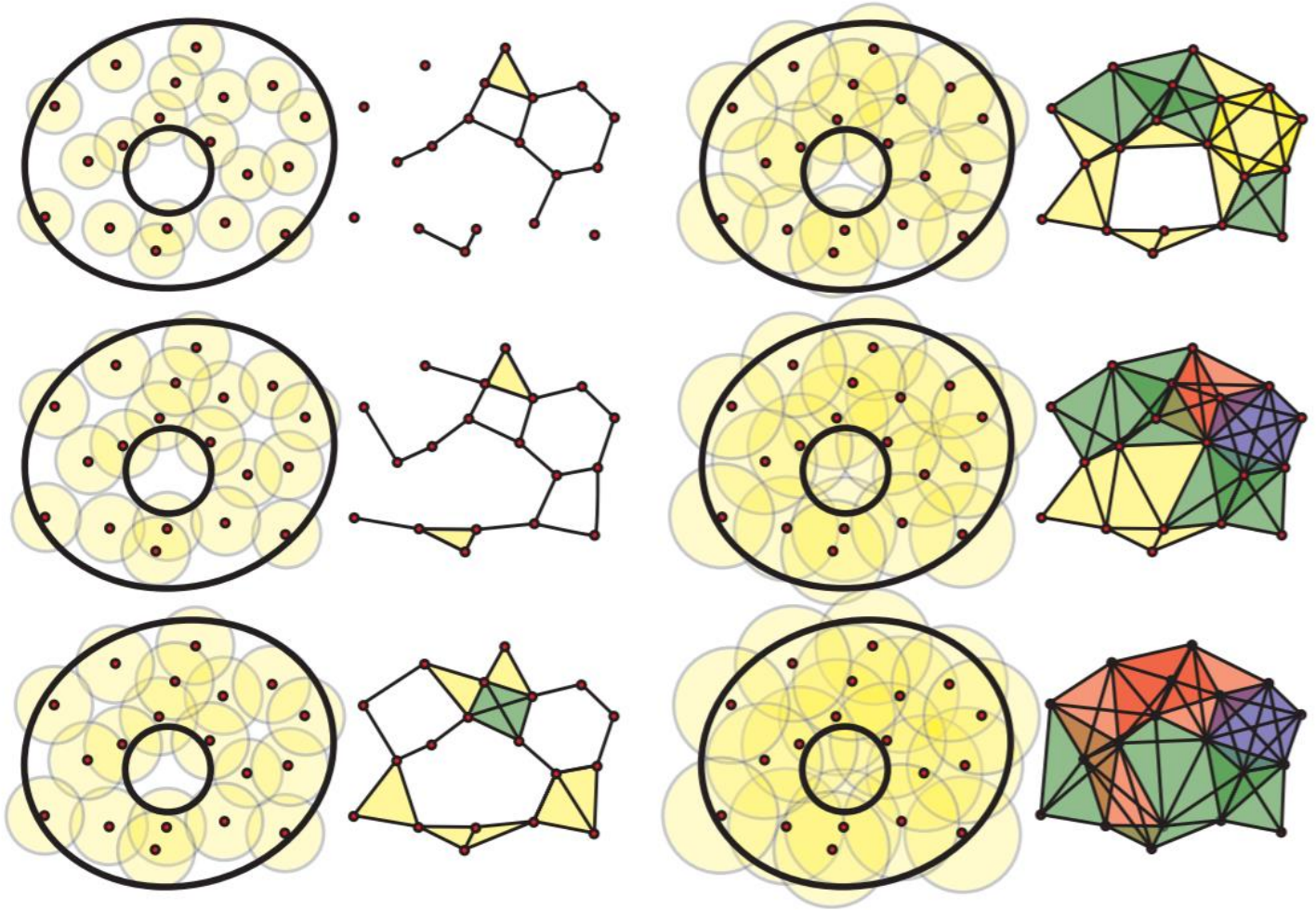
Which “ε” ?





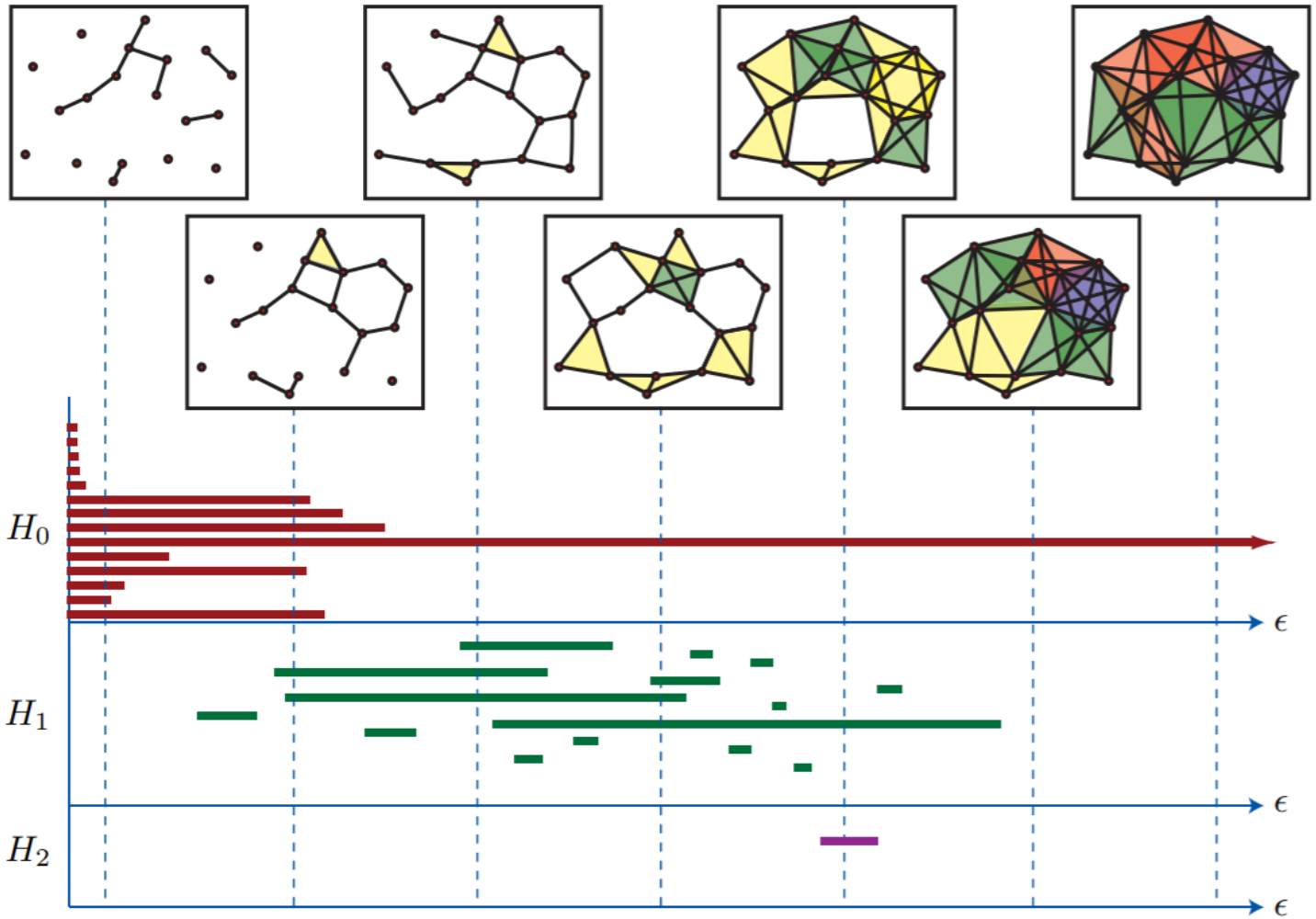
How does TDA work?

Barcode



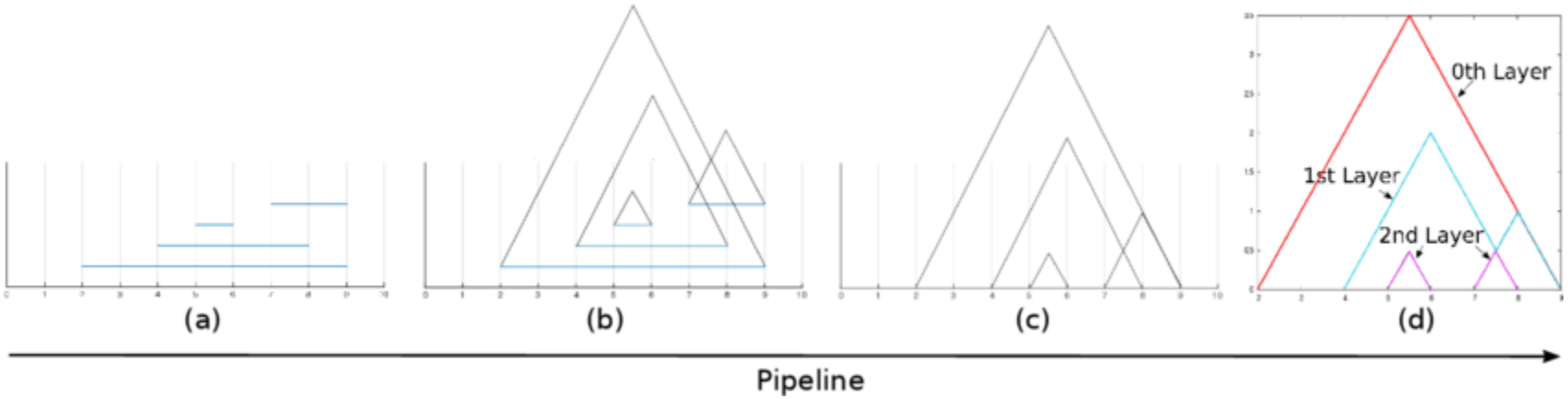
How does TDA work?

Barcode



How does TDA work?

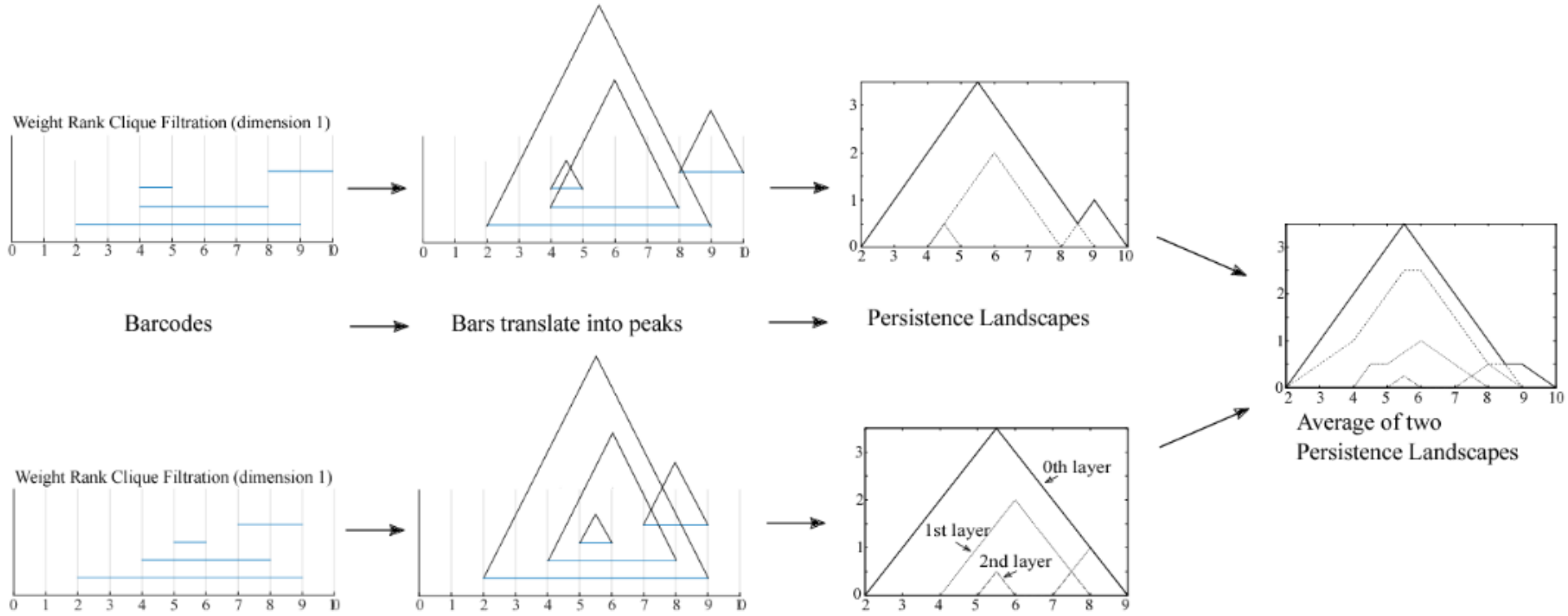
Barcode → persistence landscape



Why persistence landscape (instead of Barcode for ML)?

How does TDA work?

1. We can do statistics on it!



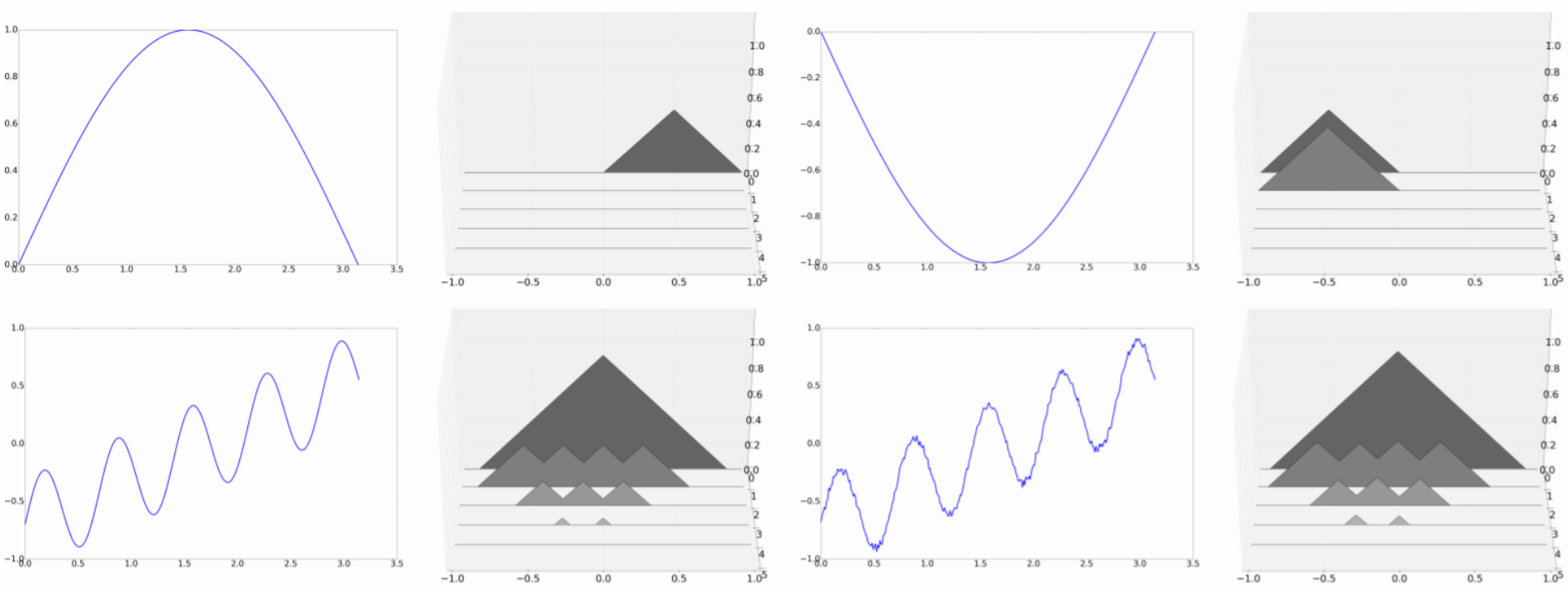
How does TDA work?

2. It’s in a matrix form

→ Machine loves matrices!

How does TDA work?

3. It’s persistent and robust over noises

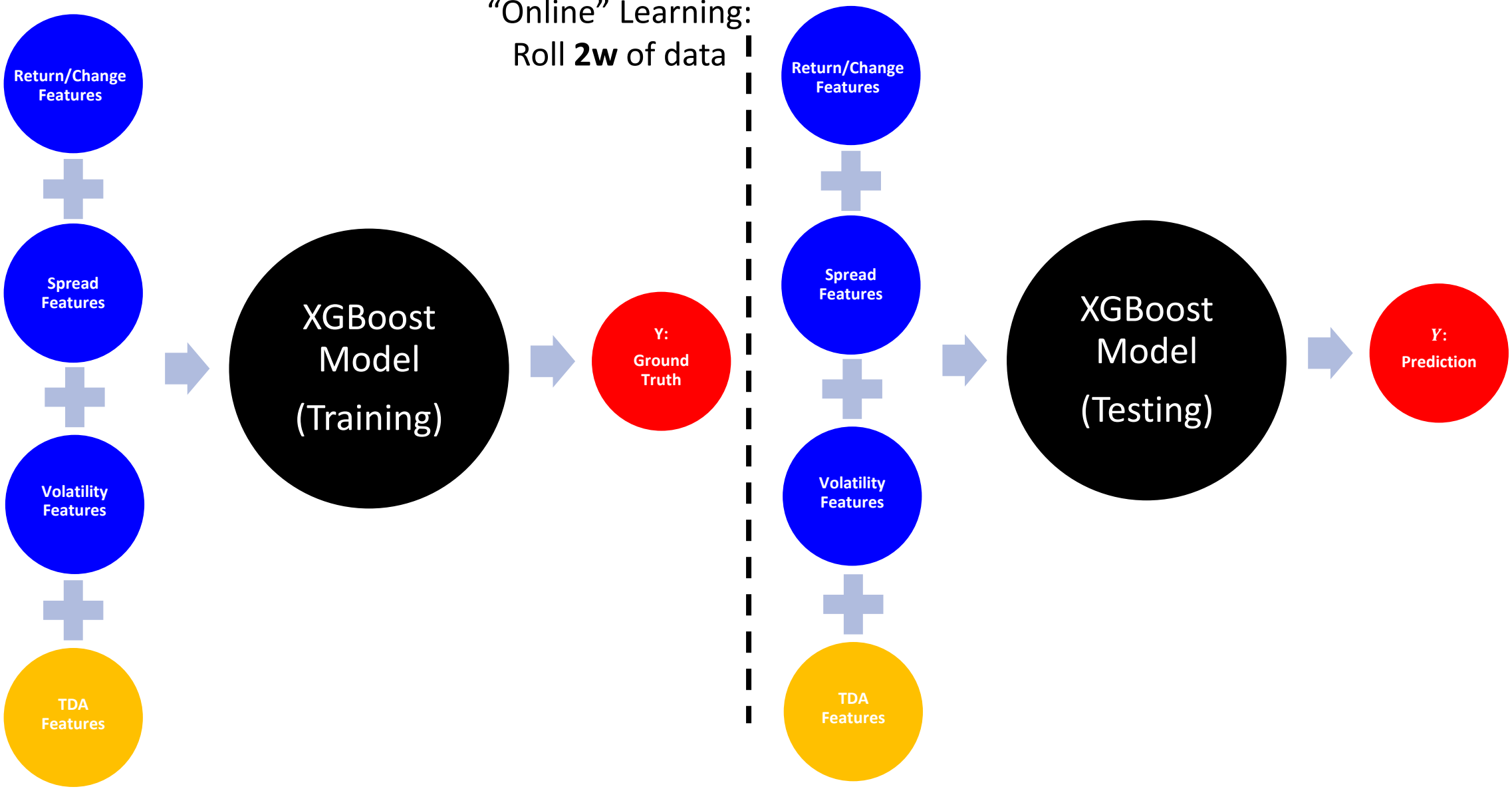


How does TDA work?

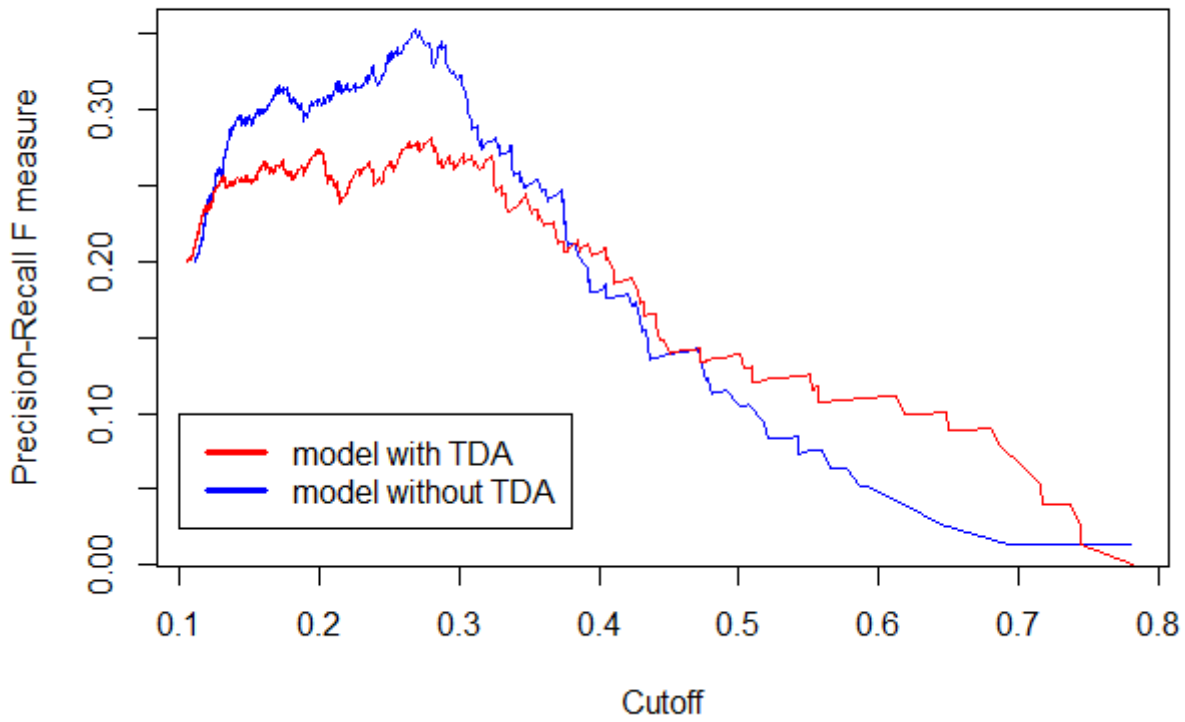
What have we seen?

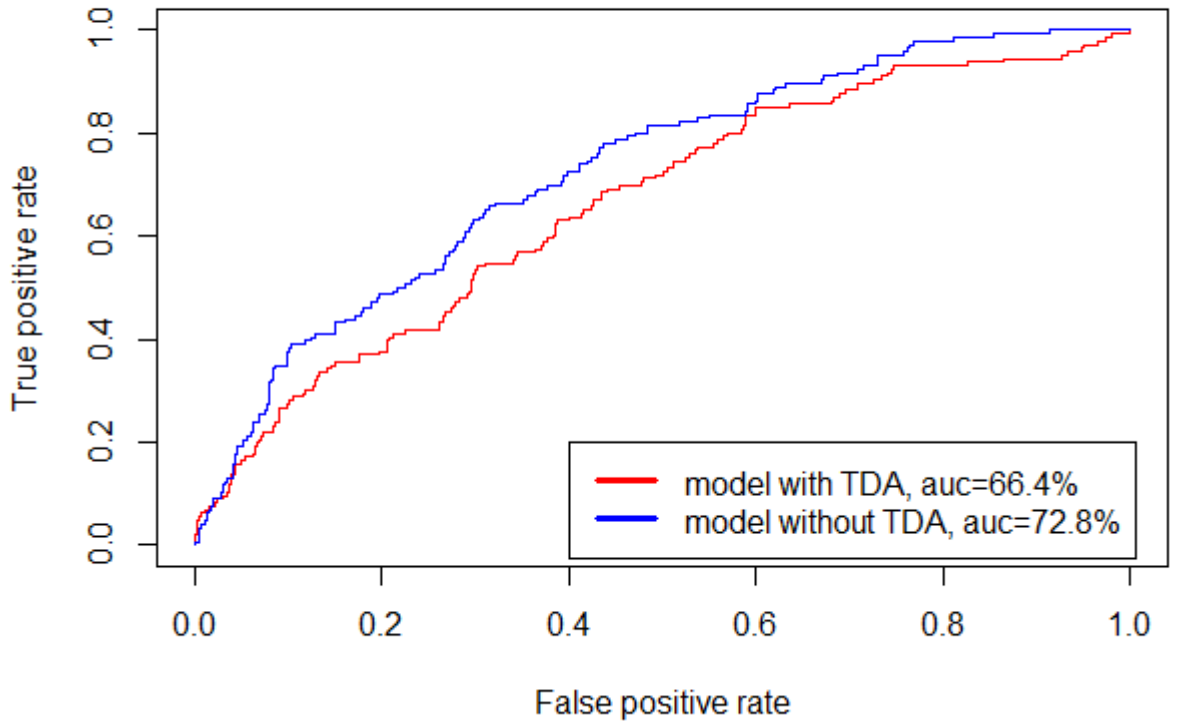
1. “Data has shape, shape matters”
2. TDA/Persistence Landscape is able to encode the shape of the data
3. Potentially novel features for supervised learning

The Model

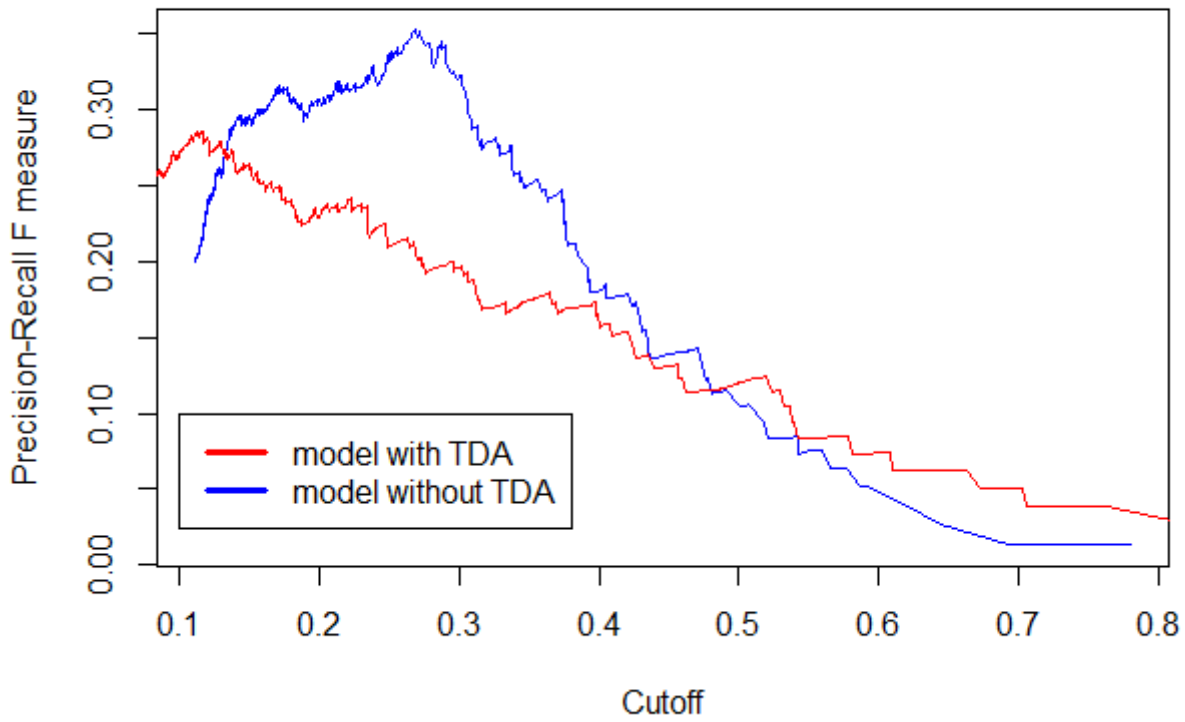

Return/Change Features
Spread Features
Volatility Features
TDA Features
XGBoost Model (Training)
Y: Ground Truth
"Online" Learning:
Roll **2w** of data
Return/Change Features
Spread Features
Volatility Features
TDA Features
XGBoost Model (Testing)
Y: Prediction

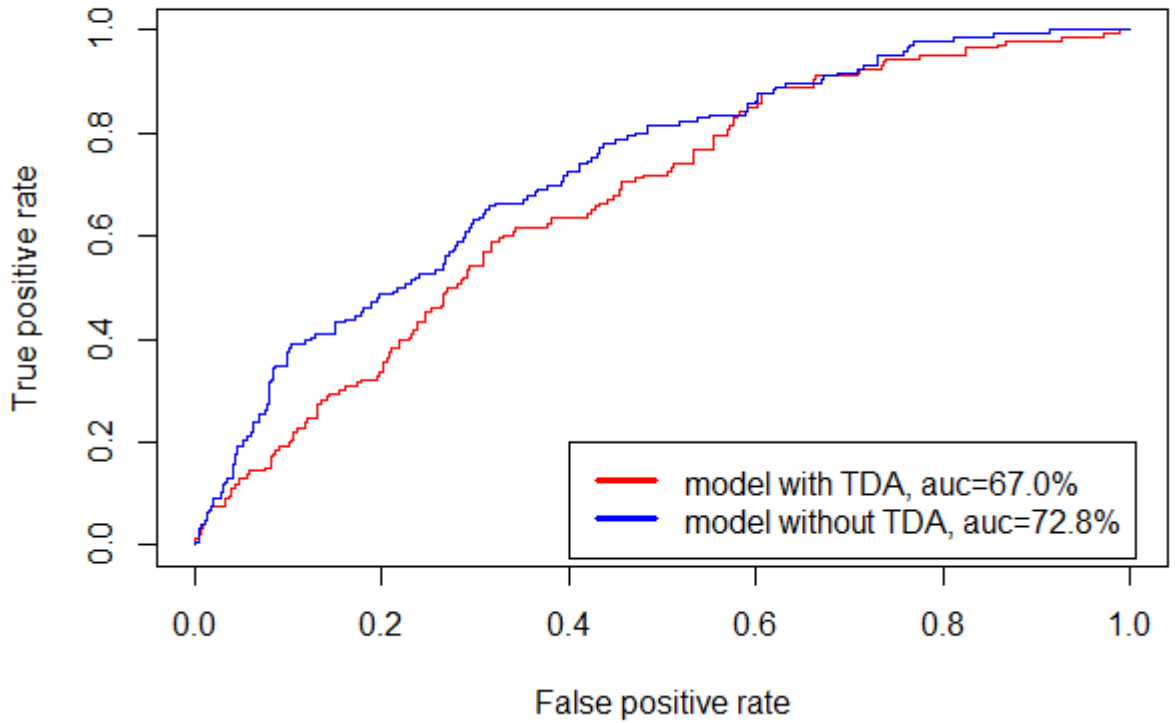
The Results 2.0 – HSI



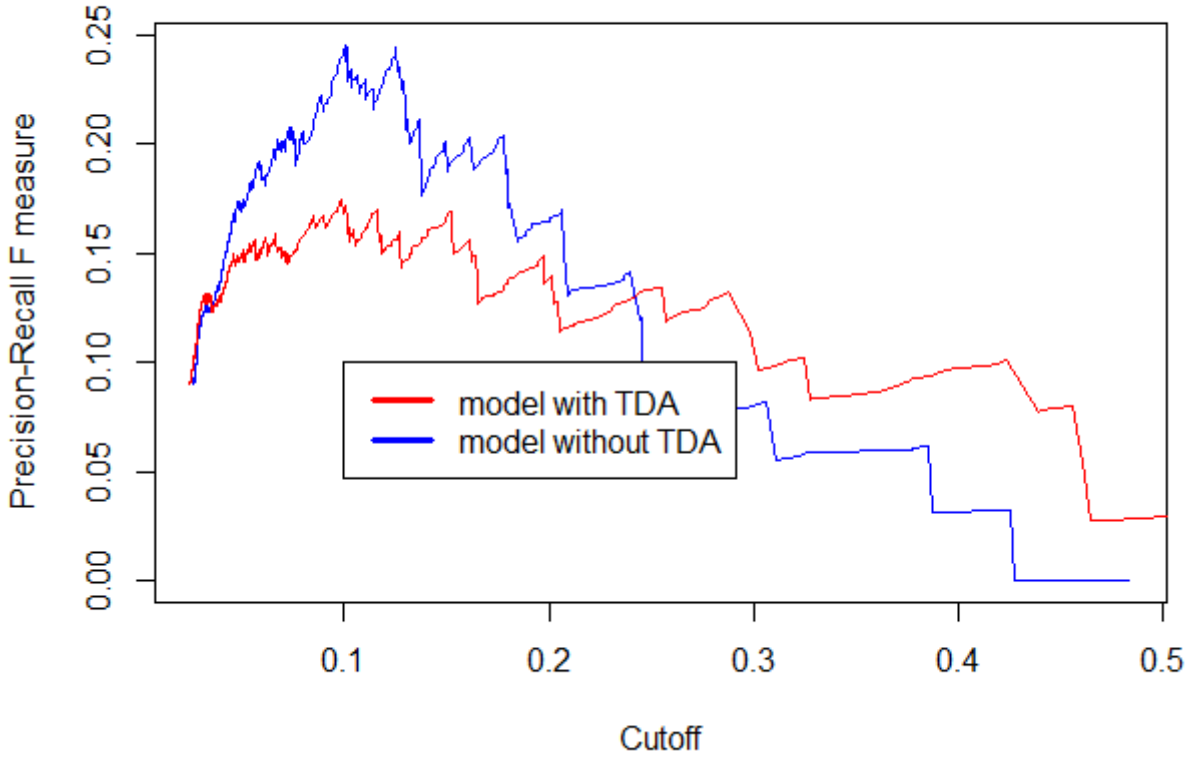


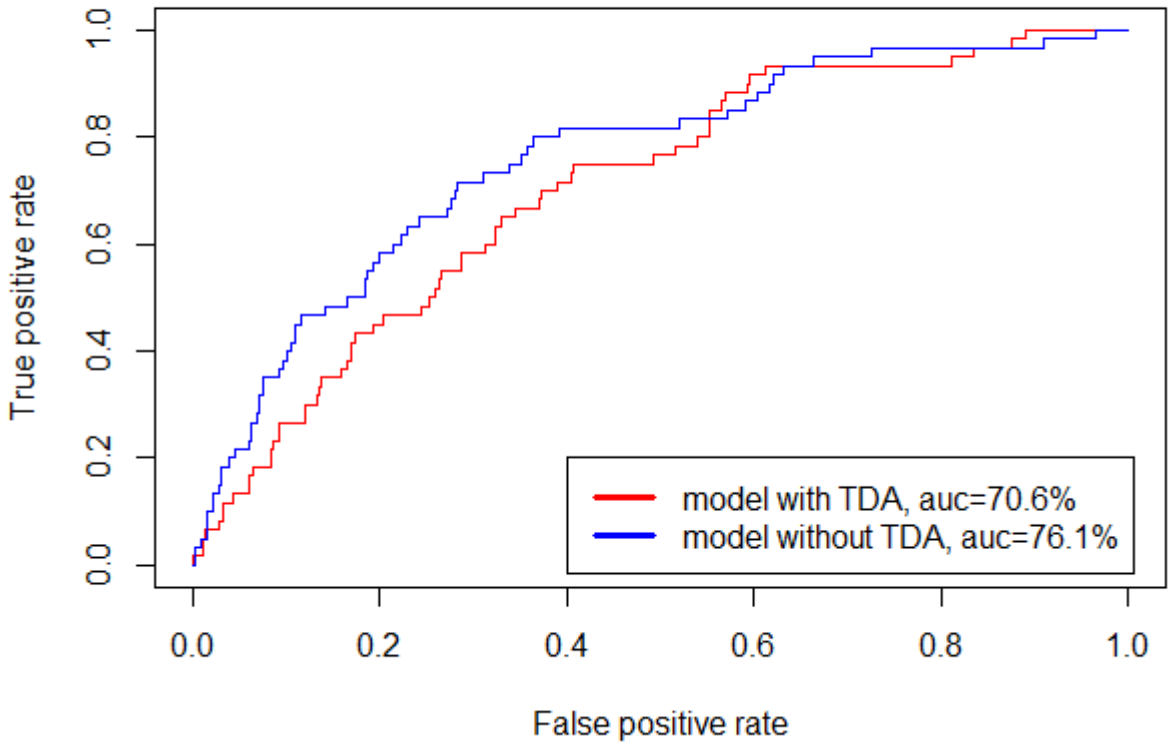
The Results 2.0 – S&P 500



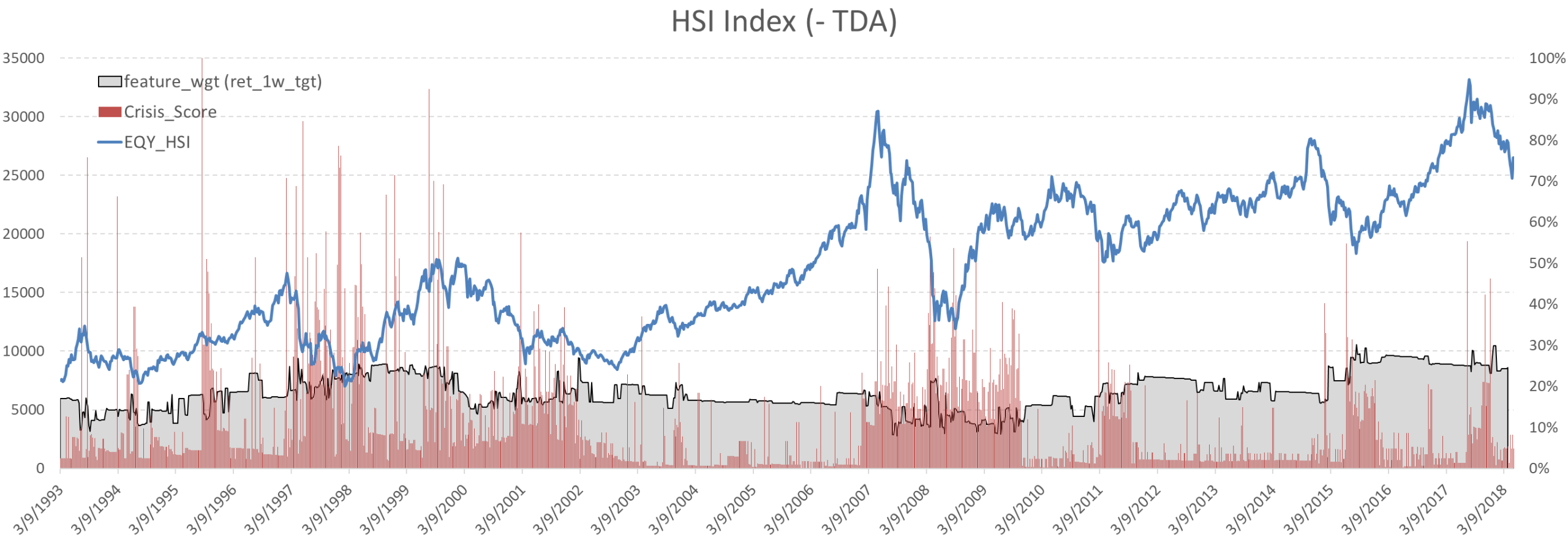


The Results 2.0 – FTSE 100

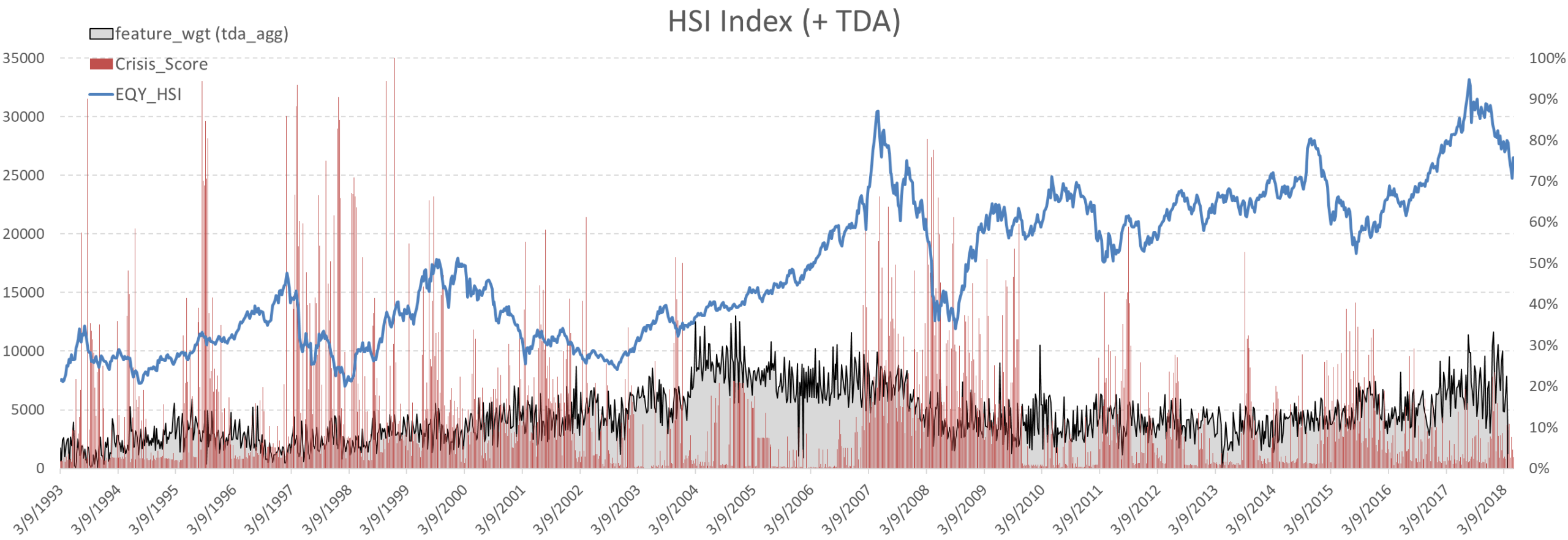




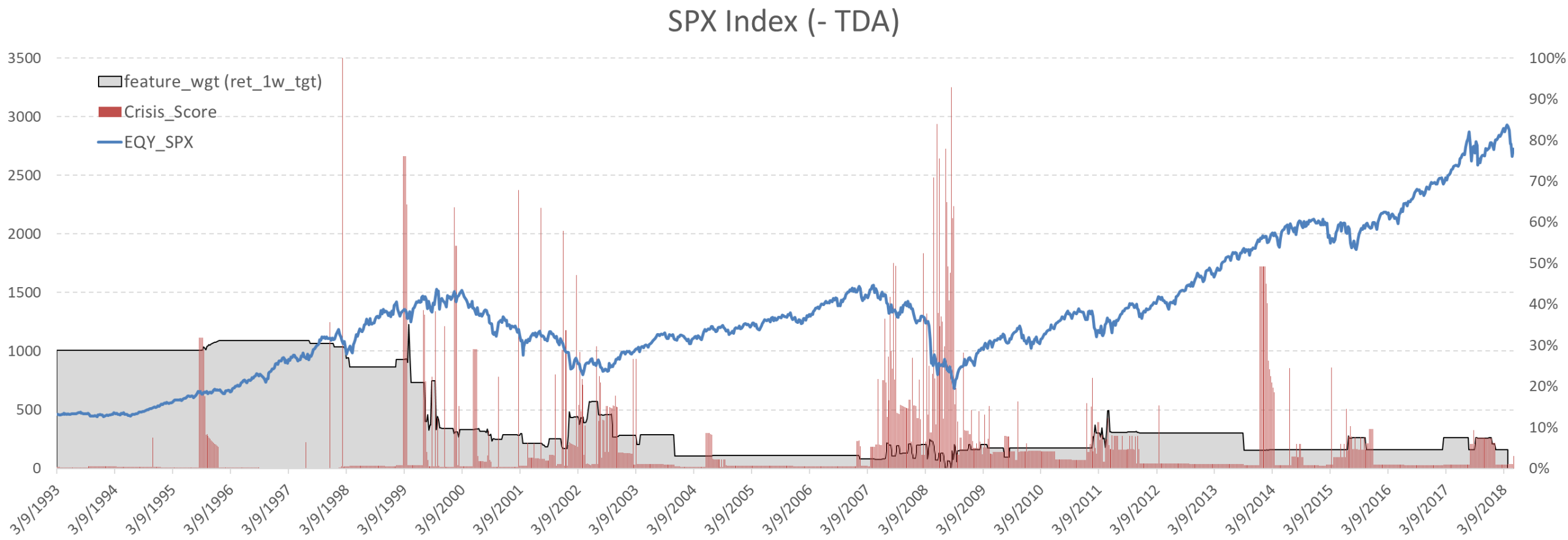
The Results in Details: HSI (-TDA)



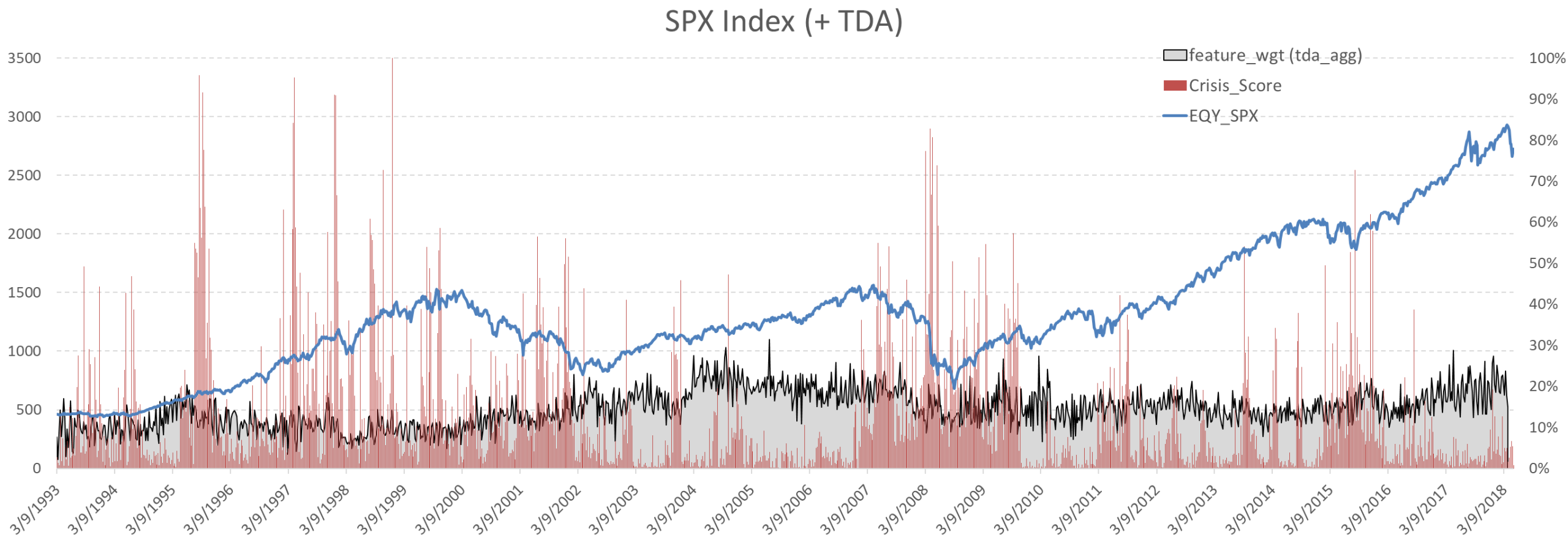
The Results in Details: HSI (+TDA)



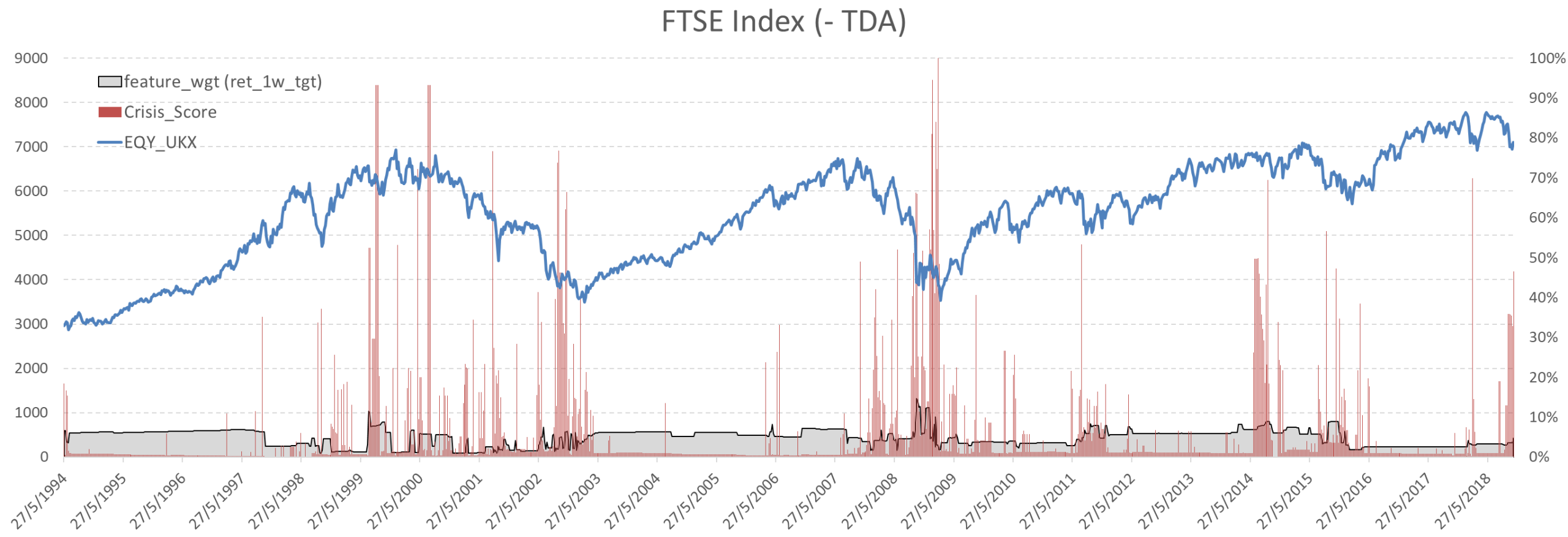
The Results in Details: S&P 500 (-TDA)



The Results in Details: S&P 500 (+TDA)



The Results in Details: FTSE 100 (-TDA)



The Results in Details: FTSE 100 (+TDA)