Persistence-Boost Quant Model with ETFs
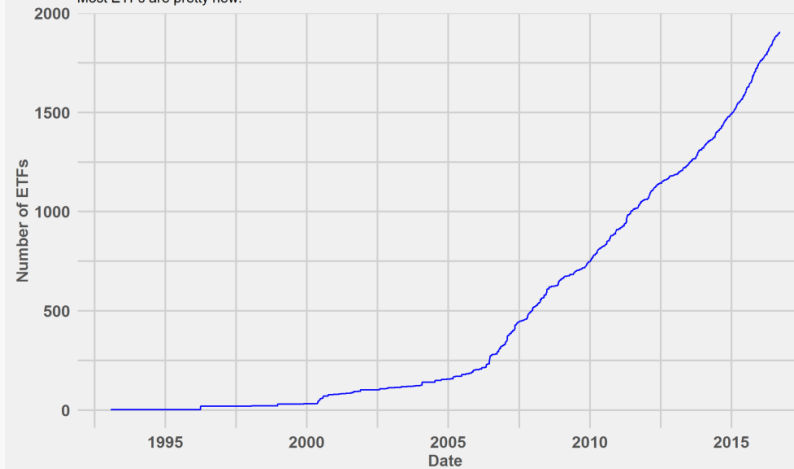
Saturday, 12 May 18

The Idea

Motivation

Number of ETFs Over Time
Most ETFs are pretty new.

Motivation

➢ Moreover, ETFs are likely to exhibit periodicity in their return cycle driven by the sectors, the investment styles or smart beta strategies whose underlying structures they are trying to replicate

Annual Returns of S&P 500 Sectors

2007	2008	2009	2010	2011	2012	2013	2014	2015	2016
32.38%	-17.66%	59.92%	25.72%	14.84%	26.26%	40.96%	24.29%	8.43%	23.65%
19.99%	-24.48%	45.22%	23.92%	10.53%	21.87%	38.74%	23.30%	5.21%	20.14%
15.81%	-31.55%	38.76%	19.92%	10.18%	15.19%	37.63%	18.19%	4.27%	17.81%
15.53%	-33.61%	17.27%	17.86%	4.41%	13.15%	33.21%	13.10%	3.77%	16.08%
11.60%	-34.72%	17.07%	12.30%	2.77%	12.49%	26.23%	12.87%	-1.73%	14.08%
9.83%	-35.93%	14.81%	10.83%	1.33%	12.46%	22.73%	8.05%	-3.48%	12.20%
8.45%	-41.52%	11.29%	10.67%	0.84%	12.24%	22.68%	7.52%	-4.72%	11.99%
5.39%	-43.68%	11.21%	9.13%	-2.92%	7.52%	22.27%	4.68%	-8.39%	4.32%
-14.32%	-47.05%	6.80%	0.85%	-11.64%	2.33%	8.75%	-1.91%	-10.36%	2.58%
-20.84%	-56.95%	2.63%	0.71%	-18.41%	-2.91%	6.49%	-9.99%	-23.55%	-4.36%

➢ These periodic signals are hard for traditional algorithms and factors（features in the machine learning world）to capture

➢ But a newly and rapidly developing set of techniques from Pure Mathematics offers the potential "killer feature"
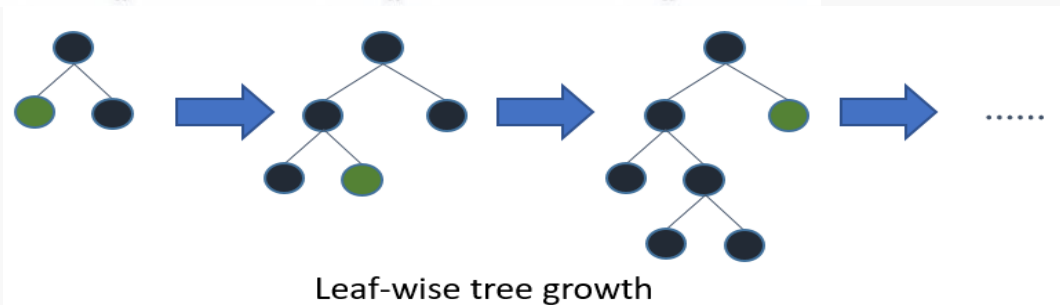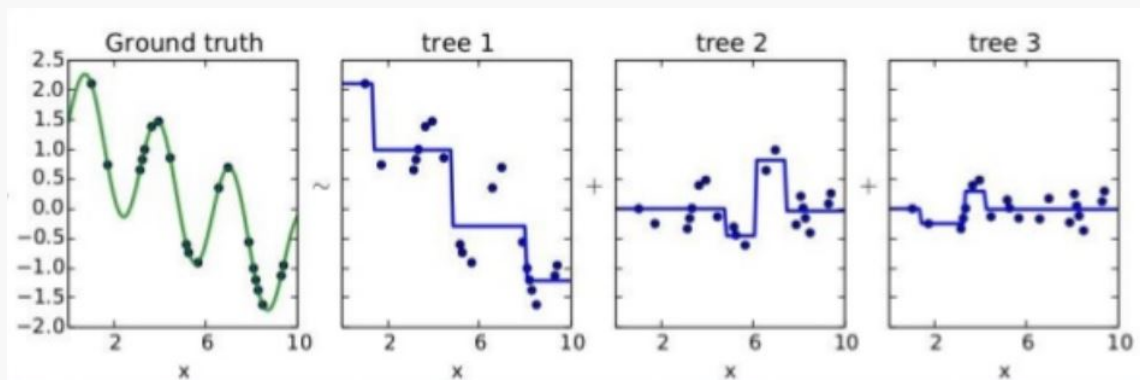
The Data
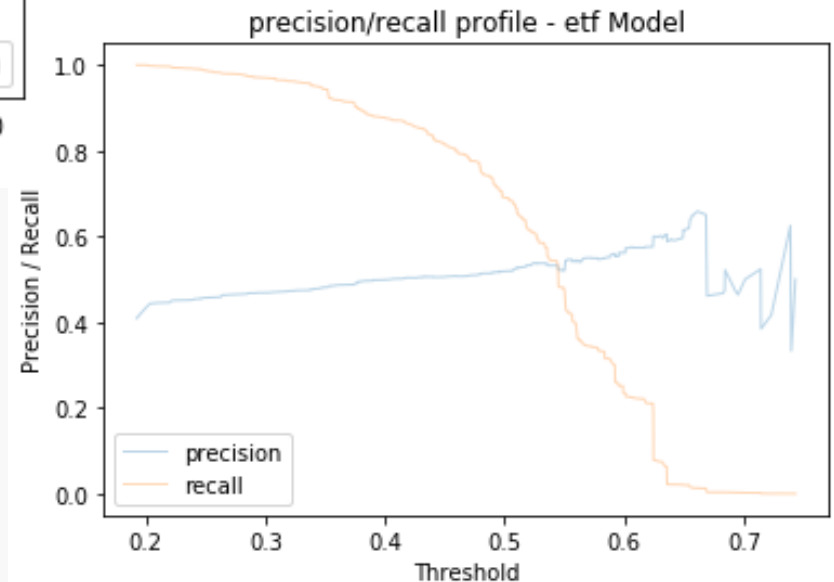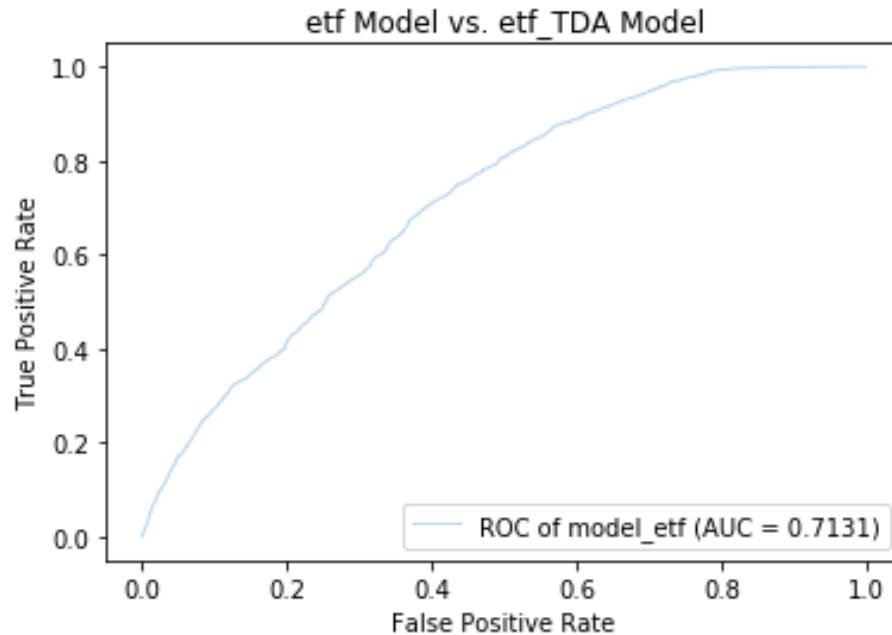
Feature Engineering
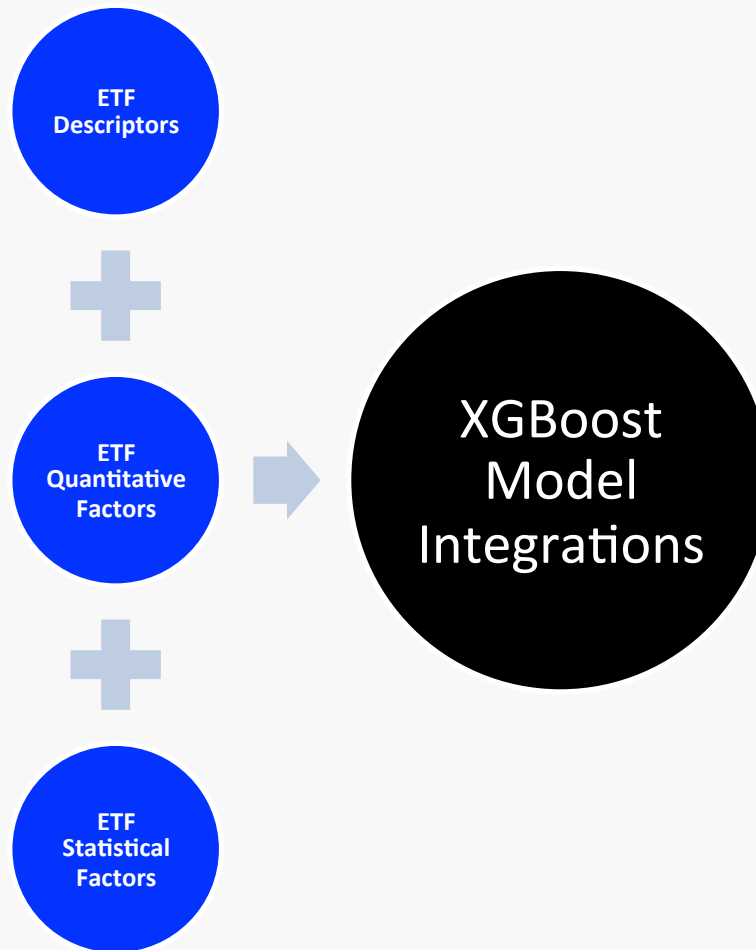
The Model

Introducing XGBoost

➤ Tree-based

    ➤ Transparent and whitebox as regression vs. deep neural networks

    ➤ Free from normalisation/standardisation of the features vs. regression
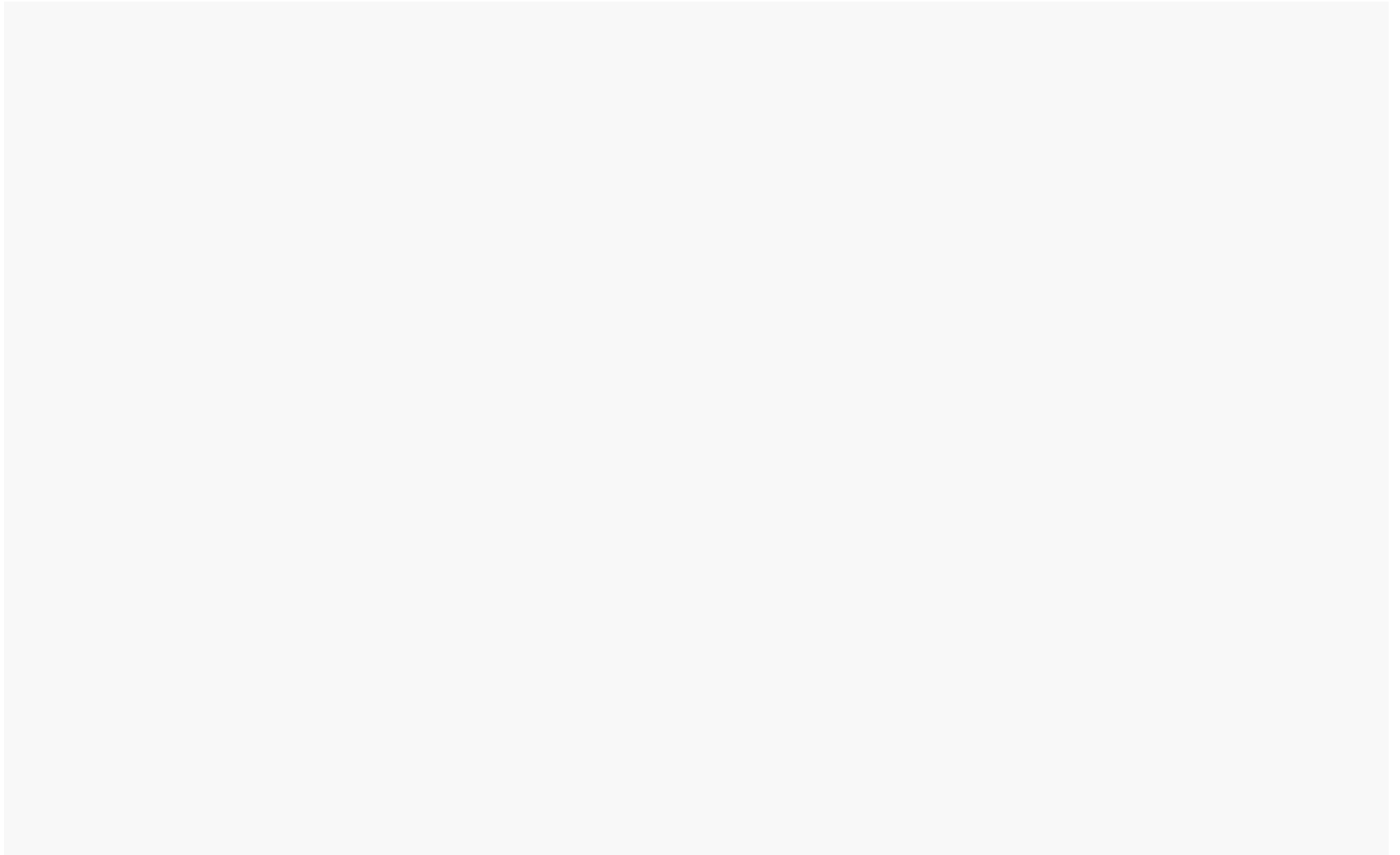
➤ Kagglers' love（used by 17 out of 29 winners）

Leaf-wise tree growth

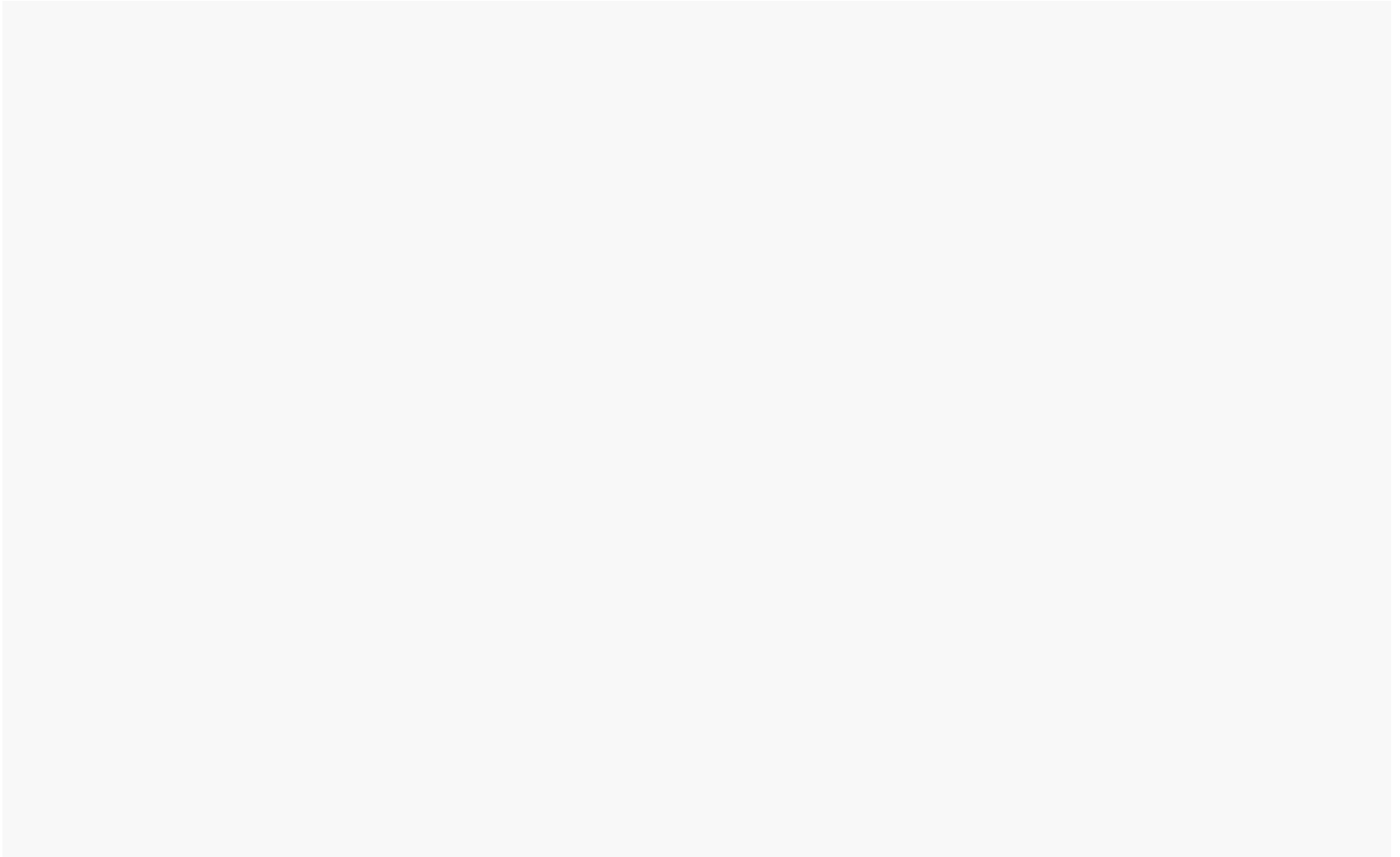# The Result

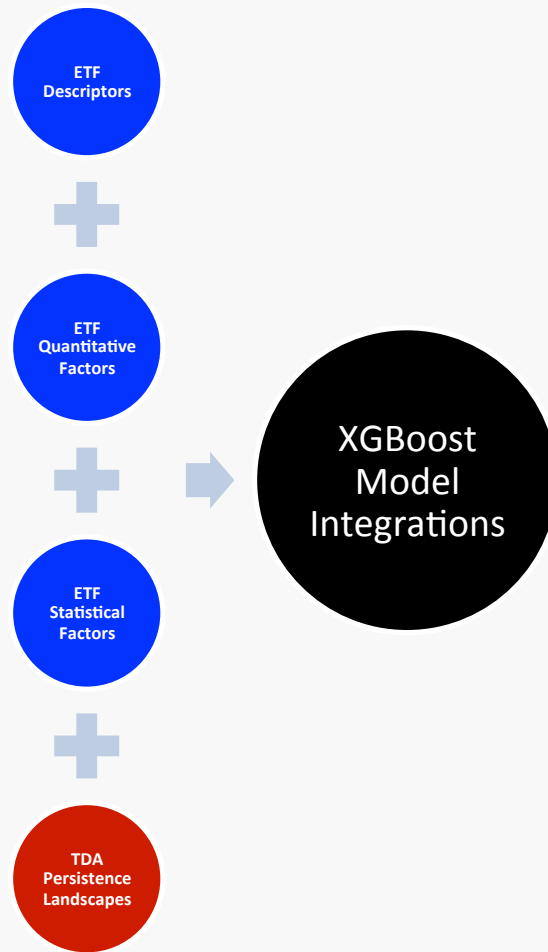# Precision, Recall, Area under Curve (auc) and Their Friends
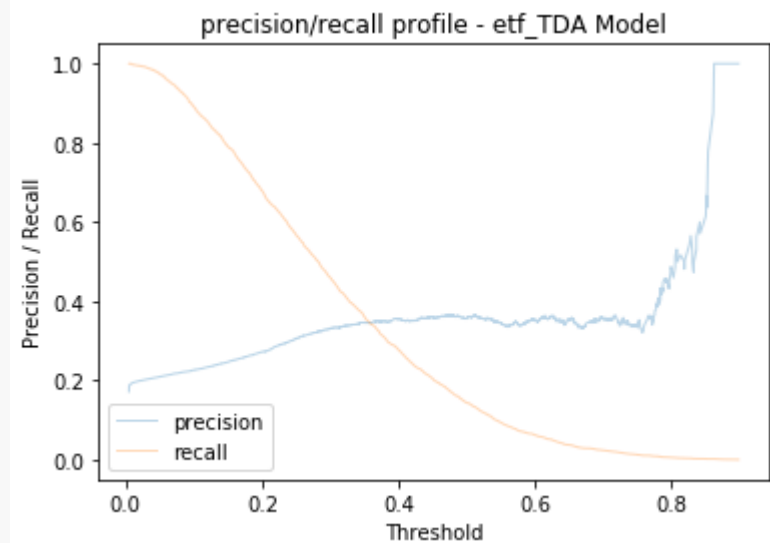
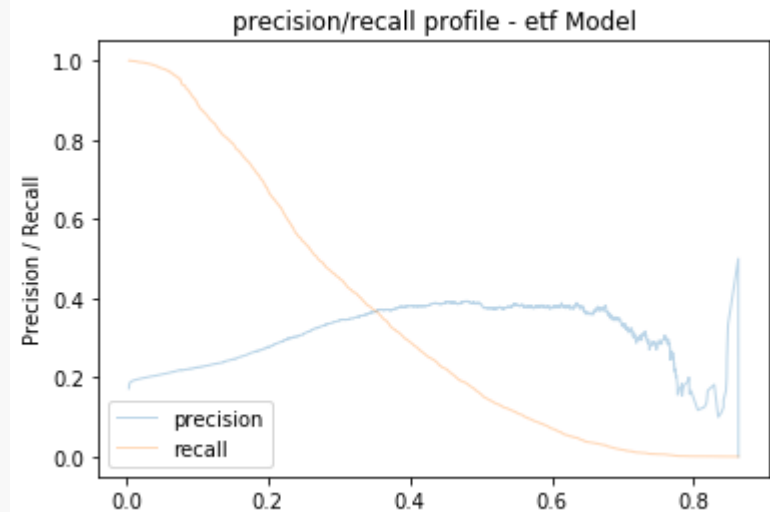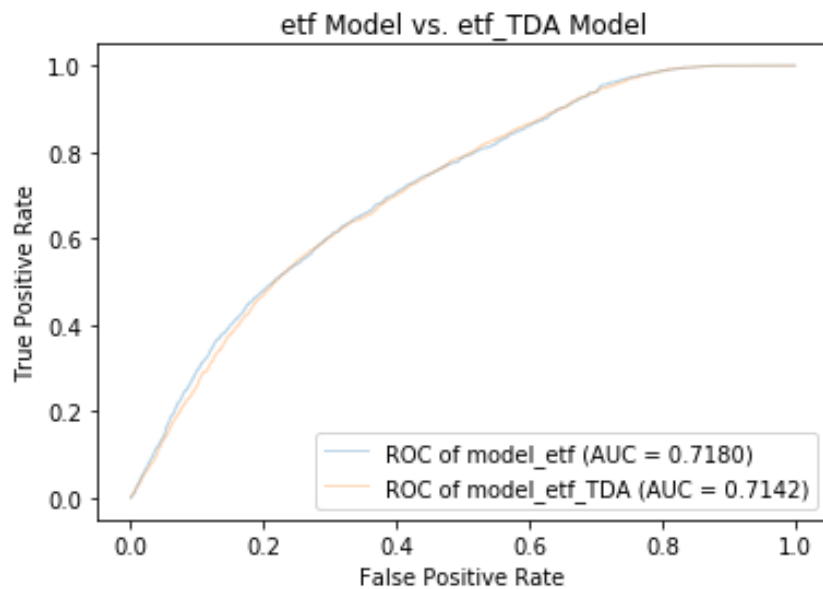# The Model 2.0

TDA in a Nutshell

TDA Applications

Feature Engineering 2.0
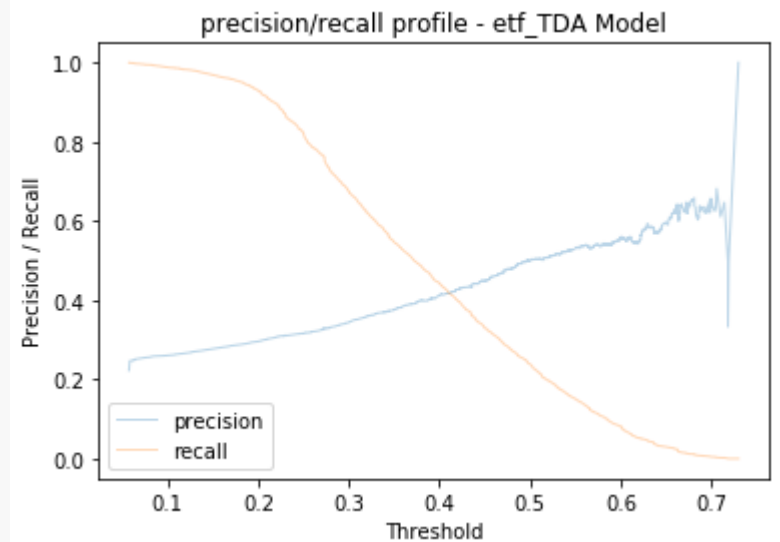
The Result 2.0

Result (score)

Result (score)

etf Model vs. etf_TDA Model

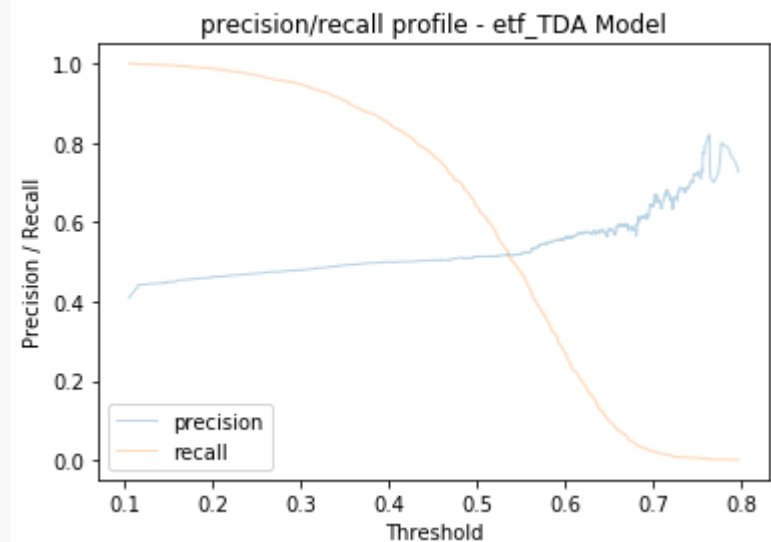precision/recall profile - etf Model

precision/recall profile - etf_TDA Model
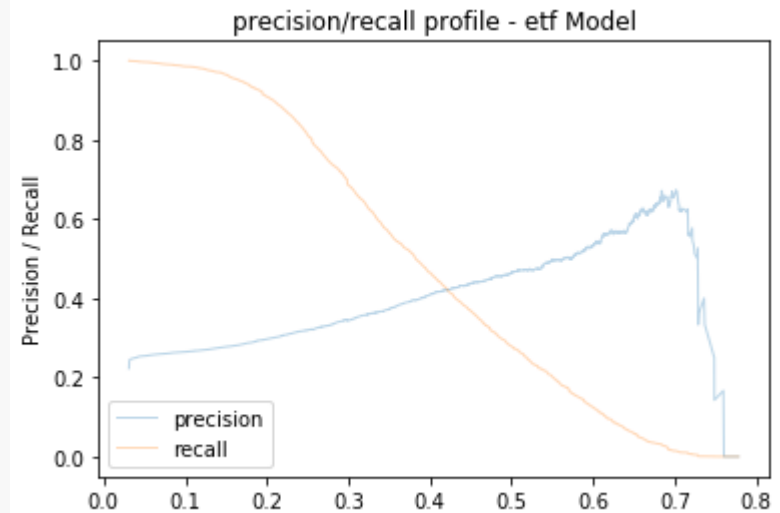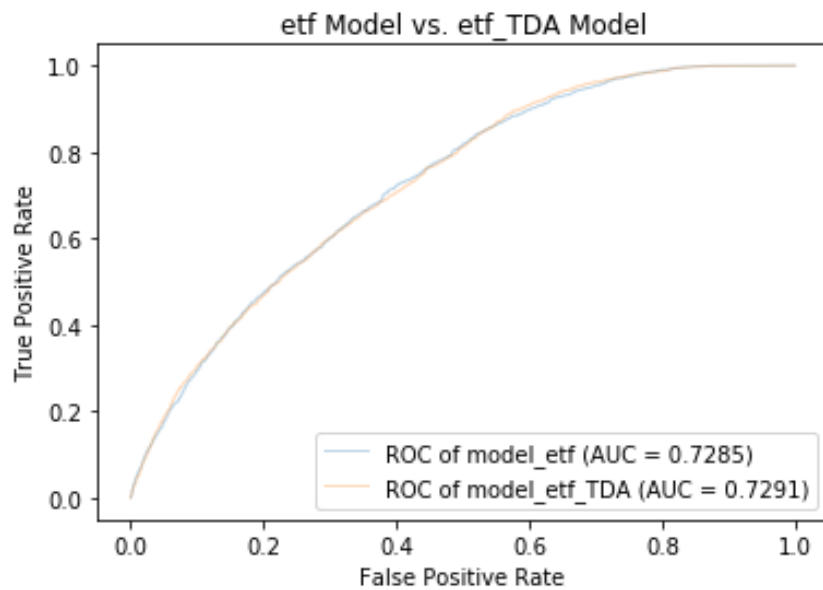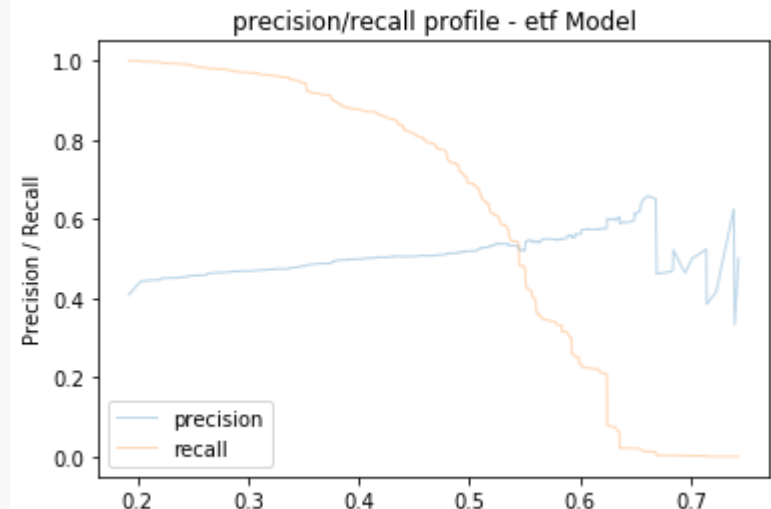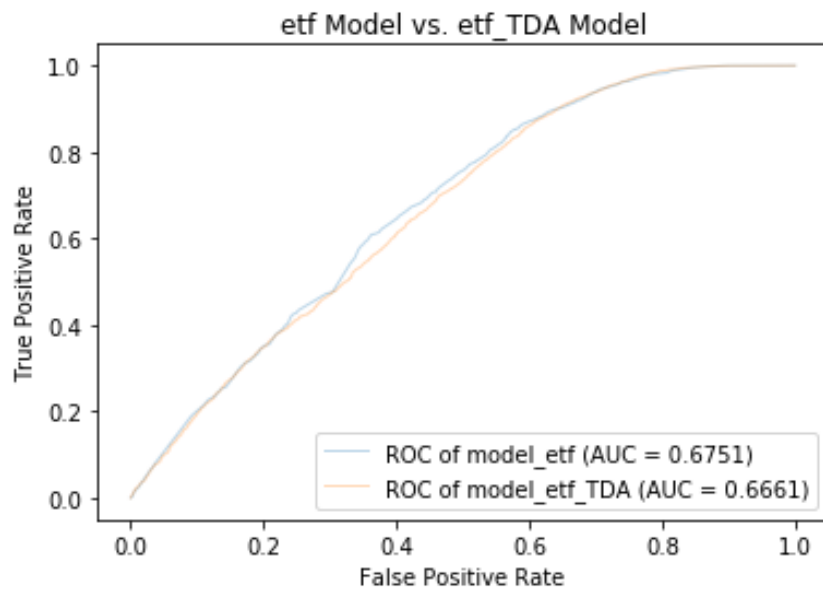
Result (score)

The Next Step

Next Step

➢ Backtesting investment strategies based on current model

  ➢ Buy and hold the ETFs with score higher than the threshold （95%） for max 1 month

  ➢ During the month, cash out whenever a holding ETF's return reaches the cut off return （4%）

  ➢ Cash out at the end of the month, buy and hold new set of ETF from the recalibrated model

➢ Data 2.0: ← fund flow data, volume data, ratio data, ETF analytics data

➢ Further iterations （reinforcement learning？）