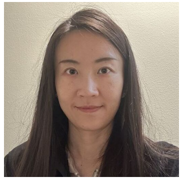
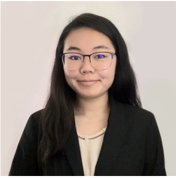
DeepLearning Sentiment Portfolio (DSP)

ML Driven Portfolio Framework to Adapt to Regime Change

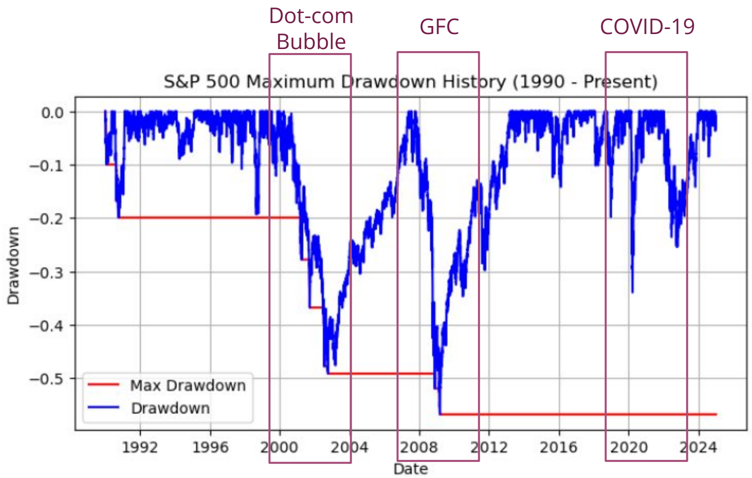
Irene Na, Elaine Keung, Mia Kobayashi, Austen Lowitz

Problem

"Do you still remember the chill those words sent down your spine? Dot-com Bubble. Great Financial Crisis (GFC). COVID-19 Pandemic…"

Every time the market dynamic undergoes a major shift, a 'regime change', billions vanish—first in financial markets, then often rippling through the real economy



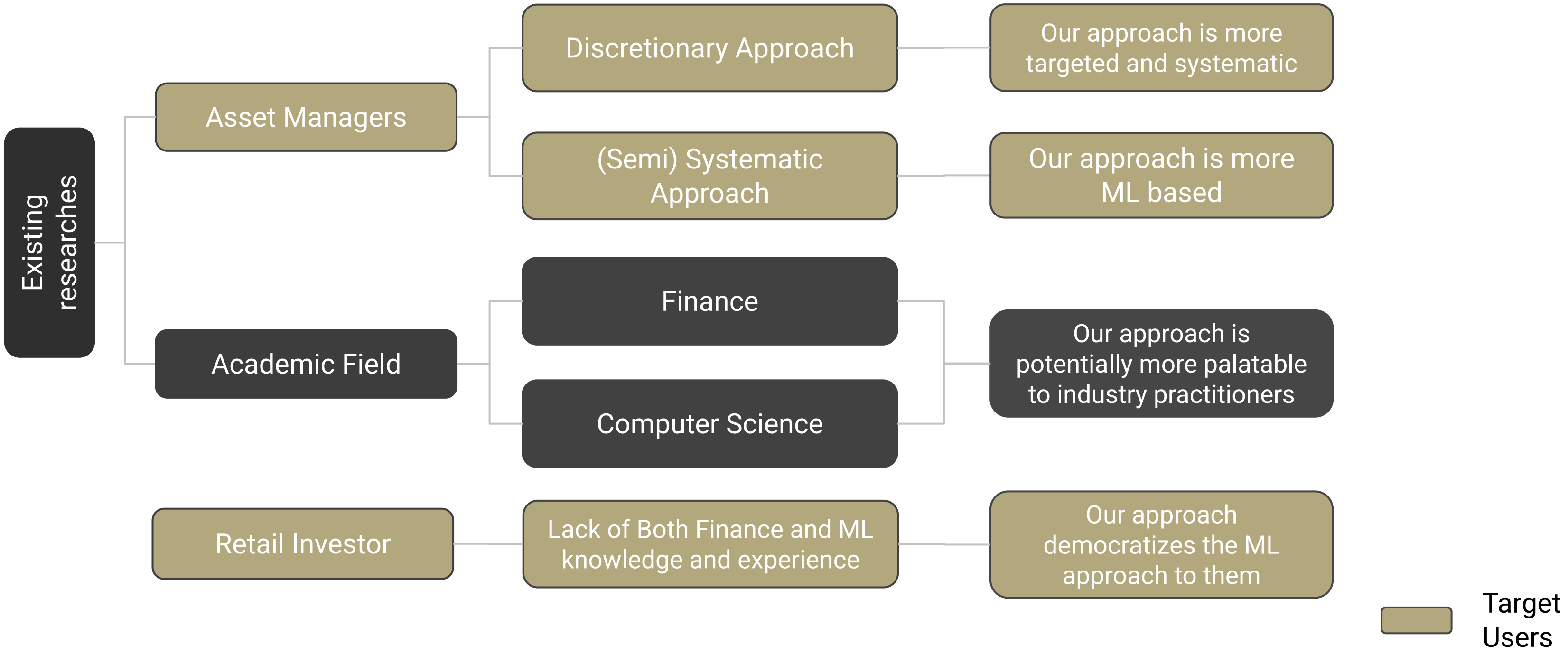
Source: Los Angeles Times

Think it doesn't affect you? **Think again.** If not personal stocks investment, how about your pension and insurance funds?

For background, AUM of Public Pension Funds (PPFs) in U.S. by Dec 2024 is ~$13 trillion

Our Goal and Target Users

Our **goal**: Build a machine learning driven portfolio framework which could help retail and institutional investors better capture market dynamics and preserve capital, especially during regime changes - **through non-linear relationship** and **text information** capturing

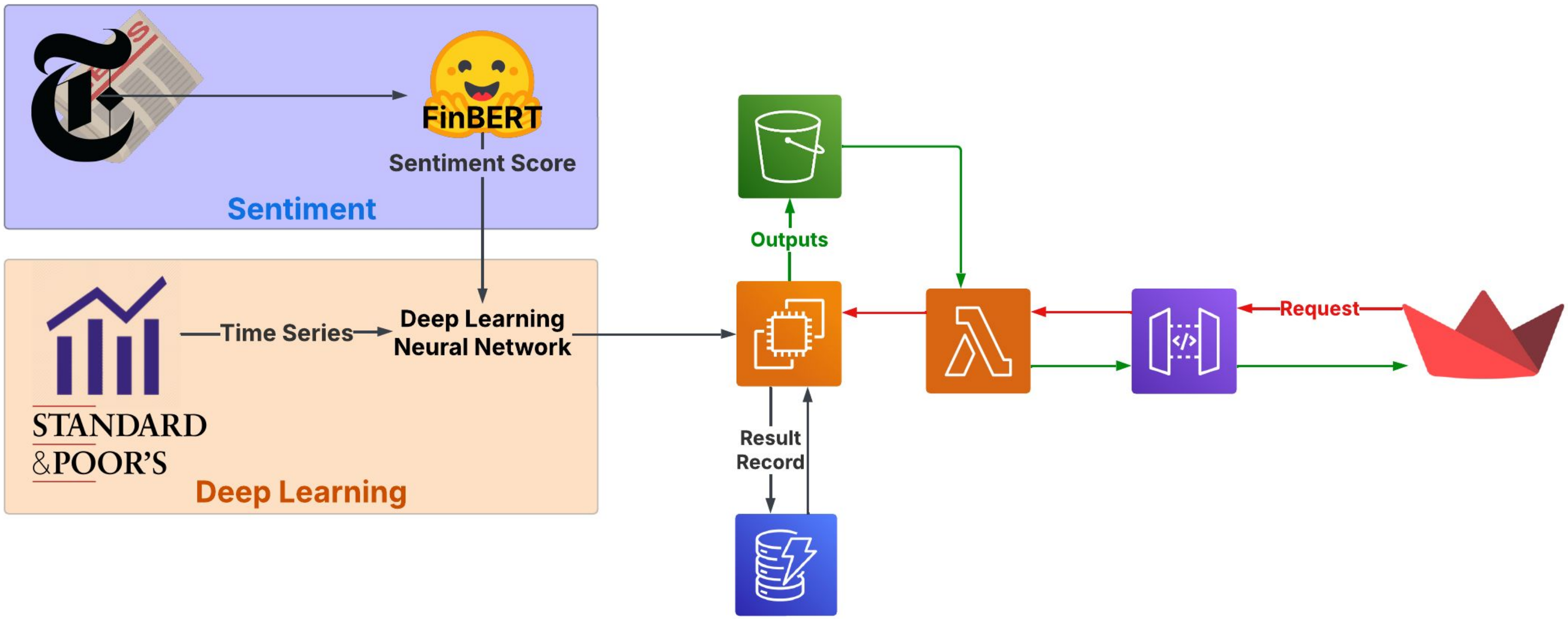


Minimum Viable Product, Demo

[Link to our landing page
(with video demo embedded)]

Technical Discussion

Overall Roadmap & Model Setup



Sentiment Model

Goal: Develop a sentiment model that extracts predictive signals from financial news to enhance market timing decisions. These signals will be integrated into a deep learning framework for security selection, enabling more informed portfolio decisions during market regime shifts

Datasets for Sentiment Model

- **Research Question:**
  - Can we use news sentiment to enhance our market timing by predicting tomorrow's market returns?

LIVE Updated 11 minutes ago

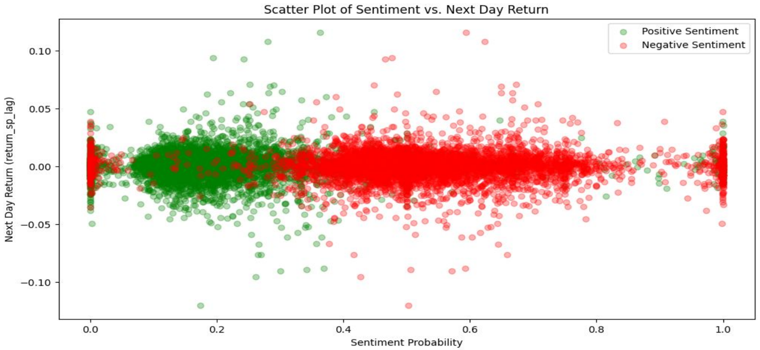
Tariff Live Updates: Trump Backs Down on Reciprocal Tariffs for 90 Days

Stocks immediately made steep gains after President Trump's reversal. But he said China would not be included, raising tariffs on its exports to 125 percent after Beijing announced a new round of retaliation.

- ~2.5M Finance and economic oriented news from 2001-2022
  - FNSPID Collection (2.3M articles)
  - NYT (173k articles)
  - Web Scraping (3.8k articles)

Linear Models Miss the Signal – Deep Learning Finds It

Initial Approach: Explore Linear Modeling

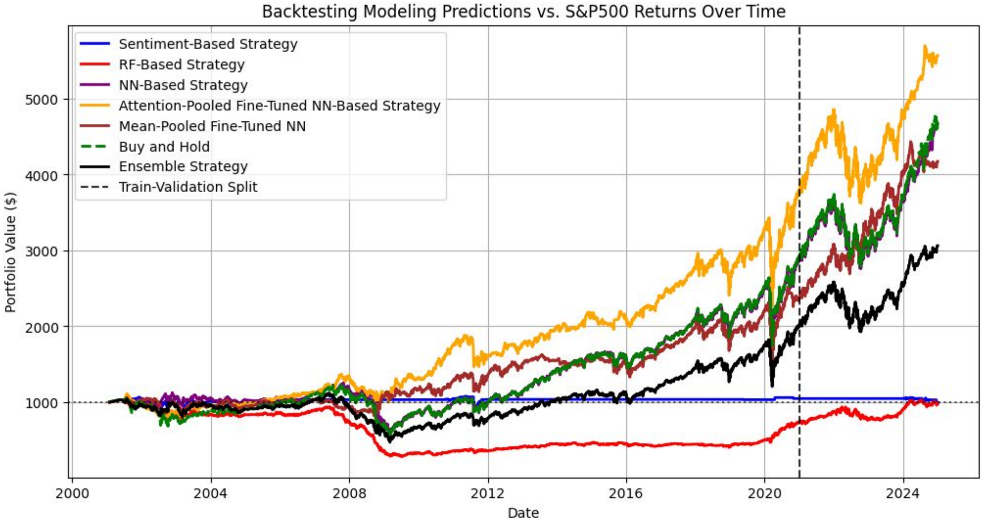


Sentiment probabilities have **NO CORRELATION** with *next day's market returns*

Revised Approach: Leverage Deep Learning

1. Extract features (**sentiment probabilities** and **embeddings**) from ~2.5M articles
2. Address challenge of aggregating one embedding per day via **pooling strategies**
3. Assess efficacy of **fine-tuning** FinBERT's last two layers
4. Run **classification models** to predict **next day's market return**

Attention-Pooled Fine-Tuned NN Outperforms Market



Model	Accuracy	Sharpe Ratio	Max Drawdown %	Backtesting Ranking
RF Classifier	50%	-0.09	-73%	6
NN (MLP) Based	56%	0.36	-57%	3
Attention-Pooled Fine-Tuned NN	56%	0.44	-34%	1
Mean-Pooled Fine-Tuned NN	56%	0.36	-35%	4
Ensemble Model	62%	0.25	-58%	5
Buy & Hold	NA	0.34	-57%	2

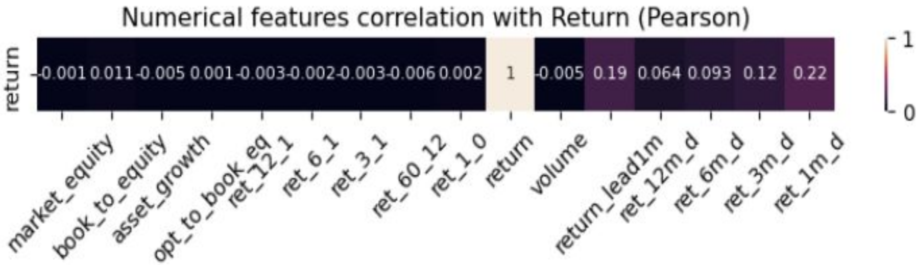
- **Benchmark**: *Buy & Hold Strategy (**Green Line**)*
- **Overall Top Performing Sentiment Model**: *Attention-Pooled Fine-Tuned NN (**Gold Line**)*

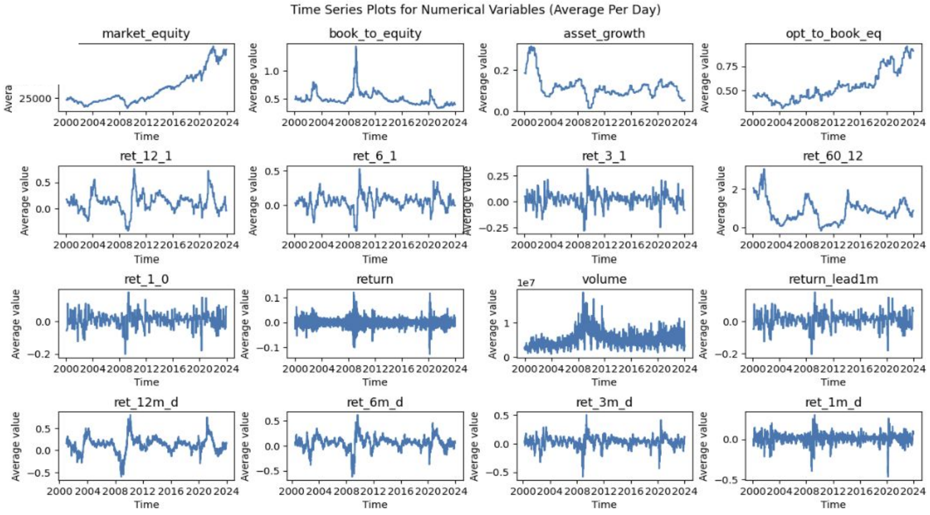
Deep Learning Model

Goal: Check if deep learning can be used to improve portfolio risk/return profile through security selection

Data & EDA - Deep Learning

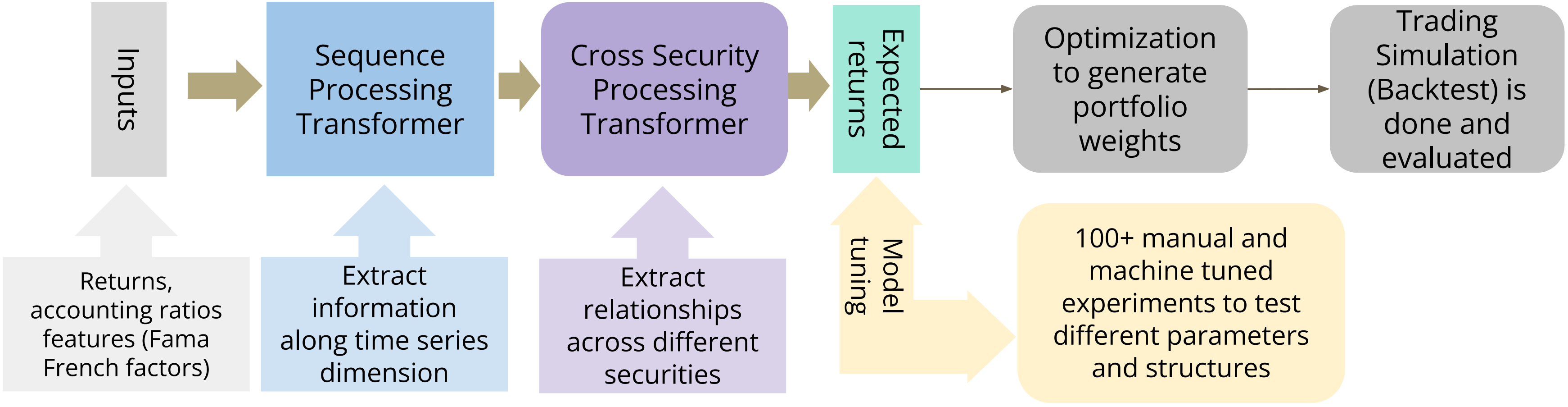
- **Source:** WDRS (3 datasets), Yahoo Finance
- **Universe:** S&P500 (data: return, accounting ratios)
- **Date:** 2000 - 2024, daily (2000-2016/2017-2020/2021-2024: train/validation/test)
- **Size:** ~1GB in size, 3mm rows and 32 columns





- **Data checking:** null values - reasonable, but we spot outliers
- **EDA:** After cleaning, the overall distributions and time series look reasonable
- **Pearson correlation - immaterial linear correlation** vs the target variable ('return'), inviting for non-linear relationships checks

DL Model - Model Structure and Parameter Exploration



Takeaways:

- Both transformers are important
- Transfer-learning could be beneficial (>1 targets)
- **Batch size is critical,** followed by **look back window**
- **L1 Loss is used to regulate over-complication**

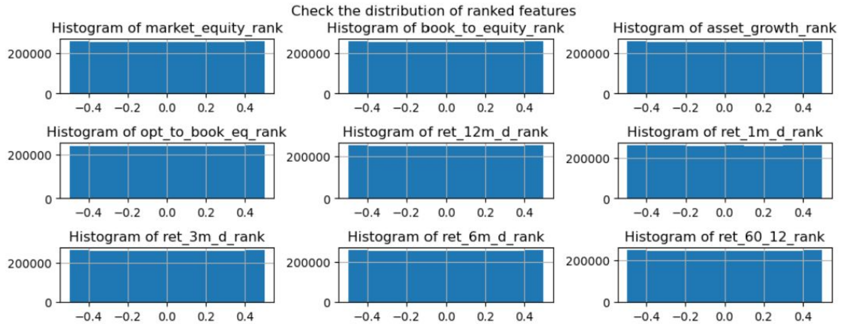
DL Model - Model Selection

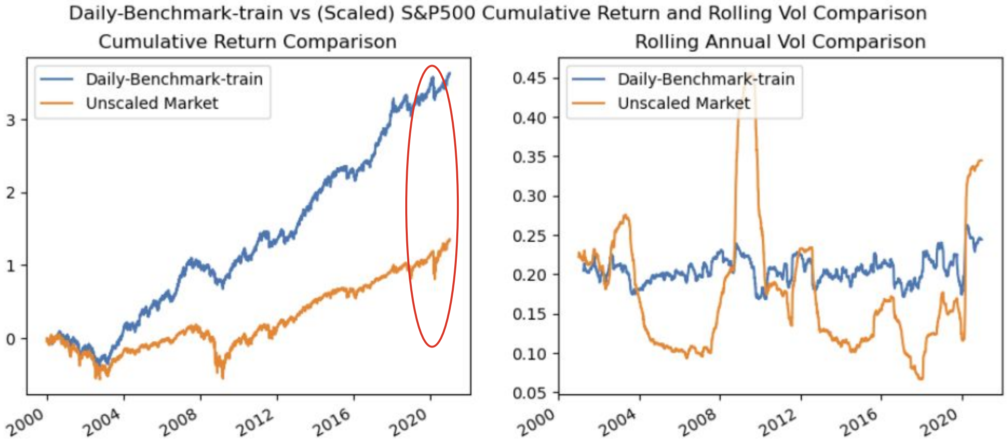
Top 10 models	Features size	Target size	embed_dim	transformer_layers_seq	num_heads_seq	forward_dim_seq	transformer_layers_cs	num_heads_cs	forward_dim_cs	learning_rate	batch_size	window_days	train_loss	val_loss	Sharpe (w vol targeting)	Max_drawdown (w vol targeting)
1	13	6	256	2	4	64	1	2	3	0.000050	3	128	0.0143	0.0138	0.9538	-0.4186
2	13	6	384	2	4	96	1	4	3	0.000025	3	128	0.0144	0.0137	0.9285	-0.4136
3	13	6	128	2	4	64	1	2	3	0.000050	3	64	0.0142	0.0135	0.9196	-0.4149
4	13	6	128	2	4	64	1	2	3	0.000050	3	128	0.0144	0.0138	0.9172	-0.4068
5	13	6	64	2	4	64	1	2	3	0.000050	3	64	0.0144	0.0135	0.9168	-0.4077
6	13	6	256	2	4	64	1	2	3	0.000050	3	252	0.0157	0.0146	0.8988	-0.4231
7	13	6	448	4	2	96	2	4	128	0.000021	9	128	0.0150	0.0138	0.8780	-0.4351
8	13	6	128	2	4	64	1	2	3	0.000050	3	3	0.0141	0.0131	0.8622	-0.4177
9	13	6	128	2	4	64	1	2	3	0.000050	3	252	0.0149	0.0143	0.8489	-0.4094
10	13	6	128	2	4	64	1	2	3	0.000050	3	3	0.0140	0.0132	0.8262	-0.4115

Model Selection Takeaways:

- The ultimate goal is **maximize return per unit of risk (Sharpe ratio)**, and **minimize risk**, which we used as **ultimate model selection criteria, with validation loss as a reference**
- Smaller train/val loss is a good, but lower loss != good backtest results
- Over complicated models could compromise the performance
- At the end, we chose a model with **the most desirable in-sample profile: 0.95 Sharpe ratio, -41.9% maximum drawdown, and low validation loss**

Benchmark Model - Raised Our Bar



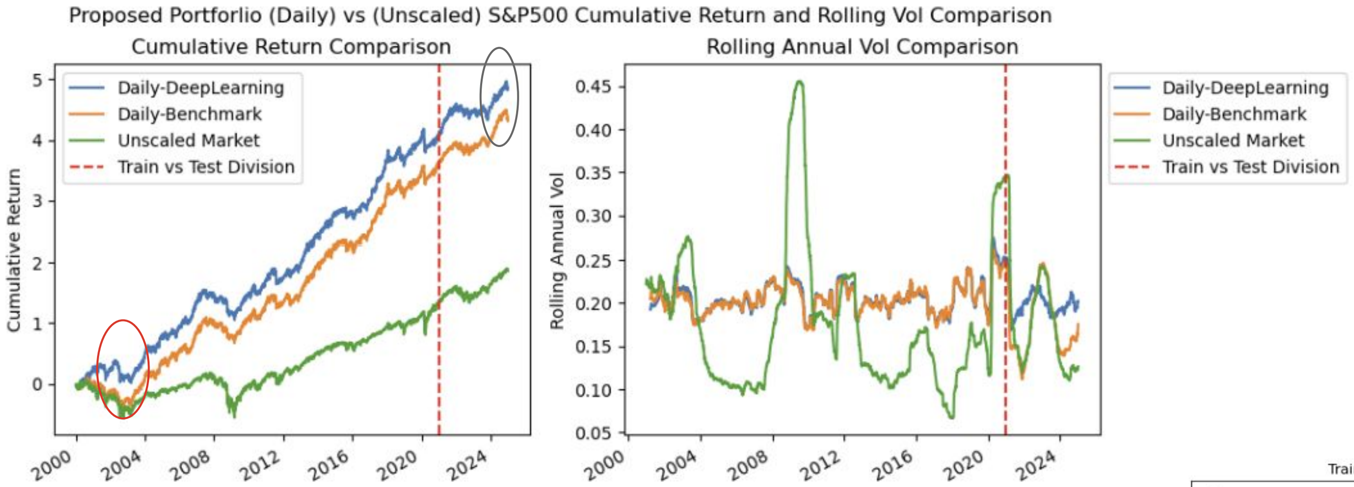


About the Benchmark Model:

- **Raised our bar** by using a benchmark portfolio based on traditional statistical method **that beats the market by large**
- Using similar returns and accounting ratio as key inputs (left hand side)
- We decide on daily rebalancing, max-sharpe optimization goal, with targeted rolling vol

	Daily-Benchmark-train	Unscaled Market
avg_rtn_ann	0.175136	0.064591
vol_ann	0.205825	0.198926
sharpe_ann	0.850899	0.324698
max_drawdown	-0.42506	-0.567754

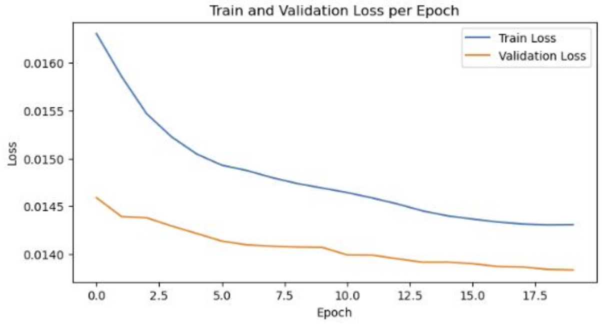
DL Model - Chosen Model Outperforms Benchmark



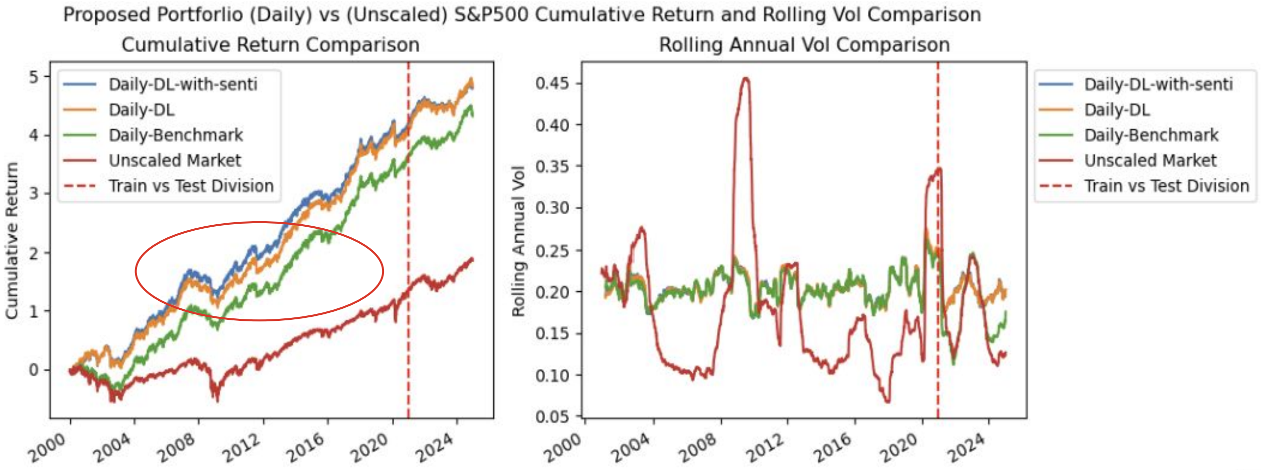
	Daily-DL-Max-Sharpe	Daily-Benchmark	Unscaled Market
avg_rtn_ann	0.195387	0.174167	0.074665
vol_ann	0.205625	0.201538	0.193884
sharpe_ann	0.950213	0.864193	0.385098
max_drawdown	-0.418583	-0.42506	-0.567754

Chosen Model Takeaways:

- **Vs Benchmark**, deep learning model avoids the dip around 2004, reduced max-drawdown
- **Consistent** out of sample performance
- **Caveat**: we can’t guarantee DL model to outperform in every environment, such as extreme bull market (e.g. 2024)



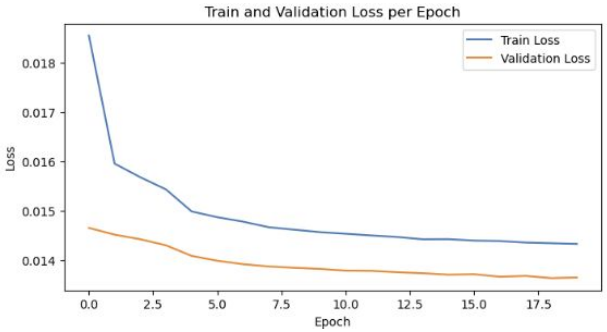
Sentiment + Deep Learning - Further Mitigates Drawdown



	Daily-DL-Max-Sharpe	Daily-DL-Max-Sharpe w senti	Daily-Benchmark	Unscaled Market
avg_rtn_ann	0.195387	0.193873	0.174167	0.074665
vol_ann	0.205625	0.205485	0.201538	0.193884
sharpe_ann	0.950213	0.943488	0.864193	0.385098
max_drawdown	-0.418583	-0.409979	-0.42506	-0.567754

Final Model Takeaways:

- **Incorporate** sentiment **in two ways**: as inputs and scaler
- Sentiment incorporation further **mitigates large sudden loss**
- Although incremental, it **improves the profile during GFC**, reduces the max drawdown
- Verified with SME (subject matter expert), the result is reasonable



Key Takeaways and Next Steps

Future Work

Extend Data Collection

- Expand early 2000s sentiment data
- Further expand time period of all data

Reinforcement Learning

- Align gradient descent with multi-period portfolio performance goal (e.g. Max Sharpe)
- Boost robustness across market environments

Deepen Financial Analysis

- Turnover analysis
- Transaction cost modeling
- Implementation study

Mission Statement

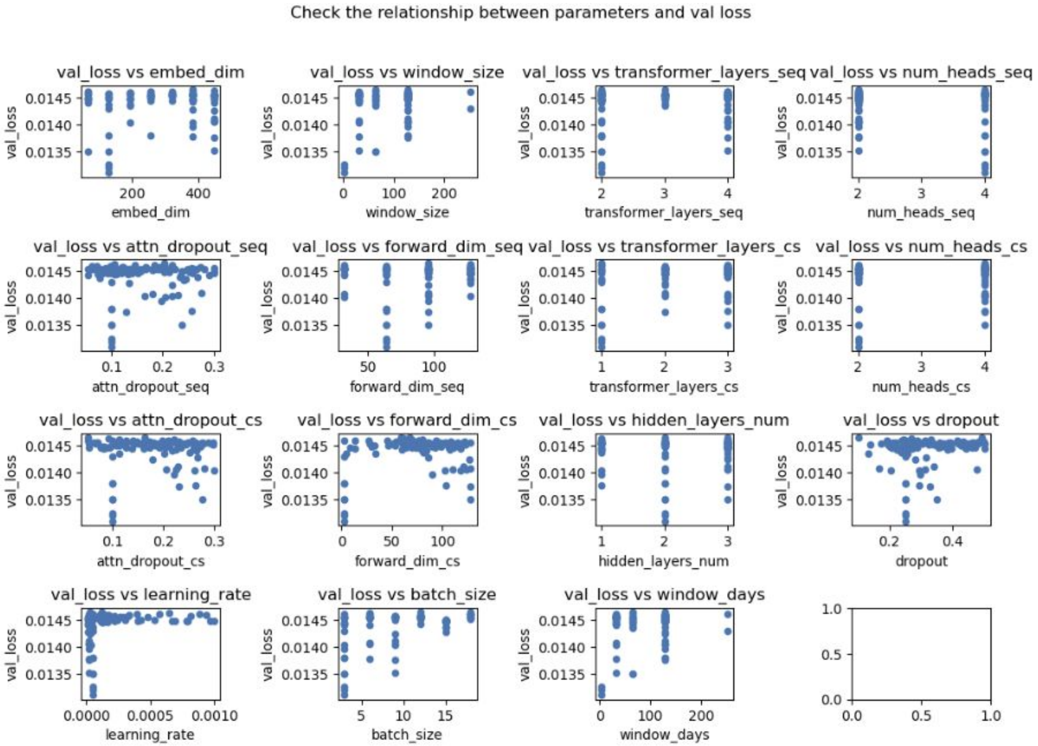
Build a machine learning–driven portfolio framework that empowers retail and institutional investors to navigate market regime shifts, capture dynamic trends, and preserve capital during periods of uncertainty

Appendix

Acknowledgements

- Tarun Sanghi: Senior Quant Expert at London Stock Exchange
  - Key contributions:
    - Validated our modeling approach and confirmed results are sensible

Deep Learning Modeling - Further Tuning



Comprehensive Hyperparameter Optimization (HPO)

- Two rounds of comprehensive HPO were done to search broader parameter space
- Smaller train/val loss is an indicator for good candidate solutions, but lower loss != good backtest results
- HPO prones to choose complicated model However they may diminish the security differentiation, take longer to train, and not necessarily a good candidate.
- At the end, our manually tuned cases excelled with simpler structure, better results