Predicting Financial Markets: New Modeling Paradigms

Wilson Kung

September 2015

Goal

Predict the financial market direction using macroeconomic and financial time series data

Data

- Predicting Next Day's Movement:
  - US Treasury Yields
  - S&P 500
  - GLD ETF
  - OIL ETF
- Features
  - 500 daily financial and macroeconomic time series
  - Source: St. Louis Federal Reserve FRED database

- 10 years of data: 2006 – Present

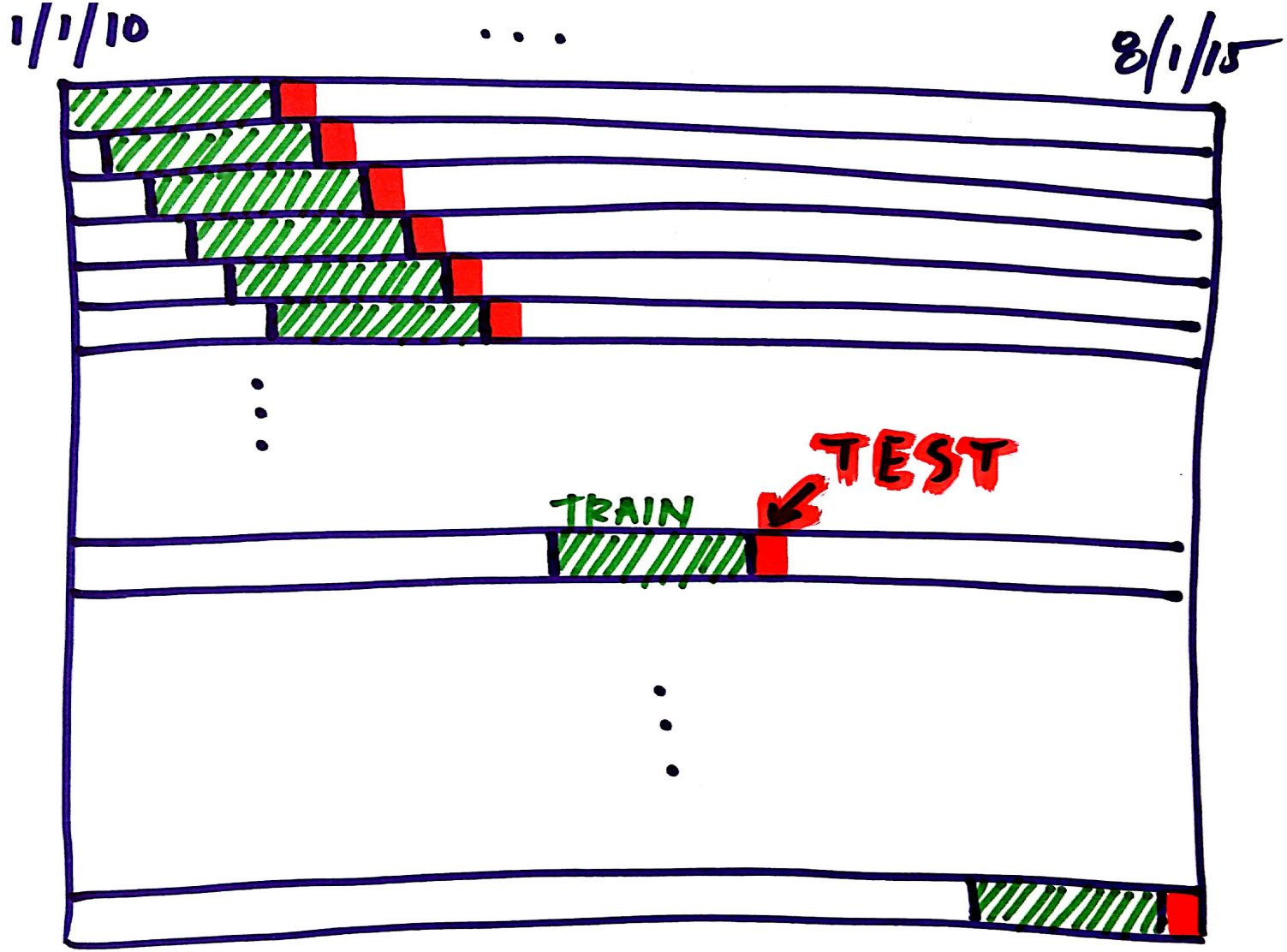
Data Issues and Preprocessing

- Avoid Look-Ahead Bias:
  - Shifted market movement by 1 day to predict market movement by day before's feature movements

- Trained on various rolling windows and tested on next one day only
  - mimics a trading strategy that updates model on a daily basis to make trading decision for the next day

- Standardized data
  - Subtracting mean and dividing by standard deviation

Training and Testing Methodology



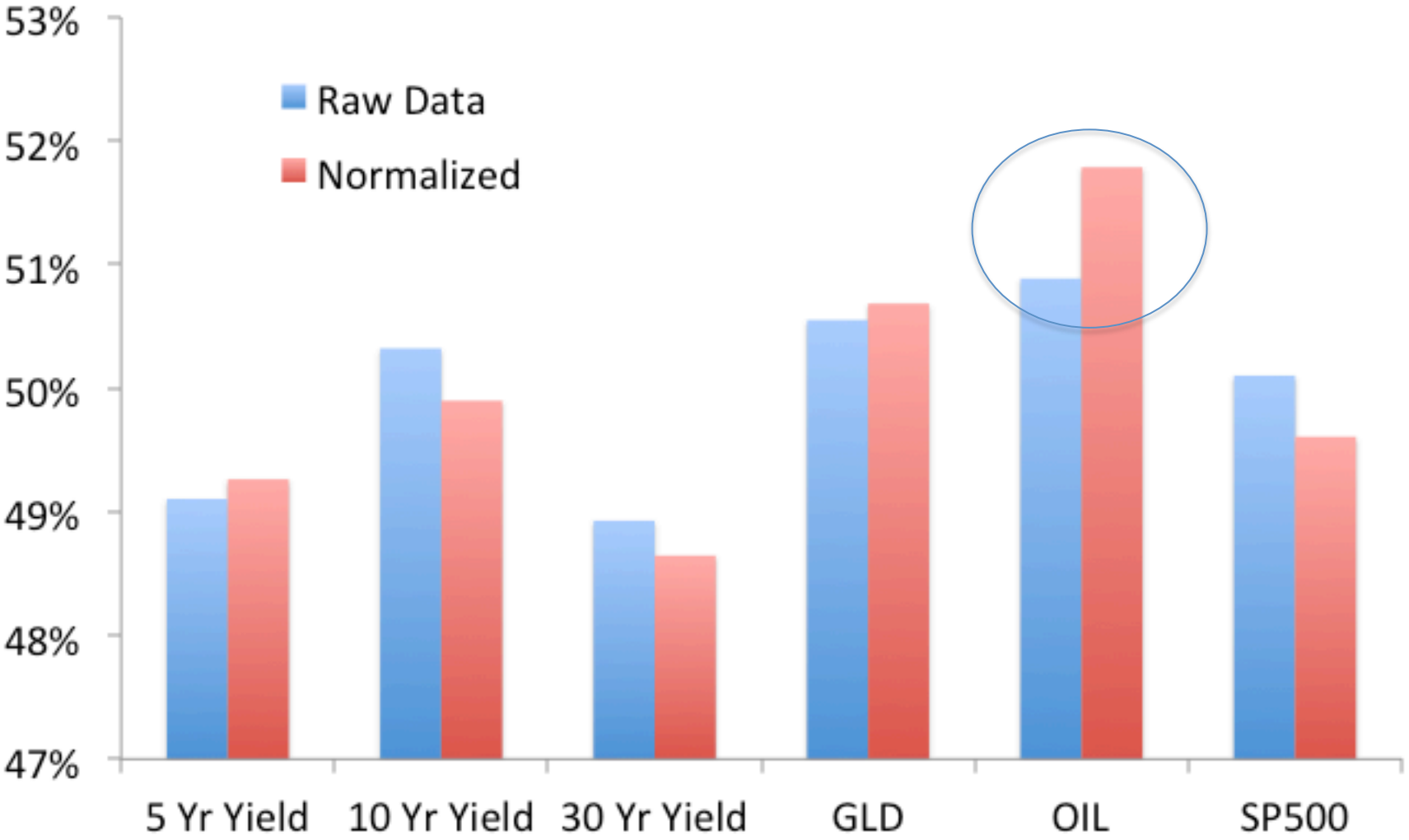
Methodology

- Find the market that is most predictable using standard classification algorithms
- Calculate accuracy of best performing algorithm on most predictable market
- Compare to accuracy of using various forms of neural nets

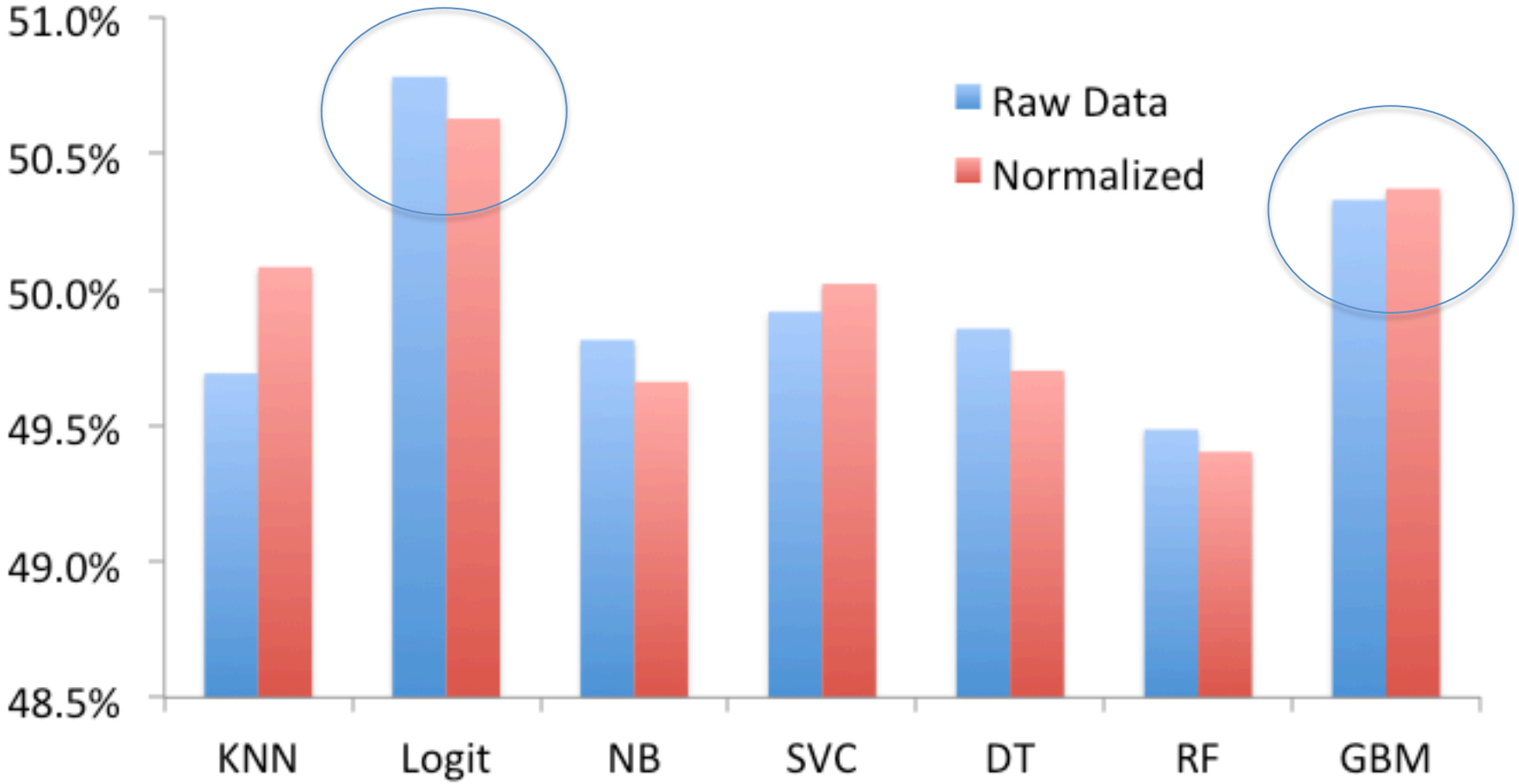
Supervised Learning Prediction Accuracies

		5 Yr Yield	10 Yr Yield	30 Yr Yield	GLD	OIL	SP500	Average
Raw Data								
	KNN	50.5%	51.2%	48.3%	49.0%	49.3%	49.8%	**49.7%**
	Logit	50.9%	51.2%	52.2%	50.8%	49.7%	49.9%	**50.8%**
	NB	47.0%	49.8%	48.7%	50.6%	51.8%	50.9%	**49.8%**
	SVC	47.8%	49.3%	47.8%	50.7%	51.4%	52.5%	**49.9%**
	DT	48.1%	49.7%	47.9%	52.6%	51.1%	49.7%	**49.9%**
	RF	49.8%	48.9%	49.6%	49.6%	51.5%	47.4%	**49.5%**
	GBM	49.5%	52.1%	47.9%	50.5%	51.4%	50.5%	**50.3%**
	Average	**49.1%**	**50.3%**	**48.9%**	**50.6%**	**50.9%**	**50.1%**	
Normalized Data								
	KNN	49.7%	50.1%	48.8%	50.7%	52.0%	49.2%	**50.1%**
	Logit	49.7%	50.0%	51.4%	49.9%	54.0%	48.9%	**50.6%**
	NB	47.7%	49.1%	47.8%	51.1%	52.1%	50.2%	**49.7%**
	SVC	49.1%	49.8%	48.6%	50.8%	50.9%	50.9%	**50.0%**
	DT	48.5%	49.0%	47.6%	52.3%	50.8%	50.0%	**49.7%**
	RF	50.1%	49.0%	48.7%	49.2%	51.2%	48.2%	**49.4%**
	GBM	50.0%	52.3%	47.6%	50.8%	51.6%	50.0%	**50.4%**
	Average	**49.3%**	**49.9%**	**48.6%**	**50.7%**	**51.8%**	**49.6%**	

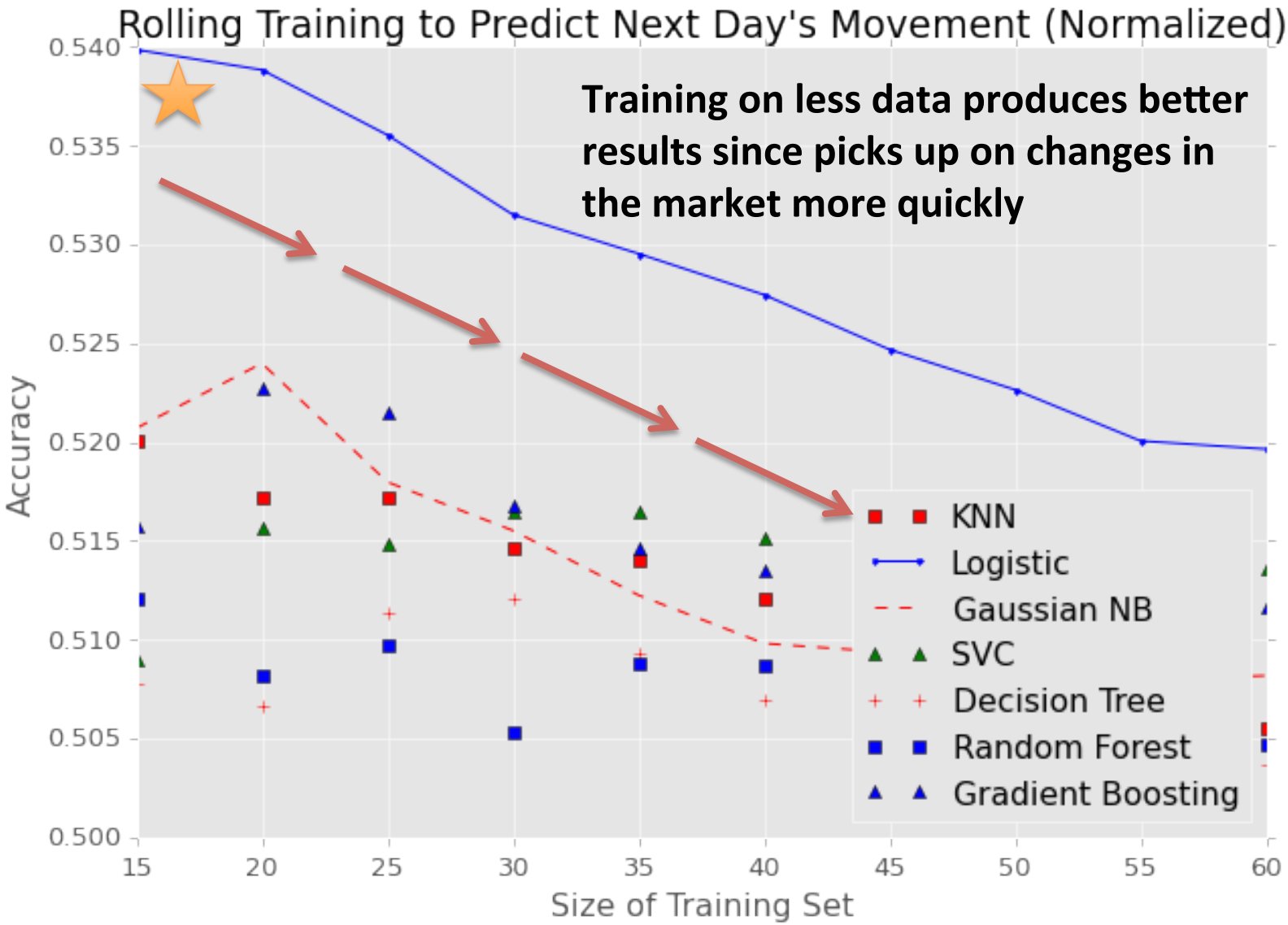
Average Accuracy By Market Across All Algorithms



Average Accuracy By Algorithm Across All Markets

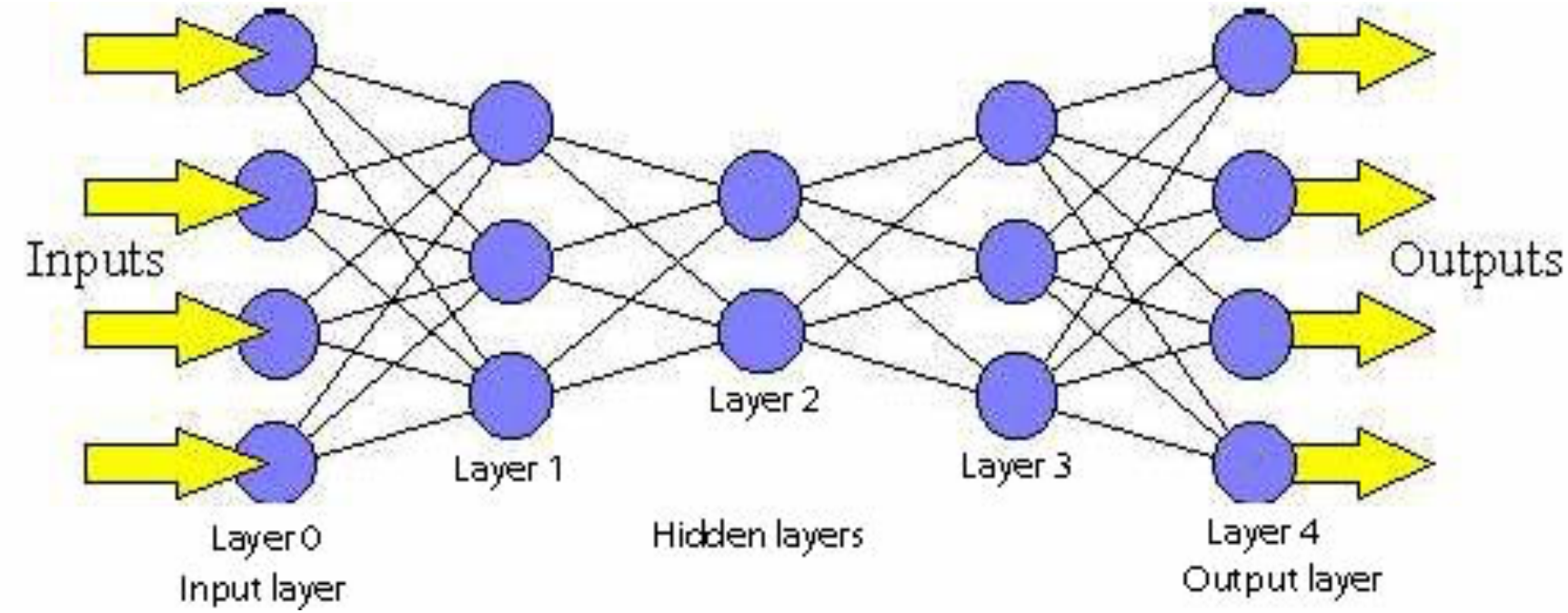


Prediction Accuracy for OIL ETF



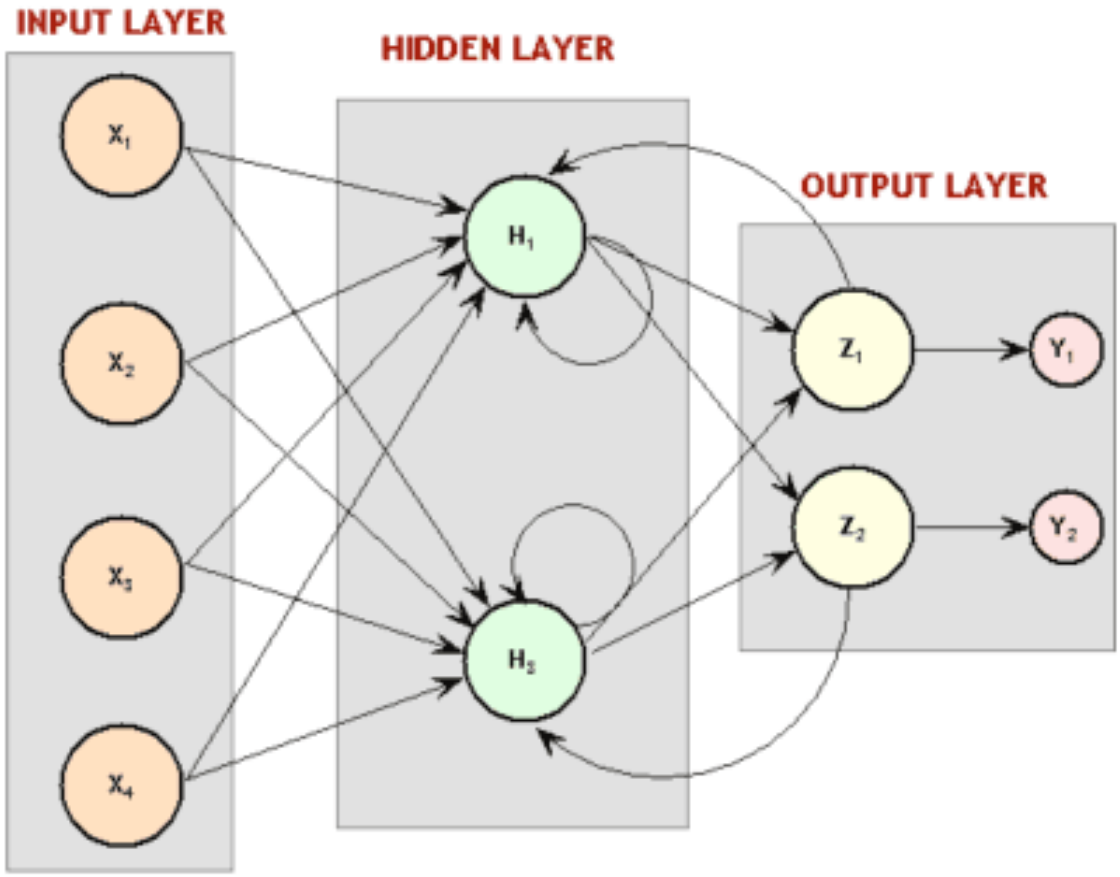
Can We Do Better with Neural Networks?

Feed Forward Neural Networks



Used Nolearn to implement Feed Forward Neural Networks

Recurrent Neural Networks



- Used PyBrain to implement Recurrent Neural Networks
- Reputation for good performance with time series data

Neural Network Accuracies for OIL prediction

- Feed Forward Neural Networks:
  - One, two, and three hidden layers
  - 200 through 800 nodes per layer
- 84 different structures
- Results:
  - They all performed about the same
  - Highest accuracy was 54.48% vs. 54% for Logistic Regression

Recurrent Neural Network Accuracies for OIL prediction

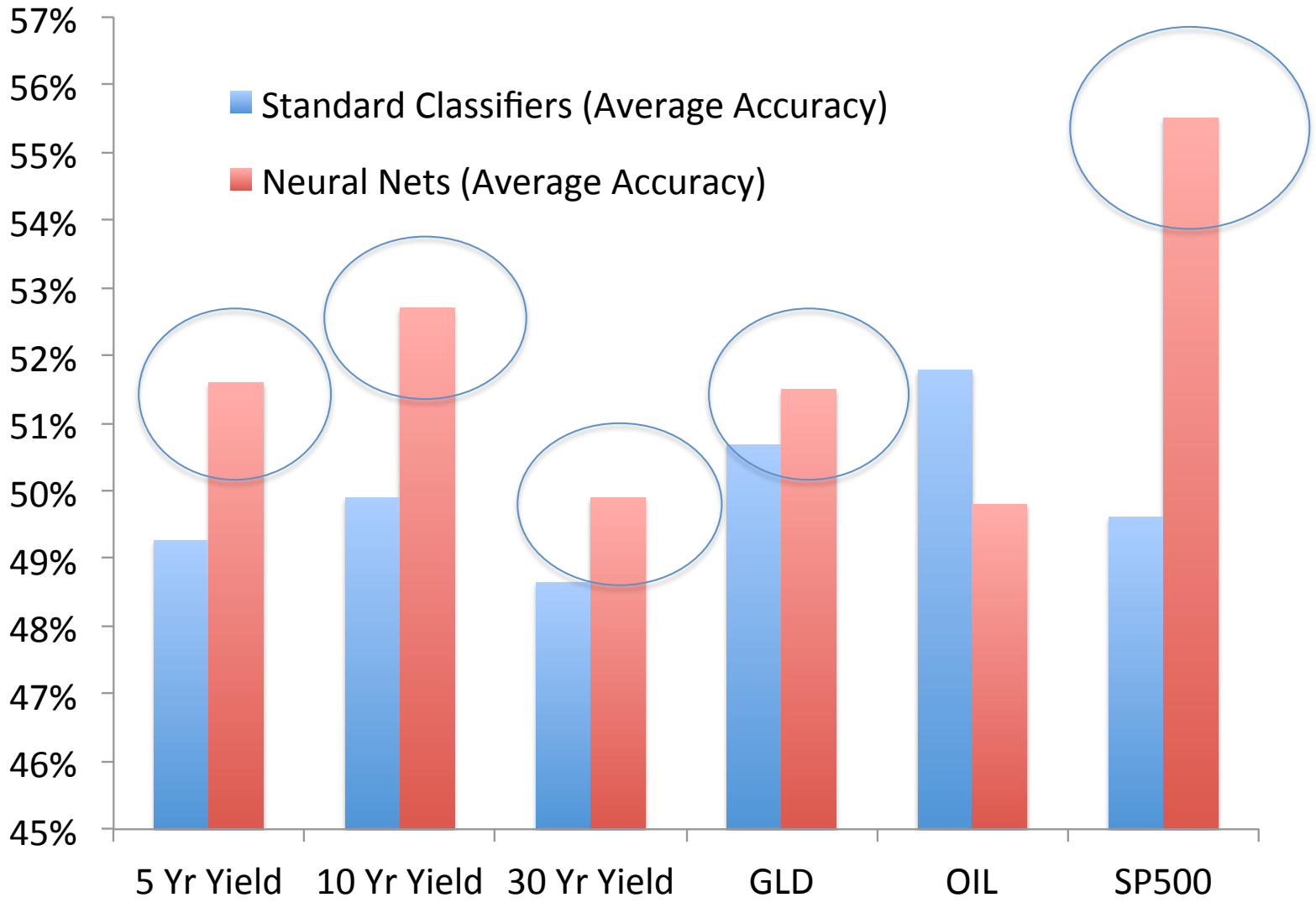
- Recurrent Neural Networks:
  - Two hidden layers
  - 25 through 400 nodes per layer
- 28 different structures
- Results:
  - They all performed about the same
  - Highest accuracy was 54.28% vs. 54.48% for Feed Forward Neural Networks

Neural Network* Accuracies for Different Markets

	Standardized		Raw		
	250, 250	**500, 50**	**250, 250**	**500, 50**	**Average**
5 Yr Yield	53.4%	52.1%	50.4%	50.4%	**51.6%**
10 Yr Yield	50.4%	50.9%	53.6%	56.1%	**52.7%**
30 Yr Yield	52.0%	50.2%	48.7%	48.6%	**49.9%**
GLD	52.7%	52.4%	50.2%	50.8%	**51.5%**
OIL	48.9%	50.9%	49.7%	49.7%	**49.8%**
S&P 500	52.6%	53.7%	54.7%	61.0%	**55.5%**
Average	**51.7%**	**51.7%**	**51.2%**	**52.8%**	

* Feed Forward: 2 Layers

Average Accuracy By Market Across All Algorithms



Average Accuracy By Algorithm Across All Markets