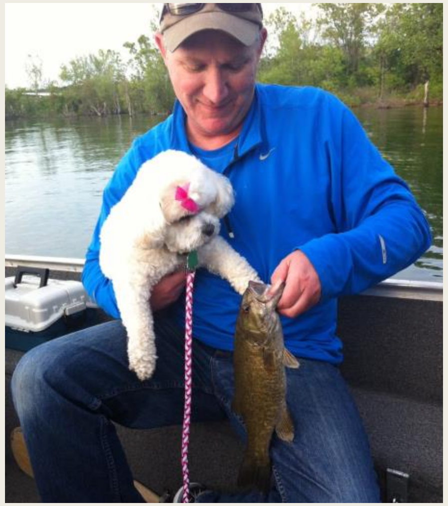
STOCK PRICE TREND PREDICTION USING EVOLUTIONARY LEARNING

Steven Wilson DSA 5303

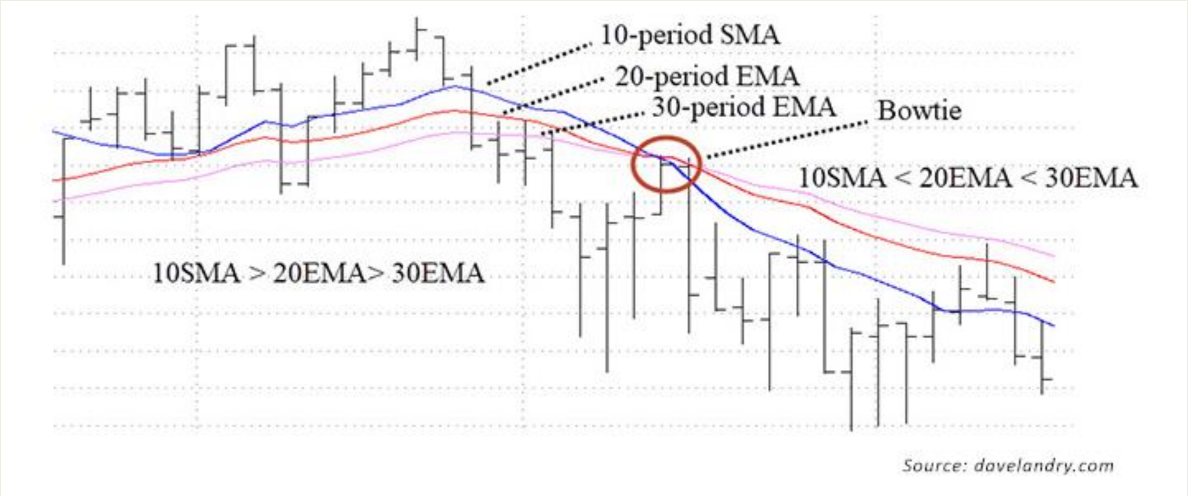
About Me

- 2nd year Data Science Master's Program
- Software Developer at Devon Energy
- BS Chemistry, MS Chemical Engineering
- Married 30 years, 2 adult children, 1 granddaughter
- Enjoy fishing and walking my dog on campus



Why This Project?

- 150% increase in retail investment since 2019
- Trend following is an investing approach that attempts to identify price trend changes and position investments to ride current trends for maximum gain.
- Common strategies
  - *Moving average crossovers*
  - Moving average convergence divergence (MACD)
  - Relative Strength Index (RSI)
- Weaknesses
  - *Lagging indicators*
  - *Rapid price changes*
  - *Support and resistance levels*

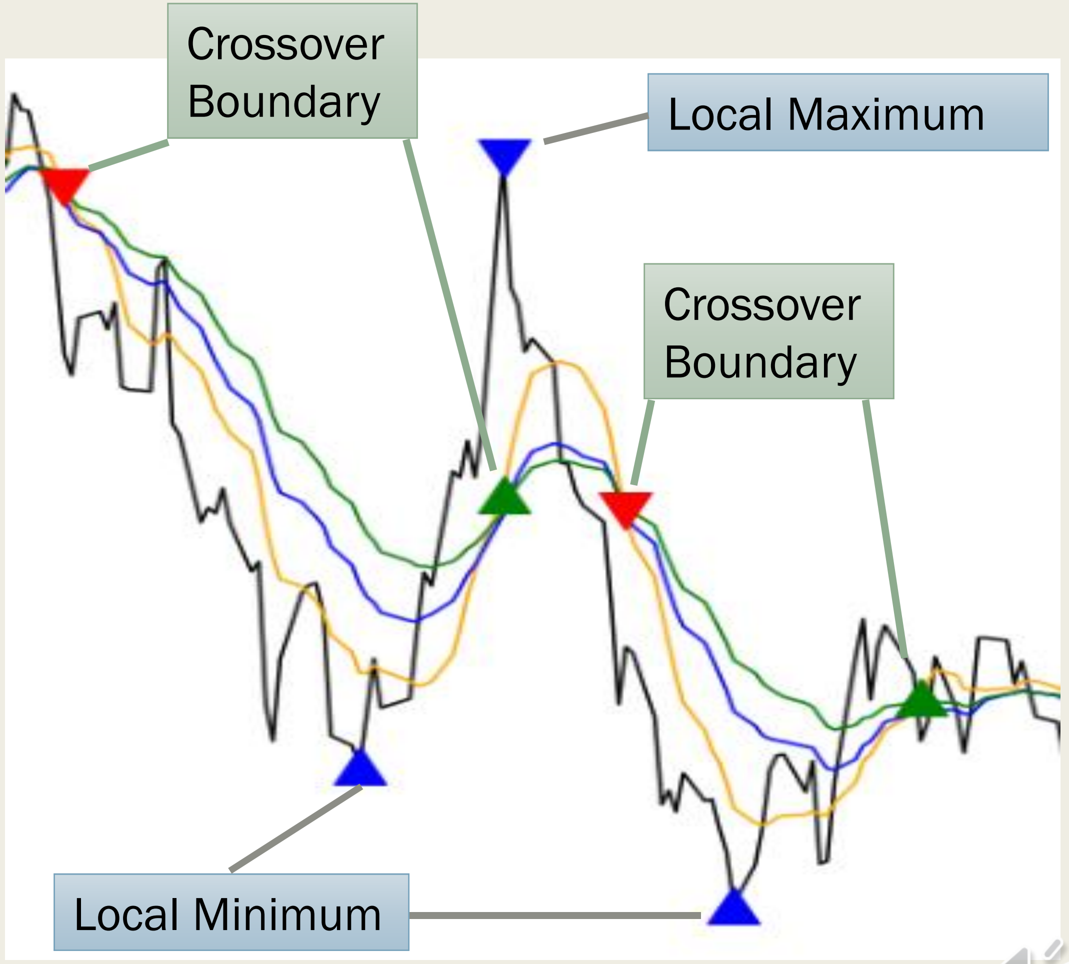


Project Goals

- Use an evolutionary learning process to:
  - ***Develop a model to predict stock price trends*** *earlier and more accurately than the bowtie moving average crossover strategy.*
  - ***Evaluate if the model generalizes*** *across industry sectors.*
  - ***Construct an optimum portfolio*** *based on average model returns.*

Data Collection and Preparation

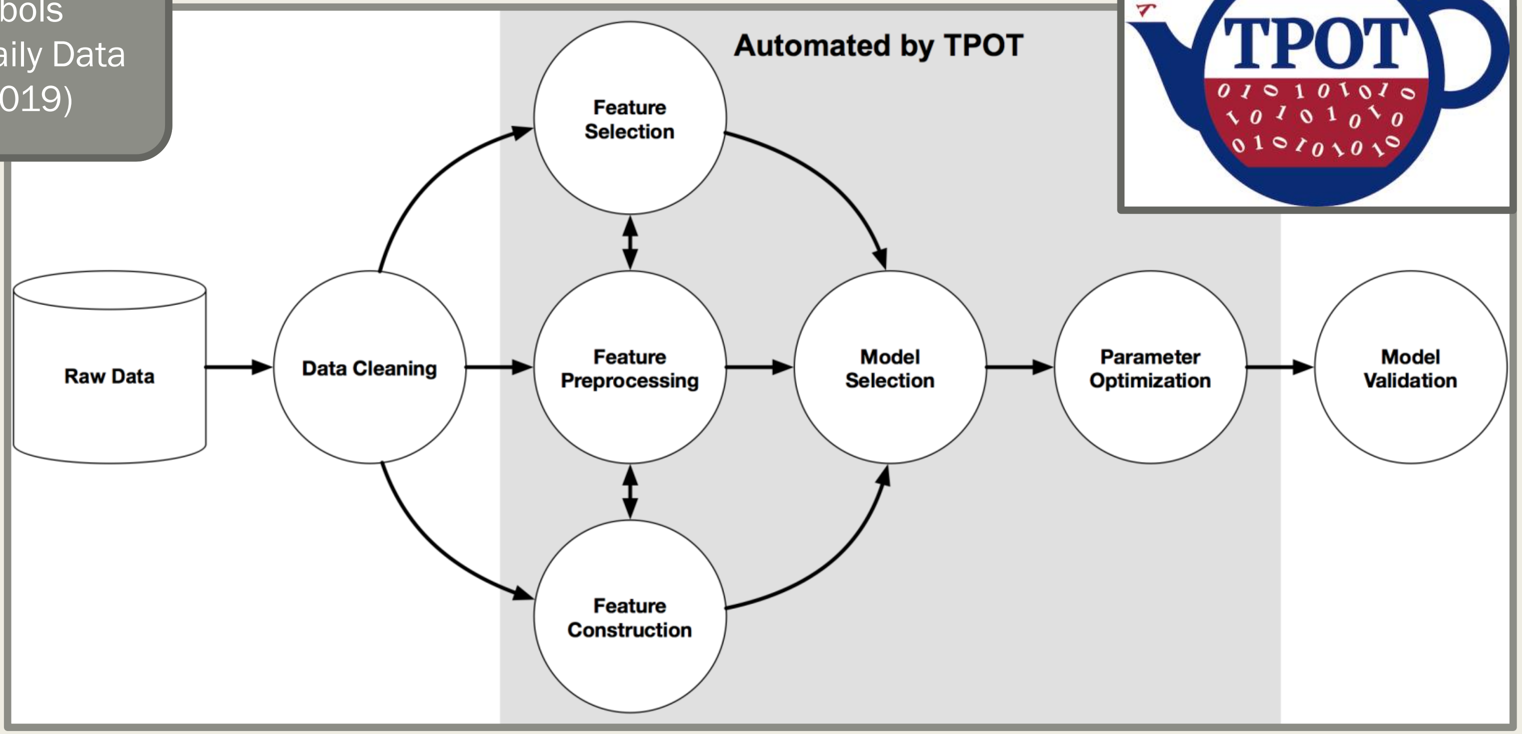
- Collected 10 years of daily price data for S&P 500 companies (2009-2019)
  - *Symbol, Date, High, Low, Open, Close, Volume*
- Added calculated values for trending
  - *SMA10, EMA20, EMA30, MACD, RSI*
- Added binary labels for target values
  - *0 = downtrend (max to min)*
  - *1 = uptrend (min to max)*

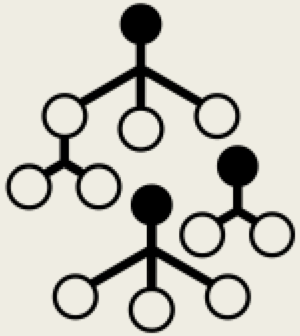


Model Selection

"TPOT is a python Automated Machine Learning tool that optimizes machine learning pipelines using genetic programming." (source TPOT)

20 Symbols
S&P 500 Daily Data
(2009-2019)

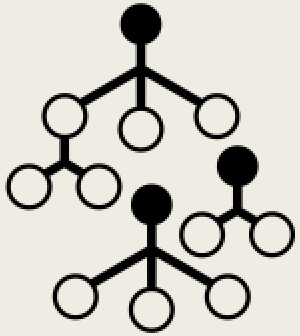




Random Forest
Classifier

Model Validation

477 Symbols
S&P 500 Daily Data
(2009-2019)



Random Forest
Classifier

```
RandomForestClassifier
     (bootstrap=True,
     criterion="gini",
     max_features=0.65,
     min_samples_leaf=10,
     min_samples_split=16,
     n_estimators=100)
```

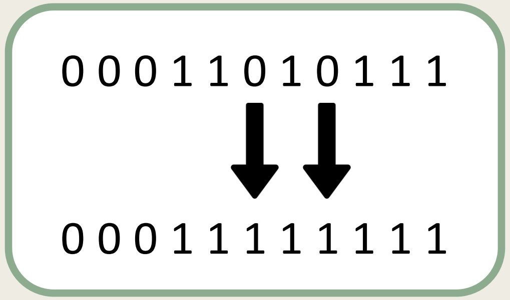
Accuracy 0.77

	Predicted Down	Predicted Up
Actual Down	89906	47470
Actual Up	30905	176941

	precision	recall	f1-score
0	0.74	0.66	0.69
1	0.79	0.84	0.81

Signal Smoothing Factor

- Model initially suffered from excessive trading
  - *Sensitive to short term trend changes inside of longer-term trends*
  - *Negatively impacted performance*
- Added signal smoothing factor to overcome this problem
  - *Prevents trading on two consecutive days when opposite signal occurs inside of longer-term pattern*
  - *Improved average model returns by 13.4%*

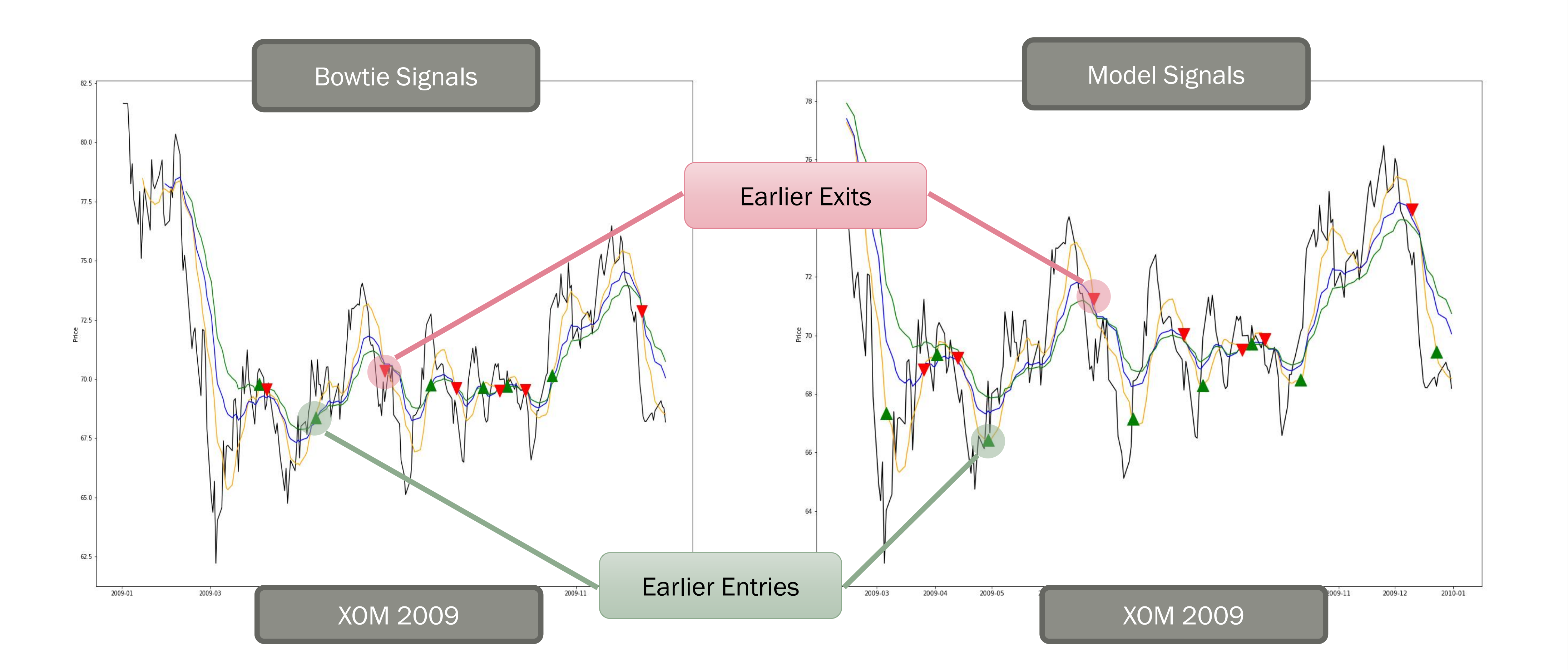


Experimentation

- 23 symbols from 11 sectors
  - *All excluded from training symbols to prevent data leakage*
- Daily price data (2009-2019)
- Assumptions
  - *Based on 100 share trades*
  - *Trades executed at closing prices*
  - *Short selling was not allowed*
  - *All positions closed at year end*
  - *Positions bought if first signal of year was a sell signal*
- Backtests on the bowtie strategy vs model comparing annual returns

Sector	Symbol	Company Name
Energy	XOM	Exxon Mobil
Energy	CVX	Chevron
Utilties	DUK	Duke Energy
Utilties	ED	Consolidated Edison
Utilties	AEP	American Electric Power
Materials	SHW	Sherwin-Williams
Materials	DD	DuPont
Industrials	BA	Boeing
Industrials	UNP	Union Pacific
Consumer Discretionary	AMZN	Amazon
Consumer Discretionary	MCD	McDonald's
Consumer Staples	KO	Coca-Cola
Consumer Staples	PG	Proctor & Gamble
Healthcare	UNH	UnitedHealth Group
Healthcare	JNJ	Johnson & Johnson
Financials	BRK-A	Berkshire Hathaway
Financials	JPM	JPMorgan Chase
Information Technology	AAPL	Apple
Information Technology	MSFT	Microsoft
Communication Services	FB	Facebook
Communication Services	GOOG	Google
Real Estate	AMT	American Tower
Real Estate	SPG	Simon Property Group

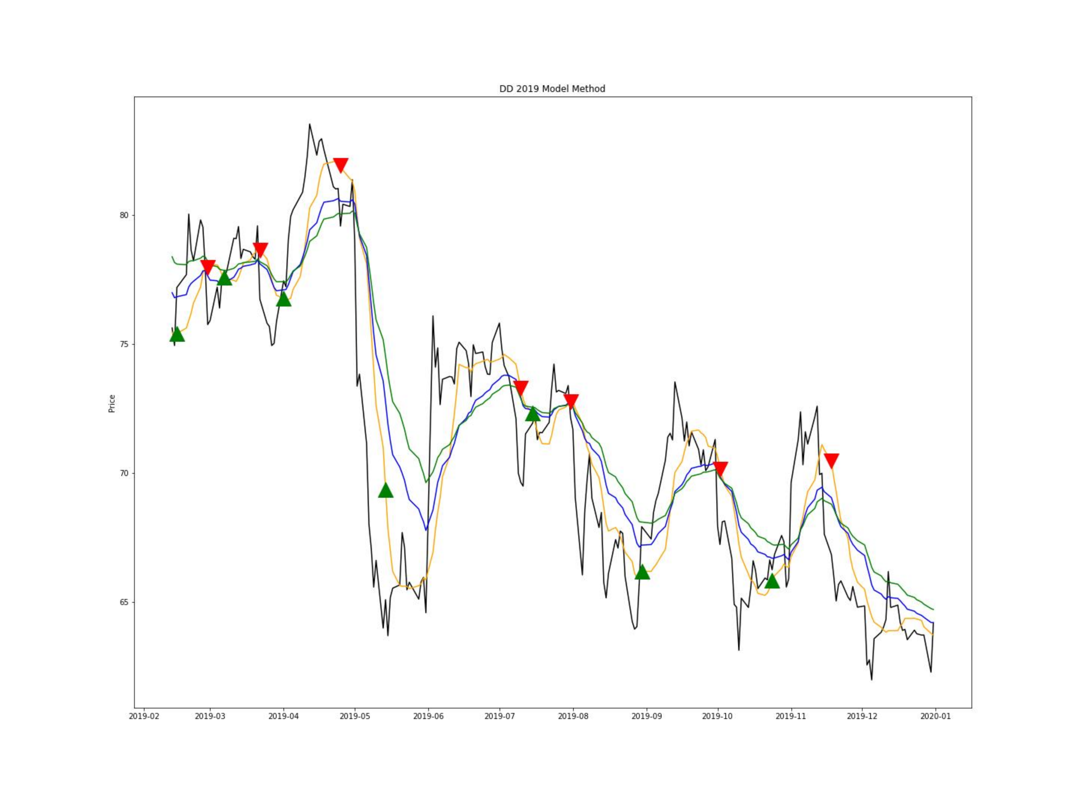
Signal Plots



Quantitative Results

- Sample Results shown from 2019

Sample	Sector	Symbol	Year	Bowtie	Model	Delta
10	Energy	XOM	2019	-12.57	3.11	15.68
21	Energy	CVX	2019	-10.69	3.67	14.36
32	Utilities	DUK	2019	0.44	9.23	8.79
43	Utilities	ED	2019	13.30	20.17	6.88
54	Utilities	AEP	2019	15.00	17.07	2.07
65	Materials	SHW	2019	28.27	20.65	-7.62
76	Materials	DD	2019	-37.52	5.95	43.47
87	Industrials	BA	2019	-44.79	-11.46	33.33
98	Industrials	UNP	2019	-6.72	19.42	26.14
109	Consumer Discretionary	AMZN	2019	5.63	-0.78	-6.41
120	Consumer Discretionary	MCD	2019	15.29	24.66	9.37
131	Consumer Staples	KO	2019	6.52	10.42	3.90
142	Consumer Staples	PG	2019	13.81	12.91	-0.91
153	Healthcare	UNH	2019	11.96	18.30	6.34
164	Healthcare	JNJ	2019	2.57	6.41	3.84
175	Financials	BRK-A	2019	5.42	8.67	3.24
186	Financials	JPM	2019	12.23	39.70	27.47
197	Information Technology	AAPL	2019	70.27	66.66	-3.61
208	Information Technology	MSFT	2019	47.33	45.80	-1.53
216	Communication Services	FB	2019	10.24	34.63	24.39
224	Communication Services	GOOG	2019	19.13	13.69	-5.44
232	Real Estate	AMT	2019	24.58	28.60	4.03
240	Real Estate	SPG	2019	-16.04	-10.77	5.26
					Average	9.26
					Deviation	13.18



Signals plotted for each symbol and year

Hypothesis Testing

Hypothesis 1: A predictive model exists that can identify trend changes earlier and more accurately than trend trading strategies using the bowtie moving average crossover.

Null	Hypothesis
$H_0: \delta \leq 0$	$H_1: \delta > 0$

$$where\ \delta = R_M - R_B$$

$$R_M = Return\ from\ predictive\ model$$

$$R_B = Return\ from\ bowtie\ method$$

Hypothesis 2: The predictive model generalizes across different sectors of the market.

Null	Hypothesis
$H_0: \delta_S \leq 0$	$H_1: \delta_S > 0$

$$where\ \delta_S = R_M - R_B$$

$$R_M = Return\ from\ predictive\ model\ for\ the\ sector$$

$$R_B = Return\ from\ bowtie\ method\ for\ the\ sector$$

Sector	n	$\bar{\delta}$	σ	Z score	P value	Test Result
All	241	11.69	15.99	11.33	0.000	$Reject\ H_0$
Energy	22	11.54	6.08	8.90	0.000	$Reject\ H_0$
Utilities	33	8.70	5.15	9.70	0.000	$Reject\ H_0$
Materials	22	16.64	36.54	2.14	0.016	$Reject\ H_0$
Industrials	22	17.59	17.94	4.60	0.000	$Reject\ H_0$
Consumer Discretionary	22	10.09	14.26	3.32	0.000	$Reject\ H_0$
Consumer Staples	22	8.81	7.50	5.51	0.000	$Reject\ H_0$
Healthcare	22	13.37	14.13	4.44	0.000	$Reject\ H_0$
Financials	22	10.45	13.38	3.66	0.000	$Reject\ H_0$
Information Technology	22	12.94	10.65	5.70	0.000	$Reject\ H_0$
Communication Services	16	7.10	14.19	2.00	0.023	$Reject\ H_0$
Real Estate	16	11.59	9.31	4.98	0.000	$Reject\ H_0$

Portfolio Optimization

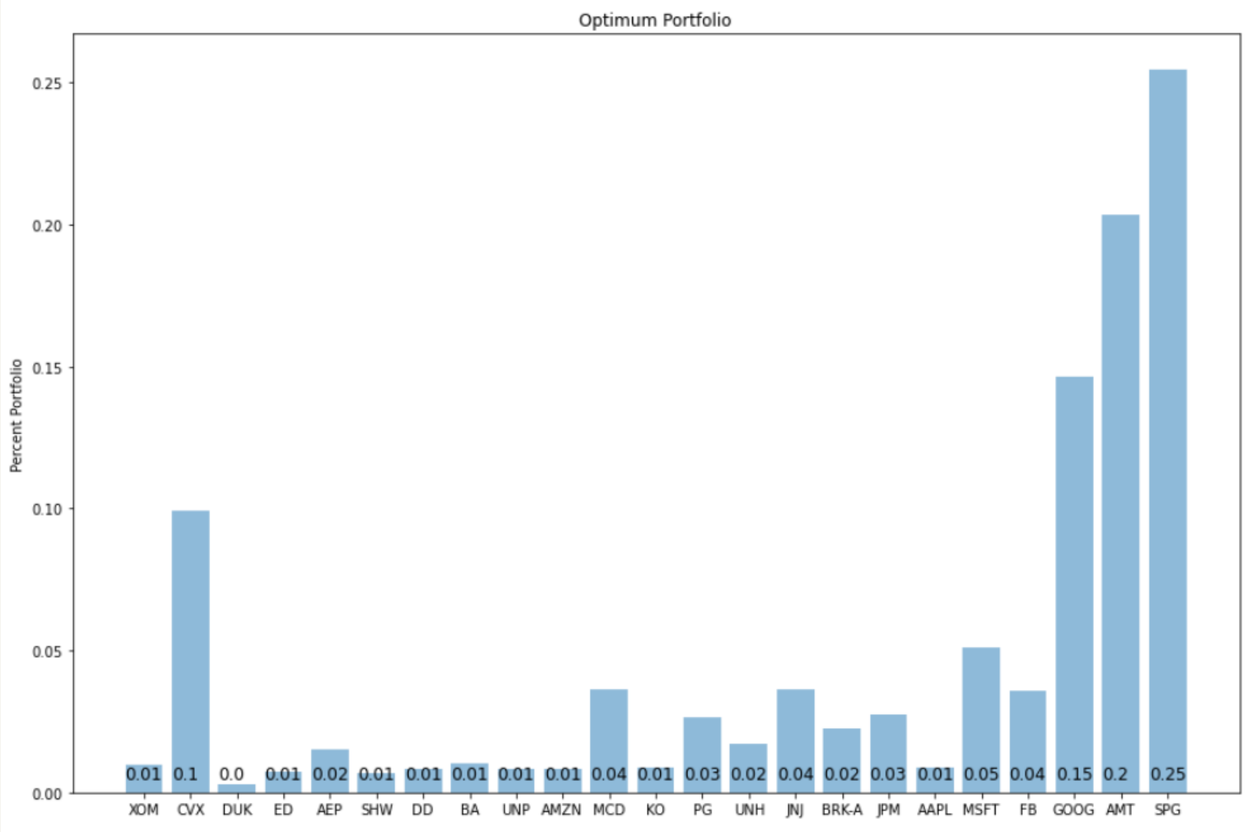
Array-Based Genetic Algorithm

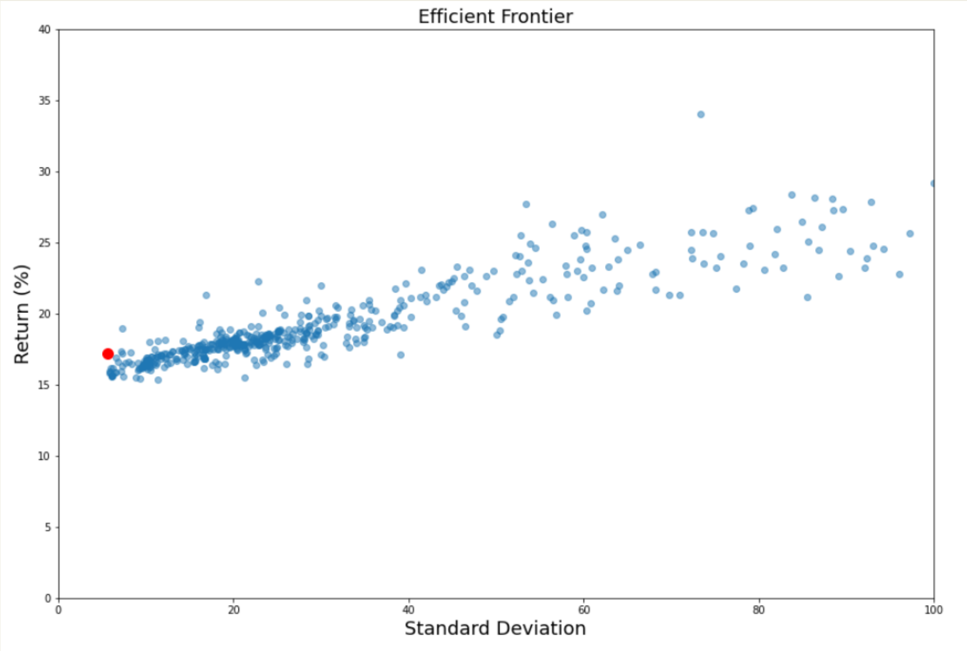
Chromosome Structure	One dimensional array of real numbers
Evolutionary Operators	Linear mutation and arithmetic crossover
Selection Method	Tournament selection
Fitness Evaluator	Sharpe ratio

Parameter Settings

Population	50
Generations	10
Crossover Rate	0.8
Mutation Rate	0.2
Elitism	None

Optimum Portfolio





Conclusions

- A genetic algorithm successfully found a random forest classifier that outperformed the bowtie strategy for 23 stocks from 11 different industry sectors with 99% confidence.
- Signal smoothing following classification significantly reduced model sensitivity to short term trends inside of longer-term trends.
- An array-based genetic algorithm can successfully optimize a portfolio using the model returns and Sharpe ratio as the fitness function.
- Potential future study
  - *Include short sales to determine the impact on model performance.*
  - *Evaluate the genetic algorithm approach on other financial instruments (commodities, futures, options)*

Resources

- Final project files including notebooks, input files and results
  - *https://github.com/nosliwes/dsa5303*
- LinkedIn
  - *https://www.linkedin.com/in/steven-wilson-0072166/*
- Questions or Comments
  - *steven.wilson@ou.edu*