Data based stock portfolio construction using Computational Intelligence

Asimina Dimara and
Christos-Nikolaos Anagnostopoulos

Data Economy workshop:
How online data change economy and business

Introduction

Investment process for constructing a stock portfolio is divided into three basic stages:

- Stock selection.
- Portfolio Management.
- Portfolio Optimization.

Objective

- Exploit published data.
- New full scale model for all stages using Computational Intelligence.
- Easy approachable.
- Adjustable to investor’s preferences.

Basic Stock valuation Methods

- **Fundamental Analysis** examines the basics of a company.
- **Technical Analysis** focuses on charts and past price behaviors.

Technical Vs Fundamental

Technical Analysis	Fundamental Analysis
External and internal factors are already priced into the stock. Price moves in trends and history tends to repeat itself.	The health and performance of a company is determined by looking economic indicators.
Patterns on charts to predict price movements.	Predict future performance using financial statements.
Short term	Long term
Trading	Investing

Predict price movements

Stock selection

How to eliminate disadvantages

Technical Analysis	Fundamental Analysis
Time consuming to gather data. Pattern recognition based on the opinion of the observer.	Time consuming to gather, calculate and compare fundamental data for each company.
Exploit published data CSS,XML	**Exploit published data Excel,XML**
Artificial Neural Networks to recognize past price patterns.	Genetic Algorithm to compare data fast.

Stage 1-Stock Selection using GA

Stock selection problem: Decide which combination of stocks will give the better return with the less possible risk.

$$\binom{2000}{8} = \frac{2000!}{8!(2000-8)!} = 626082701855652272 4750$$

Stock Valuation

- Boost portfolio returns by picking promising stocks.
- Maintain a proper degree of diversification.
- Investment using valuation ratios generates higher returns.

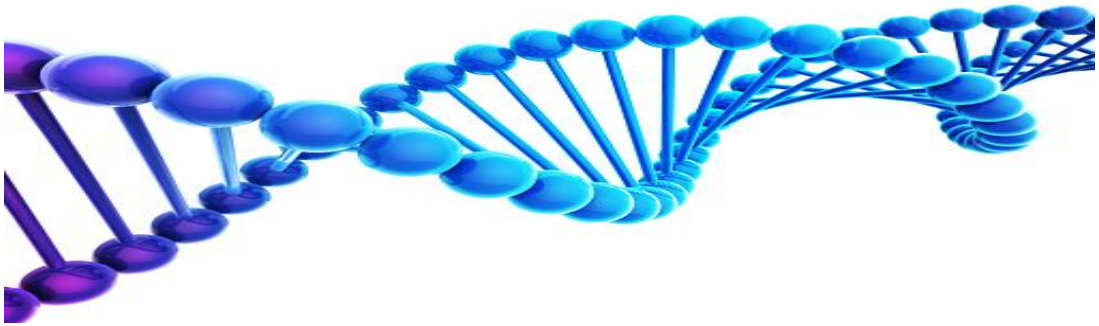
(Chin, Jim Y. F., Andrew K. Prevost, and Aron A. Gottesman. 2002. "Contrarian Investing in a Small Capitalization Market: Evidence from New Zealand).

- Exploit published data.
- Take account of investor's need and expectations.

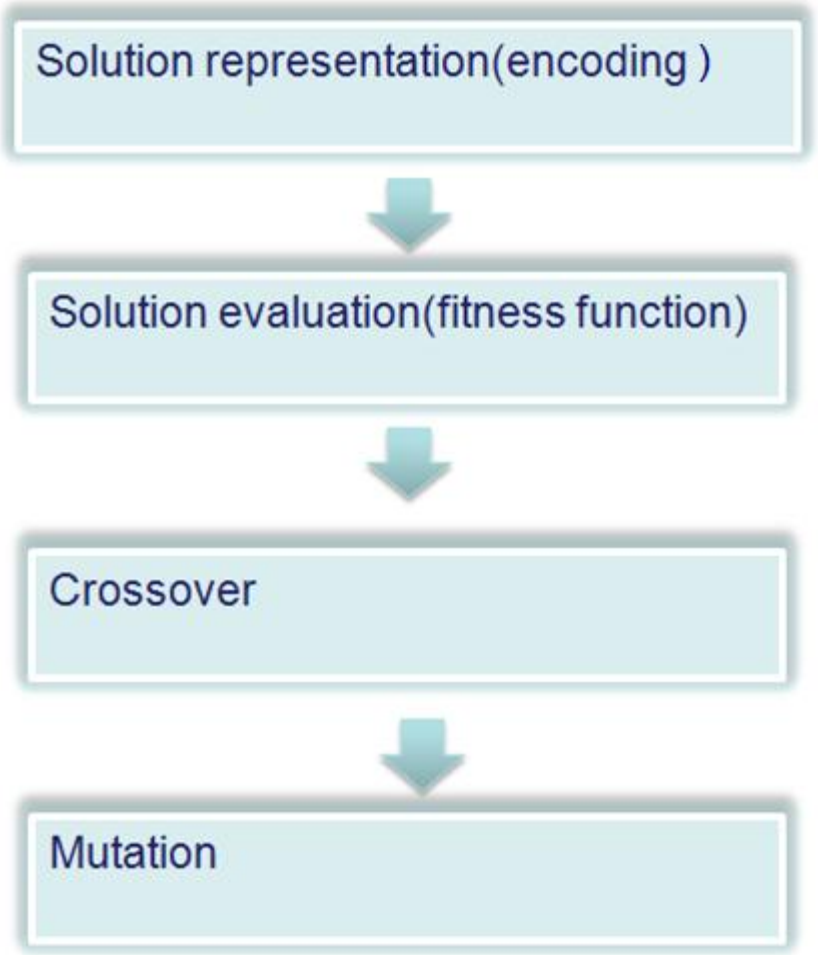
Stock evaluation using Genetic Algorithm

We can include all the precedent with the **Genetic Algorithms (GA)** through fitness function.

Genetic algorithms are a heuristic method inspired by the principles of natural selection and genetics. They attempt to find the optimal solution to a problem by handling a population of potential solutions.



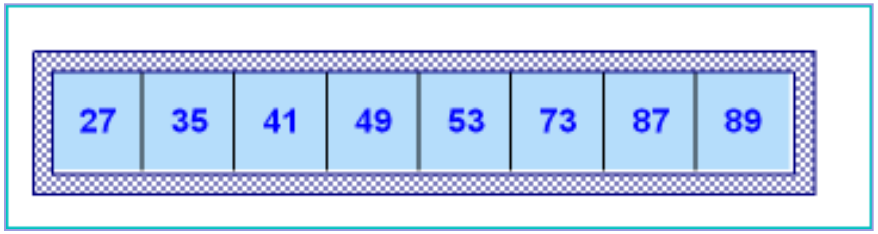
Basic steps of Genetic Algorithms



Encoding of the stock selection problem

We want to choose which company's stock we will select. A gene must represent a company.

chromosome

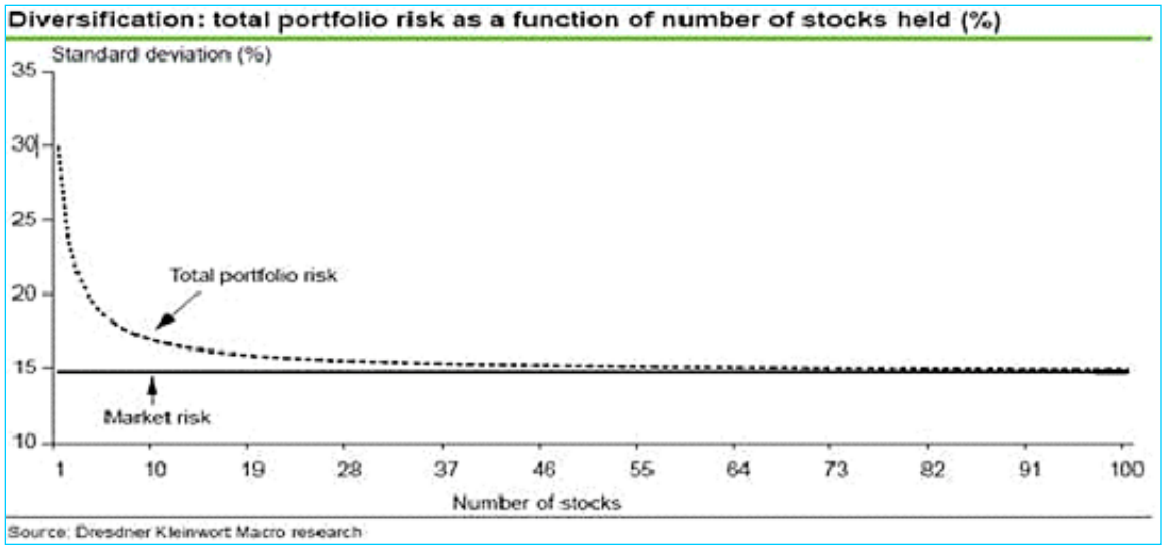


represents

code	Company
27	3I GROUP
35	ADMIRAL GROUP PLC
41	ANGLO AMERICAN
49	ANTOFAGASTA
53	ARM HLDGS
73	ASHTEAD GROUP
87	ASSOCIATED BRITISH FOODS
89	ASTRAZENECA PLC

Chromosome's length

Chromosome's length is the number of company's stocks we want to include in our portfolio. Depend on research to decide how many or select the size by investor's preference. We propose 15.



(John L.Evans and Stephen H. Archer(1968): Diversification and the reduction of dispersion: an empirical analysis. The Journal of Finance.

Evaluation Function

- Company evaluation based on provided ratios by FTSE100.
- Diversification.

FTSE 100

SUMMARY | INTERACTIVE CHART | CONSTITUENTS | NEWS

London Stock Exchange

Page 1 of 6 | 1 | 2 | 3 | 4 | 5 | 6 | Next

Code	Name	Cur	Price	+/-	%+/-
III	3I GRP.	GBX	898.50	-6.00	-0.66
ABF	A.B.FOOD	GBX	2,970.00	-27.00	-0.90
ADM	ADMIRAL GRP	GBX	2,031.00	+13.00	+0.64
AAL	ANGLO AMERICAN	GBX	1,051.00	-5.00	-0.47
ANTO	ANTOFAGASTA	GBX	815.50	+0.50	+0.06

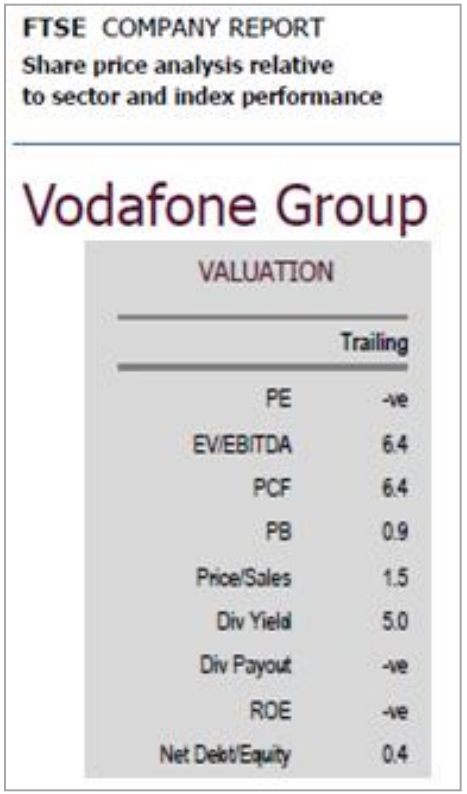
Company Evaluation using ratios

- Ratio prices provided by Analytic report.
- Each ratio's prices are grouped at intervals.

$$\text{int}\,erval_{ratio} = \frac{\max_{ratio} - \min_{ratio}}{10}$$

- Each ratio is characterized as low or high and evaluated depending on the interval.
- All ratios are summed.

$$ratioEval = \sum_{i=1}^{9} ratio_i$$



FTSE COMPANY REPORT

Share price analysis relative to sector and index performance

Vodafone Group

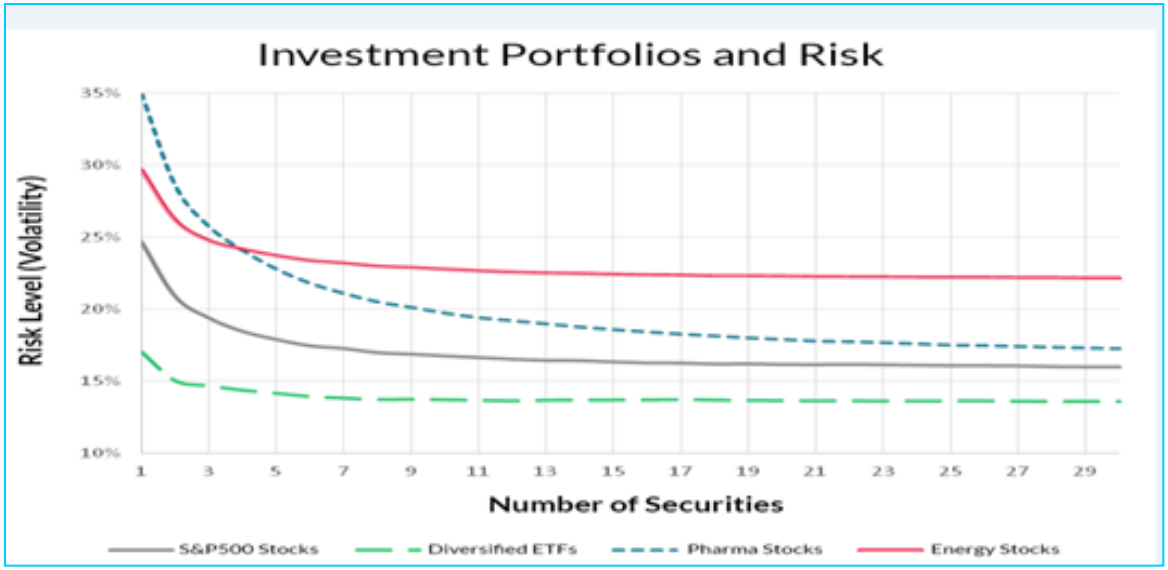
VALUATION

	Trailing
PE	-ve
EV/EBITDA	6.4
PCF	6.4
PB	0.9
Price/Sales	1.5
Div Yield	5.0
Div Payout	-ve
ROE	-ve
Net Debt/Equity	0.4

Evaluation of diversification

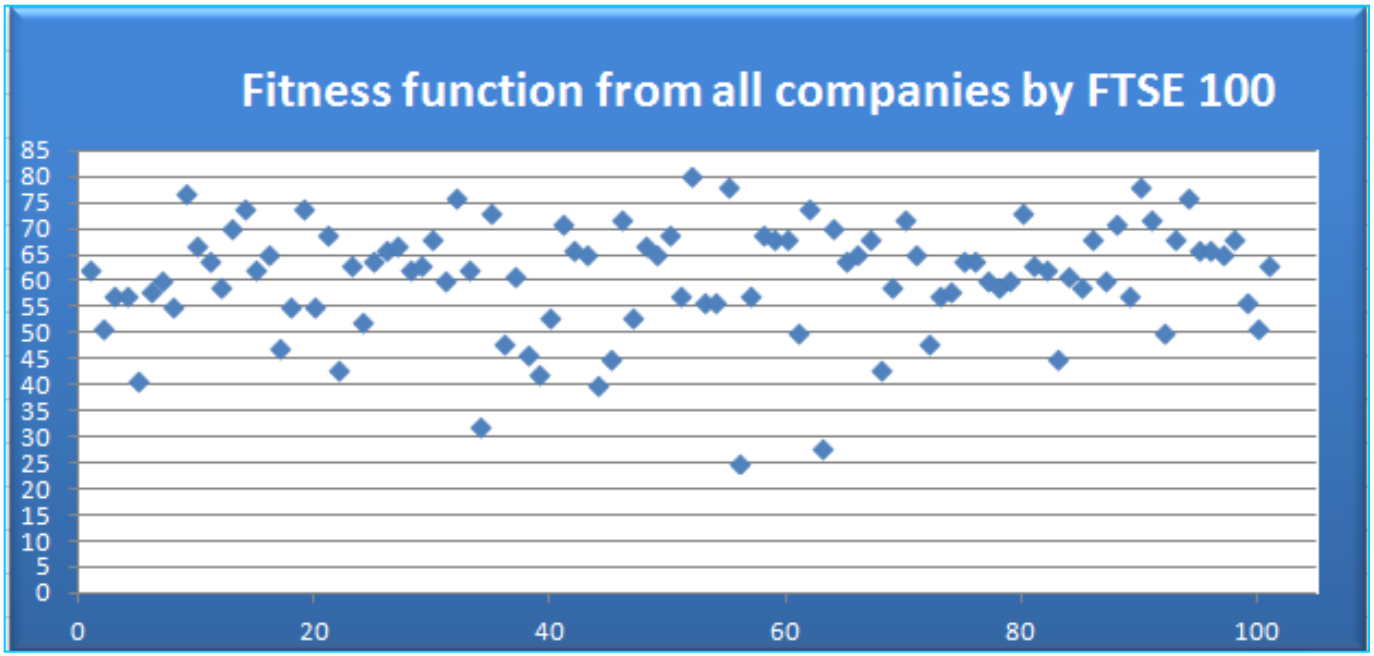
Avoid investing in securities with high covariances among themselves. It is best to have companies that belong to different sectors at a portfolio thus in the chromosome. In case all the companies belong to a different sector then the chromosome is given the best score.

(H.Markowitz," Portfolio Selection - Efficient diversification of Investments, 1959)



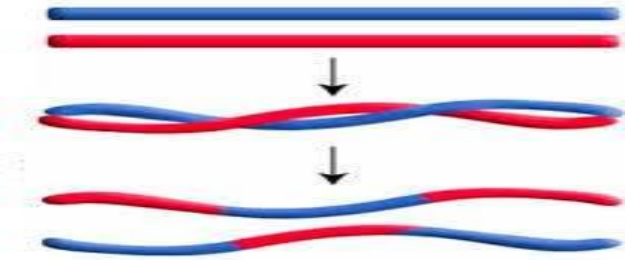
Final fitness function

$$f = ratioEval + diversification$$

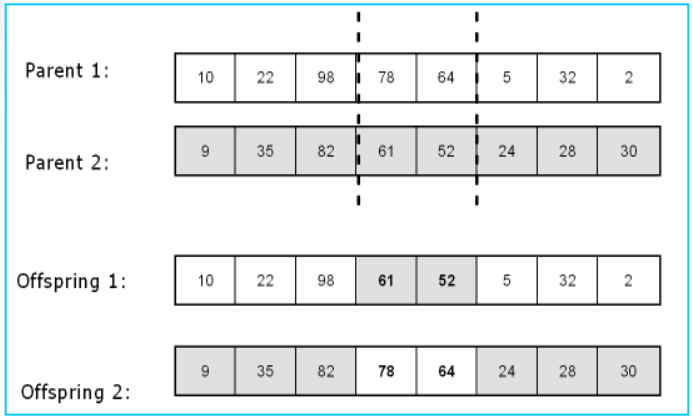


Crossover

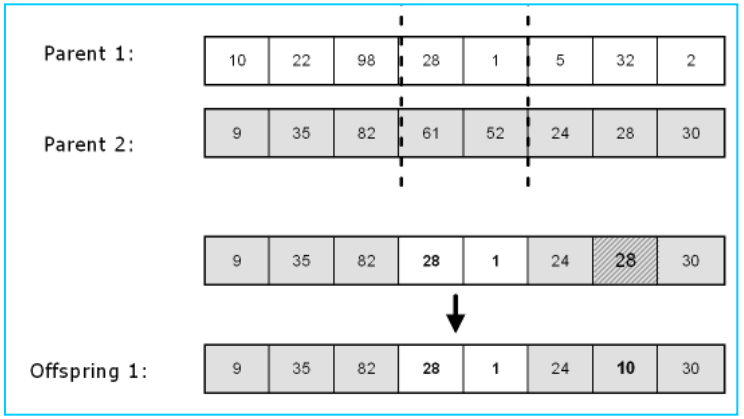
Randomized exchange of genetic material between solutions, with the possibility that "good" solutions can generate "better" ones.



Two point crossover

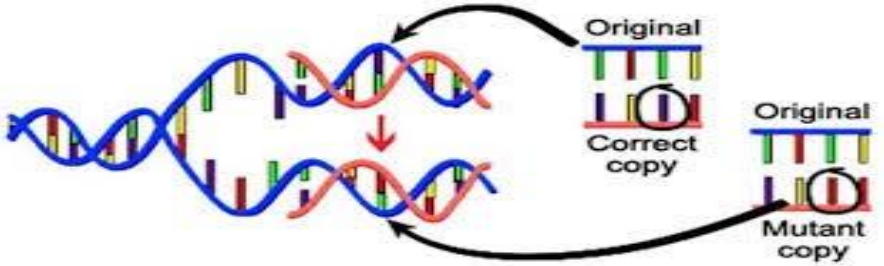


Modification of OX1 crossover



Mutation

Randomly disturbing genetic information for recovering the lost genetic materials.



Random number change

before mutation	10	22	98	28	1	5	32	2
after mutation	10	**45**	98	28	1	5	32	2

Stage 2- Portfolio Management

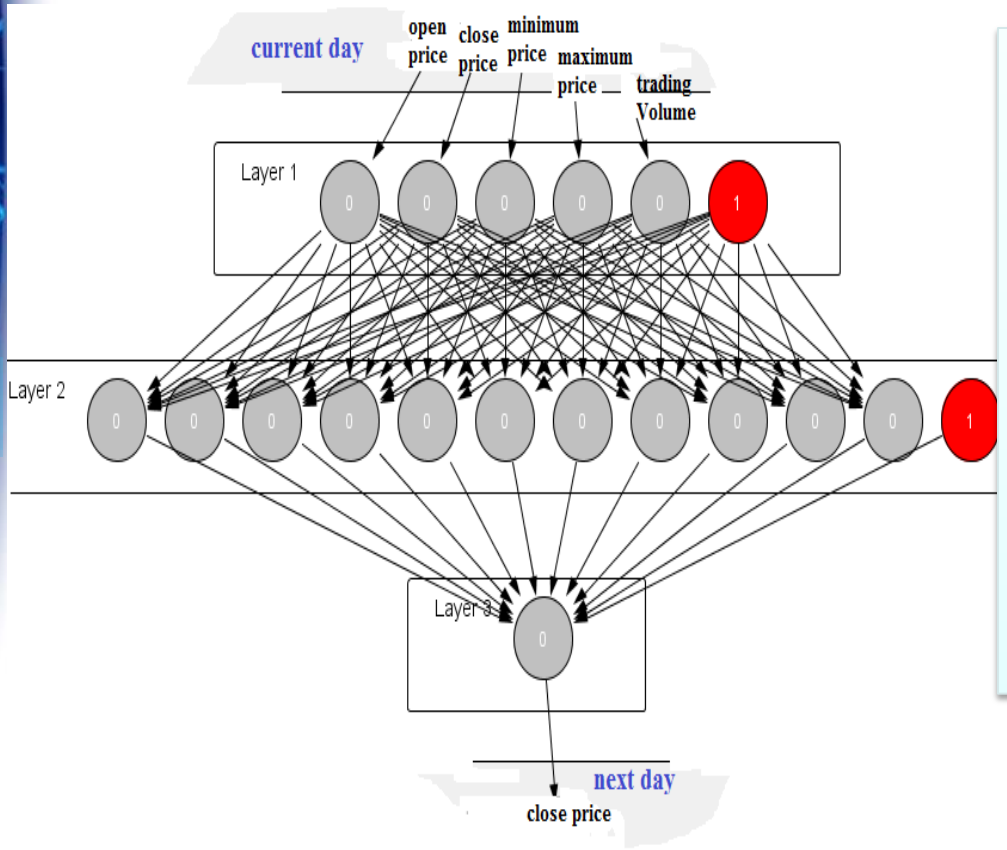
The process of making decisions about determining opportunities and risks encountered in the attempt to reduce risk and maximize the rate of return. Need to forecast stock prices to reduce uncertainty.

Basic stock prediction theories

	Technical Analysis	Chaos Theory	Artificial Neural Networks (ANNs)
Data used	Past stock prices	Past stock prices	Past stock prices
Prediction method	Chart patterns	Chart patterns (fractals)	Price patterns

ANNs for predicting stock prices

The network is given a set of past stock prices, as an input, along with the related output prices, which are the forthcoming prices, until it is able to recognize the pattern.



- Feed Forward Multi Layer Perceptron
- Back propagation
- 5 inputs
- 1 output
- 2 hidden layers

MLP Topology

If the data is non linearly separable we need at least two hidden layers. Trials with 2 models, two hidden layers:

- if *n* is the number of input nodes then the hidden layer must have exactly

$$2n+1 \quad \text{neurons} \quad \textit{(Hecht Nielsen 1987)}$$

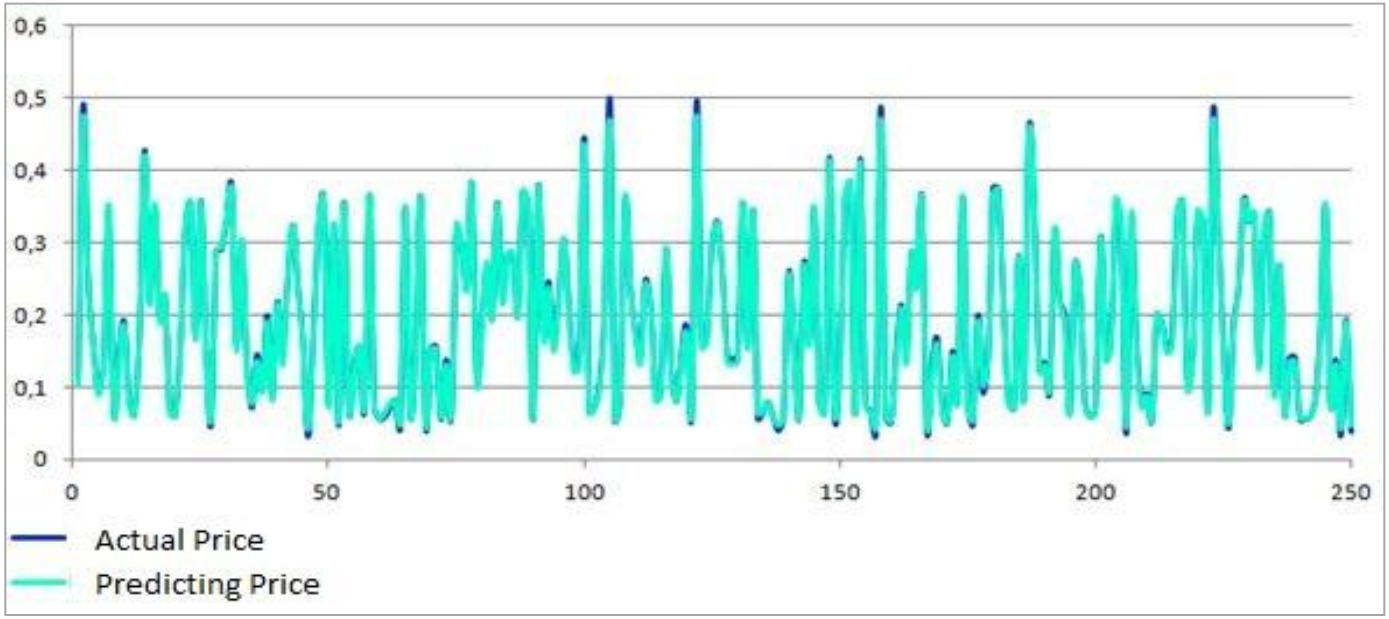
- if the number of inputs is *m* and *N* is the number of the learning samples the sufficient number of neurons for the first hidden layer is

$$\sqrt{(m+2)N} + 2\sqrt{N/(m+2)}$$

and the sufficient number of neurons for the second hidden layer is

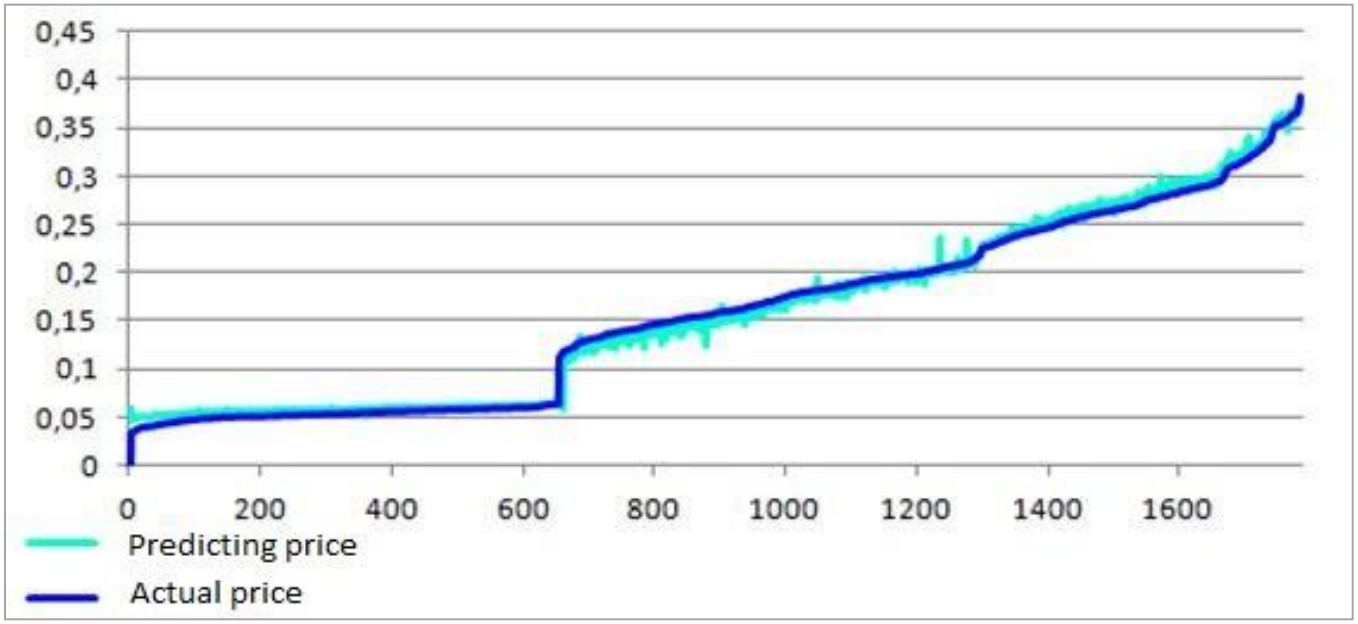
$$m\sqrt{N(m+2)} \qquad \textit{(Huang 2003)}$$

Results from 1 year trial



Graphical representation of predicting and actual price of British American stock

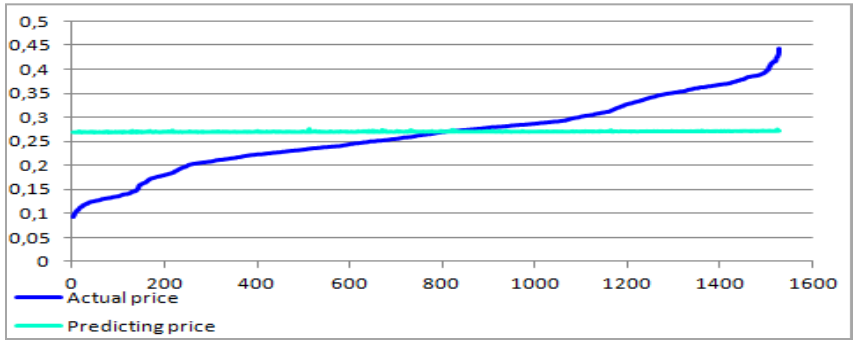
Results from 5 year trial

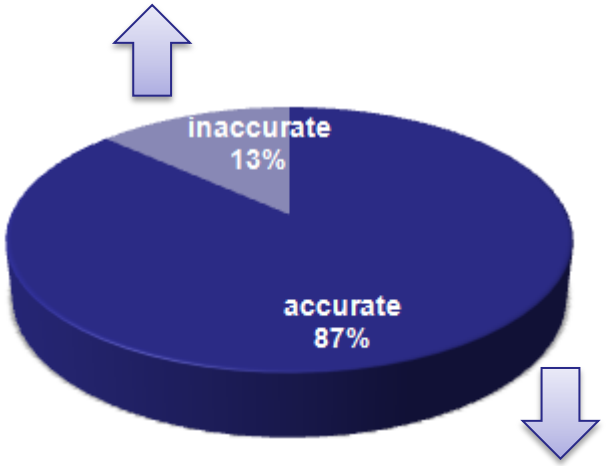


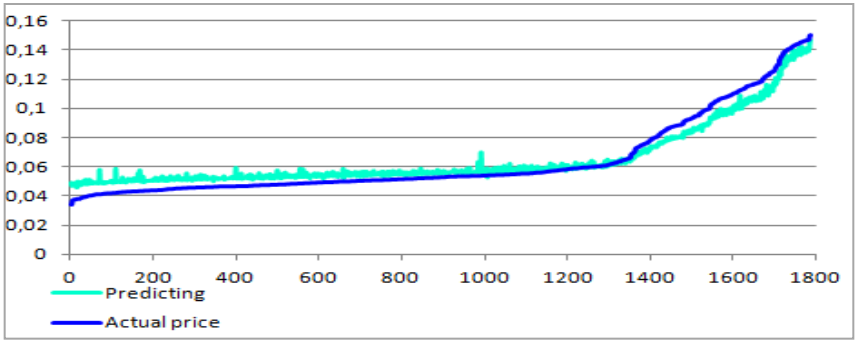
Graphical representation of predicting and actual price of Uniliver stock

Results from all trials: 2x15

STOCK	MODEL	ACCURATE
ANGLO AMERICAN	Hecht Nielsen	Yes
	Huang	Yes
BRITISH AMERICAN	Hecht Nielsen	Yes
	Huang	Yes
EXPERIAN PLC	Hecht Nielsen	Yes
	Huang	Yes
JOHNSON MATTHEY	Hecht Nielsen	NO
KINGFISHER	Huang	NO
MONDI PLC	Hecht Nielsen	Yes
	Huang	Yes
MORRISON (WM.)	Hecht Nielsen	Yes
	Huang	Yes
OLD MUTUAL PLC	Hecht Nielsen	Yes
	Huang	Yes
RELX PLC	Hecht Nielsen	Yes
	Huang	Yes
ROYAL DUTCH SHELL B	Hecht Nielsen	Yes
	Huang	Yes
SMITHS GROUP	Hecht Nielsen	Yes
	Huang	Yes
STANDARD CHARTERED	Hecht Nielsen	Yes
	Huang	Yes
UNILEVER	Hecht Nielsen	Yes
	Huang	Yes
VODAFONE GROUP	Hecht Nielsen	Yes
	Huang	Yes
3I GROUP	Hecht Nielsen	Yes
	Huang	Yes
TOTAL ACCURACY	**26/30**	







Stage 3- Portfolio Optimization

Decide the proportions of each stock to be held in the selected portfolio in such a way that the portfolio has the maximum rate of return.

- If the initial price of a stock is *Pi* and its final value is *Pf* then the logarithmic return is

$$r_{\Delta t} = \log\left(\frac{P_f}{P_i}\right)$$

- For a portfolio compound by *n* companies in which *xi* equals the weight of each stock *i* in the portfolio then the portfolio return is

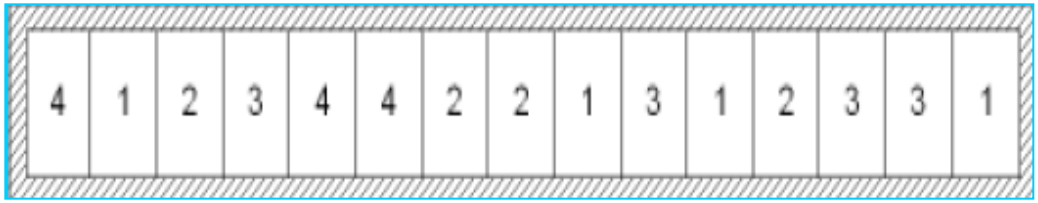
$$r_{p\Delta t} = \sum_{i=1}^{n} x_i r_{i\Delta t}$$

Portfolio return Optimization Using GA

- The fitness function is a function which produces as an output how good is a solution for the problem we want to take in consideration. Therefore, fitness function *f* is given by

$$f = \sum_{i=1}^{n} x_i r_{i\Delta t}$$

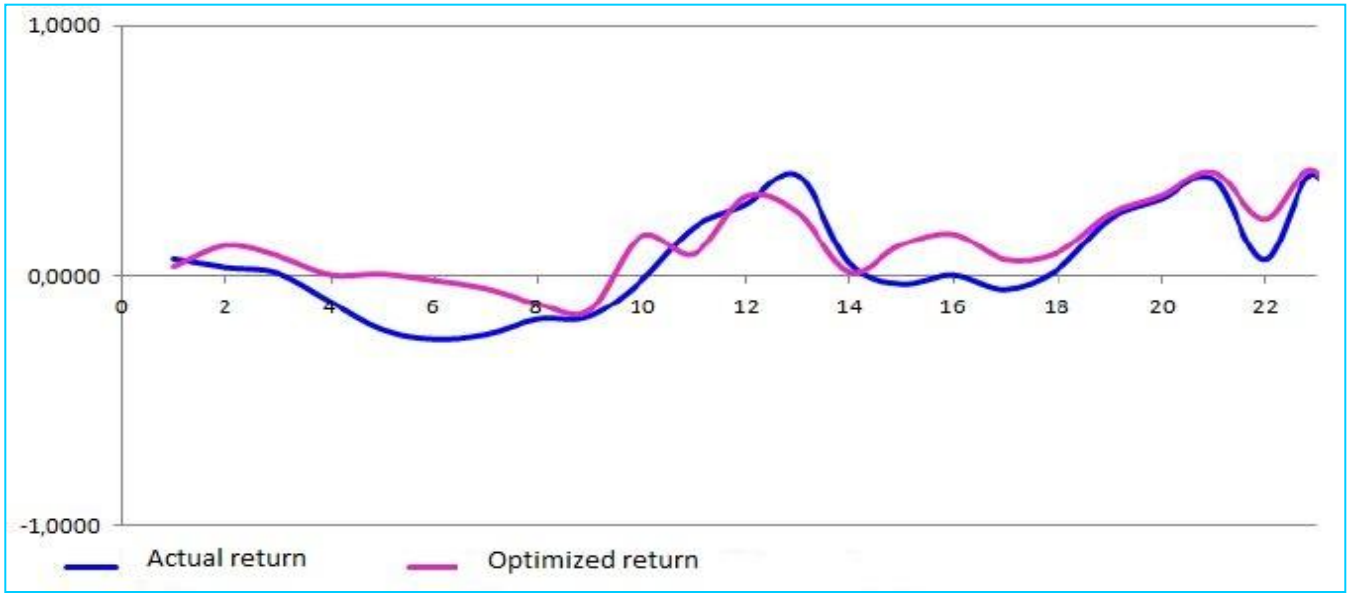
- Each gene must represent the proper number of stocks *xi. We* assume that we can buy up to 4 stocks from each company



Representation of a 15 companies portfolio

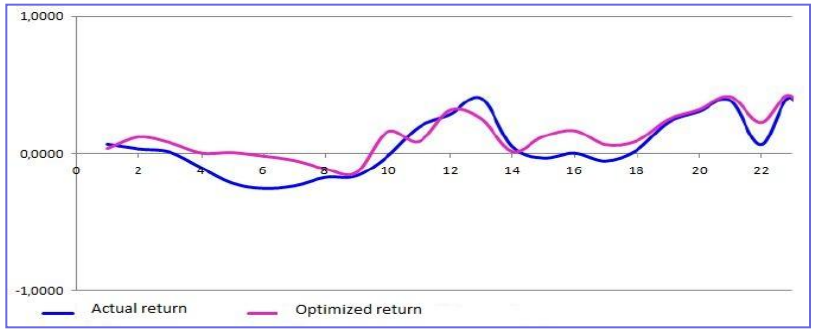
Portfolio rate of return

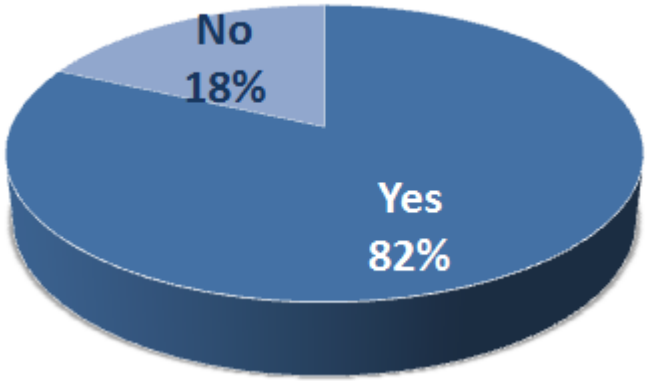
Actual return is calculated based on real prices and optimized return is calculated on predicted prices from MLPs, during March 2017.



Optimization results

Actual return	Optimized return	Optimized > Actual
0,0682	0,0350	NO
0,0343	0,1191	Yes
0,0113	0,0767	Yes
-0,1019	0,0029	Yes
-0,2149	0,0058	Yes
-0,2525	-0,0192	Yes
-0,2349	-0,0507	Yes
-0,1734	-0,1120	Yes
-0,1626	-0,1405	Yes
-0,0182	0,1587	Yes
0,1959	0,0848	NO
0,2871	0,3140	yes
0,4067	0,2530	Yes
0,0519	0,0135	NO
-0,0303	0,1220	NO
0,0048	0,1666	Yes
-0,0528	0,0631	Yes
0,0245	0,0893	Yes
0,2283	0,2448	Yes
0,3079	0,3174	Yes
0,3915	0,4149	Yes
0,0646	0,2250	Yes
0,4067	0,4149	Yes
Total optimization	**18/22**	





Conclusions

- The GAs model for stock evaluation can be easily adjusted to produce a stock portfolio based on different investing preferences. The portfolio that came up as a solution was profitable at the end of the month.
- The suggested MLPs are considered to be sufficient in forecasting stock prices. From the 2x15 MLPS tested, the 87% was accurate in prediction. As a result, MLPs are a good tool for portfolio management.
- The optimization process based on predicted values using GAs accomplished to increase stock portfolios return and to absorb the prediction error from the MLPs.

Future work

- The GAs model for stock evaluation can be easily adjusted to test different characterization in ratios in order to find the most profitable portfolio.

- The suggested MLPs must be tested in various time periods to find how periods affect prediction.

Thank you for your attention

Data Economy workshop:
How online data change economy and business

22 November 2017