OTTO VON GUERICKE UNIVERSITÄT MAGDEBURG

WW FACULTY OF ECONOMICS AND MANAGEMENT

Master Seminar:

Computational Finance and Financial Management

Submitted By:

Md Mehedi Hasan (230319)

Md Najmul Sharkar (229940)

Under the supervision of:

Prof. Dr. Elmar Lukas

Dipl.-Ing. Gordon Briest

Management Summary

Prediction of the stock market using historical trend is noticeably challenging, and it has been the area of research for many decades. The report analyzes technical aspects of stock price movements based on a machine learning model, particularly using Logistic Regression (LR) and Naive Bayes (NB) algorithms. The study will examine the effectiveness of model prediction in stock price movement in trading strategies considering with and without trading cost.

Technical analysis opposes the fundamental theory of the Efficient Market Hypothesis (EMH) as well as Random walk Hypothesis as market direction cannot be estimated through historical data analysis. However, in recent years some statistical model integrated in machine learning algorithm has some ability to anticipate the movement of stock prices (Dongdong Lv et. el 2019).

Logistic regression is a classification model to classify the binary or multi-class label based on multiple independent variables. It uses the sigmoid activation function (appendix 1) to calculate each class's probability, where a predefined threshold determines the class label. It can be applied to predict the stock market movement (e.g., up or down) using single or numerous features as independent variables. The naive Bayes is another type of classification model based on the Bayes theorem that assumes features independence on each other and estimates features prior probability and likelihood probability with each attribute for each class to calculate the posterior probability using the Bayes formula. The higher probability predicts the class label.

Different indices, stocks, and commodities/currencies have been extracted from yahoo finance and onvista into python, and the closing price is taken as the base price. Logarithmic return of assets closing price assures the data stationarity. Jarque-Bera test validates the dataset's log return is normally distributed (see appendix 2). The binary value of log return eliminates the seasonality and trend from the datasets. Class label as Upward [1] and Downward [-1] movement has been obtained from log return where the threshold is 0. Different binary lag variables, e.g., five lags, five digitized lags, five ROC lags, have been created to perform the test within the model. From Appendix 3, Table:1, it has been observed that digitize lag features produce higher cumulative returns while conducting vectorized backtesting with relatively similar volatility. Therefore, findings of digitized lag feature employed in different trading strategies.

Performance evaluations have been conducted with two different portfolios to $100,000 and $10,000, respectively, while considering the trading cost. From Appendix 4, Table:3 shows that eleven assets beat traditional buy and hold strategy return while considering no trading cost. With considering the trading cost, yields are sensitive to different portfolios. For Example, Dow Jones produces a 276% cumulative return in the higher portfolio amount in the Logistic Regression model, whereas only 18% in the lower portfolio cannot beat the benchmark return of 87%. This trend holds almost for all the asset categories. Therefore it shows lower portfolio amount yields relatively very low compare to the higher portfolio amount.

Another type of performance evaluation has been considered with the python library "Backtrader" that helps traders develop their trading strategies and evaluate the performance using python library "pyfolio" developed by quantipion. The Initial Portfolio amount is $100,000 with a 0.09% commission on each transaction without considering short selling. The strategy is, when the next day prediction according to the models is higher than 0, it will buy the shares with the next day opening price as much as possible within portfolio amount and sell owned shares when the prediction is negative. Appendix 5 demonstrates that the trading strategy using backtrader beat the usual strategy with higher return without considering the trading cost. However, returns are relatively low, even negative in some cases, including the trading cost. Appendix 5 Table 4.3 indicates that Ahold's performance is very satisfactory without any trading cost, but with trading cost, no transaction occurs during the backtesting period.

In a nutshell, from the above studies, it can be perceived that technical analysis beat the conventional approach among all asset classes without considering the trading cost. However, it produces inadequate yields while commission added in transactions, even worst in a lower portfolio amount. In most cases, both model's accuracy exceeds 51%, slightly better than random guessing.

References:

LV, D., Yuan, S., Li, Meizi., Xiang, Yang. (2019). An Empirical Study of Machine Learning Algorithms for Stock Daily Trading Strategy. Hindawi. Retrieved from https://www.hindawi.com/journals/mpe/2019/7816154/

Navlani, A. (2018). Naive Bayes Classification using Scikit-learn. Datacamp. Retrieved from https://www.datacamp.com/community/tutorials/naive-bayes-scikit-learn

Amidi, A., Amidy, S.(N.D). CS 229 - Machine Learning. Retrieved from https://stanford.edu/~shervine/teaching/cs-229/cheatsheet-supervised-learning#linear-models

Logistic Regression. Wikipedia. Retrieved from https://en.wikipedia.org/wiki/Logistic_regression

Yahoo Finance. https://finance.yahoo.com/

Appendix 1:

Equation of Logistics Regression and Naïve Bayes

Sigmoid Activation Function:

$$g(z) = \frac{1}{1+e^{-z}}$$

Where,

$g(z)$ = *Output between 0 and 1 (Probability Estimate)*

z = *input to the function*

e= *base of natural log*

Naïve Bayes:

$$y = argmaxP(y)\prod_{i=1}^{n} P(x_i|y_i)$$

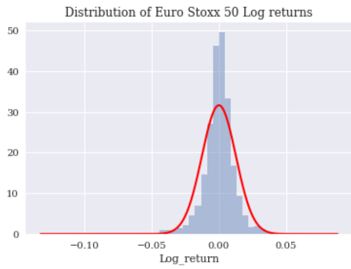
Where,

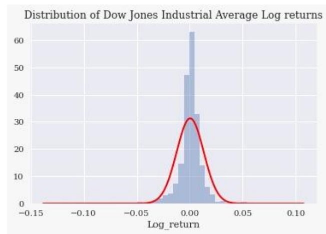
$P(y) = Class\ Probabilit$

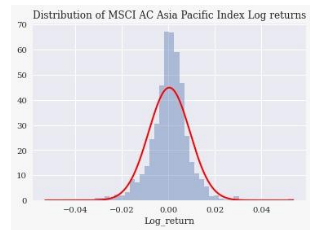
$P(x_i|y_i) = Conditional\ Probability$

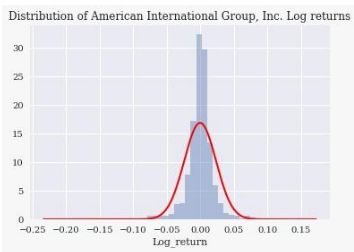
Appendix 2:

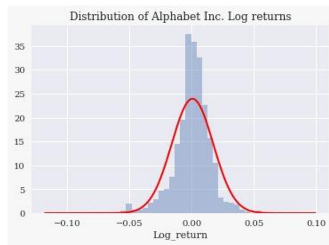
Normal Distribution Graph

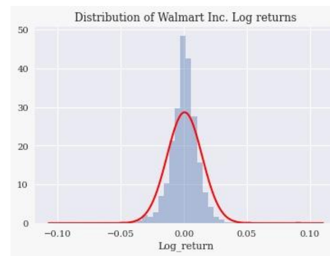


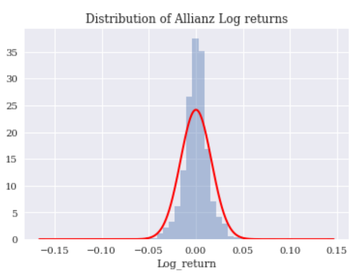


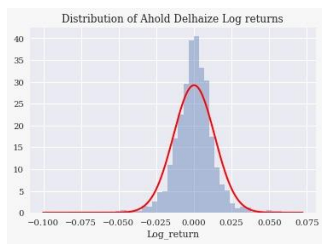


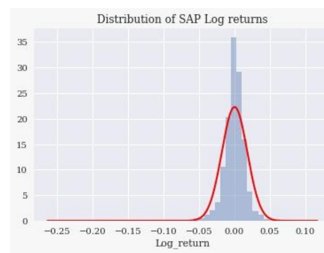


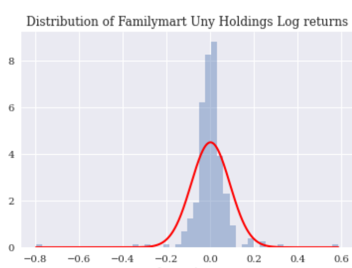


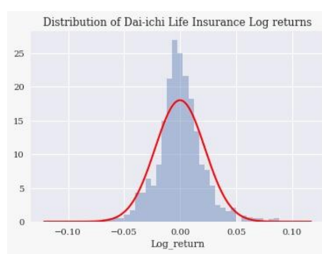


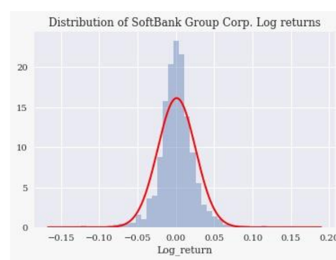


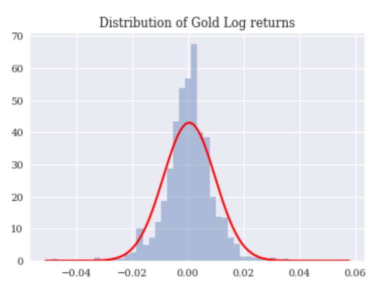


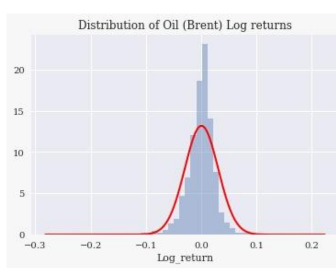


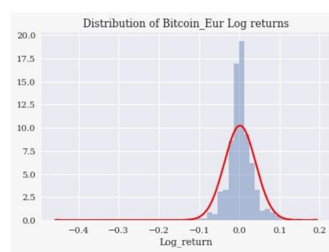












Appendix 3:

Table 1: Vectorized back-testing cumulative return of different lag feature variables.

Asset's	Return	2 Lags		3 Lags		4 Lags		5 Lags		Digitize Lags	
		LR	NB	LR	NB	LR	NB	LR	NB	LR	NB
Dow Jones	1.31	1.31	1.31	1.31	1.31	1.31	1.31	1.31	1.31	2.64	2.65
Euro Stoxx	1.18	1.18	1.18	1.18	1.18	1.42	1.42	1.42	1.42	1.62	1.93
MSCI	1.37	1.37	1.37	1.37	1.37	1.37	1.37	1.31	1.31	1.27	1.25
Alphabet	1.64	1.64	1.64	1.64	1.64	1.64	1.64	1.64	1.64	1.70	2.04
AIG	0.94	1.07	1.07	1.86	1.86	3.30	3.30	2.20	3.05	4.55	1.46
Walmart	1.52	2.05	2.05	2.34	2.34	2.20	2.36	2.40	2.30	2.19	2.11
SAP	1.32	1.32	1.32	2.09	2.09	1.52	1.52	1.33	1.27	1.36	2.17
Allianz	1.32	1.32	1.32	1.98	1.98	1.12	1.12	2.26	3.24	3.14	1.70
Ahold	1.05	1.24	1.24	1.19	1.19	1.36	1.18	1.19	1.19	0.98	1.74
SoftBank	2.36	0.85	0.85	2.28	2.28	1.38	1.38	1.06	0.82	1.14	2.16
Dai-ichi	0.94	3.89	3.89	3.89	3.89	2.48	4.20	3.10	2.49	4.63	2.73
Familymart	0.77	0.77	1.50	0.40	0.40	0.40	0.40	5.60	5.12	0.56	1.29
Gold	1.48	1.48	1.48	1.40	1.40	1.62	1.62	1.74	1.74	2.18	1.75
Oil	2.29	2.29	2.29	2.29	2.29	2.29	2.29	3.17	3.17	4.08	4.17
Bitcoin-Eur	2.37	2.33	2.33	2.33	2.31	2.33	2.31	2.46	3.01	1.67	4.11

Table 2: Annual Volatility of each for Digitized lag features.

Assets's	Digitize Lags		Volatility Digitized lags	
	LR	NB	LR	NB
Dow Jones	2.64	2.65	0.2020	0.2022
Euro Stoxx 50	1.62	1.93	0.2005	0.2004
MSCI	1.27	1.25	0.1410	0.1411
Alphabet	1.70	2.04	0.2644	0.2641
AIG	4.55	1.46	0.3760	0.3764
Walmart	2.19	2.11	0.2207	0.2205
SAP	1.36	2.17	0.2842	0.2838
Allianz	3.14	1.70	0.2614	0.2617
Ahold	0.98	1.74	0.2166	0.2160
SoftBank	1.14	2.16	0.3933	0.3930
Dai-ichi	4.63	2.73	0.3507	0.3506
Familymart	0.56	1.29	1.4056	1.4057
Gold	2.18	1.75	0.1469	0.1471
Oil	4.08	4.17	0.4801	0.4801
Bitcoin-Euro	1.67	4.11	0.6196	0.6178

Appendix 4

Performance Evaluation of Simple back-testing

Table 3: Simple Back-testing result with different portfolio amount

Categories	Asset's Name	Benchmark Cumulative Return	Without Trading Cost		With Trading Cost			
			Cum. Return		Cum. Return		Cum. Return	
			LR	NB	LR	NB	LR	NB
			Cumulative Return		Portfolio $ 100,000		Portfolio $ 10,000	
Indices	Dow Jones	1.87	4.04	3.51	3.76	3.24	1.18	0.81
	Euro Stoxx	1.18	1.62	1.93	1.49	1.80	0.35	0.60
	MSCI	1.71	1.79	1.63	1.65	1.50	0.37	0.28
Large Cap USA	Alphabet	2.45	2.27	3.58	2.09	3.33	0.43	1.13
	AIG	0.66	5.90	3.54	5.41	3.32	1.00	1.33
	Walmart	2.24	3.72	4.75	3.44	4.44	0.93	1.71
Large Caps Europe	SAP	1.66	2.62	4.73	2.46	4.46	1.01	2.08
	Allianz	1.32	3.14	1.70	2.89	1.54	0.64	0.12
	Ahold	1.11	1.12	3.45	1.01	3.22	-0.01	1.22
Large Caps Asia Pacific	SoftBank	3.25	0.98	3.71	0.88	3.41	-0.07	0.65
	Dai-ichi	0.87	8.36	9.55	7.71	9.10	1.88	5.12
	Familymart	1.72	1.32	0.93	1.30	0.91	1.07	0.72
Commodities & Currencys	Gold	1.73	3.21	2.58	2.97	2.39	0.79	0.67
	Oil	2.29	4.08	4.17	3.84	4.01	1.70	2.54
	Bitcoin-Euro	2.37	1.67	4.11	1.62	4.04	1.09	3.34

Performance graph of simple back-testing

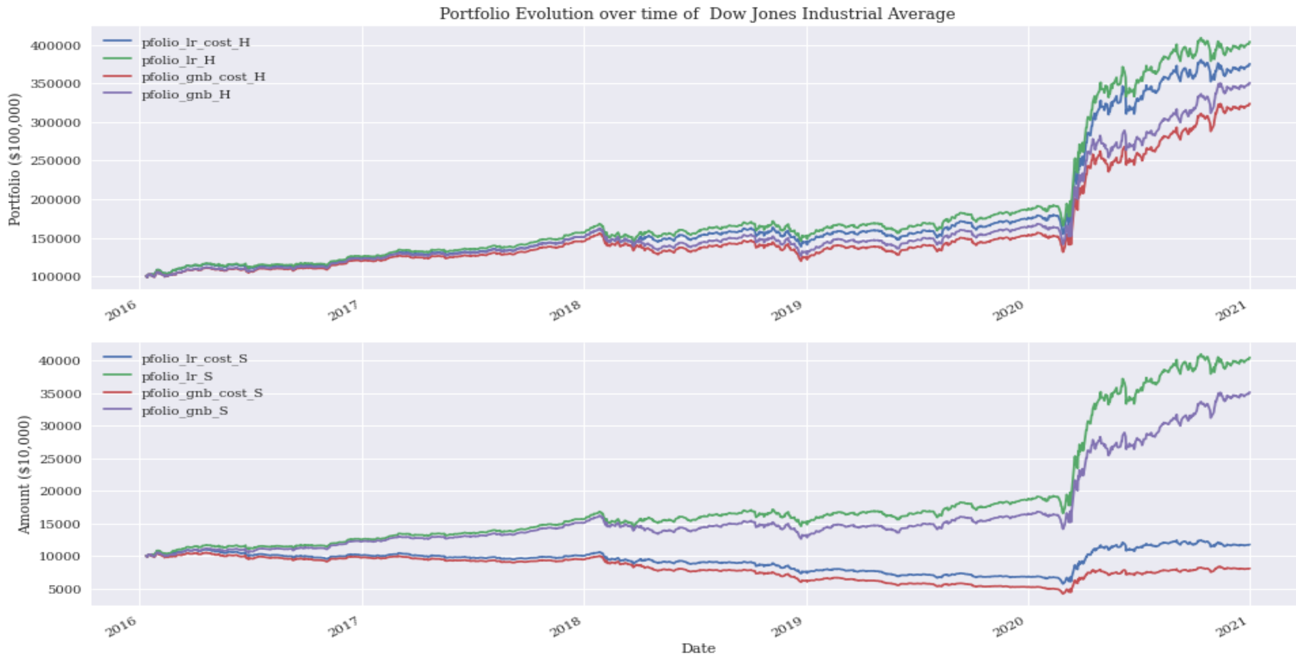


Figure 1: Back-testing Performance of Dow Jones Industrial Average

Figure 2: Back-testing Performance of Alphabet

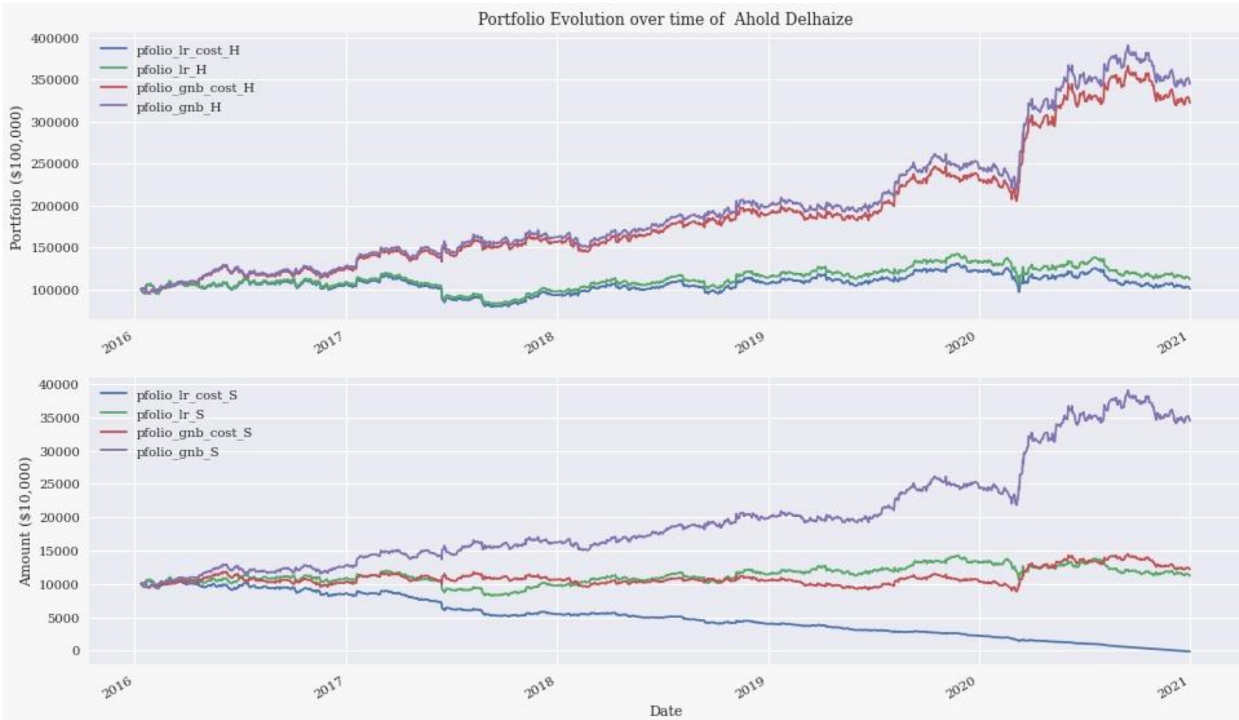


Figure 3: Back-testing Performance of Ahold Delhaize

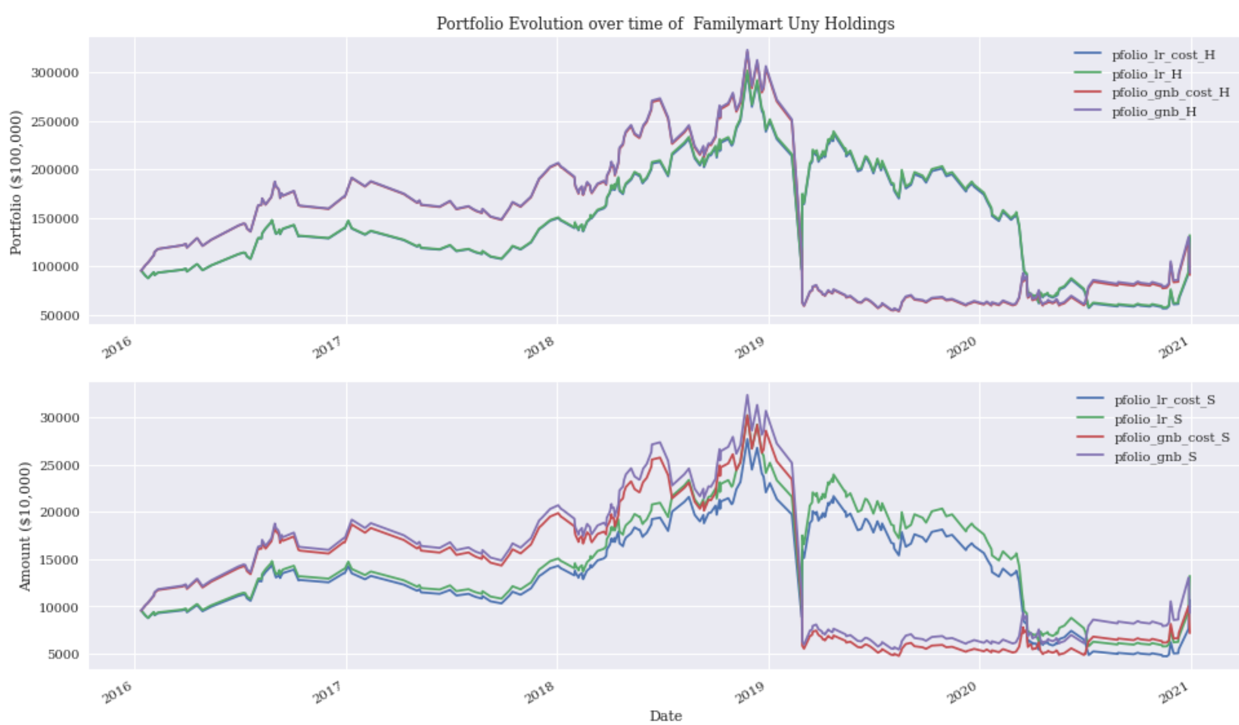


Figure 4: Back-testing Performance of FamilyMart

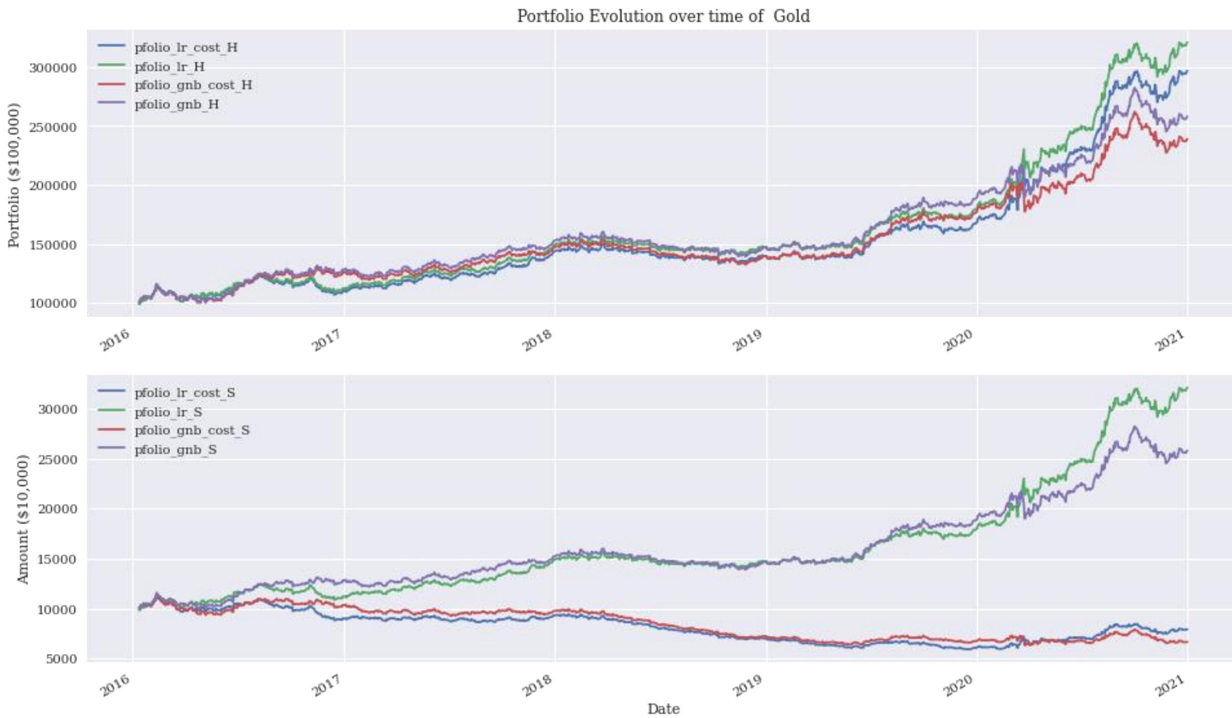


Figure 5: Back-testing Performance of Gold

Appendix 5:

Performance Evaluation of Backtrader back-testing using pyfolio

Table 4.1: Performance of Indices

Assets	Date Range	Months	Particulars		Bench. Return	Annual return	Cum. returns	Volatility	Sharpe ratio	Max drawdown
DowJones	1/12/2016-12/31/2020	59	LR	w-c	1.87	51.2%	679.9%	13.7%	3.08	-22.6%
				w/o-c		69.8%	1288.6%	13.9%	3.89	-22.5%
			NB	w-c		54.0%	753.6%	13.6%	3.25	-22.9%
				w/o-c		72.1%	1383.3%	13.6%	4.07	-22.5%
EuroStoxx 50	1/12/2016-12/31/2020	59	LR	w-c	1.18	49.8%	642.0%	11.7%	3.5	-5.9%
				w/o-c		69.6%	1271.0%	12.6%	4.27	-8.5%
			NB	w-c		39.0%	412.0%	13.0%	2.6	-8.8%
				w/o-c		59.9%	922.7%	13.3%	3.59	-8.7%
MSCI	1/12/2016-12/31/2020	58	LR	w-c	1.71	-0.2%	-1.0%	9.0%	0.02	-13.3%
				w/o-c		8.7%	50.6%	9.7%	0.91	-11.6%
			NB	w-c		-0.7%	-3.4%	8.4%	-0.04	-13.8%
				w/o-c		10.3%	61.8%	9.4%	1.1	-11.8%

Table 4.2: Performance of Large Cap USA

Assets	Date Range	Months	Particulars		Bench. Return	Annual return	Cum. Returns	Volatility	Sharpe ratio	Max drawdown
Alphabet	1/12/2016-12/31/2020	59	LR	w-c	2.45	75.8%	1554.4%	12.9%	443.0%	-8.0%
				w/o-c		163.3%	12214.1%	16.9%	584.0%	-9.2%
			NB	w-c		69.3%	1272.1%	14.8%	364.0%	-13.3%
				w/o-c		125.9%	5649.0%	18.6%	449.0%	-22.4%
American Group	1/12/2016-12/31/2020	59	LR	w-c	0.66	0.0%	0.0%	0.0%	0.0%	0.0%
				w/o-c		43.2%	491.5%	25.9%	151.0%	-26.0%
			NB	w-c		0.0%	0.0%	0.0%	0.0%	0.0%
				w/o-c		47.9%	595.3%	26.8%	159.0%	-33.1%
Walmart	1/12/2016-12/31/2020	59	LR	w-c	2.24	4.0%	21.5%	9.0%	48.0%	-14.2%
				w/o-c		48.0%	600.7%	17.6%	232.0%	-14.8%
			NB	w-c		4.1%	21.8%	5.3%	78.0%	-6.3%
				w/o-c		60.3%	942.0%	16.8%	290.0%	-13.0%

Table 4.3: Performance of Large cap Europe

Assets	Date	Month	Particulars		Bench. Return	Annual return	Cum. Returns	Volatility	Sharpe ratio	Max drawdown
SAP	1/12/2016-12/31/2020	59	LR	w-c	1.66	1.4%	7.1%	13.1%	18.0%	-25.2%
				w/o-c		27.5%	234.6%	21.9%	122.0%	-31.0%
			NB	w-c		4.5%	24.2%	11.3%	44.0%	-18.7%
				w/o-c		28.2%	242.7%	22.7%	121.0%	-33.2%
Allianz	1/12/2016-12/31/2020	58	LR	w-c	1.32	12.7%	79.8%	7.7%	160.0%	-11.1%
				w/o-c		70.2%	1253.1%	17.0%	322.0%	-21.4%
			NB	w-c		11.2%	68.2%	9.6%	116.0%	-18.2%
				w/o-c		59.7%	890.5%	16.9%	285.0%	-22.6%
Ahold	1/12/2016-12/31/2020	59	LR	w-c	1.11	0.0%	0.0%	0.0%	0.0%	0.0%
				w/o-c		94.5%	2634.8%	14.6%	463.0%	-9.7%
			NB	w-c		0.0%	0.0%	0.0%	0.0%	0.0%
				w/o-c		52.8%	724.4%	15.6%	280.0%	-14.5%

Table 4.4: Performance of Large Cap Asia Pacific.

Assets	Date	Month	Particulars		Bench. Return	Annual return	Cum. Returns	Volatility	Sharpe ratio	Max drawdown
Soft Bank	1/12/2016-12/31/2020	59	LR	w-c	3.25	0.0%	0.0%	0.0%	0	0.0%
				w/o-c		48.4%	598.7%	29.2%	149.0%	-22.8%
			NB	w-c		0.0%	0.0%	0.0%	0	0.0%
				w/o-c		2.5%	13.2%	30.2%	23.0%	-61.4%
Dai Ichi	1/12/2016-12/31/2020	57	LR	w-c	0.87	-39.9%	-91.3%	25.2%	-189.0%	-91.3%
				w/o-c		-32.7%	-85.0%	25.7%	-141.0%	-85.1%
			NB	w-c		-48.9%	-96.0%	25.8%	-247.0%	-96.0%
				w/o-c		-44.1%	-93.8%	26.4%	-207.0%	-93.8%
Family Mart	1/12/2016-12/31/2020	57	LR	w-c	1.72	0.0%	0.0%	0.0%	0	0.0%
				w/o-c		17.2%	18.2%	85.4%	60.0%	-44.5%
			NB	w-c		0.0%	0.0%	0.0%	0	0.0%
				w/o-c		-20.7%	-21.8%	87.0%	16.0%	-68.8%

Table 4.5: Commodities and Currencies

Assets	Date	Months	Particulars		Bench. Return	Annual return	Cum. Returns	Volatility	Sharpe ratio	Max drawdown
Gold	1/12/2016-12/31/2020	58	LR	w-c	1.73	25.4%	203.2%	11.0%	212.0%	-8.1%
				w/o-c		42.5%	465.2%	11.1%	325.0%	-7.5%
			NB	w-c		17.9%	124.2%	11.1%	154.0%	-16.9%
				w/o-c		33.1%	304.5%	11.0%	264.0%	-16.1%
Oil	1/12/2016-12/31/2020	58	LR	w-c	2.29	0.0%	0.0%	0.0%	0	0.0%
				w/o-c		516.4%	701124.2 %	33.9%	555.0%	-35.7%
			NB	w-c		0.0%	0.0%	0.0%	0	0.0%
				w/o-c		513.2%	683778.4 %	31.9%	586.0%	-30.6%
Bitcoin-Euro	9/24/2019-12/31/2020	21	LR	w-c	2.37	88.4%	207.7%	37.4%	188.0%	-14.7%
				w/o-c		108.9%	269.5%	37.5%	215.0%	-12.7%
			NB	w-c		40.6%	83.0%	45.6%	99.0%	-38.2%
				w/o-c		55.2%	118.2%	46.6%	120.0%	-38.9%