Forecasting the movements of Bitcoin prices: an application of Machine Learning Algorithms

Mainack Paul (MD2111)

May 13, 2023

For this assignment, first we will give a broad overview of the numerical findings of all the Machine Learning models used in this paper for prediction of Bitcoin prices and then, we are going to extend this using some traditional time series models.

1 Data preparation

1.1 Continous data

In this study, for the output, we considered the changes of up and down movements of closing prices of bitcoin from previous days. We coded these as $+1$ and -1 for ups and downs, respectively. Hence, it is a classification problem. We used the same output for continuous and discrete datasets. Closing, high and low prices were used for computing technical indicators and output as reported in Table 1.

Table 1: Selected Technical Indicators

Indicators	Formula
Simple 14 days moving average (MA)	$\frac{(C_t+C_{t-1}+.....+C_{t-14})}{14}$
Simple 14 days weighted moving average (WMA)	$\frac{n*C_t+(n-1)*C_{t-1}+.....+C_{t-14}}{n+(n-1)+...+1}$
Momentum (Mom)	$C_t - C_{t-n}$
Stochastic K% (K%)	$\frac{C_t-LL_{t-n}}{HH_{t-n}-LL_{t-n}}*100$
Stochastic D% (D%)	$\sum\limits_{i=0}^{n-1} K_{t-i}/n$
Relative strength index (RSI)	$100 - \frac{100}{1+(\sum\limits_{i=0}^{n-1} Up_{t-i}/n)/(\sum\limits_{i=0}^{n-1} Dw_{t-i}/n)}$
Moving average convergence/divergence (MACD)	$MACD(n)_{t-1} + \frac{2}{n+1} * (DIFF_t - MACD(n)_{t-1})$
Larry William's R% (LW)	$\frac{H_n-C_t}{H_n-L_n} * 100$
Accumulation/distribution oscillator (A/D)	$\frac{H_t-C_{t-1}}{H_t-L_t}$

n is the number of days accepted as 10 here, C_t:closing price, L_t:low price,H_t:High price. $DIFF_t : EMA(12)_t - EMA(26)_t$. EMA is exponential moving average, $EMA(k)_t : EMA(k)_{t-1}+$

$\alpha * (C_t - EMA(k)_{t-1})$, α is correction factor. LL_t is the lowest low, HH_t is the highest high for the last t days. $M_t = (H_t + L_t + C_t)/3$, $SM_t = \sum_{i=0}^{n} M_{t-i+1}/n$, $D_t = (\sum_{i=0}^{n} |M_{t-i+1} - SM_t|/n$, Up_t and Dw_t are upward and downward price change at time t respectively.

1.2 Discrete Data

For creating the discrete dataset, the continuous dataset was converted to -1 or $+1$ by applying the discretization process. $+1$ and -1 indicate upward and downward movements, respectively.

2 Empirical Findings

Table 2: Descriptive Statistics for Selected Technical Indicators

Indicator	Minimum	Maximum	Mean	Standard Deviation
MA	158.407	16866.037	2501.552	3395.42
WMA	176.498	17802.757	2507.79	3402.815
Mom	-5578	8212.55	23.237	920.879
K%	0	100	54.624	29.336
D%	6.337	93.153	54.343	22.571
RSI	10.954	93.491	52.549	14.209
MACD	-1479.221	2520.715	12.44	292.976
LW	-100	0	-45.376	29.336
A/D	-0.879	1.521	0.399	0.185

2.1 Findings of continous data

The best parameter combinations are determined by means of experiments for each forecasting algorithm.

2.1.1 Artificial Neural Network

Table 3: Best 3 parameter combinations for ANN

	Learning Rate	Iteration	Momentum Constant	Hidden neuron	Accuracy	MAE	RMSE	RAE
1	0.3	500	0.2	6	0.843	0.203	0.341	0.409
2	0.3	500	0.2	8	0.841	0.201	0.360	0.405
3	0.3	500	0.2	7	0.835	0.201	0.349	0.404

Comment:

Test results in Table 3 indicate that the accuracy levels and error statistics calculated are within acceptable levels. The best accuracy level is determined as 0.843 for the ANN. This means that we will be able to forecast the movements of Bitcoin prices at a high degree of accuracy.

2.1.2 Support Vector Machine

Table 4: Best 3 parameter combinations for SVM

	Kernel Function	d	γ	c	Accuracy	MAE	RMSE	RAE
1	Polynomial	2	-	100	0.808	0.192	0.438	0.387
2	Polynomial	1	-	30	0.804	0.196	0.443	0.395
3	Polynomial	2	-	20	0.802	0.198	0.445	0.339
4	RBF(Gaussian)	-	0.1	20	0.733	0.266	0.516	0.538
5	RBF(Gaussian)	-	0.1	10	0.729	0.271	0.520	0.546
6	RBF(Gaussian)	-	0.1	40	0.717	0.283	0.532	0.571

d, γ and c represents respectively the Kernel function degree, Kernel function Gamma coefficient and Regularization parameter.

2.1.3 Naive Bayes Classifier

Table 5: NB classification parameters

	Accuracy	MAE	RMSE	RAE
1	0.626	0.368	0.572	0.743
2(Gaussian)	0.717	0.283	0.461	0.571

Comment:

Test results indicate that the accuracy level is determined to be 0.717 for Gaussian NB classifiers as the best forecasting algorithm.

2.1.4 Random Forest Classifier

Table 6: Best 3 parameter combinations for RF

	Feature	Number of Tree	Accuracy	MAE	RMSE	RAE
1	3	297	0.884	0.191	0.297	0.384
2	8	251	0.882	0.180	0.293	0.362
3	6	267	0.880	0.184	0.293	0.370

Comment:

A number of trees are selected as parameter for the RF. It ranges from 50 to 300 during the best parameter selection process and it uses 1 to 10 features to train the trees. The best accuracy level is selected as 0.884 for RF with 3 features and 297 trees.

2.1.5 Comparision

Table 7: Comparision of the best models

Model	TP	FP	ROC	F-stat	Rank
ANN	0.843	0.149	0.910	0.843	2
SVM	0.808	0.191	0.809	0.808	3
NB	0.717	0.278	0.826	0.717	4
RF	0.884	0.118	0.949	0.884	1
LR	0.781	0.832	0.828	0.562	(Benchmark)

Comment:

Test results in Table 7 indicate that, while the Gaussian process NB model presents the lowest performance at 0.717, the RF model has the highest at 0.884 value of F statistic. The performance differences of the ANN, RF, SVM and NB algorithms with the LR model are statistically significant and they provide better performances compared to the LR model.

Table 8: t test results of model comparisons in terms of benchmark

Model	Mean(Accuracy)	N	Standard Deviation	t
LR	0.551	10	0.719	
ANN	0.826	10	0.017	-13.658*
RF	0.854	10	0.056	-9.467*
SVM	0.754	10	0.064	-11.809*
NB	0.657	10	0.040	-4.262**

Note: *shows the statistical significance at level 0.01 ** shows the statistical significance at level 0.05

2.2 Findings of discrete data

The best parameter combinations are determined by means of experiments for each forecasting algorithm.

2.2.1 Artificial Neural Network

Table 9: Best 3 parameter combinations for ANN

	Learning Rate	Iteration	Momentum Constant	Hidden neuron	Accuracy	MAE	RMSE	RAE
1	0.3	500	0.2	20	0.9483	0.072	0.206	0.480
2	0.1	500	0.1	20	0.9463	0.077	0.207	0.512
3	0.1	500	0.1	20	0.9395	0.086	0.214	0.546

Comment:

Test results in Table 9 indicate that the accuracy levels and error statistics calculated are within acceptable levels. The best accuracy level is determined as 0.9483 for the ANN. This means that we will be able to forecast the movements of Bitcoin prices at a high degree of accuracy.

2.2.2 Support Vector Machine

Table 10: Best 3 parameter combinations for SVM

	Kernel Function	d	γ	c	Accuracy	MAE	RMSE	RAE
1	Polynomial	3	-	1	0.9463	0.054	0.232	0.358
2	Polynomial	3	-	2	0.9442	0.056	0.236	0.372
3	Polynomial	2	-	1	0.9421	0.058	0.240	0.386
4	RBF(Gaussian)	-	0.2	1	0.9483	0.052	0.227	0.344
5	RBF(Gaussian)	-	0.2	100	0.9463	0.054	0.232	0.358
6	RBF(Gaussian)	-	0.1	10	0.9442	0.056	0.236	0.372

d, γ and c represents respectively the Kernel function degree, Kernel function Gamma coefficient and Regularization parameter.

2.2.3 Naive Bayes Classifier

Table 11: NB classification parameters

	Accuracy	MAE	RMSE	RAE
1	0.8822	0.136	0.310	0.905
2(Gaussian)	0.8822	0.134	0.309	0.905

Comment:

Test results indicate that the accuracy level is determined to be 0.8822 for Gaussian NB classifiers as the best forecasting algorithm.

2.2.4 Random Forest Classifier

Table 12: Best 3 parameter combinations for RF

	Feature	Number of Tree	Accuracy	MAE	RMSE	RAE
1	10	79	0.9462	0.076	0.205	0.509
2	8	71	0.9438	0.078	0.208	0.513
3	10	69	0.9390	0.212	0.212	0.538

Comment:

A number of trees are selected as parameter for the RF. It ranges from 50 to 300 during the best parameter selection process and it uses 1 to 10 features to train the trees. The best accuracy level is selected as 0.946 for RF with 10 features and 79 trees.

2.2.5 Comparision

Table 13: Comparision of the best models

Model	TP	FP	ROC	F-stat	Rank
ANN	0.948	0.557	0.931	0.914	1
SVM	0.948	0.610	0.669	0.938	3
NB	0.882	0.167	0.901	0.902	4
RF	0.946	0.557	0.923	0.939	2
LR	0.858	0.873	0.681	0.854	(Benchmark)

Comment:

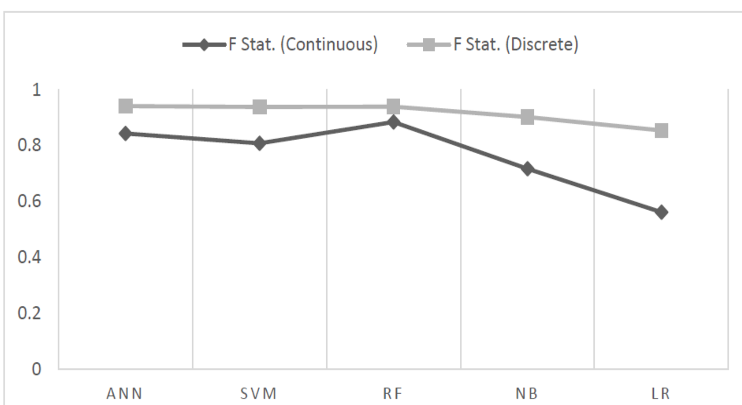
Test results in Table 13 indicate that, while the LR model presents the lowest performance at 0.858, the ANN model has the highest at 0.914 value of F statistic.

Table 14: t test results of model comparisons in terms of benchmark

Model	Mean(Accuracy)	N	Standard Deviation	t
LR	0.623	10	0.032	
ANN	0.850	10	0.037	-13.208*
RF	0.835	10	0.012	-16.582*
SVM	0.786	10	0.053	-7.398*
NB	0.674	10	0.027	-2.908*

Note: *shows the statistical significance at level 0.01

Figure 1: F statistics with continuous and discrete data for all models



3 Extension

In the earlier section, we can see that we have used various types of Machine Learning models for predicting the prices of Bitcoin. In this section, I want to explore and see how the traditional time series models are performing. I will be working with closing price of Bitcoin from July 2012 to December 2021 and hence, it will be a regression problem. While working with the data, we found out that closing price itself is not stationary using the Augmented Dickey Fuller Test and hence, we will be working with the Log Return of the Bitcoin price (y_t) and for t^{th} time point, it can be computed as –

$$y_t = ln(p_t) - ln(p_{t-1})$$

where, p_t denotes the price of the Bitcoin for t^{th} day and it has come out to be stationary by using the Augmented Dickey Fuller Test.

A suitable univariate model is fitted which not only model the conditional variances but can also describe the volatility transmission effects therein. The following model is considered for the returns:

$$y_t = c + u_t$$

where, y_t is the vector of price returns of Bitcoin, c is the parameter vector with means of return series and u_t is the residual vector. We have modelled u_t using various types of GARCH models.

A residual process u_t is called GARCH of order m and r if

$$u_t = \sqrt{h_t}\varepsilon_t \;\; where \;\; \varepsilon_t iid(0,1)$$

$$h_t = \alpha_0 + \sum_{j=1}^{m} \alpha_j u_{t-j}^2 + \sum_{i=1}^{r} \delta_i h_{t-i}, \alpha_0 > 0, \alpha_j \geq 0, \delta_i \geq 0, \sum_{j=1}^{m} \alpha_j + \sum_{i=1}^{r} \delta_i < 1$$

.

So, we will use various GARCH models for modelling and compare those using AIC. We will also check the residual plots to justify the usage of these models.

3.1 Various GARCH Models

3.1.1 Standard GARCH

The standard GARCH model can be written as:

$$\sigma_t^2 = w + \sum_{j=1}^{m} \alpha_j v_{jt} + \sum_{j=1}^{q} \alpha_j \varepsilon_{t-j}^2 + \sum_{j=1}^{p} \delta_j \sigma_{t-j}^2$$

with σ_t^2 denoting the conditional variance, ω the intercept and $\varepsilon_t^2$are the residuals from the mean filtration process discussed previously.

3.1.2 Exponential GARCH

$$log\sigma_t^2 = w + \sum_{j=1}^{m} \alpha_j v_{jt} + \sum_{j=1}^{q} (\alpha_j z_{t-j} + \gamma_j(|z_{t-j}| - E|z_{t-j}|)) + \sum_{j=1}^{p} \delta_j log\sigma_{t-j}^2$$

3.1.3 GJR GARCH

$$\sigma_t^2 = w + \sum_{j=1}^{m} \alpha_j v_{jt} + \sum_{j=1}^{q} (\alpha_j \varepsilon_{t-j}^2 + \gamma_j I_{t-j} \varepsilon_{t-j}^2) + \sum_{j=1}^{p} \delta_j \sigma_{t-j}^2$$

where γ_j now represents the 'leverage'.

3.1.4 Asymmetric power ARCH

$$\sigma_t^\partial = w + \sum_{j=1}^{m} \alpha_j v_{jt} + \sum_{j=1}^{q} \alpha_j(|z_{t-j}| - \gamma_j \varepsilon_{t-j})^\partial + \sum_{j=1}^{p} \delta_j \sigma_{t-j}^\partial$$

3.2 Results

All the parameters of all the models has came out to be significant and hence, we can assume that the modelling done is correct. For further verification, we have used a residual plot and the residual plot of all 4 models seems to be moderately good which implies that the modelling done is corect. Here, we have only shown the residual plot of sGARCH model and the rest are similar to it.

Figure 2: Residual Plot of sGARCH(1,1) model

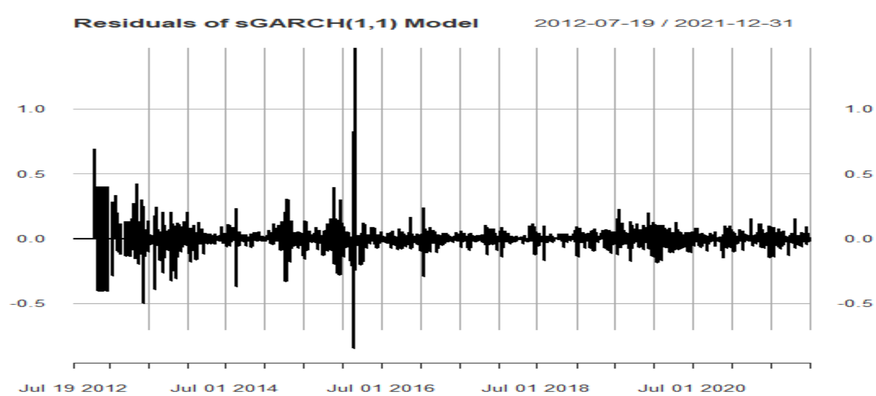


Table 15: Comparision of the best models

Model	AIC
sGARCH(1,1)	-1.8670
eGARCH(1,1)	-4.2543
gjrGARCH(1,1)	-1.5734
apARCH(1,1)	-1.8625

Exponential GARCH has the lowest AIC and hence it is the best model among the 4 models defined above.