ORTA DOĞU TEKNİK ÜNİVERSİTESİ

CENG790 – BIG DATA ANALYTICS

PROJECT

CANDLESTICK PATTERNS AND THEIR EFFECTS ON STOCK PRICE

Selim ÖZTÜRK - 2480408

Ridhiwan MSEYA – 1999989

Contents

Table of Figures

1. INTRODUCTION

In the current state of the world acquiring a lot of money is not simple and living a life without the concern of economic welfare is not reached by the majority of people. As we observe individuals that are better off economically, they tend to have investments that are held valuably. The main goal of such investors is to have their investment return them profits fairly reliably to have their retirements early or make life easier for themselves when they retire.

Therefore, we tried to build a model that would try to support investors on their investments to make better decisions. Due to its high volatility and the technical analysis on stocks (generally) not working on high intervals, we chose Bitcoin as our reference stock. We tried to implement an aspect of Technical Analysis (candlestick patterns).

2. DATA

The data that we are using is Bitcoin Price Chart. What we have in this chart are candlesticks that are representing the Time, Open, High, Low, Close and Volume of that particular moment in the market.

We have all the historical data available in Binance database for Bitcoin. We are using the 5-minute intervals to have more data to train and test on.

Index	Date-Time	Open	High	Low	Close	Volume
450000	2021-12-02 15:35:00	56403.9	56486.2	56321.8	56414	124.748
450001	2021-12-02 15:40:00	56417	56428	56349.7	56360	102.706
450002	2021-12-02 15:45:00	56360	56380.5	56268.4	56268.5	95.6247
450003	2021-12-02 15:50:00	56268.5	56550	56255.1	56520.5	127.799
450004	2021-12-02 15:55:00	56520.5	56590	56460	56521.1	152.675
450005	2021-12-02 16:00:00	56521.1	56591	56450	56477.6	124.761
450006	2021-12-02 16:05:00	56477.6	56629	56450	56581.2	113.128

Figure 1 : Stock Price Data Example

For a candle, open means the price at the point of the start of time interval. High means the maximum value of the stock in that time interval. Similarly, low means the minimum value of the stock in that time interval. Close means the price at the point of the end of the time interval. Finally, volume means how many of total Bitcoin was traded in that time interval. Please see Figure 2 for representation of Bullish (stock price increasing) and Bearish Candles (stock price decreasing).
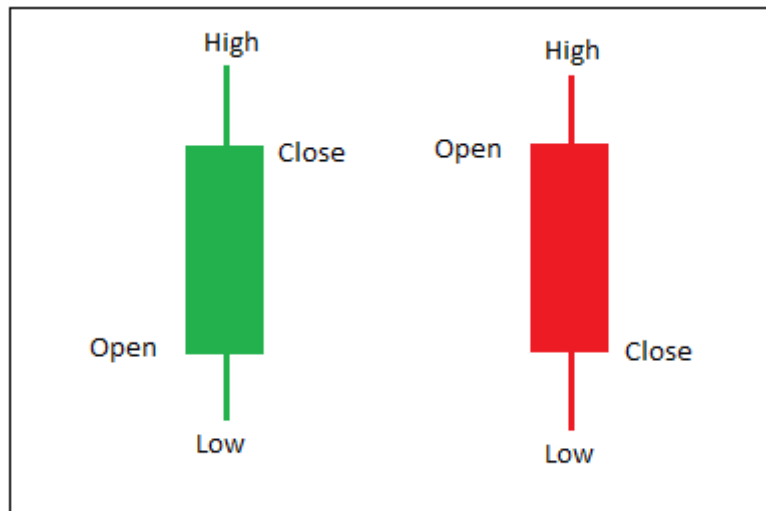
Figure 2: Candlestick Bars

3. CANDLESTICK PATTERNS

Candlestick patterns can be summarized by generalized movements of candles. They seem to occur every once in a while, and when they are seen, they are treated as a support or resistance indicators due to accumulated and learned knowledge. So, in a sense, they are learned behavior. There are lots of different usages of the candlestick patterns, but we are going to use them as they are implemented on TA-Lib package of python.

Figure 3: CandleStick Patterns

TA-Lib Library offers different functions on Technical Analysis, but we are only interested on candlestick patterns. We are going to use all the candlestick patterns that are available on TA-Lib. Please see http://mrjbq7.github.io/ta-lib/ for more information.

For our purposes, candlestick functions return 100,200, -100,100 values which represent a buying signal if return is bigger than zero and selling signal if return is smaller than zero. Buying will be expressed as Long and selling will be expressed as Short.

4. DATA ANALYSIS

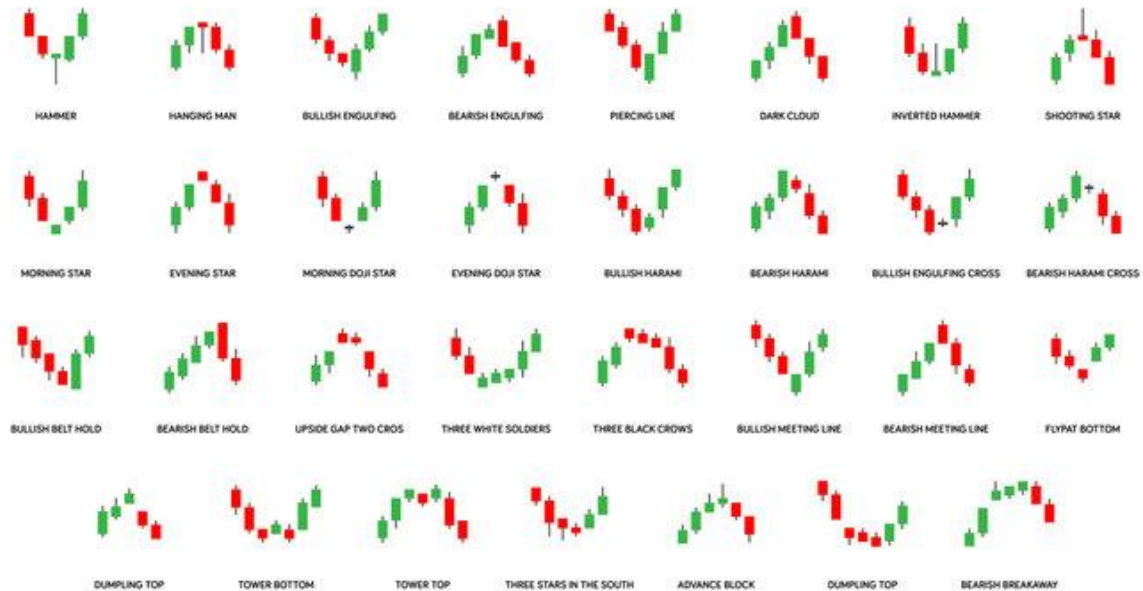We have the data between 17-08-2017 and 31-01-2022 for all 5-minute intervals. In total there are 466993 5-minute candles. However, we are not going to use all these data at once. Rather we are going to use a small proportion of this data because stock prices tend to move in a trend and trying to make sense out of the whole data for a small test part does not make sense for our case, since the method will over fit the trend and the results will be worse thereof.

For ease of explanation, we are going to use the data between 2-12-2021 and 20-12-2021 and try to explain the model according to this data. In the conclusion part, we are going to explain the general results for the tests that we have done.

There are 61 candlestick patterns that are available on TA-Lib. But not all occur on the data that we have chosen. However, sometimes more than one of the candle patterns occur. Please see Appendix 1 to see the individual pattern analysis. In the patterns analysis part, we are trying to find how many times the patterns occurred individually, what are their accuracy on their suggestions, how well they work with position taking mechanism and the accuracy on their positions.

We create an accuracy tester that checks, if the price moved %0.5 percent in the next 14 bars as we desired. And we used a decision mechanism that, basically has properties below.

- Open Position When Signal Comes
- Create a Stop Loss when the price moves %1 the way we wanted and update the stop loss as the price moves the way we desired
- Close position when either.
    - I. Reverse position is offered
    - II. We gain %20 profit
    - III. We have %10 loss

Before heading more into technical analysis part let's view the data.

Index	Long	Short	Total
MoreThanOneLongPattern	771	442	1213
MoreThanOneShortPattern	322	529	851
BothLongandShortPattern	433	1	442

Figure 4: Occurrence Analysis

From Figure 4 what we understand is that in 1213 rows of total 5000 rows, that we chose for experimenting for this report, there are more than one long suggestions from the candlestick patterns. However, we can see that 442 of 1213 rows also has patterns that are suggesting short. Similarly, for Short, there are 851 rows that have more than one short suggestion but 322 of them also has long suggestions as well. What we understand from here is that there is no certainty on suggestions as expected.

To add more detail, if we look at the Appendix 1, we can see that some of patterns are only giving signals for one position which makes their accuracy to drop on decision making part. Thus, we divided the patterns into three separate groups: giving both signals, only giving long signals, and only giving short candles. Please check Appendix 2.1,2.2,2.3.

5. METHODS

As mentioned in the Data Analysis part, we use an accuracy tester and a decision tester to check how well our model will work.

We'll combine these testers with 3 methods to get to a model which will give us support on our decisions.

First, combine all the one-way suggester patterns with each other (one long suggesting pattern and one short suggesting pattern) to see if they work well and represent them as a+b. We do this because pattern suggestions lose sense when they can't close the position. To ensure that the position is profitable we must make a decision to close the position. We tried to ensure that the positions are closing making the assumption that the combination of two pattern will work well together. We check the results and use the ones that seem to work.

Second, we'll use apriori to check if the decision gets stronger when two of patterns offer the same position and represent them as a*b. We do this because some patterns occur way too much and they tend lose its accuracy in the long run. Since the time series problem does not have a certain outcome it is very hard to predict what will happen but trying to get to a stronger connection between patterns can have more significant results in a sense make stronger decisions.

And finally, we'll use a ranking mechanism to make our decisions. We do this do have a stronger background on making decisions. Some of the patterns are working better on the given time interval and they should effect the decision making more. We decided to use the returns as their ranking to get to a better model.

And finally run the new model on the train and test data to see how well the model worked for the chosen time interval.

6. IMPLEMENTATION

First, we get the groups that are mentioned in the data analysis part and check their accuracy on both decision making and opening position. Both position suggester group runs individually. Then we combine the only long suggester and the short suggester group, to see which combinations work great together. For our case, we assumed that the patterns which are gaining more than %10 as the successful patterns and we put them into a database. We chose them as significant patterns. Please see Appendix 3.

Then, for the left-over patterns, we ran Apriori on these patterns to see if they get some useful information when used together. We get the patterns that occur in the same row (long or short does not matter for our case), we tried to get to a level of confidence where they worked well. Please see

Appendix 4 for Apriori results. We put the patterns which are successful (return bigger than %10) into significant pattern list.

Finally, we basically ordered the significant patterns by their gained money values to create a ranking and deducted the ones that have less than 0.75 position accuracy and also deducted the ones that have the same results since apriori returned decisions that have the same properties from other patterns.

With the rankings, we made assumption that we'll open the positions that are offered by the significant patterns that are left. Please see Appendix 5.

7. WHAT WE ALSO TRIED

We also came up with the idea of checking if seeing more than one position is enough to say it is a correct position. However, this idea did not seem to make sense after testing.

We tried to run Apriori by dividing the positions. This applied to very little part of the pattern set. However, in this position, there was also a need on combining short and long positions together again and this led to very high computation times, which is not desired considering this project is very time sensitive.

8. RESULTS

Now, we will check the model that we created for training data on the test data. We chose the next 1000 (almost 3.5 days) rows as the test data. What this means is that (our first time interval was 5000 rows) we get a 5 to 1 train/test ratio.

When we check the significant candlestick patterns individually (on test data), we see that their accuracy is great but they are not opening as many positions, thus we get less money. (considering our assumptions)

Index	Long	orrectLon	alseLong	Short	orrectShc	alseShor	alOccure	Accuracy	loney Gaine	itionAccur	sitionCo
CDLXSIDEGAP3METHODS	nan	nan	nan	nan	nan	nan	nan	nan	33.7918	1	1
CDLTRISTAR+CDLSTALLEDPATTERN	nan	nan	nan	nan	nan	nan	nan	nan	10.9856	1	1
CDLHARAMICROSS	nan	nan	nan	nan	nan	nan	nan	nan	39.5459	0.833333	6
CDLSTALLEDPATTERN	nan	nan	nan	nan	nan	nan	nan	nan	10.9856	1	1
CDLGAPSIDESIDEWHITE+CDLSTALLEDPATTERN	nan	nan	nan	nan	nan	nan	nan	nan	10.9856	1	1
CDLHANGINGMAN	nan	nan	nan	nan	nan	nan	nan	nan	28.2987	1	3
CDLTRISTAR+CDLHANGINGMAN	nan	nan	nan	nan	nan	nan	nan	nan	28.2987	1	3
CDLGAPSIDESIDEWHITE+CDLHANGINGMAN	nan	nan	nan	nan	nan	nan	nan	nan	28.2987	1	3
CDLMORNINGSTAR+CDLSTALLEDPATTERN	nan	nan	nan	nan	nan	nan	nan	nan	-0.970899	0.5	4

Figure 5: Significant Pattern Tested Individually on Test Data

When we used all the patternw given above together, we get the results below.

Index	Long	orrectLon	alseLong	Short	orrectShc	alseShor	alOccure	Accuracy	oney Gain	itionAccur	sitionCou
TrainData	41	26	15	108	72	36	149	0.657718	48.0808	0.650602	83
TestData	9	2	7	20	9	11	29	0.37931	49.7911	0.642857	14

Figure 6: First Test Results

Comparing on train data, these results does not seem that good since, the individual results look way better than the table results above. However, when we look at the test data, we see that we have more profit. The statement here is acceptable since we know the train data and we used that to train the model. So saying that it worked better is normal, but getting better results for unknown periods is what we wanted to achieve and in a sense it seems like we could do that here.

For further analysis, we tried to use only the first three significant patterns and tested again.

Index	Long	orrectLor	alseLonc	Short	orrectShc	alseShor	alOccure	Accuracy	oney Gain	itionAccur	sitionCou
TrainData	27	17	10	34	23	11	61	0.655738	178.799	0.72093	43
TestData	5	2	3	6	2	4	11	0.363636	37.551	0.714286	7

Figure 7: Second Test Results

As we can see, the model works much better on train data. However, it works worse for the test data according to returns. Although, this seems like a worse solution when we look at the position accuracy, we can see that it improved and that would be a much desirable outcome for us.

Please check Appendix 6 to see the trading mechanism results (for the results of a different test data)

9. CONSLUSION

To summarize, our goal was to create a system that would support us to make decisions on how the stock price will move. Using the candlestick patterns, we tried to achieve that goal. We tried different methods, but this was the best model that we could reach. So, the model that we are using can be explained as trying to understand how we can combine the candlestick patterns so that we can create a profitable indicator. We checked their combinations and tried to make sense of the candlestick charts that does not give reliable results. Overall, we have a momentum indicator that realizes the trend and try to come up with the combination of candlestick patterns that could give support to decisions.

Although, the example results were not bad, there may be some improvements with the model. After lots of testing, our takeaways are listed below.

- When the effects of significant candlestick patterns are not sensible on train data, model does not work on test data. So, it means that we will not make any decisions according to this model on that time interval
- Apriori helped very little due to usage on unsignificant patterns. Maybe more extensive studies can be made on apriori (such as working on whole model) but since the significant patterns are already working goof, we did not think that using it for whole data set would make sense since apriori created duplicates of significant patterns
- This model acted as a momentum indicator and should be combined with other technical analysis tools. It seems like it is not sensible to only depend on this model although it works good on some scenarios.
- A better method can be reached by combining with shifting the rows. (if a pattern occurs and if the other occurs in 15 bars it supports it.) However, this seems to be very time consuming.
- Sine the Long Suggesting patterns are more than Short Suggesting Patterns, sometimes changing positions can be problematic (for example, no short suggesting patterns in significant patterns), thus leading to loss of money

10. APPENDIX

Appendix 1: All The Candlestick Patterns That Occurred

Index	Long	orrectLor	alseLonc	Short	orrectShc	alseShor	alOccure	Accuracy	oney Gain	sitionAccur	sitionCou
CDL3INSIDE	18	8	10	15	8	7	33	0.484848	-32.5469	0.692308	26
CDL3LINESTRIKE	10	4	6	7	5	2	17	0.529412	-165.589	0.466667	15
CDL3OUTSIDE	64	26	38	92	46	46	156	0.461538	-84.7092	0.451613	93
CDL3WHITESOLDIERS	2	0	2	0	0	0	2	0	-204.706	0	2
CDLADVANCEBLOCK	0	0	0	65	31	34	65	0.476923	83.916	0.875	16
CDLBELTHOLD	445	200	245	414	206	208	859	0.472643	-377.528	0.307143	420
CDLCLOSINGMARUBOZU	302	127	175	311	150	161	613	0.451876	-375.771	0.311798	356
CDLDOJI	617	249	368	0	0	0	617	0.403566	-154.697	0.764706	17
CDLDOJISTAR	15	4	11	22	9	13	37	0.351351	243.044	0.586207	29
CDLDRAGONFLYDOJI	86	30	56	0	0	0	86	0.348837	-153.815	0.6	10
CDLENGULFING	141	58	83	177	103	74	318	0.506289	-23.8848	0.494048	168
CDLEVENINGDOJISTAR	0	0	0	5	2	3	5	0.4	26.0814	1	3
CDLEVENINGSTAR	0	0	0	13	7	6	13	0.538462	49.6294	0.833333	6
CDLGAPSIDESIDEWHITE	5	4	1	0	0	0	5	0.8	-110.887	0.666667	3
CDLGRAVESTONEDOJI	65	19	46	0	0	0	65	0.292308	-152.994	0.714286	7
CDLHAMMER	135	56	79	0	0	0	135	0.414815	-181.783	0.666667	12
CDLHANGINGMAN	0	0	0	78	51	27	78	0.653846	144.267	0.75	12
CDLHARAMI	84	40	44	89	48	41	173	0.508671	-18.4782	0.528302	106
CDLHARAMICROSS	15	9	6	13	8	5	28	0.607143	156.306	0.791667	24
CDLHIGHWAVE	292	121	171	268	138	130	560	0.4625	-522.797	0.333333	291
CDLHIKKAKE	320	141	179	393	208	185	713	0.489481	-486.303	0.388715	319
CDLHIKKAKEMOD	3	0	3	5	4	1	8	0.5	18.317	0.833333	6
CDLIDENTICAL3CROWS	0	0	0	6	2	4	6	0.333333	47.0176	1	6
CDLINVERTEDHAMMER	17	4	13	0	0	0	17	0.235294	-104.863	0.666667	6
CDLLONGLEGGEDDOJI	616	249	367	0	0	0	616	0.404221	-147.142	0.777778	18
CDLLONGLINE	466	213	253	451	226	225	917	0.478735	-387.649	0.2846	513
CDLMARUBOZU	126	55	71	93	51	42	219	0.484018	360.567	0.547619	126
CDLMATCHINGLOW	72	31	41	0	0	0	72	0.430556	-143.926	0.727273	11
CDLMORNINGDOJISTAR	3	2	1	0	0	0	3	0.666667	7.75787	0.666667	3
CDLMORNINGSTAR	10	5	5	0	0	0	10	0.5	56.0266	0.833333	6
CDLRICKSHAWMAN	436	173	263	0	0	0	436	0.396789	-167.82	0.736842	19
CDLRISEFALL3METHODS	2	2	0	1	0	1	3	0.666667	-39.4215	0.666667	3
CDLSEPARATINGLINES	17	7	10	17	11	6	34	0.529412	78.8687	0.75	24
CDLSHOOTINGSTAR	0	0	0	30	15	15	30	0.5	66.328	0.875	8
CDLSHORTLINE	361	155	206	325	174	151	686	0.479592	-591.442	0.272464	345
CDLSPINNINGTOP	507	211	296	477	247	230	984	0.465447	-562.129	0.328629	496
CDLSTALLEDPATTERN	0	0	0	10	5	5	10	0.5	155.965	0.875	8
CDLTAKURI	87	30	57	0	0	0	87	0.344828	-153.815	0.6	10
CDLTRISTAR	1	1	0	0	0	0	1	1	6.7405	1	1
CDLXSIDEGAP3METHODS	11	7	4	11	10	1	22	0.772727	187.268	0.857143	21

Appendix 2.1: Both Position Suggester Candles

bothpositioncandles - DataFrame

Index	Long	orrectLor	FalseLonc	Short	orrectShc	alseShor	alOccure	Accuracy	oney Gain	sitionAccur	sitionCou
CDL3INSIDE	18	8	10	15	8	7	33	0.484848	-32.5469	0.692308	26
CDL3LINESTRIKE	10	4	6	7	5	2	17	0.529412	-165.589	0.466667	15
CDL3OUTSIDE	64	26	38	92	46	46	156	0.461538	-84.7092	0.451613	93
CDLBELTHOLD	445	200	245	414	206	208	859	0.472643	-377.528	0.307143	420
CDLCLOSINGMARUBOZU	302	127	175	311	150	161	613	0.451876	-375.771	0.311798	356
CDLDOJISTAR	15	4	11	22	9	13	37	0.351351	243.044	0.586207	29
CDLENGULFING	141	58	83	177	103	74	318	0.506289	-23.8848	0.494048	168
CDLHARAMI	84	40	44	89	48	41	173	0.508671	-18.4782	0.528302	106
CDLHARAMICROSS	15	9	6	13	8	5	28	0.607143	156.306	0.791667	24
CDLHIGHWAVE	292	121	171	268	138	130	560	0.4625	-522.797	0.333333	291
CDLHIKKAKE	320	141	179	393	208	185	713	0.489481	-486.303	0.388715	319
CDLHIKKAKEMOD	3	0	3	5	4	1	8	0.5	18.317	0.833333	6
CDLLONGLINE	466	213	253	451	226	225	917	0.478735	-387.649	0.2846	513
CDLMARUBOZU	126	55	71	93	51	42	219	0.484018	360.567	0.547619	126
CDLRISEFALL3METHODS	2	2	0	1	0	1	3	0.666667	-39.4215	0.666667	3
CDLSEPARATINGLINES	17	7	10	17	11	6	34	0.529412	78.8687	0.75	24
CDLSHORTLINE	361	155	206	325	174	151	686	0.479592	-591.442	0.272464	345
CDLSPINNINGTOP	507	211	296	477	247	230	984	0.465447	-562.129	0.328629	496
CDLXSIDEGAP3METHODS	11	7	4	11	10	1	22	0.772727	187.268	0.857143	21

Appendix 2.2: Only Long Suggester Candles

onlylongcandles - DataFrame

Index	Long	orrectLor	FalseLonc	Short	orrectShc	alseShor	alOccure	Accuracy	oney Gain	sitionAccur	sitionCou
CDL3WHITESOLDIERS	2	0	2	0	0	0	2	0	-204.706	0	2
CDLDOJI	617	249	368	0	0	0	617	0.403566	-154.697	0.764706	17
CDLDRAGONFLYDOJI	86	30	56	0	0	0	86	0.348837	-153.815	0.6	10
CDLGAPSIDESIDEWHITE	5	4	1	0	0	0	5	0.8	-110.887	0.666667	3
CDLGRAVESTONEDOJI	65	19	46	0	0	0	65	0.292308	-152.994	0.714286	7
CDLHAMMER	135	56	79	0	0	0	135	0.414815	-181.783	0.666667	12
CDLINVERTEDHAMMER	17	4	13	0	0	0	17	0.235294	-104.863	0.666667	6
CDLLONGLEGGEDDOJI	616	249	367	0	0	0	616	0.404221	-147.142	0.777778	18
CDLMATCHINGLOW	72	31	41	0	0	0	72	0.430556	-143.926	0.727273	11
CDLMORNINGDOJISTAR	3	2	1	0	0	0	3	0.666667	7.75787	0.666667	3
CDLMORNINGSTAR	10	5	5	0	0	0	10	0.5	56.0266	0.833333	6
CDLRICKSHAWMAN	436	173	263	0	0	0	436	0.396789	-167.82	0.736842	19
CDLTAKURI	87	30	57	0	0	0	87	0.344828	-153.815	0.6	10
CDLTRISTAR	1	1	0	0	0	0	1	1	6.7405	1	1

Appendix 2.3: Only Short Suggester Candles

onlyshortcandles - DataFrame

Index	Long	orrectLor	FalseLonc	Short	orrectShc	alseShor	alOccure	Accuracy	oney Gain	sitionAccur	sitionCou
CDLADVANCEBLOCK	0	0	0	65	31	34	65	0.476923	83.916	0.875	16
CDLEVENINGDOJISTAR	0	0	0	5	2	3	5	0.4	26.0814	1	3
CDLEVENINGSTAR	0	0	0	13	7	6	13	0.538462	49.6294	0.833333	6
CDLHANGINGMAN	0	0	0	78	51	27	78	0.653846	144.267	0.75	12
CDLIDENTICAL3CROWS	0	0	0	6	2	4	6	0.333333	47.0176	1	6
CDLSHOOTINGSTAR	0	0	0	30	15	15	30	0.5	66.328	0.875	8
CDLSTALLEDPATTERN	0	0	0	10	5	5	10	0.5	155.965	0.875	8

Appendix 3: Significant Candle Pattern List

Index	Long	orrectLor	FalseLong	Short	orrectSho	alseShor	alOccure	Accuracy	oney Gain	sitionAccur	sitionCou
CDLDOJISTAR	15	4	11	22	9	13	37	0.351351	243.044	0.586207	29
CDLHARAMICROSS	15	9	6	13	8	5	28	0.607143	156.306	0.791667	24
CDLMARUBOZU	126	55	71	93	51	42	219	0.484018	360.567	0.547619	126
CDLXSIDEGAP3METHODS	11	7	4	11	10	1	22	0.772727	187.268	0.857143	21
CDLGAPSIDESIDEWHITE+CDLHANGINGMAN	4	3	1	77	51	26	81	0.666667	130.114	0.76	25
CDLGAPSIDESIDEWHITE+CDLSTALLEDPATTERN	5	4	1	10	5	5	15	0.6	152.638	0.785714	14
CDLINVERTEDHAMMER+CDLSTALLEDPATTERN	17	4	13	10	5	5	27	0.333333	111.068	0.5	16
CDLMORNINGDOJISTAR+CDLADVANCEBLOCK	3	2	1	65	31	34	68	0.485294	211.115	0.740741	27
CDLMORNINGSTAR+CDLADVANCEBLOCK	10	5	5	65	31	34	75	0.48	137.692	0.558824	34
CDLMORNINGSTAR+CDLSTALLEDPATTERN	10	5	5	10	5	5	20	0.5	129.043	0.785714	14
CDLTRISTAR+CDLHANGINGMAN	1	1	0	78	51	27	79	0.658228	136.533	0.789474	19
CDLTRISTAR+CDLSTALLEDPATTERN	1	1	0	10	5	5	11	0.545455	163.757	0.888889	9
CDLHANGINGMAN	0	0	0	78	51	27	78	0.653846	144.267	0.75	12
CDLSTALLEDPATTERN	0	0	0	10	5	5	10	0.5	155.965	0.875	8

Index	Long	orrectLon	alseLon	Short	orrectSho	alseShor	alOccure	Accuracy	oney Gain	sitionAccur	sitionCou
CDLDOJI	617	249	368	0	0	0	617	0.403566	-154.697	0.764706	17
CDLENGULFING	141	58	83	177	103	74	318	0.506289	-23.8848	0.494048	168
CDLHARAMI	84	40	44	89	48	41	173	0.508671	-18.4782	0.528302	106
CDLHIGHWAVE	292	121	171	268	138	130	560	0.4625	-522.797	0.333333	291
CDLHIKKAKE	320	141	179	393	208	185	713	0.489481	-486.303	0.388715	319
CDLLONGLEGGEDDOJI	616	249	367	0	0	0	616	0.404221	-147.142	0.777778	18
CDLLONGLINE	466	213	253	451	226	225	917	0.478735	-387.649	0.2846	513
CDLRICKSHAWMAN	436	173	263	0	0	0	436	0.396789	-167.82	0.736842	19
CDLSHORTLINE	361	155	206	325	174	151	686	0.479592	-591.442	0.272464	345
CDLSPINNINGTOP	507	211	296	477	247	230	984	0.465447	-562.129	0.328629	496
CDLBELTHOLD*CDLCLOSINGMARUBOZU	126	55	71	93	51	42	219	0.484018	360.567	0.547619	126
CDLBELTHOLD*CDLHIKKAKE	16	8	8	25	16	9	41	0.585366	-74.5534	0.545455	33
CDLBELTHOLD*CDLLONGLINE	269	123	146	239	121	118	508	0.480315	-24.4561	0.394464	289
CDLLONGLINE*CDLCLOSINGMARUBOZU	224	97	127	193	97	96	417	0.465228	-147.329	0.340081	247
CDLDOJI*CDLHIGHWAVE	238	97	141	0	0	0	238	0.407563	-140.155	0.8125	16
CDLDOJI*CDLLONGLEGGEDDOJI	616	249	367	0	0	0	616	0.404221	-147.142	0.777778	18
CDLDOJI*CDLRICKSHAWMAN	436	173	263	0	0	0	436	0.396789	-167.82	0.736842	19
CDLDOJI*CDLSPINNINGTOP	284	114	170	0	0	0	284	0.401408	-146.48	0.769231	13
CDLHIGHWAVE*CDLLONGLEGGEDDOJI	238	97	141	0	0	0	238	0.407563	-140.155	0.8125	16
CDLHIGHWAVE*CDLRICKSHAWMAN	199	83	116	0	0	0	199	0.417085	-139.974	0.8125	16
CDLHIGHWAVE*CDLSPINNINGTOP	292	121	171	268	138	130	560	0.4625	-522.797	0.333333	291
CDLHIKKAKE*CDLLONGLINE	20	10	10	36	24	12	56	0.607143	124.513	0.681818	44
CDLLONGLEGGEDDOJI*CDLRICKSHAWMAN	436	173	263	0	0	0	436	0.396789	-167.82	0.736842	19
CDLSPINNINGTOP*CDLLONGLEGGEDDOJI	284	114	170	0	0	0	284	0.401408	-146.48	0.769231	13
CDLSPINNINGTOP*CDLRICKSHAWMAN	233	97	136	0	0	0	233	0.416309	-142.681	0.8	15
CDLBELTHOLD*CDLLONGLINE*CDLCLOSINGMARUBOZU	126	55	71	93	51	42	219	0.484018	360.567	0.547619	126
CDLDOJI*CDLHIGHWAVE*CDLLONGLEGGEDDOJI	238	97	141	0	0	0	238	0.407563	-140.155	0.8125	16
CDLDOJI*CDLHIGHWAVE*CDLRICKSHAWMAN	199	83	116	0	0	0	199	0.417085	-139.974	0.8125	16
CDLDOJI*CDLHIGHWAVE*CDLSPINNINGTOP	238	97	141	0	0	0	238	0.407563	-140.155	0.8125	16
CDLDOJI*CDLLONGLEGGEDDOJI*CDLRICKSHAWMAN	436	173	263	0	0	0	436	0.396789	-167.82	0.736842	19
CDLDOJI*CDLSPINNINGTOP*CDLLONGLEGGEDDOJI	284	114	170	0	0	0	284	0.401408	-146.48	0.769231	13
CDLDOJI*CDLSPINNINGTOP*CDLRICKSHAWMAN	233	97	136	0	0	0	233	0.416309	-142.681	0.8	15
CDLHIGHWAVE*CDLLONGLEGGEDDOJI*CDLRICKSHAWM...	199	83	116	0	0	0	199	0.417085	-139.974	0.8125	16
CDLHIGHWAVE*CDLSPINNINGTOP*CDLLONGLEGGEDDOJI	238	97	141	0	0	0	238	0.407563	-140.155	0.8125	16
CDLHIGHWAVE*CDLSPINNINGTOP*CDLRICKSHAWMAN	199	83	116	0	0	0	199	0.417085	-139.974	0.8125	16
CDLSPINNINGTOP*CDLLONGLEGGEDDOJI*CDLRICKSHA...	233	97	136	0	0	0	233	0.416309	-142.681	0.8	15
CDLDOJI*CDLHIGHWAVE*CDLLONGLEGGEDDOJI*CDLRIC...	199	83	116	0	0	0	199	0.417085	-139.974	0.8125	16
CDLDOJI*CDLHIGHWAVE*CDLSPINNINGTOP*CDLLONGL...	238	97	141	0	0	0	238	0.407563	-140.155	0.8125	16
CDLDOJI*CDLHIGHWAVE*CDLSPINNINGTOP*CDLRICKSH...	199	83	116	0	0	0	199	0.417085	-139.974	0.8125	16
CDLDOJI*CDLSPINNINGTOP*CDLLONGLEGGEDDOJI*CDL...	233	97	136	0	0	0	233	0.416309	-142.681	0.8	15
CDLHIGHWAVE*CDLSPINNINGTOP*CDLLONGLEGGEDDO...	199	83	116	0	0	0	199	0.417085	-139.974	0.8125	16
CDLHIGHWAVE*CDLRICKSHAWMAN*CDLSPINNINGTOP*...	199	83	116	0	0	0	199	0.417085	-139.974	0.8125	16

Appendix 5: Significant Candles from All Models

Index	Long	orrectLor	FalseLonc	Short	orrectShc	alseShor	alOccure	Accuracy	ney Gain	sitionAccur	sitionCou
CDLXSIDEGAP3METHODS	11	7	4	11	10	1	22	0.772727	187.268	0.857143	21
CDLTRISTAR+CDLSTALLEDPATTERN	1	1	0	10	5	5	11	0.545455	163.757	0.888889	9
CDLHARAMICROSS	15	9	6	13	8	5	28	0.607143	156.306	0.791667	24
CDLSTALLEDPATTERN	0	0	0	10	5	5	10	0.5	155.965	0.875	8
CDLGAPSIDESIDEWHITE+CDLSTALLEDPATTERN	5	4	1	10	5	5	15	0.6	152.638	0.785714	14
CDLHANGINGMAN	0	0	0	78	51	27	78	0.653846	144.267	0.75	12
CDLTRISTAR+CDLHANGINGMAN	1	1	0	78	51	27	79	0.658228	136.533	0.789474	19
CDLGAPSIDESIDEWHITE+CDLHANGINGMAN	4	3	1	77	51	26	81	0.666667	130.114	0.76	25
CDLMORNINGSTAR+CDLSTALLEDPATTERN	10	5	5	10	5	5	20	0.5	129.043	0.785714	14

Appendix 6: Trading on Different Test Data (Rows:420000-425000 for Train, 425000-426000 for Test)

Index	Long	orrectLor	FalseLonc	Short	orrectShc	alseShor	alOccure	Accuracy	ney Gain	sitionAccur	sitionCou
TrainData	91	40	51	0	0	0	91	0.43956	104.404	1	8
TestData	19	12	7	0	0	0	19	0.631579	16.2157	0.666667	3

Index	tartCapita	sition Tak	startprice	HowManv	StartTime	EndTime	Duration	endprice	ndCapita	ProfitLoss	Perc
0	1000	1	51671.4	0.0193234	2021-09-06 21:05:00	2021-09-07 03:50:00	81	52493	1013.33	13.3304	0.0133304
1	1013.33	1	46797.2	0.021645	2021-09-07 20:50:00	2021-09-08 05:15:00	101	47287.1	1022.51	9.17634	0.00905562
2	1022.51	1	47055.5	0.0217202	2021-09-08 06:00:00	2021-09-10 05:30:00	571	46833.4	1016.22	-6.29104	-0.00615257

10. REFERENCES

1. Figure 1: https://stock.adobe.com/ee/search?k=candlestick+pattern
2. Figure 2: https://www.definedge.com/candlestick-patterns/