# BTC/USD Prediction

Machida Hiroaki, Yi Ting Yeh

CS699 A1 – Fall 2020 Data Mining Project Assignment

# Table of Contents

## Chapter 1 Responsibilities

Machida Hiroaki was in charge of data collection to attribute & model selection. Yi Ting Yeh was responsible for model testing and performance comparison. Although we defined responsibilities, each phase is done with close collaboration.

## Chapter 2 Data Mining Goal

To predict whether BTCUSD will go up or down on the next day.

## Chapter 3 Dataset

**Overview**

Time series price data of BTC/USD, other currency pairs, market indices, commodities, stocks, and bonds on the same day as BTC/USD and the previous day, shown as "up", "down", or "equal". For all the attributes and class attribute, "up" means the closing price of the day is higher than the closing price of the previous day. "down" means the opposite. "equal" when the two prices are equal. If the value is N/A, "equal" is put in.
To the best of my knowledge, the timings of closing are BTC/USD 18:00, Currency 18:00, Futures 17:00, Bond 17:00, and Stock 16:00.

Number of Attributes: 63
Number of Tuples: 387

**Attributes**

<u>Class attribute</u>
1. BTC/USD
Change of Bitcoin Dollar US Dollar closing price from the previous day to the day shown as "up", "down", or "equal".
Source: https://www.investing.com/crypto/bitcoin/btc-usd?cid=21

<u>Predictor attributes</u>
1. Date
The number of business days passed starting from 1 on 27-Mar-19.

2. USD/JPY
Change of US Dollar Japanese Yen closing price from the day to the next day shown as "up", "down", or "equal".
Source: https://www.investing.com/currencies/usd-jpy

3. EUR/USD
Change of EURO US Dollar closing price from the previous day to the day shown as "up", "down", or "equal".
Source: https://www.investing.com/currencies/eur-usd

4. GBP/USD
Change of British Pound US Dollar closing price from the previous day to the day shown as "up", "down", or "equal".
Source: https://www.investing.com/currencies/gbp-usd

5. AUD/USD
Change of Australian Dollar US Dollar closing price from the previous day to the day shown as “up”, “down”, or “equal”.
Source: https://www.investing.com/currencies/aud-usd

6. S&P 500
Change of S&P closing price from the previous day to the day shown as “up”, “down”, or “equal”. S&P is a stock market index that measures the stock performance of 500 large companies in the US, provided by S&P Dow Jones Indices LLC.
Source: https://www.investing.com/indices/us-spx-500

7. Dow Jones Industrial Average
Change of Dow Jones Industrial Average closing price from the previous day to the day shown as “up”, “down”, or “equal”. Dow Jones Industrial Average is a stock market index that measures the stock performance of 30 large companies in the US, provided by S&P Dow Jones Indices LLC.
Source: https://www.investing.com/indices/us-30

8. Nasdaq 100
Change of Nasdaq 100 closing price from the previous day to the day shown as “up”, “down”, or “equal”. Nasdaq 100 is a stock market index that measures the stock performance of 100 large non-financial companies in the US, provided by NASDAQ.
Source: https://www.investing.com/indices/nq-100

9. Nikkei 225
Change of Nikkei 225 closing price from the previous day to the day shown as “up”, “down”, or “equal”. Nikkei 225 is a stock market index that measures the stock performance of 225 large companies in Japan, provided by Nihon Keizai Shimbun.
Source: https://www.investing.com/indices/japan-ni225

10. Gold Futures
Change of Gold Futures closing price from the previous day to the day shown as “up”, “down”, or “equal”.
Source: https://www.investing.com/commodities/gold

11. Silver Futures
Change of Silver Futures closing price from the previous day to the day shown as “up”, “down”, or “equal”.
Source: https://www.investing.com/commodities/silver

12. Copper Futures
Change of Copper Futures closing price from the previous day to the day shown as “up”, “down”, or “equal”.
Source: https://www.investing.com/commodities/copper

13. Crude Oil WTI Futures
Change of Crude Oil WTI Futures closing price from the previous day to the day shown as “up”, “down”, or “equal”. This is a benchmark for US oil prices.
Source: https://www.investing.com/commodities/crude-oil

14. Brent Oil Futures
Change of Brent Oil Futures closing price from the previous day to the day shown as “up”, “down”, or “equal”. This is a benchmark used by the Organization of Petroleum Exporting Countries (OPEC).
Source: https://www.investing.com/commodities/brent-oil

15. Natural Gas Futures
Change of Natural Gas Futures closing price from the previous day to the day shown as “up”, “down”, or “equal”.
Source: https://www.investing.com/commodities/natural-gas

16. US Cotton #2 Futures
Change of US Cotton #2 Futures closing price from the previous day to the day shown as “up”, “down”, or “equal”. US Cotton #2 Futures are the ones on the New York Board of Trade.
Source: https://www.investing.com/commodities/us-cotton-no.2

17. US Coffee C Futures
Change of US Coffee C Futures closing price from the previous day to the day shown as “up”, “down”, or “equal”. C just stands for Coffee.
Source: https://www.investing.com/commodities/us-coffee-c

18. Apple
Change of Apple stock closing price from the previous day to the day shown as “up”, “down”, or “equal”.
Source: https://www.investing.com/equities/apple-computer-inc

19. Alphabet A
Change of Alphabet Inc Class A stock closing price from the previous day to the day shown as “up”, “down”, or “equal”.
Source: https://www.investing.com/equities/google-inc

20. Facebook
Change of Facebook stock closing price from the previous day to the day shown as “up”, “down”, or “equal”.
Source: https://www.investing.com/equities/facebook-inc

21. NVIDIA
Change of NVIDIA stock closing price from the previous day to the day shown as “up”, “down”, or “equal”.
Source: https://www.investing.com/equities/nvidia-corp

22. Citigroup
Change of Citigroup stock closing price from the previous day to the day shown as “up”, “down”, or “equal”.
Source: https://www.investing.com/equities/citigroup

23. AT&T
Change of AT&T stock closing price from the previous day to the day shown as “up”, “down”, or “equal”.

Source: https://www.investing.com/equities/at-t

24. 3M
Change of 3M stock closing price from the previous day to the day shown as “up”, “down”, or “equal”.
Source: https://www.investing.com/equities/3m-co

25. BP
Change of BP stock closing price from the previous day to the day shown as “up”, “down”, or “equal”.
Source: https://www.investing.com/equities/bp

26. Tesla
Change of Tesla stock closing price from the previous day to the day shown as “up”, “down”, or “equal”.
Source: https://www.investing.com/equities/tesla-motors

27. Amazon.com
Change of Amazon stock closing price from the previous day to the day shown as “up”, “down”, or “equal”.
Source: https://www.investing.com/equities/amazon-com-inc

28. Japan Government Bond Futures
Change of Japan Government Bond Futures closing price from the previous day to the day shown as “up”, “down”, or “equal”.
Source: https://www.investing.com/rates-bonds/japan-govt.-bond

29. US 10 Year T-Note Futures
Change of US 10 Year T-Note Futures closing price from the previous day to the day shown as “up”, “down”, or “equal”.
Source: https://www.investing.com/rates-bonds/us-10-yr-t-note

30. US 30 Year T-Bond Futures
Change of US 30 Year T-Bond Futures closing price from the previous day to the day shown as “up”, “down”, or “equal”.
Source: https://www.investing.com/rates-bonds/us-30-yr-t-bond

31. US 2 Year T-Note Futures
Change of US 2 Year T-Note Futures closing price from the previous day to the day shown as “up”, “down”, or “equal”.
Source: https://www.investing.com/rates-bonds/us-2-yr-t-note

32. USD/JPY T-1
The closing price of USD/JPY on the previous day.

33. EUR/USD T-1
The closing price of EUR/USD on the previous day.

34. GBP/USD T-1
The closing price of GBP/USD on the previous day.

35. AUD/USD T-1
The closing price of AUD/USD on the previous day.

36. BTC/USD T-1
The closing price of BTC/USD on the previous day.

37. S&P 500 T-1
The closing price of S&P 500 on the previous day.

38. Dow Jones Industrial Average T-1
The closing price of Dow Jones Industrial Average on the previous day.

39. Nasdaq 100 T-1
The closing price of Nasdaq on the previous day.

40. Nikkei 225 T-1
The closing price of Nikkei 225 on the previous day.

41. Gold Futures T-1
The closing price of Gold Futures on the previous day.

42. Silver Futures T-1
The closing price of Silver Futures on the previous day.

43. Copper Futures T-1
The closing price of Copper Futures on the previous day.

44. Crude Oil WTI Futures T-1
The closing price of Crude Oil WTI Futures on the previous day.

45. Brent Oil Futures T-1
The closing price of Brent Oil Futures on the previous day.

46. Natural Gas Futures T-1
The closing price of XXX on the previous day.

47. US Cotton #2 Futures T-1
The closing price of US Cotton #2 Futures on the previous day.

48. US Coffee C Futures T-1
The closing price of US Coffee C Futures on the previous day.

49. Apple T-1
The closing price of Apple on the previous day.

50. Alphabet A T-1
The closing price of Alphabet A on the previous day.

51. Facebook T-1

The closing price of Facebook on the previous day.

52. NVIDIA T-1
The closing price of NVIDIA on the previous day.

53. Citigroup T-1
The closing price of Citigroup on the previous day.

54. AT&T T-1
The closing price of AT&T on the previous day.

55. 3M T-1
The closing price of 3M on the previous day.

56. BP T-1
The closing price of XXX on the previous day.

57. Tesla T-1
The closing price of XTeslaXX on the previous day.

58. Amazon.com T-1
The closing price of Amazon.com on the previous day.

59. Japan Government Bond Futures T-1
The closing price of Japan Government Bond Futures on the previous day.

60. US 10 Year T-Note Futures T-1
The closing price of US 10 Year T-Note Futures on the previous day.

61. US 30 Year T-Bond Futures T-1
The closing price of US 30 Year T-Bond Futures on the previous day.

62. US 2 Year T-Note Futures T-1
The closing price of XUS 2 Year T-Note Futures on the previous day.

## Chapter 4 Tools and Algorithms

**Tool**
Waikato Environment for Knowledge Analysis (Weka)[1] is used for this project. Weka is an open source machine learning software developed by University of Waikato, New Zealand.

**Algorithm**
The attribute evaluators with search methods on Weka are utilized to select attributes from the dataset, and the classifiers to train models and make predictions.

Attribute evaluators
Attribute evaluator is a function to estimate the worth of a subset of attributes.

[1] Weka 3 - Data Mining with Open Source Machine Learning Software in Java
https://www.cs.waikato.ac.nz/ml/weka/

1. CfsSubsetEval[2]
Correlation-based Feature Selection. Prefers the attributes with high correlation and low intercorrelation.

2. CorrelationAttributeEval
Measures the Pearson's correlation between it and the class.

3. OneRAttributeEval
Uses the OneR lassifier.

4. ReliefFAttributeEval[3] [4] [5]
Repeatedly samples instance and considers the value of the given attribute for the nearest instance of the same and different class.

Search methods
Search method is the way to search a better subset of attributes.

1. Best First
Searchs the subsets of attributes by greedy hilcliming augmented with a backtracking facility.

2. Ranker
Ranks attributes by individual attribute evaluators.

Classifiers
Classifier is the model that is trained by the train dataset and predicts the class attribute.

1. Bayes Net
Bayes Network with the default estimator, SimpleEstimator, and search algorithm, K2.

2. SGD
Stochastic Gradient Descent with the default loss function, Hinge loss (SVM).

3. Logit Boost[6]
A boosting algorithm with logistic loss.

4. Decision Table[7]
Builds and uses a simple decision table majority classifier.

---

[2] M. A. Hall (1998). Correlation-based Feature Subset Selection for Machine Learning. Hamilton, New Zealand.
[3] Kenji Kira, Larry A. Rendell: A Practical Approach to Feature Selection. In: Ninth International Workshop on Machine Learning, 249-256, 1992.
[4] Igor Kononenko: Estimating Attributes: Analysis and Extensions of RELIEF. In: European Conference on Machine Learning, 171-182, 1994.
[5] Marko Robnik-Sikonja, Igor Kononenko: An adaptation of Relief for attribute estimation in regression. In: Fourteenth International Conference on Machine Learning, 296-304, 1997.
[6] J. Friedman, T. Hastie, R. Tibshirani (1998). Additive Logistic Regression: a Statistical View of Boosting. Stanford University.
[7] Ron Kohavi: The Power of Decision Tables. In: 8th European Conference on Machine Learning, 174-189, 1995.

5. LMT[8] [9]
Logistic Model Trees. These are classification trees with logistic regression functions at the leaves.

## Chapter 5 Procedure

**Split dataset**
The dataset contains 387 entries from 27-Mar-19 to 17-Sep-20. It is shuffled and split into train and test by 2:1, 258 and 129 entries.
The datasets are correctly stratified. The frequency distribution of train is up 138, down 120. The frequency of test is up 65, down 64.

**Attribute selection**
All the attribute evaluators with a search method automatically selected are performed with train dataset.
First, we shall pick up four evaluators out of five attribute sets.
For best first search method, there were no evaluators that automatically choose it, so best first search method is not listed.
For greedy stepwise, only #1 GreedyStepwise + CfsSubsetEval has the eligible result, so that is chosen.
For ranker search method, the point where the scores show a sudden decrease is picked up as a threshold to separate attributes to chosen and not. The evaluators with the least number of attributes are chosen because these requires less computing power. There are three evaluators with two attributes, #4, #5, and #10, however, the attributes chosen are all the same, so #4 Ranker + CorrelationAttributeEval is randomly picked up. For the evaluators with three attributes, #6 and #7 have the same set of attributes, so #7 Ranker + OneRAttributeEval is chosen. #9 Ranker + ReliefFAttributeEval has an unique set of attributes, so this is chosen as well.
Second, we selected a set of attributes manually, that are considered to be the most relevant to the class attribute.

Attribute evaluators and search methods and attributes chosen

| # | Attribute evaluators and search methods | # of attributes | Selected |
|---|---|---|---|
| 1 | GreedyStepwise + CfsSubsetEval | 10 | * |
| 2 | Ranker + ClassifierAttributeEval | N/A | |
| 3 | GreedyStepwise + ClassifierSubsetEval | N/A | |
| 4 | Ranker + CorrelationAttributeEval | 2 | * |
| 5 | Ranker + GainRatioAttributeEval * same as #4 | 2 | |
| 6 | Ranker + InfoGainAttributeEval | 3 | |
| 7 | Ranker + OneRAttributeEval * same as #6 | 3 | * |
| 8 | Ranker + PrincipalComponents | N/A | |
| 9 | Ranker + ReliefFAttributeEval | 3 | * |
| 10 | Ranker + SymmetricalUncertAttributeEval * same as #4 | 2 | |
| 11 | GreedyStepwise + WrapperSubsetEval | N/A | |

[8] Niels Landwehr, Mark Hall, Eibe Frank (2005). Logistic Model Trees. Machine Learning. 95(1-2):161-205.
[9] Marc Sumner, Eibe Frank, Mark Hall: Speeding up Logistic Model Tree Induction. In: 9th European Conference on Principles and Practice of Knowledge Discovery in Databases, 675-683, 2005.

| | | | |
|---|---|---|---|
| 12 | attributes chosen by yourself. | 3 | * |

Attributes for each evaluators and search method

| # | Attribute evaluators and search methods | Attributes |
|---|---|---|
| 1 | GreedyStepwise + CfsSubsetEval (set1) | 5 AUD/USD<br>10 Gold Futures<br>11 Silver Futures<br>27 Amazon.com<br>29 US 10 Year T-Note Futures<br>32 USD/JPY T-1<br>42 Silver Futures T-1<br>46 Natural Gas Futures T-1<br>49 Apple T-1<br>59 Japan Government Bond Futures T-1 |
| 2 | Ranker + CorrelationAttributeEval (set2) | 10 Gold Futures<br>11 Silver Futures |
| 3 | Ranker + GainRatioAttributeEval | 10 Gold Futures<br>11 Silver Futures |
| 4 | Ranker + InfoGainAttributeEval | 10 Gold Futures<br>49 Apple T-1<br>11 Silver Futures |
| 5 | Ranker + OneRAttributeEval (set3) | 10 Gold Futures<br>11 Silver Futures<br>49 Apple T-1 |
| 6 | Ranker + ReliefFAttributeEval (set4) | 6 S&P 500<br>7 Dow Jones Industrial Average<br>10 Gold Futures |
| 7 | Ranker + SymmetricalUncertAttributeEval | 10 Gold Futures<br>11 Silver Futures |
| 8 | attributes chosen by yourself. (set5) | 1 Date<br>2 USD/JPY<br>10 Gold Futures |

**Model selection**

All the models on Weka were tested with the train data containing all the attributes. Models should be chosen from different categories to increase the chance for better performance. The categories are split into five, 1. Bayes, 2. Function, 3. Lazy, Meta, Miscellaneous, 4. Rules, and 5. Trees.

The models with the best accuracy are chosen from the categories. For the third category, #23 and #29 has the same accuracy, so #23 meta.LogitBoost is randomly chosen.

Model accuracy for train data

* Top 5 of accuracy %, Precision of up and down are highlighted.

| # | Model | Accuracy | Selected |
|---|---|---|---|
| 1 | bayes.BayesNet | 56.5891 | * |
| 2 | bayes.NaiveBayes | 55.0388 | |
| 3 | bayes.NaiveBayesMultinomialText | 53.4884 | |

| | | | |
|---|---|---|---|
| 4 | bayes.NaiveBayesUpdateable | 55.0388 | |
| 5 | functions.Logistic | 53.4884 | |
| 6 | functions.MultilayerPerceptron | 52.7132 | |
| 7 | functions.SGD | 54.2636 | * |
| 8 | functions.SGDText | 53.4884 | |
| 9 | functions.SimpleLogistic | 53.876 | |
| 10 | functions.SMO | 51.1628 | |
| 11 | functions.VotedPerceptron | 52.3256 | |
| 12 | lazy.IBk | 56.2016 | |
| 13 | lazy.KStar | 57.3643 | |
| 14 | lazy.LWL | 51.5504 | |
| 15 | meta.AdaBoostM1 | 56.9767 | |
| 16 | meta.AttributeSelectedClassifier | 52.7132 | |
| 17 | meta.Bagging | 57.3643 | |
| 18 | meta.ClassificationViaRegression | 52.3256 | |
| 19 | meta.CostSensitiveClassifier | | |
| 20 | meta.CVParameterSelection | 53.4884 | |
| 21 | meta.FilteredClassifier | 52.7132 | |
| 22 | meta.IterativeClassifierOptimizer | 50.7752 | |
| 23 | meta.LogitBoost | 57.7519 | * |
| 24 | meta.MultiClassClassifier | 53.4884 | |
| 25 | meta.MultiClassClassifierUpdateable | 54.2636 | |
| 26 | meta.MultiScheme | 53.4884 | |
| 27 | meta.RandomCommittee | 56.9767 | |
| 28 | meta.RandomizableFilteredClassifier | 52.3256 | |
| 29 | meta.RandomSubSpace | 57.7519 | |
| 30 | meta.Stacking | 53.4884 | |
| 31 | meta.Vote | 53.4884 | |
| 32 | meta.WeightedInstancesHandlerWrapper | 53.4884 | |
| 33 | misc.InputMappedClassifier | 53.4884 | |
| 34 | misc.SerializedClassifier | | |
| 35 | rules.DecisionTable | 58.5271 | * |
| 36 | rules.JRip | 54.6512 | |
| 37 | rules.OneR | 54.6512 | |
| 38 | rules.PART | 53.1008 | |
| 39 | rules.ZeroR | 53.4884 | |
| 40 | trees.DecisionStump | 53.4884 | |
| 41 | trees.HoeffdingTree | 54.6512 | |
| 42 | trees.J48 | 53.1008 | |
| 43 | trees.LMT | 55.0388 | * |
| 44 | trees.RandomForest | 54.2636 | |
| 45 | trees.RandomTree | 51.938 | |

| | | | |
|---|---|---|---|
| 46 | trees.REPTree | 51.5504 | |

**Performance test**

Now the five set of attribute selection and the five models are chosen. Each combination of a set of attribute selection and model will be trained with the train dataset, and tested with the test dataset.

# Chapter 6 Result and Evaluation

The results of performance measures are shown below.

Accuracy

| Attr set | bayes.BayesNet | | | functions.SGD | | | meta.LogitBoost | | | rules.DecisionTable | | | trees.LMT | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | avg | down | up | avg | down | up | avg | down | up | avg | down | up | avg | down | up |
| set1 | **0.597** | 0.547 | 0.646 | **0.581** | 0.469 | 0.692 | **0.581** | 0.453 | 0.708 | 0.527 | 0.375 | 0.677 | **0.581** | 0.484 | 0.677 |
| set2 | **0.581** | 0.547 | 0.615 | **0.581** | 0.547 | 0.615 | **0.581** | 0.547 | 0.615 | **0.581** | 0.547 | 0.615 | **0.581** | 0.547 | 0.615 |
| set3 | 0.574 | 0.547 | 0.6 | **0.581** | 0.547 | 0.615 | **0.581** | 0.344 | 0.815 | **0.581** | 0.547 | 0.615 | 0.566 | 0.547 | 0.585 |
| set4 | 0.574 | 0.328 | 0.815 | **0.581** | 0.547 | 0.615 | **0.581** | 0.547 | 0.615 | **0.581** | 0.547 | 0.615 | **0.581** | 0.547 | 0.615 |
| set5 | **0.589** | 0.578 | 0.6 | **0.581** | 0.547 | 0.615 | 0.519 | 0.391 | 0.646 | **0.581** | 0.547 | 0.615 | **0.581** | 0.547 | 0.615 |

bayes.BayesNet with set1 has the best accuracy 0.597, bayes.BayesNet with set2 the second 0.589. Almost half of the models comes at the third 0.581.

TP rates

| Attr set | bayes.BayesNet | | | functions.SGD | | | meta.LogitBoost | | | rules.DecisionTable | | | trees.LMT | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | avg | down | up | avg | down | up | avg | down | up | avg | down | up | avg | down | up |
| set1 | **0.597** | 0.547 | 0.646 | **0.581** | 0.469 | 0.692 | **0.581** | 0.453 | 0.708 | 0.527 | 0.375 | 0.677 | **0.581** | 0.484 | 0.677 |
| set2 | **0.581** | 0.547 | 0.615 | **0.581** | 0.547 | 0.615 | **0.581** | 0.547 | 0.615 | **0.581** | 0.547 | 0.615 | **0.581** | 0.547 | 0.615 |
| set3 | 0.574 | 0.547 | 0.6 | **0.581** | 0.547 | 0.615 | **0.581** | 0.344 | 0.815 | **0.581** | 0.547 | 0.615 | 0.566 | 0.547 | 0.585 |
| set4 | 0.574 | 0.328 | 0.815 | **0.581** | 0.547 | 0.615 | **0.581** | 0.547 | 0.615 | **0.581** | 0.547 | 0.615 | **0.581** | 0.547 | 0.615 |
| set5 | **0.589** | 0.578 | 0.6 | **0.581** | 0.547 | 0.615 | 0.519 | 0.391 | 0.646 | **0.581** | 0.547 | 0.615 | **0.581** | 0.547 | 0.615 |

The same as accuracy.

FP rates

| Attr set | bayes.BayesNet | | | functions.SGD | | | meta.LogitBoost | | | rules.DecisionTable | | | trees.LMT | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | avg | down | up | avg | down | up | avg | down | up | avg | down | up | avg | down | up |
| set1 | 0.404 | 0.354 | 0.453 | 0.42 | 0.308 | 0.531 | 0.421 | 0.292 | 0.547 | **0.475** | 0.323 | 0.625 | 0.42 | 0.323 | 0.516 |
| set2 | 0.419 | 0.385 | 0.453 | 0.419 | 0.385 | 0.453 | 0.419 | 0.385 | 0.453 | 0.419 | 0.385 | 0.453 | 0.419 | 0.385 | 0.453 |
| set3 | 0.427 | 0.4 | 0.453 | 0.419 | 0.385 | 0.453 | 0.422 | 0.185 | 0.656 | 0.419 | 0.385 | 0.453 | **0.434** | 0.415 | 0.453 |
| set4 | 0.43 | 0.185 | 0.672 | 0.419 | 0.385 | 0.453 | 0.419 | 0.385 | 0.453 | 0.419 | 0.385 | 0.453 | 0.419 | 0.385 | 0.453 |
| set5 | 0.411 | 0.4 | 0.422 | 0.419 | 0.385 | 0.453 | **0.483** | 0.354 | 0.609 | 0.419 | 0.385 | 0.453 | 0.419 | 0.385 | 0.453 |

The result seems the opposite to accuracy and TP rates, the models with the lowest accuracy and TP rates have the highest FP rates. meta.LogitBoost with set5 has the best FP rate 0.483, trees.LMT with set3 the second 0.434, rules.DecisionTable with set1 the third 0.475.

ROC Area

| Attr set | bayes.BayesNet | functions.SGD | meta.LogitBoost | rules.DecisionTable | trees.LMT |
|---|---|---|---|---|---|
| set1 | **0.634** | 0.581 | 0.61 | 0.572 | 0.621 |
| set2 | 0.59 | 0.581 | 0.59 | 0.581 | 0.581 |
| set3 | 0.602 | 0.581 | 0.606 | 0.581 | 0.603 |
| set4 | **0.627** | 0.581 | 0.614 | 0.581 | 0.581 |
| set5 | **0.636** | 0.581 | 0.583 | 0.581 | 0.581 |

Bayes.BayesNet with set1 has the highest ROC Area 0.636, Bayes.BayesNet with set 1 the second 0.634, and Bayes.BayesNet with set 4 the third 0.627.

**Justification for your selection of the best model**
Bayes.BayesNet with set1 is the best model because it has the highest accuracy. The accuracy is 0.597 in 129 business days, so if we invest 1 unit per day, the total gain is 25 unit. Although Bayes.BayesNet with set5 has a slightly better ROC area 0.636, compared with the one with set1 0.634, Bayes.BayesNet with set1 has the better expected return.

# Chapter 7 Conclusion

**Discussion**
*Most of the (class) attributes are on the same day, how can we make money from this?*
We can make money by doing the following method. Because of that the closing timing of stock, bond, future and currency are different, we can use the time difference to make some profits. Since BTC/USD closes at 18:00, people can buy BTC/USD at 16:00 based on the stock prices, or at 17:00 based on bond and futures prices, or probably a few minutes before 18:00 based on currency prices.

Closing time

| | |
|---|---|
| BTC/USD | 18:00 |
| Currency | 18:00 |
| Futures | 17:00 |
| Bond | 17:00 |
| Stock | 16:00 |

Attributes of set1
5 AUD/USD
10 Gold Futures
11 Silver Futures
27 Amazon.com
29 US 10 Year T-Note Futures
32 USD/JPY T-1
42 Silver Futures T-1
46 Natural Gas Futures T-1
49 Apple T-1
59 Japan Government Bond Futures T-1

*Is this model’s performance good?*
Model with accuracy 0.597 is good considering it is a market dataset. If people trade 100 times, they can get 19 surplus profit.

*Can we improve model performance?*
We analyzed daily data, but there might be more abnormality on the market if we analyze the hourly, minutely, or even tick data. Though there might be ways to better improve the model, unfortunately, we'd say that we might not be able to improve the model performance.

*What was the difficulty?*

1. Firstly, it was difficult to find classification problem dataset from government web sites, such as data.gov, except UCI repository or Kaggle. We were not able to find one, so we downloaded the market data from investing.com, and converted it to a classification dataset.
2. Secondly, after finishing preprocessing, we face our second difficulties when looking for the best 4 selection methods. Since we don't know all of the methods well, we decided to learn and test all methods and then to choose the best 5 models with better performance.
3. Thirdly, when we imported both train and test data we originally generated, we found the data were not compatible because the category varies. For example, group 'equal' might appear in train data but not test data. Then we modified the part and resolved this problem.
4. Last but the most important one, we tried to predict USD/JPY based on economic indicators on the previous day. The train data is taken from the old period and the test data from new period. However, in the end, we were not able to see any correlation between USD/JPY and the economic indicators on the previous day. So, we had no choice but to start every steps such as finding out the attributes that are easiest to predict, changing the class attribute, shuffling the dataset and then splitting it into train and test data, and testing the performance again.

**What we've learned from this project:**
*Check feasibility*
We needed to go over all the processes again because we found out that it was not possible to predict USD/JPY from other economic indicators. This happened all because of that we chose the dataset without confirming the feasibility. If we were to check whether we can get a good result before determining the dataset, we wouldn't face this difficulty.

*Do it early and ask feedback*
It was good that we do this project early and ask the professor for feedbacks. It allows us to have enough time reexamining our report and dealing with problems we had experienced before we finally got the best version of our project.

*Teamwork*
It works well to split tasks based on the chapters on the deliverables. We were able to do our tasks by ourselves even in this COVID19 situation.

*Data mining techniques*
We were able to get familiar with some extent with the evaluators, search methods, and classifiers used in this project.

**Conclusion**

We successfully made a practically usable model by BayesNet to predict BTC/USD from economic indicators, that are AUD/USD, Gold Futures, Silver Futures, Amazon.com, US 10 Year T-Note Futures, USD/JPY T-1, Silver Futures T-1, Natural Gas Futures T-1, Apple T-1, and Japan Government Bond Futures T-1. The model performed with 0.597 accuracy even on test data.