



Faculty of Technology
University of Sri Jayewardenepura

ITS 4202

Emerging Technologies

Assignment 2

Group Members

G.S. Chamika - ICT/20/818

Y. N. S. Dissanayake - ICT/20/837

B. G. D. T. T. Jeerasinghe - ICT/20/863

S.A. Dilanka Sandeepa - ICT/20/926

Summarization of the Research papers

Paper 1: Bitcoin price prediction using machine learning: An approach to sample dimension engineering

<https://doi.org/10.1016/j.cam.2019.112395>

Objectives

To improve the accuracy of Bitcoin price predictions using machine learning techniques while addressing the role of sample size and feature engineering. The study explores both daily and high-frequency (5-minute interval) price predictions.

Methodology

- Based on granularity, the prediction was divided into two categories: daily data (low-frequency) and 5-minute interval data (high-frequency).
- High-dimensional feature engineering was used to combine metrics (trading volume, gold prices)
- Evaluate statistical models and machine learning algorithms (random forest, XG Boost, LSTM).
- Data sets used from CoinMarketCap (daily data) and Binance (5-minute time interval data).

Findings

- For daily predictions, less complex statistical algorithms performed better, with logistic regression having an accuracy of 66%.
- For predicting the 5-minute time interval, the complex machine learning algorithm performed best, with LSTM having an accuracy of 67.2%.
- High-dimensional feature sets make up for less complex models for daily prediction, but low-dimensional datasets gain with complex machine learning algorithms.

Conclusions

The study emphasizes that model complexity and data granularity and feature dimensions should go hand in hand. Statistics works best with low-frequency, high-dimensional data and machine learning with high-frequency, low-dimensional data.

Key contributions

- Sample dimension engineering was introduced to Bitcoin price forecasting.
- It was shown how feature granularity affects the performance of machine learning techniques.
- It provided a framework that balances simplicity and complexity for different forecasting scenarios.

Limitations

- Exploration of other granularities beyond daily and 5-minute intervals is limited.
- Not all machine learning models (e.g. ARIMA or extended RNN models) were evaluated.
- Data sources and features were limited, and potential influences such as broad sentiment or macroeconomic indicators were omitted.

Paper 2: Prediction of Bitcoin Price Using Bi-LSTM Network

<https://doi.org/10.1109/ICCCI50826.2021.9402427>

Objectives

The objective of the research is to predict daily Bitcoin price variations using neural networks, specifically a Bi-LSTM model. And the main goal is to assist investors in decision making by providing a more accurate prediction of price trends in the cryptocurrency market.

Methodology

- The dataset includes Bitcoin market data from 2012 to 2020. And the dataset was downloaded from Kaggle. It includes price metrics such as opening, closing, high, and low.
- Bi-LSTM model was used for time series prediction. And the method leverages the memory capabilities to analyze temporal patterns in price trends.
- Data was transformed into a time series windows and normalized to increase prediction accuracy.
- Bi-LSTM model was compared with the forecasting techniques of XG Boost and linear regression.

Findings

- Bi-LSTM model achieved good accuracy compared to linear regression and XG Boost.
- The slope of the predicted price curve helps investors. And the rising slope suggests a good investment opportunity.
- The paper mentions difficulties in predicting Bitcoin prices because of the inherent volatility of the crypto market and the requirement for more advanced predictive models.

Conclusions

Bi-LSTEM model demonstrated strong potential in Bitcoin price prediction and aiding investment decisions. It outperformed other approaches, highlighting the effectiveness of advanced deep learning techniques for financial predictions.

Key contributions

- Bi-LSTM model for Bitcoin price prediction and decision making.
- Improved prediction accuracy (MAPE of 13%) against traditional models.
- The model helps investors make useful decisions regarding Bitcoin investments.

Limitations

- The dataset was aggregated daily, potentially losing fine-grained temporal information.
- The model may not perform well in highly volatile market conditions.
- Rely on historical data, making it less effective for unexpected events like new regulations or technology changes.

Paper 3: Bitcoin Price Prediction Using Machine Learning and Deep Learning Algorithm

<https://doi.org/10.1109/ICRITO56286.2022.9964677>

Objectives

In this paper, discuss and compare the performance of the different machine learning and deep learning models to predict the price of bitcoin. And the paper gave the Auto Regressive Integrated Moving Average model, which is the best model for predicting the bitcoin prices among other models.

Methodology

In this case various models used to predict the prices of bitcoins.

- Regression models (Linear Regression, Ridge Regression and LASSO)
- Long-Short Term Memory (LSTM)
- Gated Recurrent Unit (GRU)
- Auto-Regressive Integrated Moving Average (ARIMA)

To do this Bitcoin price information was collected for four months during training to minimize computational efforts. And used different parameters to predict the price. Those are Close price, High price, Low price, Open price and Volume. The whole data set is divided into 2 parts. The first one is 80% for training the model and other 20% for testing. These models are evaluated using Mean Absolute Error.

Findings

- ARIMA gave the best result among other models with the lowest MAE value.
- The ARIMA-based model was the only model specifically designed for time-series data, unlike other models used.
- The LASSO model gave the highest MAE value and that is the worst model among other models.

Conclusions

ARIMA model was the best model for predicting the prices of bitcoins among other models. And also, its can handle the time series data effectively.

Key contributions

- Used to comparatively analyze machine learning and deep learning models to predict the prices of bitcoins.
- Highlights the importance of data quality and quantity in improving forecast accuracy.

Limitations

- Used limited date set. (Only last four months)
- Limited price varying factors, number of features in data set and data size.
- Used traditional algorithms over using new technologies.

Identification of Research gaps

Compare the findings across the three papers

	Paper 1	Paper 2	Paper 3
Research Gaps	<ul style="list-style-type: none"> The analysis focused only on two granularities not including intermediate or any other intervals. Despite the use of high dimensional features, the analysis did not include additional, advanced sentiment analysis or macroeconomics. 	<ul style="list-style-type: none"> The study used daily aggregated data, which may not capture short term price fluctuations. Global financial news, geopolitical events, or regulatory changes factors were not incorporated into the model. 	<ul style="list-style-type: none"> Used limited dataset to get the prediction of prices of bitcoins. Not used to modern technologies and algorithms forget the result and not used hybrid models.
Findings	<ul style="list-style-type: none"> For daily predictions, less complex statistical algorithms performed better, with logistic regression having an accuracy of 66%. For predicting the 5-minute time interval, the complex machine learning algorithm performed best, with LSTM having an accuracy of 67.2%. 	<ul style="list-style-type: none"> Bi-LSTM model achieved good accuracy compared to linear regression and XG Boost. Achieved a MAPE of 13% with good alignment between predicted and actual prices. 	<ul style="list-style-type: none"> ARIMA gave the best result among other models with the lowest MAE value. LASSO model gave the highest MAE value and that is the worst model among other models

Opportunities for further studies

Paper 1	Paper 2	Paper 3
<ul style="list-style-type: none"> Expand to other granularities (e.g., hourly, 15-minutes) to see how granularity affects predictive accuracy. Apply text mining to social media, news, and forums for real-time sentiment information to enhance feature engineering. 	<ul style="list-style-type: none"> Integrate external data such as social media trends or blockchain network metrics. Develop a real time prediction model for intraday trading. Evaluate model performance during extreme market conditions. 	<ul style="list-style-type: none"> Used larger datasets to get the best results and used various factors, features to increase the accuracy level of the predictions. Explore new techniques and technologies, algorithms like GARCH.

Contribution

G.S. Chamika - ICT/20/818: Summarized the Research paper and identified Research gaps - Paper 2

Y. N. S. Dissanayake - ICT/20/837: Identified Research gaps - Paper 1 and create the document

B. G. D. T. T. Jeerasinghe - ICT/20/863: Summarized the Research paper - Paper 1

S.A. Dilanka Sandeepa - ICT/20/926: Summarized the Research paper and identified Research gaps
- Paper 3