Financial Forecasting Using Deep Learning and Text Data: A Cryptocurrency Return Prediction Case Study

DSSV-ECDA 2023, July 5th, Vincent Gurgul

Motivation and Relevance

- Motivation
  - Evaluating the influence of news and social media on cryptocurrency prices
  - Exploring various deep learning techniques for natural language processing (NLP)
  - Assessing market efficiency by leveraging more publicly available information
- Contribution to existing literature
  - Introducing a novel approach for converting textual data into numerical representations of market sentiment
  - Comparing and analysing various target variables

NLP Models for Sentiment Analysis

- Twitter-RoBERTa
  - Consists of the encoder from a Transformer model (Vaswani et al., 2017)
  - Finetuned for sentiment analysis on a manually labelled dataset of 124 million tweets
- BART MNLI
  - Uses the Encoder of the BART-Large Transformer model
  - Trained on the MultiNLI dataset, containing 433 thousand sentence pairs annotated with textual entailment information
  - Hypothesis used: “This example is bullish for Bitcoin.”
- RoBERTa Finetuned
  - Use RoBERTa-Base, a pretrained Transformer Encoder without a classification head
  - Classification head is trained on a specific task
  - In this case the daily price movement (binary) is used as the target

NLP Models for Sentiment Analysis

- Top Twitter-RoBERTa r/Bitcoin subreddit post:

  "So excited I finally own 50 btc!! | Thank you Bitcoin community!"

- Top BART MNLI r/Bitcoin subreddit post:

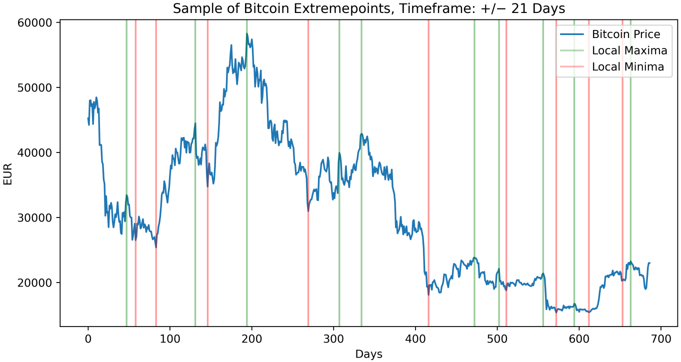
  "Foreign Exchange scandal will promote Bitcoin use! Getting screwed again by banks... | German watchdog plans to step up FX probe at Deutsche. Britain's Financial Conduct Authority began a formal investigation into possible manipulation in the $5.3 trillion-a-day global foreign exchange market. Source:"

- Top finetuned RoBERTa r/Bitcoin subreddit post

  "Ineligible to use the Coinbase platform | I tried buying some coins just to hold on to, and I got an automated email saying my transaction was cancelled for security reasons. So I contacted support and they said: "Unfortunately a manual review has determined that you are ineligible to use the Coinbase platform to purchase Bitcoin. We're sorry for any inconvenience that this may cause." Has this happened to anyone else?"

Target Variables

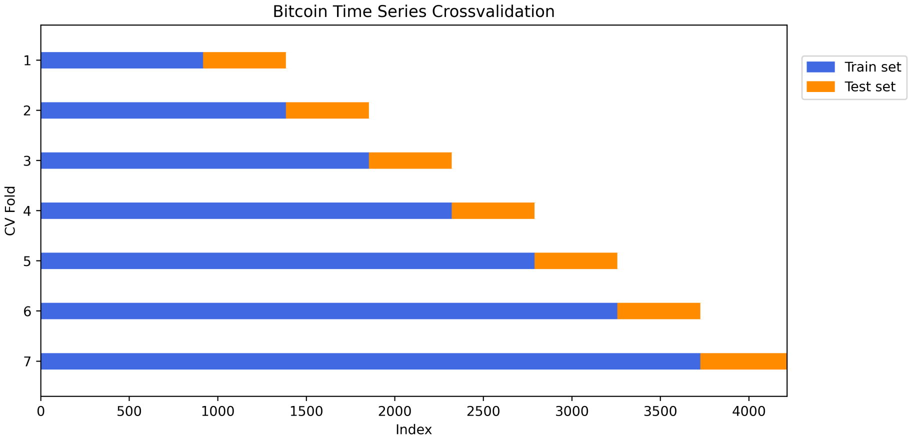
- Daily log price change
- Daily price movement as a binary variable (up or down)
- Binary variables indicating local extreme points (2, 4, and 6 week window)



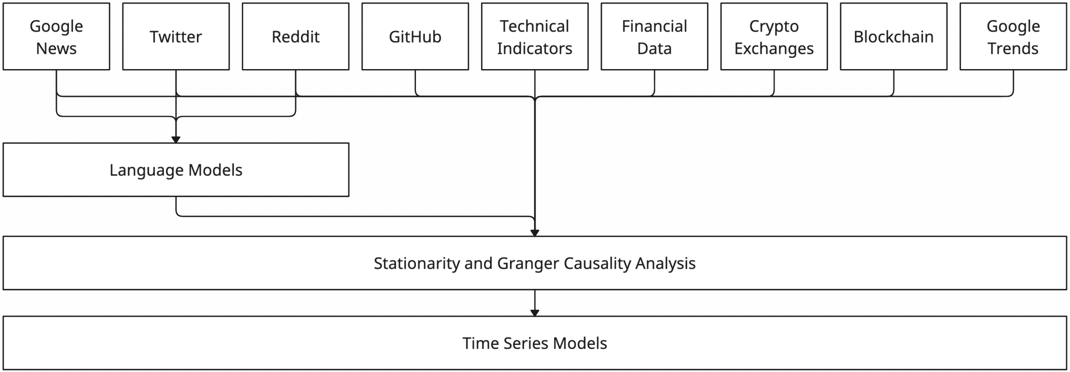
Evaluation Metrics and Trading Strategy

- Evaluation metrics used:
  - AUC ROC
  - Accuracy
  - Comparison of profit to buy-and-hold benchmark
- Trading strategy and profit calculation:
  - Assume a starting portfolio with the value of 1 EUR
  - Buy asset with entire available capital if the model predicts a price increase or a local minimum tomorrow
  - Sell all held assets if the model predicts a price decrease or a local maximum tomorrow
  - No transaction costs
- Calculation of metrics on 7-fold rolling-window cross-validation with increasing training window sizes

Evaluation Metrics and Trading Strategy



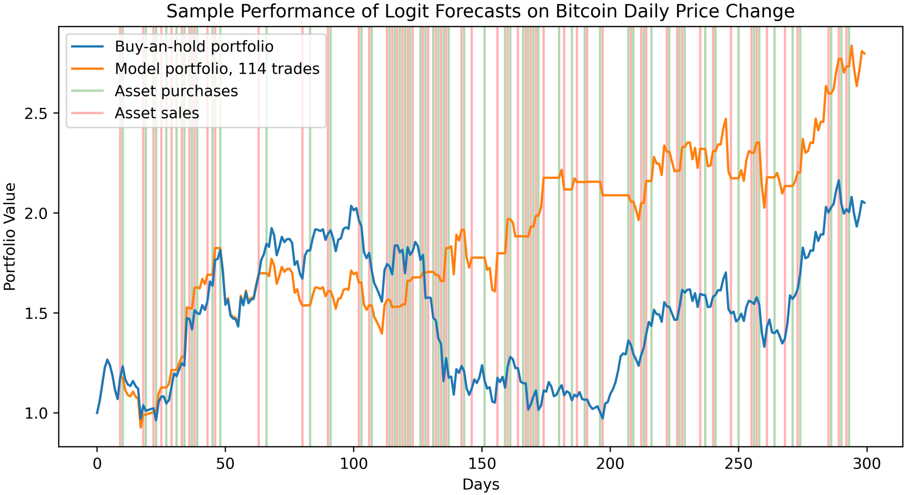
Experimental Design



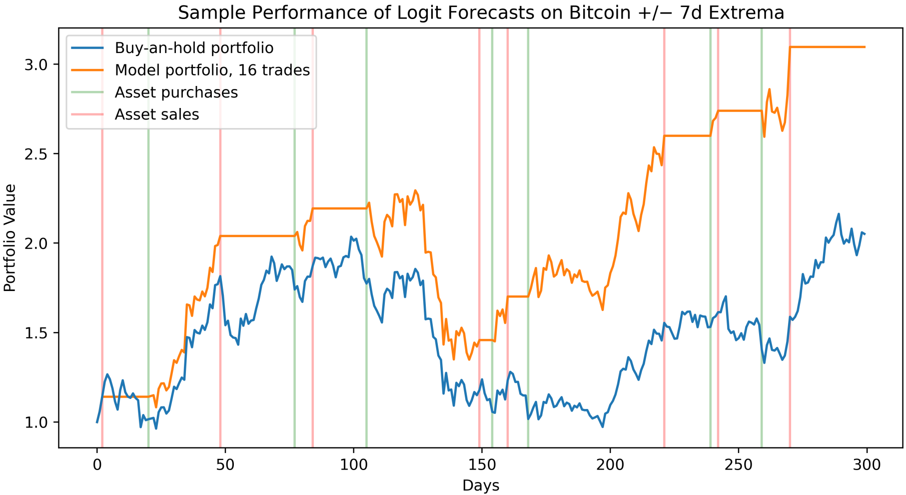
Experimental Design

- All of the numeric data was transformed using:
  - Log differencing, if it was non-stationary
  - Lagging up to a maximum lag of 14
- Variables were selected by testing Granger causality for each lag individually
- Models trained:
  - OLS-based models: Ridge Regression, Logistic Regression
  - Gradient Boosting (XGBoost)
  - Multilayer Perceptron (Feedforward Network)
- Hyperparameters were tuned using Bayesian optimisation (Akiba et al., 2019)

Example Performance



Example Performance

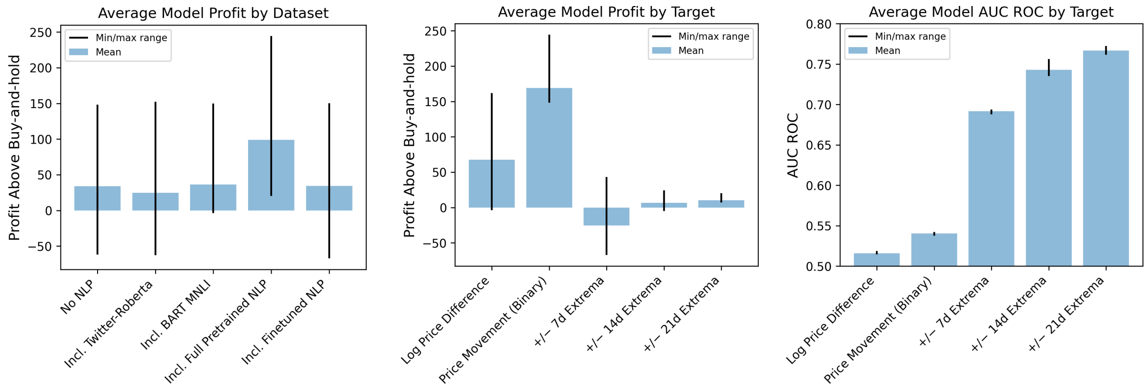


Results and Findings

- Reference Points (averages of 7-fold crossvalidation):

Trading strategy	**Buy-and-hold**		**Perfect knowledge of target**	
	Profit	Number of trades	Profit	Number of trades
BTC				
Price movement (binary)	279.56 %	2.0	151,887.02 %	247.1
Extrema (min/max) +/- 7 days	"	"	1,191.32 %	35.1
Extrema (min/max) +/- 14 days	"	"	1,266.22 %	16.6
Extrema (min/max) +/- 21 days	"	"	1,191.31 %	10.3
ETH				
Price movement (binary)	85.99 %	2.0	46,297.82 %	164.6
Extrema (min/max) +/- 7 days	"	"	327.47 %	22.6
Extrema (min/max) +/- 14 days	"	"	257.06 %	11.7
Extrema (min/max) +/- 21 days	"	"	213.23 %	8.3

Results and Findings



Conclusion

- Twitter-RoBERTa and BART MNLI exhibited promising capabilities in effectively capturing sentiment and discerning bullish / bearish signals from social media platforms, indicating their strong capability in capturing market sentiment
- Incorporating NLP data resulted in a substantial improvement in model performance
- Applying finetuning techniques on a RoBERTa model did not yield significant advantages
- Daily price change as a binary target variable consistently yielded superior outcomes compared to other targets
- However, the models were able to more accurately predict local extrema than daily price changes