GROUP - 5: PIED PIPERS

Crypto Price Forecasting

*Tirth Patel AU1940137, Nipun Patel AU1940033,
Neel Shah AU1940055, Vinay Kakkad AU1940012*

INTRODUCTION

Cryptocurrencies amounting to billions of dollars are traded every day. These currencies are extremely volatile. So, in order to minimize financial risks while trading them, one needs an approach to quantify these uncertainties and predict their future prices with appreciable accuracy.

PROBLEM STATEMENT

To devise an approach to predict the short term *(15 mins)* return of crypto currencies using machine learning algorithms so that we can minimize the financial risk which occurs while trading them

GANTT Chart

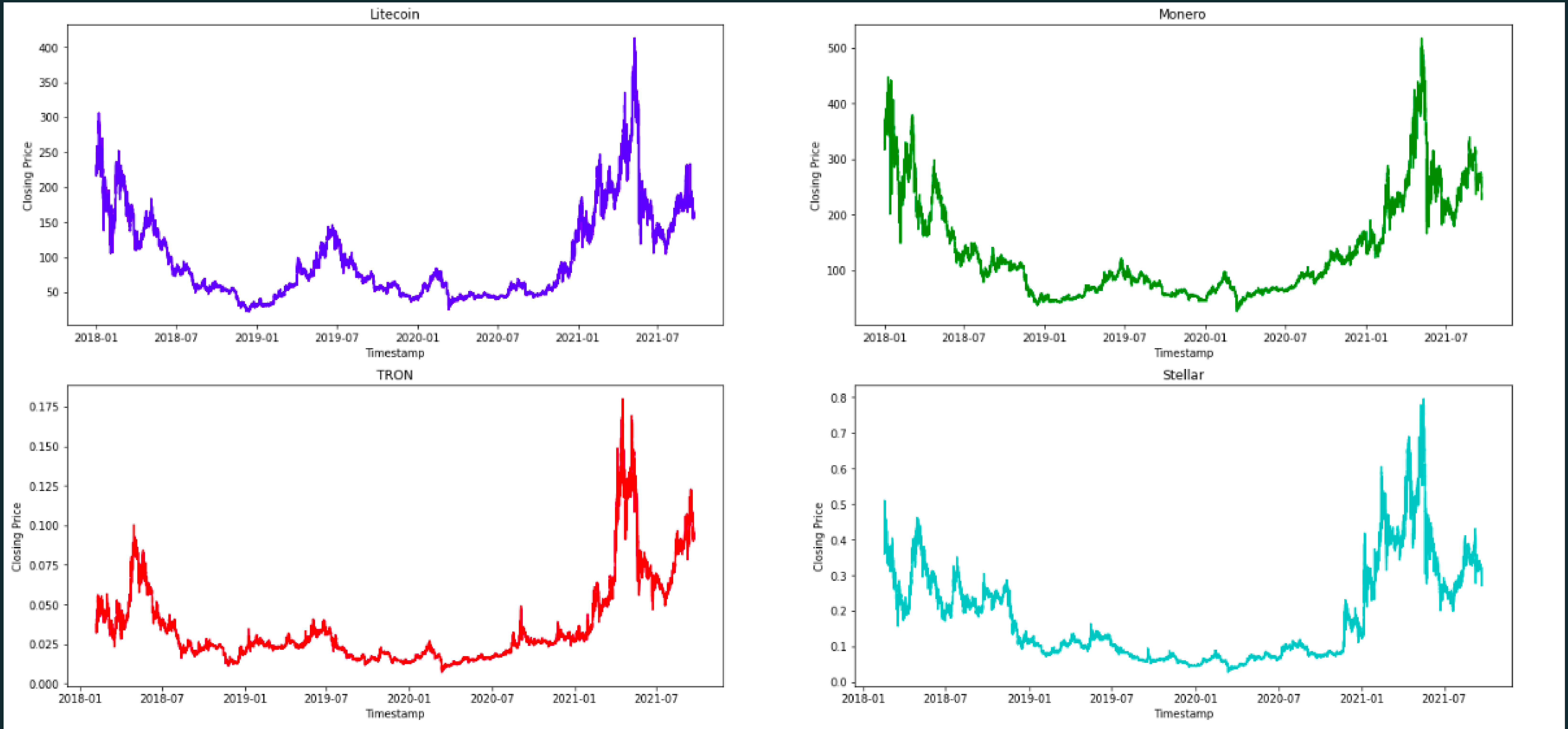

WEEK 1
WEEK 2
WEEK 3
WEEK 4
WEEK 5
WEEK 6
WEEK 7
WEEK 8
WEEK 9
Choosing a Problem Statement
Visualizing and Pre-processing data
Looking at various approaches
Model Selection
Model Training and Result Analysis
MId-Sem Presentation
Fine tuning and Result Analysis
Documentation and Presentation

Existing Body of Work

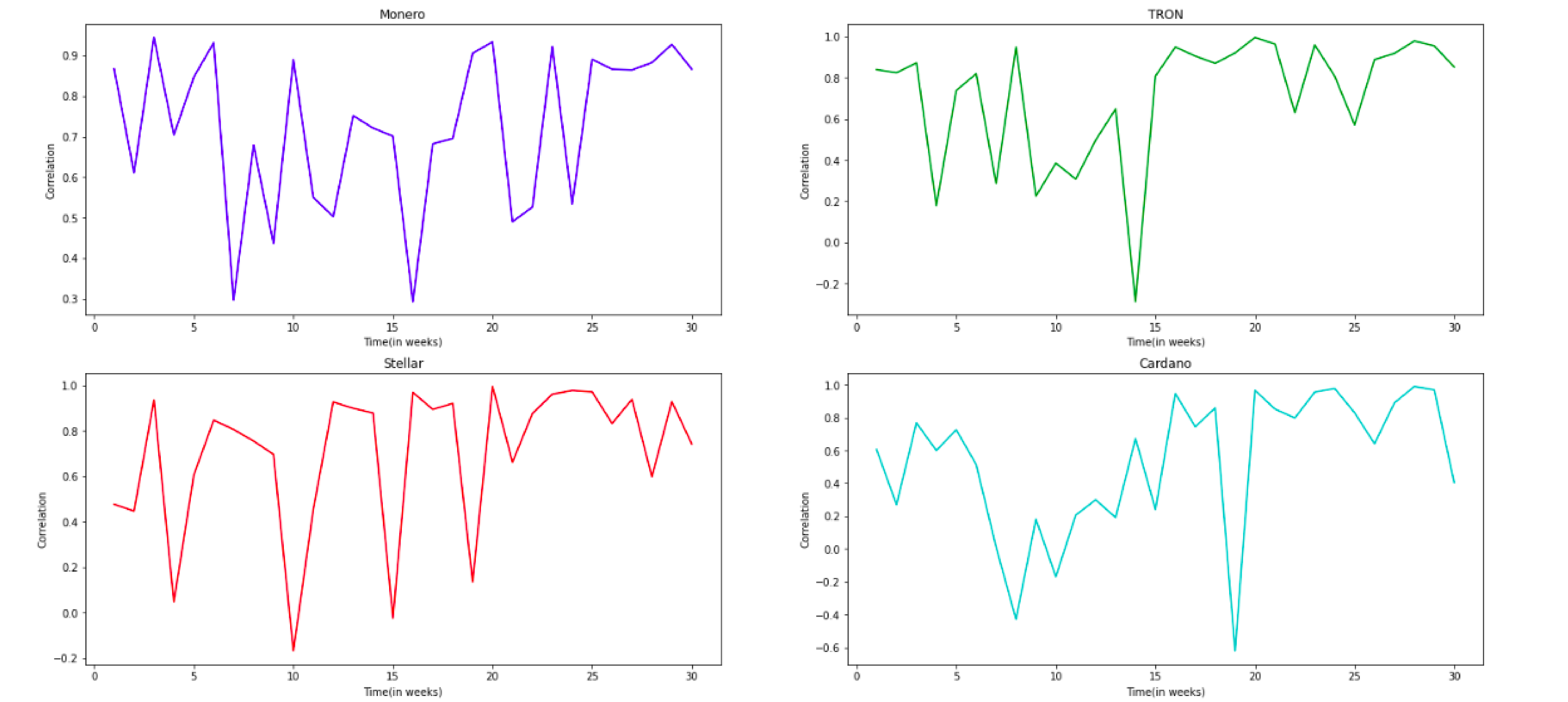
- *Bontempi G*: *[5]*
    - linear statistical models (ARIMA)
    - non-linear autoregressive models
    - decision trees, SVMs
    - ANNs
    - Role of ML in dealing with forecasting problems.
- *Velankar S*: *[3]*
    - how crypto assets are different from stocks and sales
- *Southern Methodist University: [4]*
    - using sentimental analysis of tweets
    - The research showed a high correlation of searches with cryptocurrency prices both when prices rise and when they fall, as are tweet volumes.

Our Approach

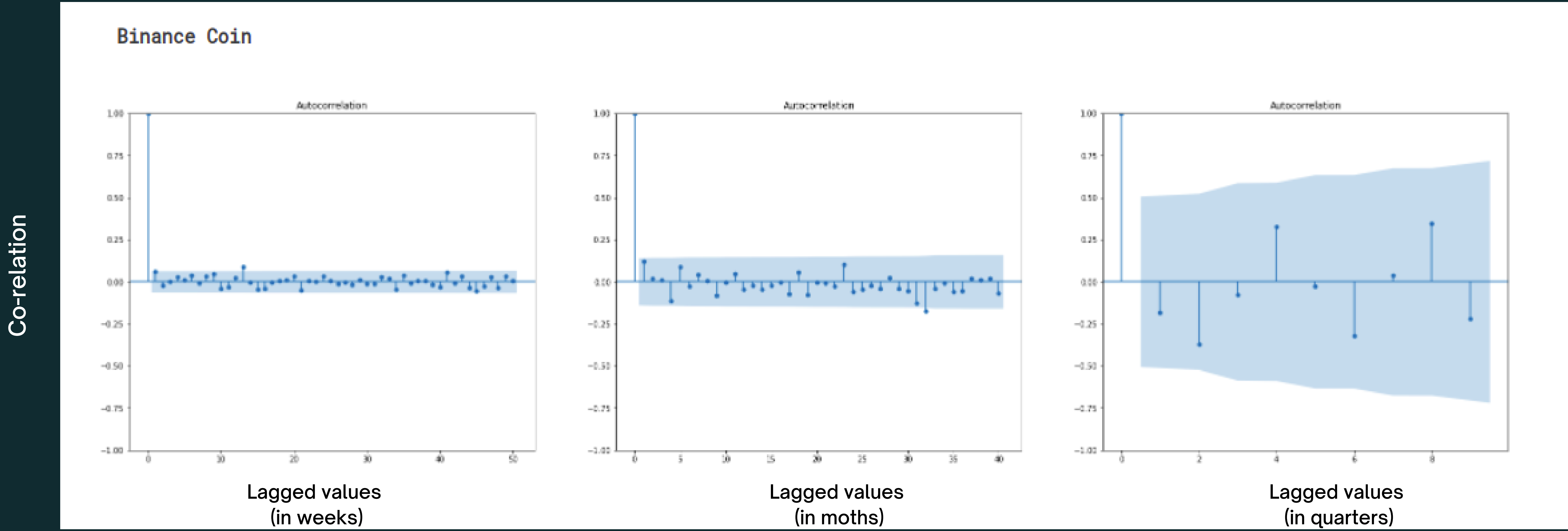
Dataset	G-Reserach's dataset providing features like opening price, closing price, timestamp, etc of 14 different crypto currencies
Data Preprocessing	Removing redundant data and handling missing and null values
Data Analysis	Various visualizations to build intutions about data
Feature Engineering	Transforming features(lag feature) to derive more information from the data and removing redundant features
Model Selection	Testing regressive, ARIMA and ensemble algorithms
Parameters Tuning	Tune the hyperparameters of the model to improve accuracy



By plotting the closing price vs time of different currencies, we could observe that in the recent past all the assets follow a very similar trend

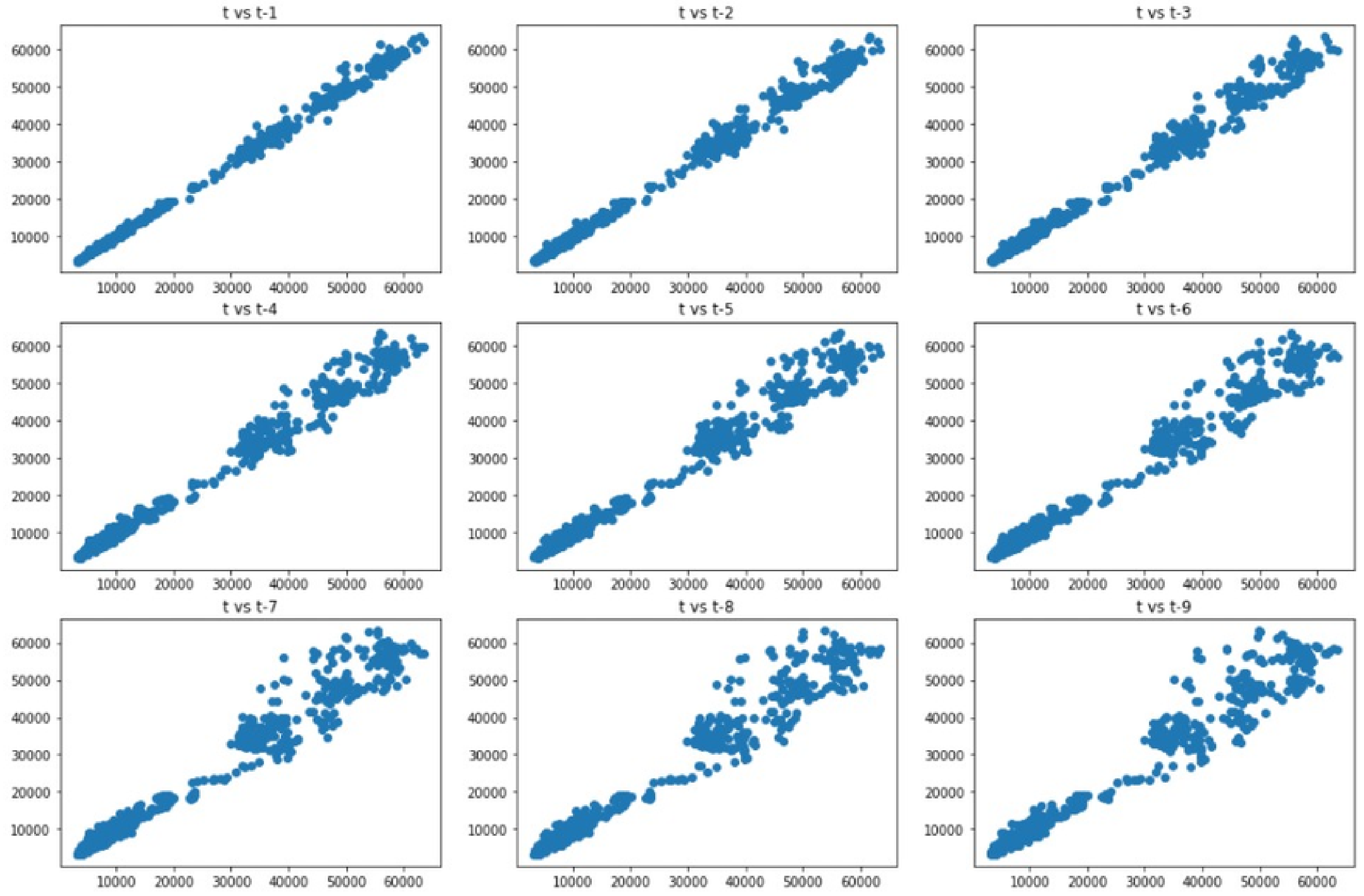


By plotting the closincorrelation vs time of different currencies(ethereum is the base asset), we could observe that the correlation is highly volatile



The autocorrelation plot does not show any strong correlation with any of the lagged values, which shows that the data is performing a random walk

- **By plotting the closing price at time *t* against the closing price at time *t + k,* could observe that those are highly correlated**
- **It motivated us to take into account the data from past while predicting the current returns**

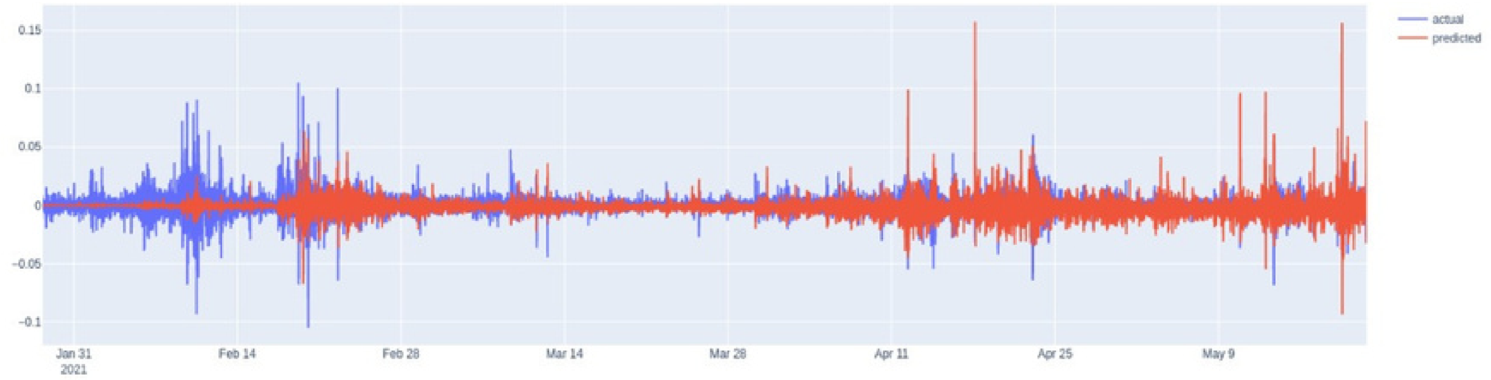


closing price at t vs *closing price at t* + k

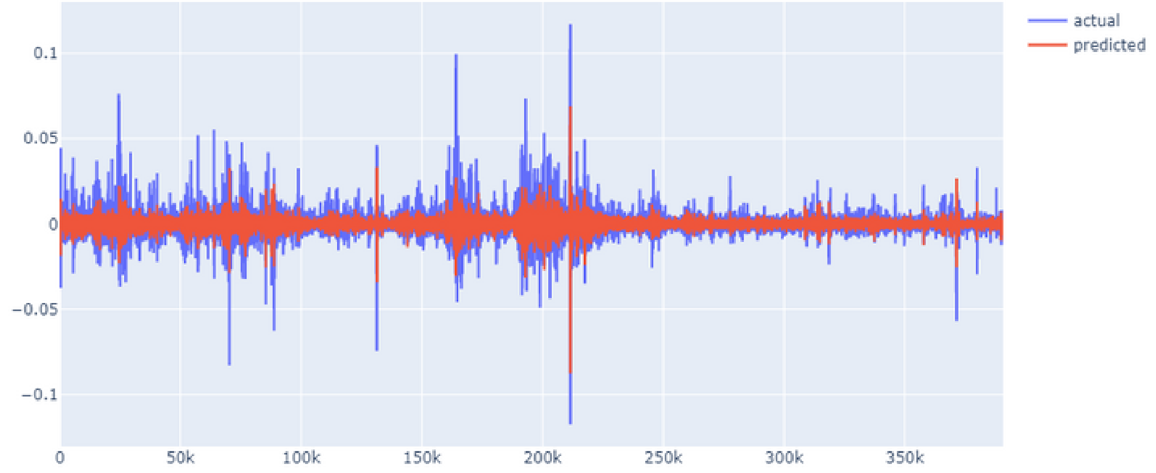
Results

Linear Regression | **ARIMA**

Bitcoin Cash



target(returns) vs time plots



target(returns) vs time plots

- For linear regression, we have performed feature engineering with 20 lagged variables
- For the ARIMA model the hyperparameters are:
  - number of lag observations - P = 4
  - number of times data is differenced - D = 1
  - size of moving average window - Q = 2
- We can see that ARIMA is better able to capture the complexity of time series data

Results

Model	Perfomance
Simple Linear Regression	0.013
Multioutput Linear Regression	0.0231
Regression(after feature Engineering)	0.0365
ARIMA	0.0423
LGBM	0.0372

Conclusions

- Classical ML algorithms cannot completely capture the volatility completely
  - The data is performing a random walk
- ARIMA gives us the best score as it takes into past values and past errors
- Adding lagged features significantly improves the performances of Linear Regression

Role of each group member

	EDA	Pre-prcessing	Feature Engineering	Linear Regression	ARIMA	LGBM
Neel	✓	✓	✓	✓		
Nipun		✓		✓	✓	✓
Tirth	✓			✓	✓	✓
Vinay	✓		✓	✓	✓	

References

[1] J. Fattah, L. Ezzine, Z. Aman, H. Moussami, and A. Lachhab, “Forecasting of demand using arima model,” International Journal of Engineering Business Management, vol. 10, p. 184 797 901 880 867, Oct. 2018. DOI: 10.1177/1847979018808673.

[2] Konradb, Ts-2: Linear vision, Mar. 2022. [Online]. Available: https://www.kaggle.com/code/konradb/ts-2-linear-vision/notebook.

[3] S. Velankar, S. Valecha, and S. Maji, “Bitcoin price prediction using machine learning,” in 2018 20th International Conference on Advanced Communication Technology (ICACT), 2018, pp. 144–147. DOI: 10.23919/ICACT.2018.8323676.

[4] J. Abraham, D. Higdon, J. Nelson, and J. Ibarra, Cryptocurrency price prediction using tweet volumes and sentiment analysis. [Online]. Available: https : / / scholar. smu.edu/datasciencereview/vol1/iss3/1.

[5] G. Bontempi, S. Ben Taieb, και Y.-A. Le Borgne, ‘Machine Learning Strategies for Time Series Forecasting’, τ. 138, 01 2013.

[6] G. P. Zhang, “Time series forecasting using a hybrid arima and neural network model,” Neurocomputing, vol. 50, pp. 159–175, 2003.