

Vijayant Mehla 121938

## MOTIVATION

- 1. Literature for ML based prediction of <u>highly volatile</u> Cryptomarket is barely available.
- 2. To gain an understanding of how cryptocurrencies work and how big and diverse the cryptocurrency market is. (170 billion US dollar)
- 3. Put conventional linear, non-linear methods to the test, as well as relatively modern machine learning models, and compare the results based on prediction error.
- Validate model efficiency in a period of unparalleled turbulence and bearish,
  Covid19-impacted markets.

## STRUCTURE

- 1. Section 1 describes the characteristics of cryptocurrencies, and explains briefly how blockchain works and how Bitcoin and Altcoins derive their value.
- 2. Section 2 involves a study of characteristics and scale of the crypto-market, and laws and regulations in place around the world.
- 3. Section 3 presents the theoretical background for predicting equity returns, conducts a literature survey and identifies how we may use the same over cryptocurrencies.
- 4. Section 4 presents the methodology, results and the conclusions

# CHALLENGES AND INITIAL QUESTIONS

- 1. Highly volatile asset, hard to maintain efficiency when market direction shifts between validation and testing.
  - At the start of this study, in 2020, BTC started at \$7000, closing at \$30000 at the end of 2020, rising up to \$66000 (Kimchi Premium) in April, 2021 and falling back to \$30000 by mid May, 2021.
  - Multiple (under-studied/hard to interpret) factors influencing the price movement US elections, [Tesla/Elon Musk, WallstreetBets, Robinhood, Microstrategy], making it hard for feature selection.
  - -What value classifies as an outlier?
  - -Lack of historical data

**Solution**: Feature engineering (SMA, instead of return based prediction)

2. Lack of literature for ML based prediction of cryptocurrencies.

<u>Solution</u>: Base study on the backbone of asset pricing - Fama and French models; Borrow literature from the equity market.

## METHODOLOGY

- 1. **Gathering relevant data** <talib>; feature selection
- Preparing data tackle NA, feature engineering < SMA, pct\_change >,
  Standardization- Box Cox Transform
- 3. Identifying models
  - a. Unsupervised Clustering
  - b. Supervised- Regression and Classification
  - c. Reinforcement- Deep learning -LSTM
  - d. Traditional-ARIMA
- 4. **Training**-Testing split. Run training of models
- 5. **Evaluation** R<sup>2</sup>,
- 6. **Hyperparameter tuning** Max\_depth, Max\_Features, Over/Underfitting, Penalty
- 7. Prediction
- 8. **Comparison** MAPE, RMSE, MSE

# **RESULTS - TRAINING DATA**

#### Table 4. Comparison of models based on R<sup>2</sup>

Source: Own elaboration

R <sup>2</sup>	Bitcoin	0.7328274481727902	
ARIMA	0.7695850519022557		
SARIMA	0.7582203696011817	0.6448775319021616	
SVM	0.11375536775069983	0.996761022177551	
LSTM	0.8395718745088385	0.6202738814923254	
Random Forest	0.9985596214560466	0.9928324507549465	

Comparison of models based on MAE -TESTING DATA	MAE	Bitcoin	Ethereum
	ARIMA	1031.7950201104456	75.09275958712588
	SARIMA	915.300164477418	85.82106880734867
	SVM	5886.77201942282	12.033184763959394
	LSTM	2967.93352707158	349.6346036923733
	Random Forest	141.3157807922258	12.453203903937544

# **RESULTS - TESTING DATA**

#### Table 6. Comparison of models on basis of RMSE

Source: Own elaboration

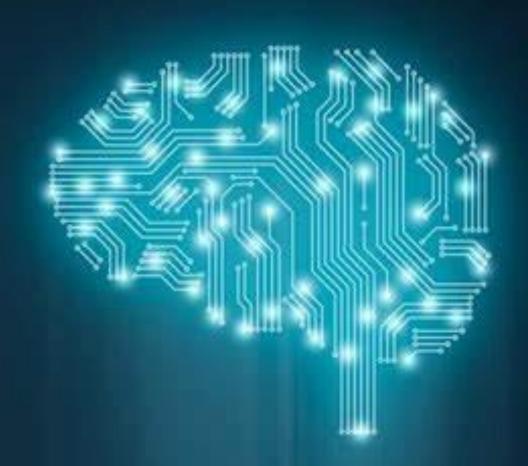
RMSE	Bitcoin	Ethereum	
ARIMA	1963.16	140.02	
SARIMA	2114.81	87.4	
SVM	120604438.8228459	942.9413510204602	
LSTM	3528.12	412.9	
Random Forest	381.94	44.19	

## <u>CONCLUSION</u>

- Ensemble Methods offer the best predictive performance.
- Seconded and Verified the results obtained by other research papers. (Jaquart P.,
  Dann D., Weinhardt C., 2021), and (Vaddi L., Neelisetty V., Vallabhanneni B.,
  Prakash K., 2020).

## NEXT STEPS

- Include more data points.
- Include more Features.
- Hyper-parameter tuning
- Introduce custom-loss function to add penalty.



THANK YOU