

# Predicting ETH's price with SARIMA.



02/07/2019

Author: Jacopo Cecchi

# AGENDA

I. CONTEXT AND PROBELM STATEMENT

II. METHODOLOGY

III. RESULTS

IV. APPENDIX

# AGENDA

I. CONTEXT AND PROBELM STATEMENT

II. METHODOLOGY

III. RESULTS

IV. APPENDIX

# CONTEXT AND PROBLEM STATEMENT 1/2

- Predicting the price of an asset can have multiple applications, for instance, profit making.
- Cryptocurrencies are a relative new type of asset – a great reference on the topic [here](#) – that have raised the attention of mainstream public in 2017-2018 following the extreme growth - followed by the extreme fall! – of their prices. ‘Cryptos’ also established big reputation for their high price volatility.
- At the time of writing, Ethereum (ETH) is the second cryptocurrency by market capitalization (>\$12B), Bitcoin (BTC) being the first (>\$64B).
- With the work described throughout the document, **the author explores the performance of ARIMA family models trained on the endogenous variable – “past ETH price” – in predicting:**
  1. ETH price
  2. ETH price change direction

# CONTEXT AND PROBLEM STATEMENT 2/2

The author is aware of the fact that:

- Exogenous variables likely affect the price of ETH.
- Other models beyond ARIMA should be evaluated in the application of predicting ETH price.

Nevertheless, he considers the exploration of the predictive power of ARIMA family models trained on the endogenous variable “past ETH price” a first step towards developing more complex models.

# AGENDA

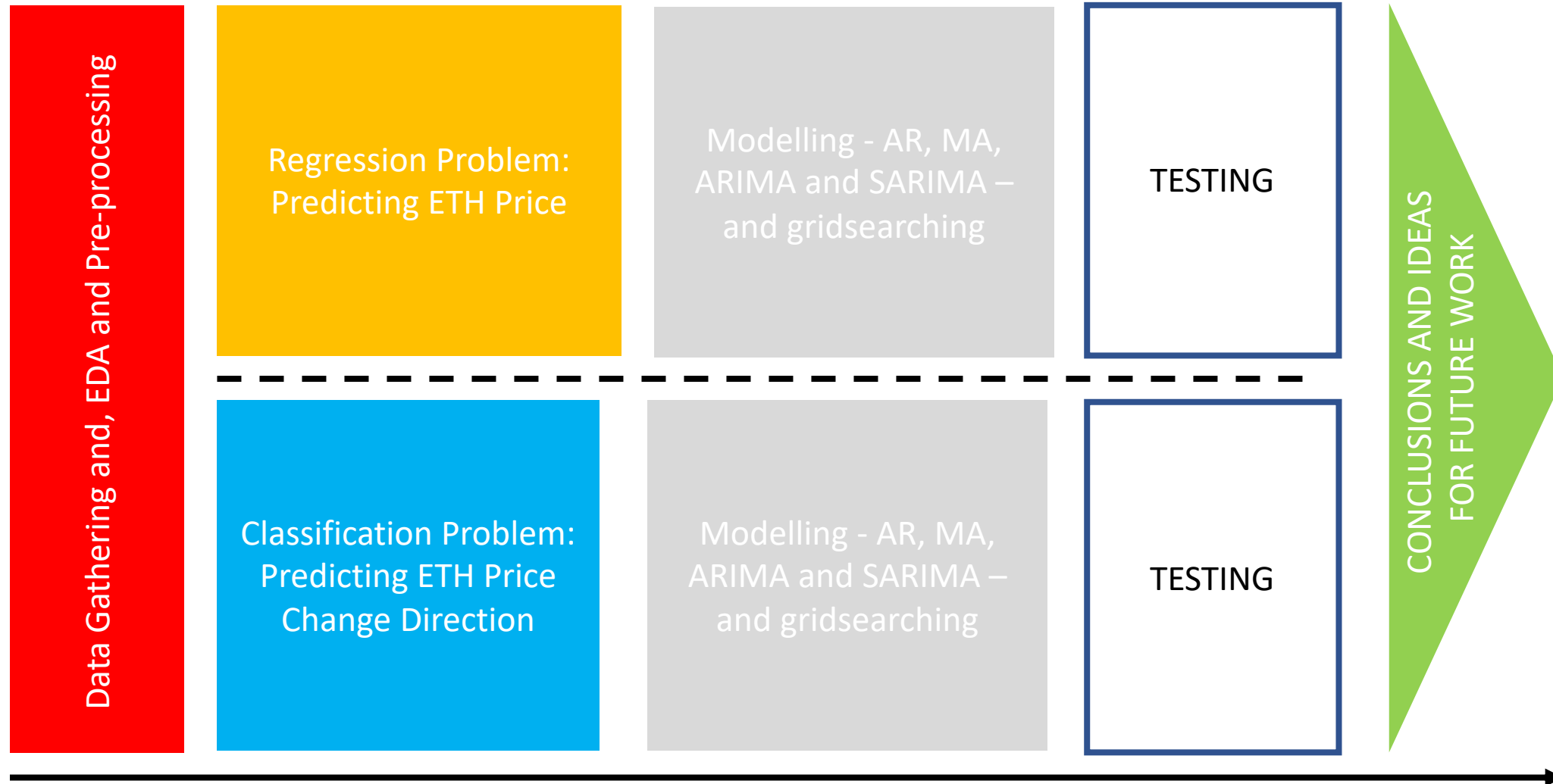
I. CONTEXT AND PROBELM STATEMENT

II. METHODOLOGY

III. RESULTS

IV. APPENDIX

# METHODOLOGY



# AGENDA

I. CONTEXT AND PROBELM STATEMENT

II. METHODOLOGY

III. RESULTS

IV. APPENDIX

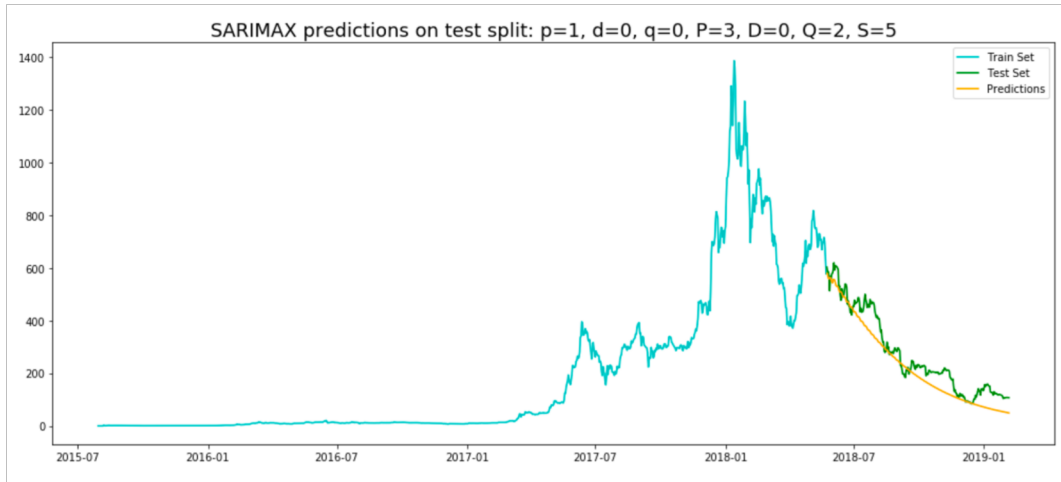
# RESULTS 1/3

- We tried to predict ETH price based - only - on past ETH price's performance.
- For the purpose we tried to fit: AR, MA, ARMA, ARIMA, SARIMA models.
- While we obtained some apparently good result with SARIMA models - R2 scores above 90%! - with best parameters obtained through grid search, we quickly demonstrated that our model does not generalize well at all and clearly overfits the training set.
- Therefore, our models have no practical value: they are not good at their job and they will likely make us lose money if we act based on their predictions.
- Possible next steps relying on the same family of models could be: trying transforming the input data - e.g. with log function or power transformations - and / or fitting SARIMAX models, with the X standing for Exogenous variable(s). A SARIMAX model leveraging the 'right' exogenous variables could perform much better than our current models. Of course, to identify the exogenous variables is not trivial and constitutes the scope of an entire and more time consuming work!

# RESULTS 2/3

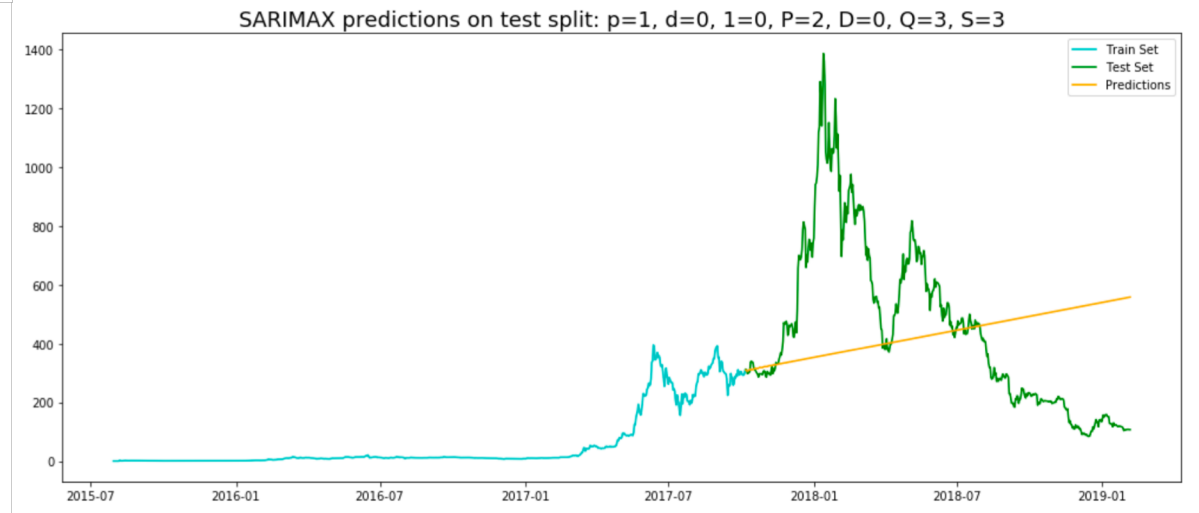
- We developed models for the classification problem – to predict direction of ETH price change – following the same methodology observed for the regression problem.
- Results are the same: our models do not have any practical application: an accuracy score of 50% on predicting the direction of ETH price change means that to flip a coin has the same predictive capabilities than our model!
- Possible improvements to the model are the same that we recommend for the regression models: open the model to exogenous variables, transform our features.

# RESULTS 3/3



Our best model

Clearly does not  
generalize to new data



# AGENDA

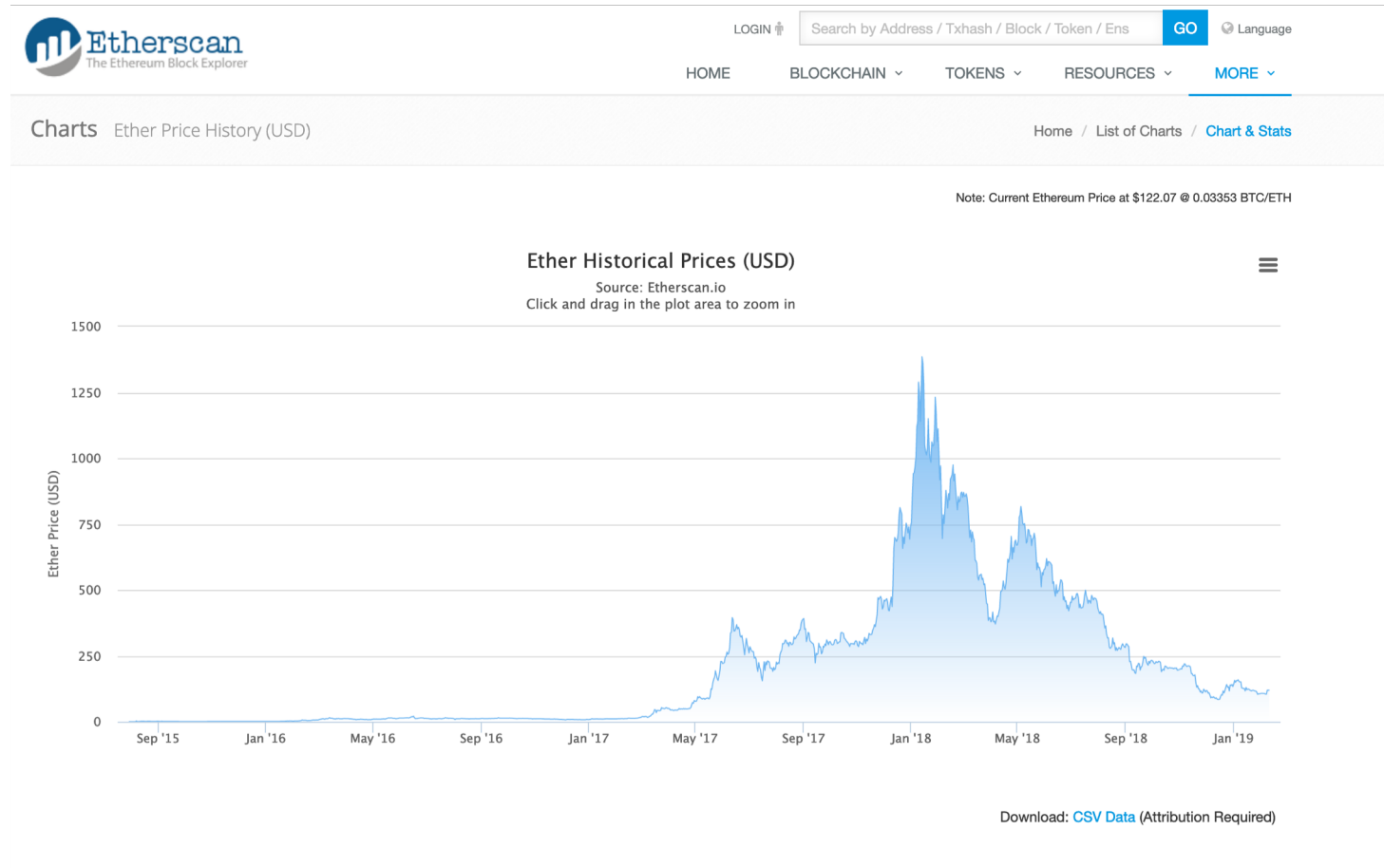
I. CONTEXT AND PROBELM STATEMENT

II. METHODOLOGY

III. RESULTS

IV. APPENDIX

# ETH' HISTORIC DAILY PRICES HAVE BEEN ACQUIRED THROUGH ETHERSCAN



[Link](#) to ETHERSCAN website