MSc in Artificial Intelligence

Deep learning

Cryptocurrency price prediction

Adamantidis Theodoros

mtn 2001

Papadopoulos Georgios

mtn 2025

July 05, 2021

Introduction

In the recent years, cryptocurrencies have been very popular because of their values change over time in a great extend.

In this project we tried to predict:

- Cryptocurrency's price over time (Regression problem)
- Cryptocurrency's trend (Classification problem)

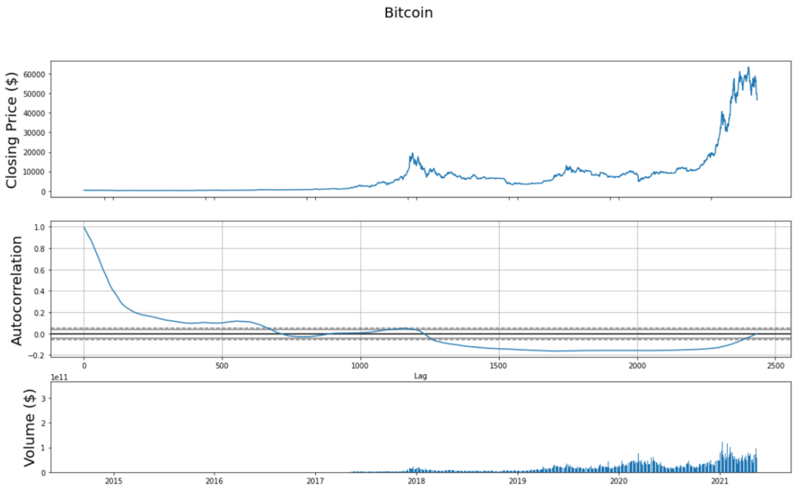
Using:

- Cryptocurrencies historical data
- Social media feed

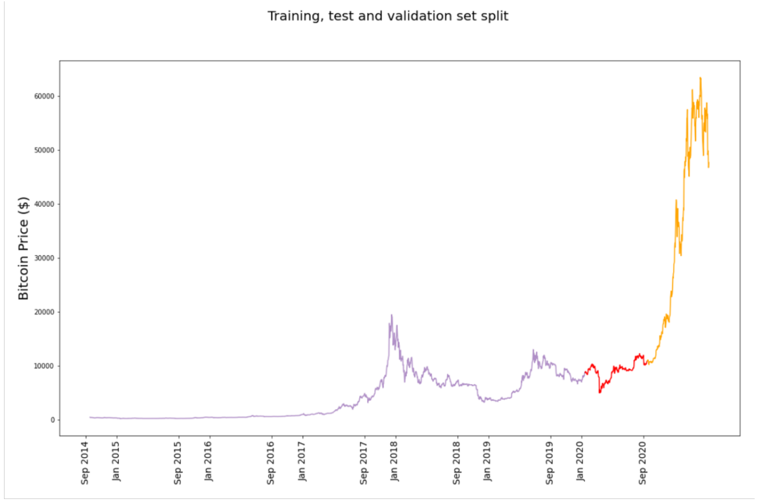
Worked with Tensorflow library using layers:

- LSTM
- GRU

Data preprocessing



- 80% train set
- 20% validation, testing
  - 50% each one
- Differ among the experiments

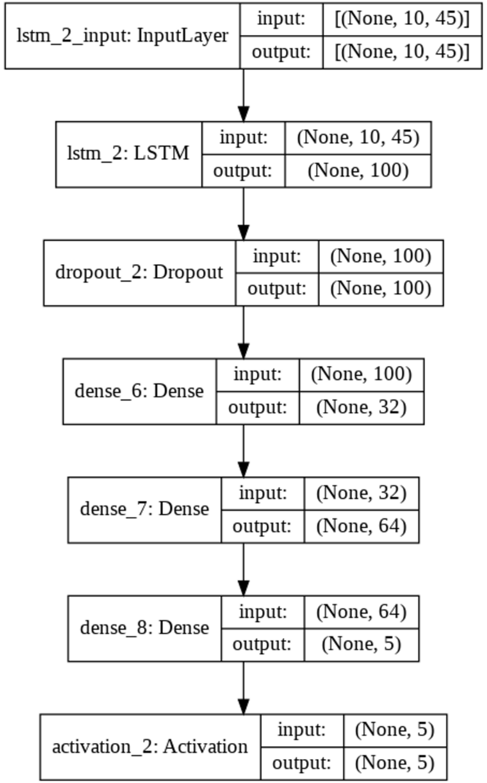


- Autocorrelation changes through time
- Price increases over time
- **Solution?**
  - Used windows normalization of 10 days

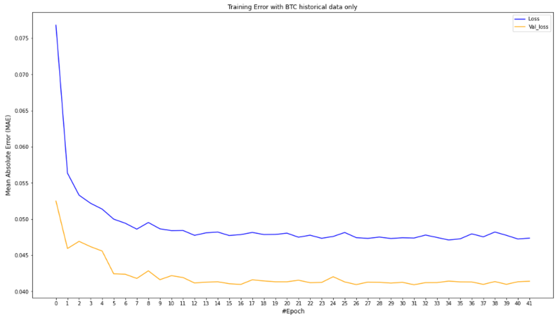
Regression with historical data - Hypertuning

After tuning (Hyperband alg.) the model, we ended up in the following:

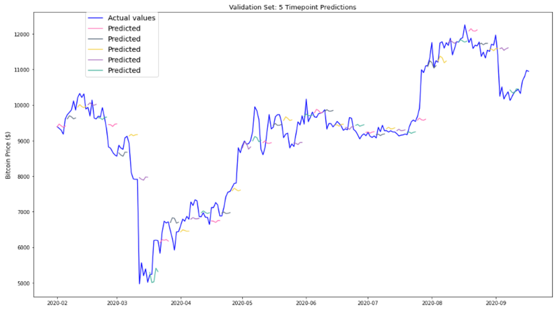
- Combinations of different coins increases performance.
- Batch size 64.
- LSTM output units 100.
- Adam optimizer
- MAE loss
- Feature set:
    - Close
    - Open
    - High

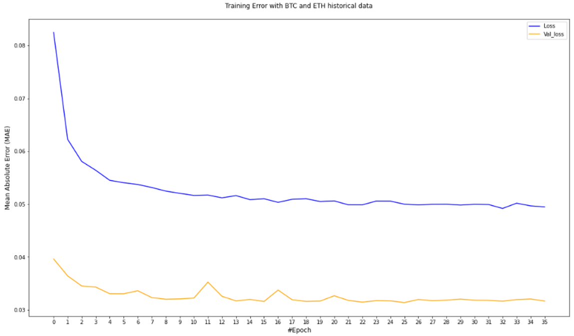


Regression with historical data - Results

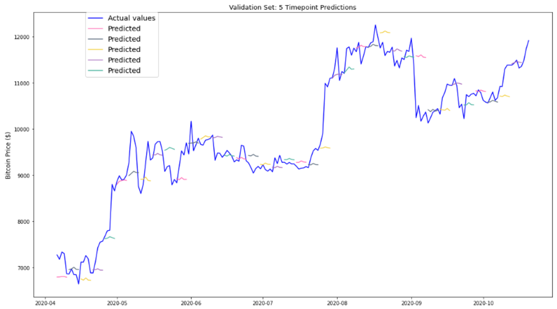


Only BTC dataset





BTC and ETH dataset

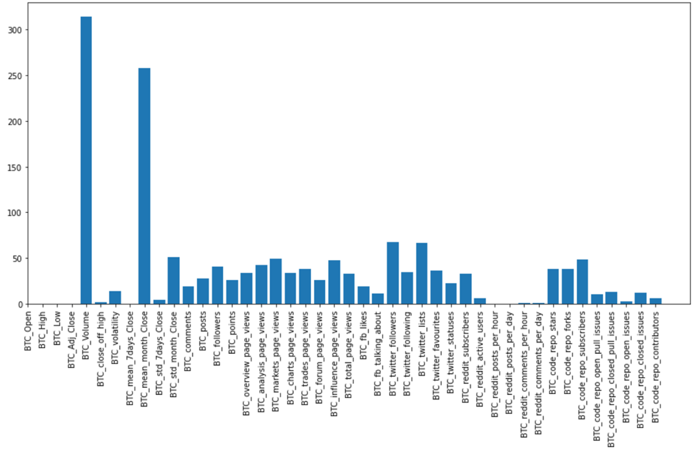


Test set average MAE:

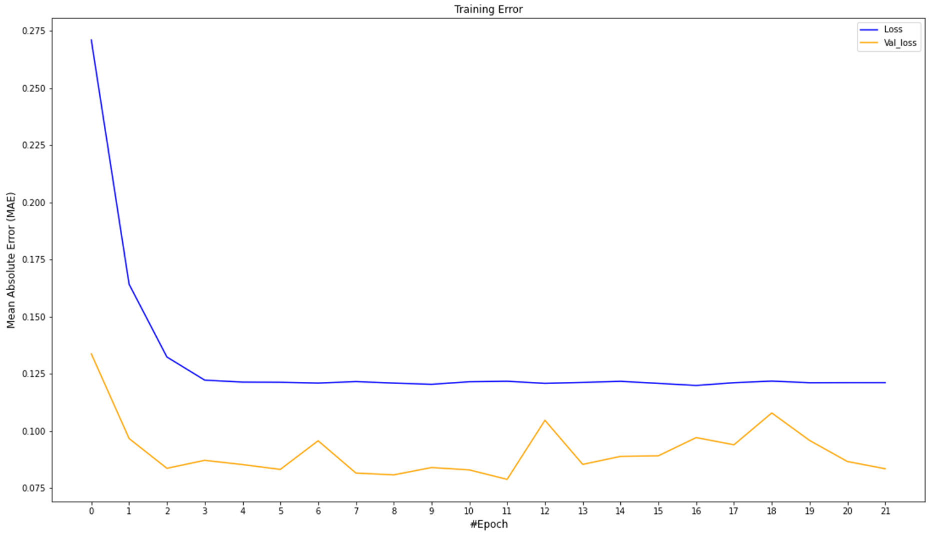
- BTC Data, 0.056
- BTC and ETH Data, 0.066

Regression with historical data and Social media feed

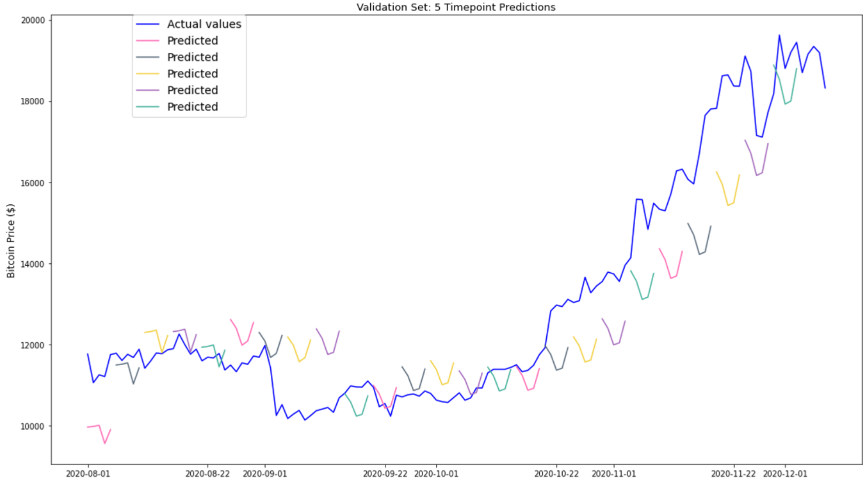
- **Social media feed data starts from 2017.**
- **From the 33 available chose 5 after data preprocessing:**
  - Twitter followers
  - code repo subscribers
  - Twitter favourites
  - Twitter following
  - Reddit subscribers



Regression with historical data and Social media feed - Results



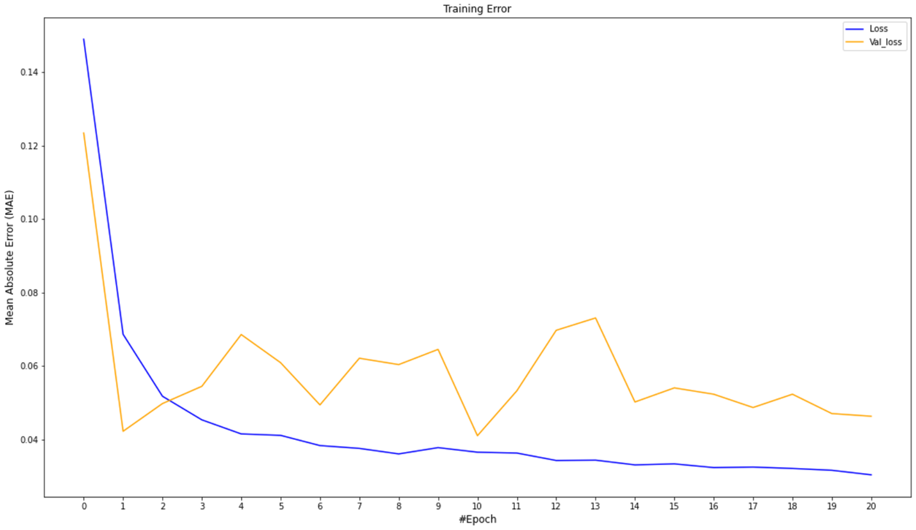
BTC, ETH and with social media dataset, with specified features

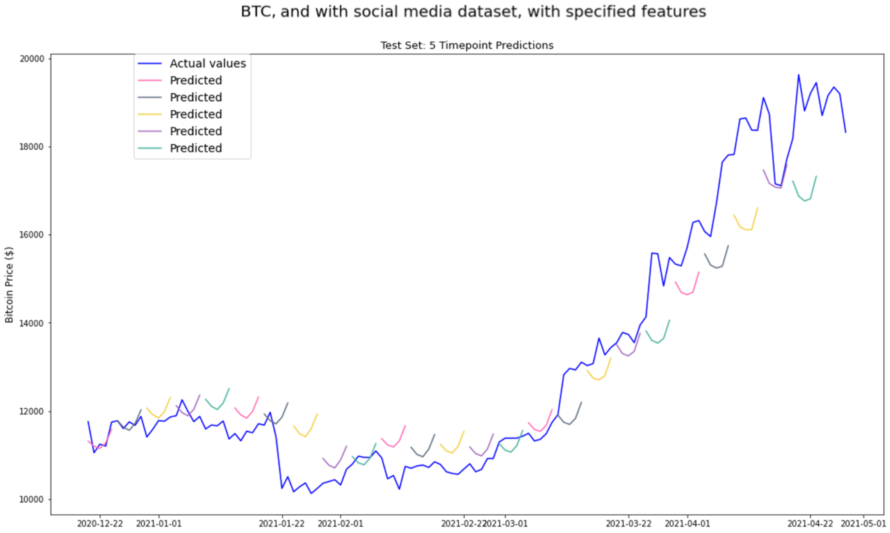


Test set average MAE:

- 0.22

Regression with historical data and Social media feed - Results
(MinMax normalization in whole dataset, social media included)



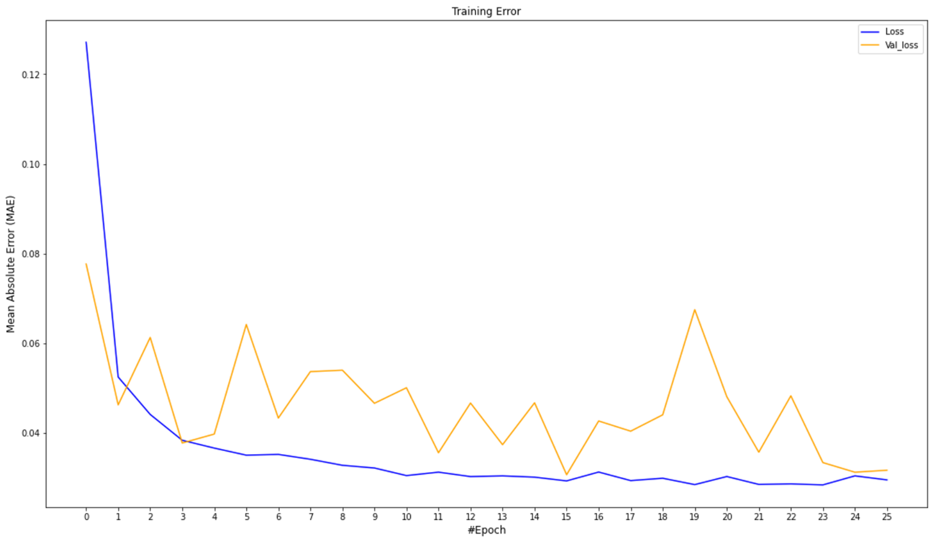


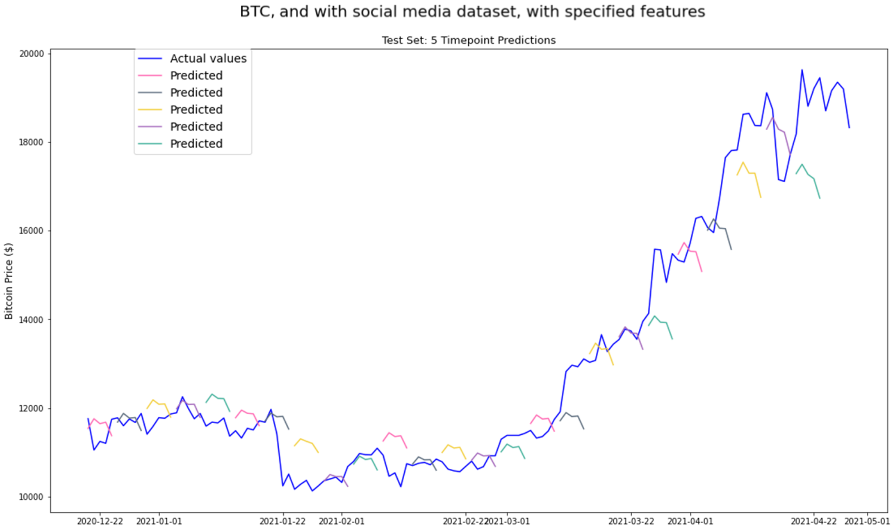
Test set average MAE:

- 0.90

Validation loss above training loss here!

Regression with historical data and Social media feed - Results
(MinMax normalization in whole dataset, social media not included)





Test set average MAE:

- 0.28

Seems social media actually decrease performance!

Classification with historical data

- Even if we couldn't achieve to predict the actual prices, the predictions follow the trend.
- Instead of trying to predict the actual prices we will try to predict whether we will have a close price **increase** or **decrease**.

P_t: close value, **μ_{P_t}**: Last 30 days average

$$p_t \geq \mu_{p_t} - 1\%\,\mu_{p_t}, \qquad \mu_{p_t} = \frac{1}{30}\sum_{i=-30}^{-1} p_{t-i}$$

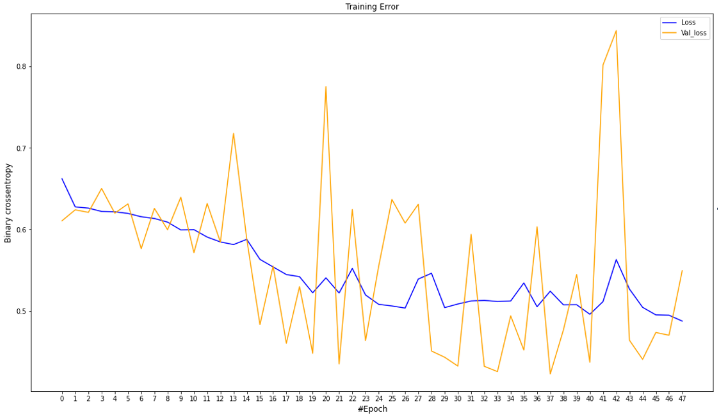
positive (1), else negative (0)

Model architecture:

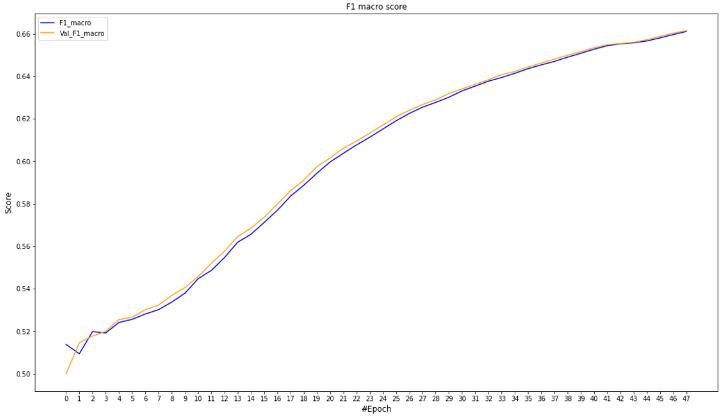
- 20 input GRU neurons
- 256 first dense neurons
- 40% percentage of first dropout
- 128 second dense neurons
- 25% percentage of second dropout

Loss function: Binary cross entropy

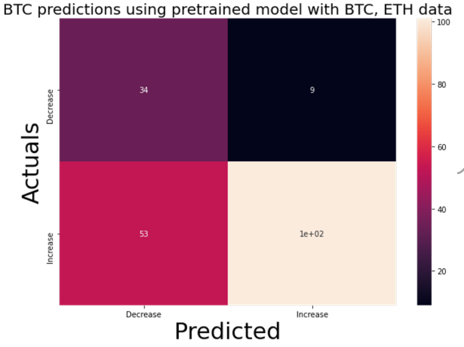
Classification with historical data - Results (1/2)



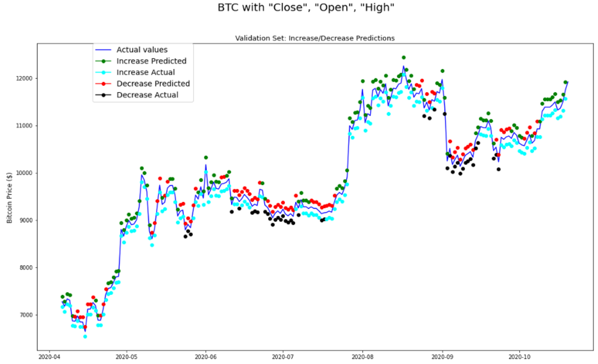
Validation loss decreases with strong oscillation!



Both train and validation F1 increases smoothly!



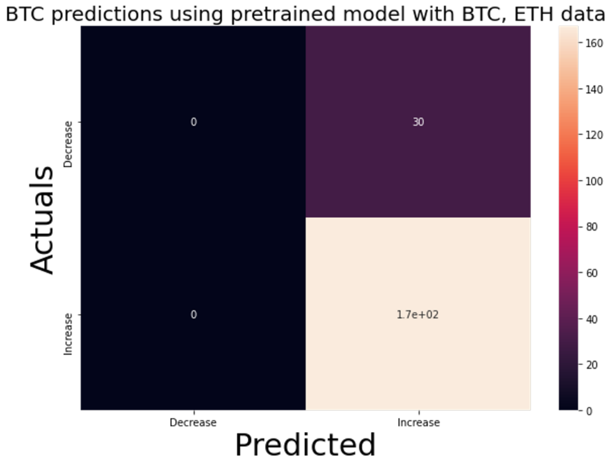
- False **more than half** predicted as *Decrease*
- Almost half of actual *Increase* misclassified
- But, **accuracy = 68%**



Classification with historical data - Results (2/2)

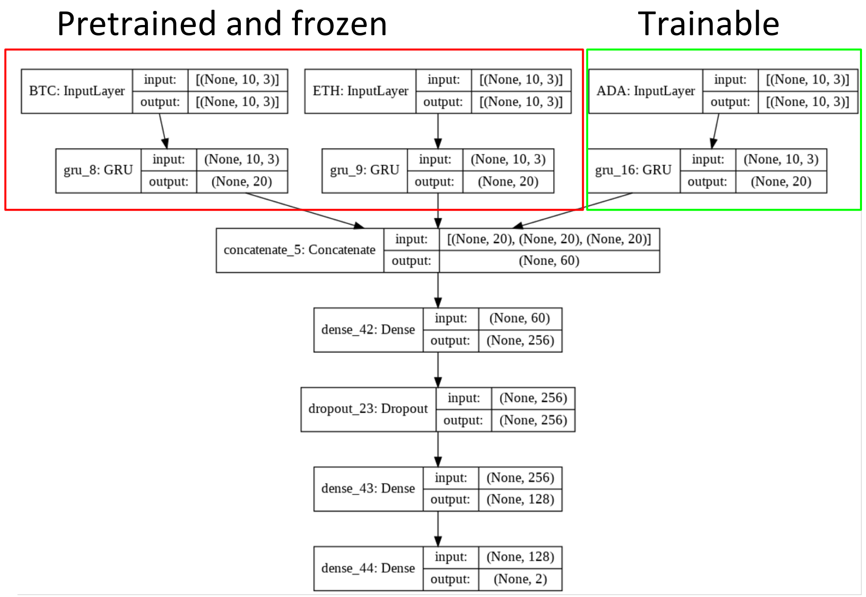
- Even if the validation set's results are somehow good, test set is biased to the ***increase*** class

- ***test f1 score:*** *0.46 (<< validation set's f1 score)*

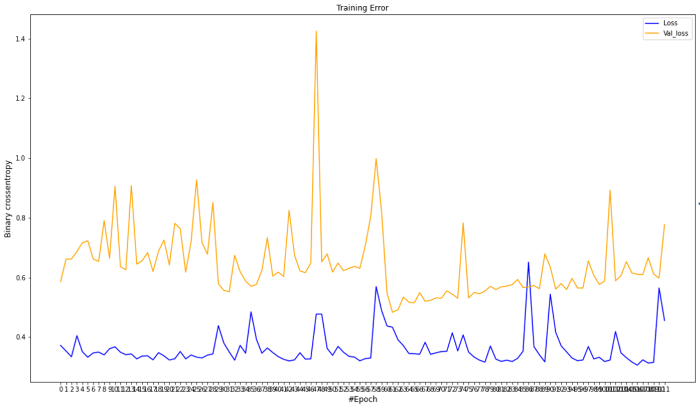


Concatenated models and transfer learning

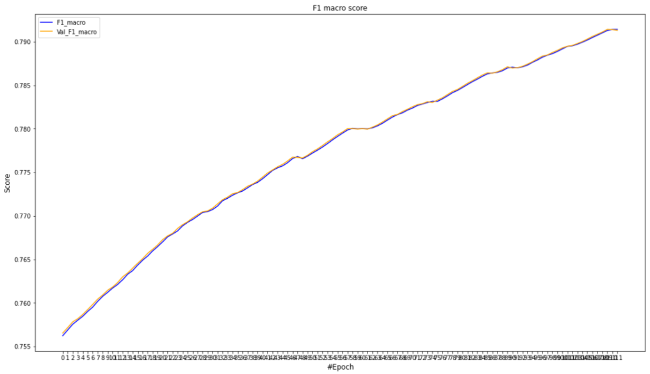
- Train multiple sequential models with different coins
- Combine them into one model using ***concatenated layer***
- Tried in regression problem but the results were not as good.
- Passed trained weights in the concatenated model
- Trained concatenated model by having pretrained ⅔ layers.
- Target coin ADA.



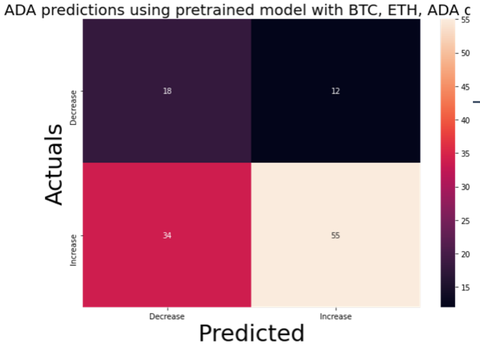
Concatenated models and transfer learning - Results



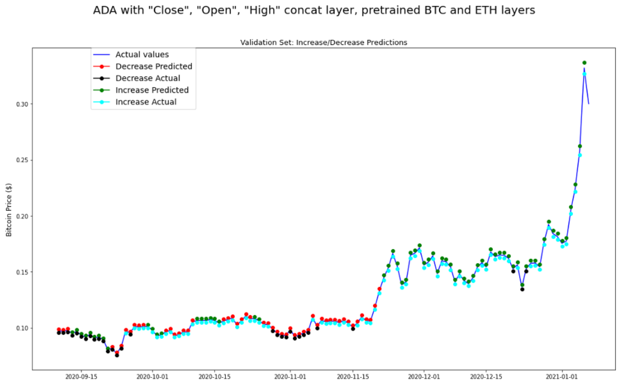
Losses do not decrease



But F1’s still increase and reach higher than before!



Similar results as without transfer learning both for validation and test



Conclusion

- Unfortunately, the results are not as good. But, maybe **there is an explanation!**

- Test *stationarity* with Augmented Decay Fuller statistical test:

ADF for original BTC Time Series:

ADF Statistic: 0.923242

p-value: 0.993386

Future work

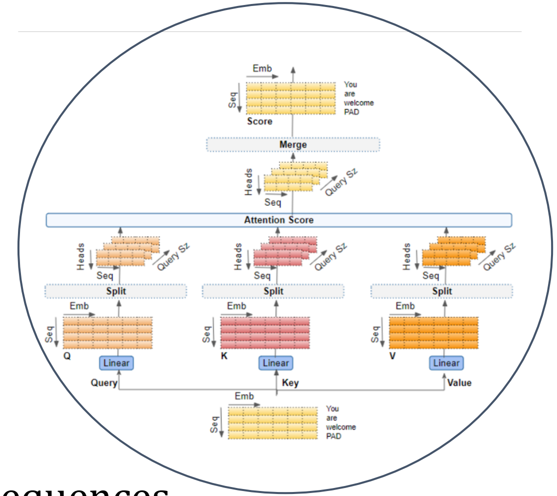


Image taken from [4]

- Perform experiments using ***Transformers*** architecture.

- ***Multi-head attention*** concept is indicated for such complex sequences.

References

1. I. E. Livieris, N. Kiriakidou, S. Stovroyiannisn and P. Pintelas, "An Advanced CNN-LSTM Model for Cryptocurrency Forecasting," *Electronics*, vol. 10, no. 3, pp. 287, Jan. 2021, doi: 10.3390/electronics10030287.
2. E. Christoforou, Z. I. Emiris and A. Florakis, "Neural Networks for Cryptocurrency Evaluation and Price Fluctuation Forecasting," in *Mathematical Research for Blockchain Economy*, P. Pardalos, I. Kotsireas, Y. Guo, W. Knottenbelt, Eds. Springer, Cham, 2020, pp. 133-149, doi: 10.1007/978-3-030-37110-4_10.
3. I. E. Livieris, S. Stavroyiannis, E. Pintelas and P. Pintelas, "A novel validation framework to enhance deep learning models in time-series forecasting," *Neural Computing and Applications*, vol. 32, pp. 17149–17167, 08 Jul. 2020, doi: 10.1007/s00521-020-05169-y.
4. K. Doshi, “Transformers Explained Visually (Part 3): Multi-head Attention, deep dive,” Towards data science [Online]. Available: https://towardsdatascience.com/transformers-explained-visually-part-3-multi-head-attention-deep-dive-1c1ff1024853 [Accessed Jul. 4, 2021].