CRYPTOCURRENCY PRICE PREDICTION

By Luigi Fiori

OUR COMPANY

LF-invest is a financial company mainly specialised in real estate and stocks market.

Considering the new opportunities deriving from the cryptocurrencies market and the potential future enlargement of it we have been asked to develop a trading model applicable in a real world scenario.

We will be using the following methodology:

1. EDA
2. Time series analysis
3. Machine Learning
4. Neural Network.

Finally we will be comparing the different methods to see which one performed the best.

01. EDA

The main website currently used for crypto-trading is **binance.com**. For our scope we:

1. Get through Binance API hourly closing prices and volume for a single cryptocurrency
2. Obtained 2500 rows for a single cryptocurrency that we can choose
3. Repeated the same process above for 500 cryptocurrencies currently traded
4. Check for missing or bad data
5. Plot cryptocurrency trend

CORRELATION BETWEEN CRYPTOCURRENCIES

Now the question is: how plausible is to get a prediction based only on a single cryptocurrency price?
Considering that is quite difficult if not impossible our approach is gonna be:

1. Create a main table with all the cryptocurrencies hourly closing prices
2. Get a correlation coefficient for all the cryptocurrencies in the table
3. Plot the correlation table using an interactive map
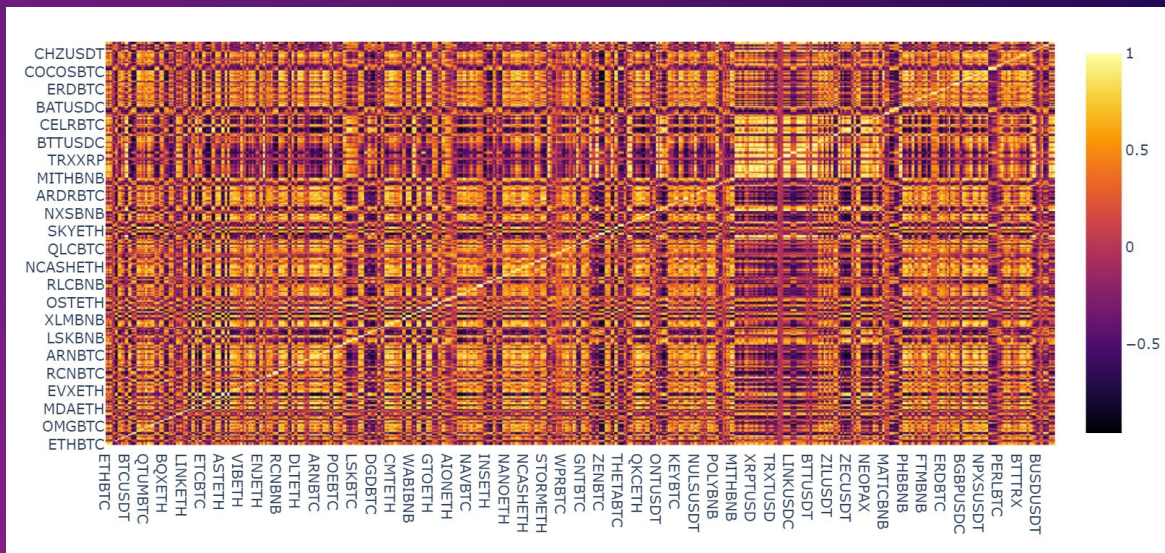
Now we are able to see **positively, negatively or neutrally correlated cryptocurrencies.**
If for example 2 cryptocurrencies are positively related when one is increasing the other might do the same but in a slightly lagged period of time.
Could also happen the opposite and our intention is to build a machine taking in consideration this thought.

02. Time series analysis

Before diving into machine learning we want to do time series analysis for a single cryptocurrency and see the result.
We will be:

1. Applying a specific model
2. Check the one is giving us the smaller error
3. Visualizing our prediction
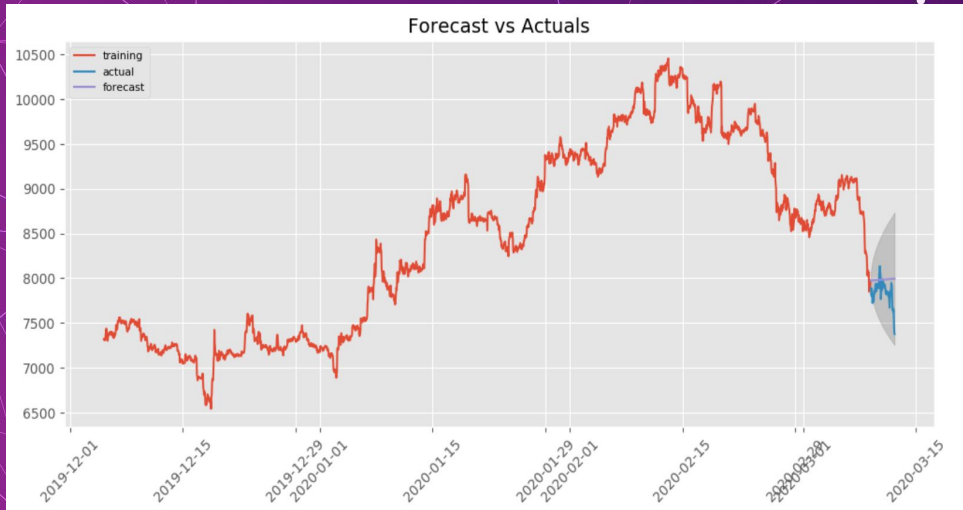
Time series analysis results:

1. Both the forecast and associated confidence interval that we have generated can now be used to further understand the time series and foresee what to expect

2. Our forecast shows that the price is expected to continue increasing at a steady pace as of 2020-03-12 09:00:00.

3. As we forecast further out in the future it is natural for us to become less confident in our values.

Next we will applying a machine learning to see how accurate we can be in terms of price prediction.

Forecast vs Actuals

03. Machine learning

Our main step will be to get 3 labels buy, sell, hold based on the future % change of the prices and train a model on it.
In order to do so we will:

1. Preprocess the data creating new columns with the prices of the cryptocurrencies lagged in the future
2. Create labels buy, sell, hold based on % change between our actual values and column above
3. Create our feature sets
4. Perform machine learning
5. Evaluate the scores obtained and visualize the most important features

Top 30 Feature Importances
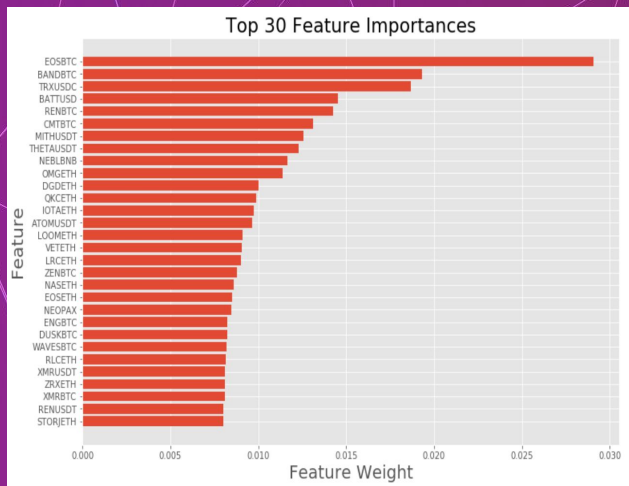
Machine learning results

Algorithm	Accuracy
Logistic Regression	0.34
Decision Tree	0.38
Random Forest	0.48

Comments

Our best model accuracy is 48% that is 12% than random guessing (33%).
A pretty decent result overall.

In the next slide we will see if we can do any better by implementing a neural network.

04. Neural Network

We will using the 3 labels buy, sell, hold based on the future % change of the prices as before but with some changes.
In order to do so we will:

1. Create sequences of data formed by 72 values, that correspond to 72 hours
2. Preprocess the data creating new columns with the prices of the cryptocurrencies lagged in the future
3. Create labels buy, sell, hold based on % change between our actual values and column above
4. Create our feature sets
5. Perform LSTM neural network
6. Evaluate the scores obtained

Comments

Surprising our best model accuracy is **60%** that is **26%** than random guessing (33%).

An extremely good result.

In the next slide we will draw our conclusions.

CONCLUSIONS AND RECOMMENDATIONS

Algorithm	Accuracy
Logistic Regression	0.34
Decision Tree	0.38
Random Forest	0.48
Neural network	0.60

1. Our LSTM neural network performed as we see from the table above performed extremely well and we are able to advice the management on implementing it through a trading bot in order to obtain the highest profit out of it

2. One thing we need to mention is that the as we now the issue we are seeing about the virus is drastically making a big impact on the market with lot of volatility, for this reason I decided to exclude from the model the past 2 weeks

3. I think that overall we need to operate with extreme caution being this sector so susceptible to sudden changes

FUTURE WORK

Having more time to spend on the project it would have been interesting to analyse more in depth the actual situation with the virus and see if there were some signals able to predict a so sudden fall.

Spending some time on selecting other features that are not the prices could lead to some interesting results.

Finally to have a more accurate prediction we could spend more time tweaking the machine learning and neural network parameters.

THANKS

Does anyone have any questions?

luigi.fiori.lf@gmail.com

github.com/illumi91

linkedin.com/in/luigi-fiori