MSc. in Computing

Practicum Approval Form

Section 1: Student Details

Project Title:	Price movement prediction in Bitcoin using traditional indicators and on-chain metrics
Student ID:	21269249
Student name:	Graham Comerford
Student email	graham.comerford4@mail.dcu.ie
Chosen major:	Blockchain
Supervisor	Irina Tal
Date of Submission	25/11/2022

Section 2: About your Practicum

Please answer all questions below. Please pay special attention to the word counts in all cases.

What is the topic of your proposed practicum? (100 words)

Bitcoin fundamental and technical analysts use market and price data to calculate indicators which are used to inform trading decisions. With recent events in the cryptocurrency space on-chain analysis has come to the foreground. On-chain analysis is the process of collecting information about a currency using its transaction history. It can include monitoring transaction count, size, active addresses, age of coins transacted, exchange inflows and outflows, and monitoring tagged addresses. For my practicum I will gather information from the bitcoin blockchain and examine to what extent it can be used to more accurately make predictions about bitcoin price movements.

Please provide details of the papers you have read on this topic (details of 5 papers expected).

1. Gkillas, K., Tantoula, M. and Tzagarakis, M., 2021. Transaction activity and bitcoin realized volatility. *Operations Research Letters*, *49*(5), pp.715-719.
2. Hau, L., Zhu, H., Shahbaz, M. and Sun, W., 2021. Does transaction activity predict Bitcoin returns? Evidence from quantile-on-quantile analysis. *The North American Journal of Economics and Finance*, *55*, p.101297.

3. Koutmos, D., 2018. Bitcoin returns and transaction activity. *Economics Letters*, *167*, pp.81-85.
4. Gkillas, K., Gupta, R. and Pierdzioch, C., 2021. Forecasting realized volatility of bitcoin returns: tail events and asymmetric loss. *The European Journal of Finance*, *27*(16), pp.1626-1644.
5. Inder, S. and Sharma, S., 2021, October. Predicting the Movement of Cryptocurrency "Bitcoin" Using Random Forest. In *International Conference on Information Processing* (pp. 166-180). Springer, Cham.
6. Basher, S.A. and Sadorsky, P., 2022. Forecasting Bitcoin price direction with random forests: How important are interest rates, inflation, and market volatility?. *Machine Learning with Applications*, p.100355.

How does your proposal relate to existing work on this topic described in these papers? (200 words)

In (Inder et al, 2021) a HAR-RV model is trained using bitcoin volatility data. In (Gkillas et al, 2021) a random forest model is trained using bitcoin volatility and transaction count. These models are used to predict future volatility. It is noted that future research could include other blockchain features in the training data. I intend to follow this path by using random forests while incorporating several sets of blockchain data into the training data.

In (Hau et al, 2021) and (Koutmos, 2018) the predictive power of transaction activity on volatility and returns is analysed. Both papers suggest the inclusion of more bitcoin microstructure indicators in future research. I intend to follow this suggestion by calculating various on-chain indicators and including them in the training data.

(Basher et al, 2022) uses tree based methods like bagging and random forests to predict bitcoin price direction. A focus of the paper is the relative importance of the various indicators used to train the models. In a similar fashion I will train models using traditional trading indicators and a combination of traditional and on-chain indicators in order to assess the impact the on-chain indicators have on the accuracy of the model.

What are the research questions that you will attempt to answer? (200 words)

Can the inclusion of on-chain data, for example coin days destroyed in transactions, lead to more accurate price or trend prediction in Bitcoin?

To what extent machine learning models provide more accurate results when on chain data is included?

Can a machine learning model provide actionable trading indicators?

How will you explore these questions? (Please address the following points. Note that three or four sentences on each will suffice.)

- What software and programming environment will you use?

I will use SQL to query the bitcoin blockchain data.

I will use Python to calculate indicators from the SQL queries output.

I will use Python to develop and test the machine learning software..

- What coding/development will you do?

Write SQL queries to gather data from the bitcoin blockchain.

Run these queries using Google BigQuery.

Develop software in python to analyse the on-chain data and calculate indicators

On chain indicators that might be included:

- Coin Days Destroyed
- Transaction volume
- Unique addresses
- Spent Output Profit Ratio
- Net Unrealized Profit or Loss
- Concentration by Large Holders
- Transactions greater than a certain threshold

Use python to calculate traditional indicators from price data.

Use the above indicators to train a machine learning model using python.

Write software to test the above model.

- What data will be used for your investigations?

The entire transaction set of bitcoin and indicators calculated from this data.

Bitcoin open, close, high, low price information.

Traditional trading indicators calculated from the bitcoin price data.

- Is this data currently available, if not, where will it come from?

There is a Google BigQuery public dataset which contains the entire bitcoin blockchain.

The bitcoin price history is available from Nasdaq Quandl.

All indicators will be calculated from the above.

If there are issues calculating the on-chain indicators there are sources for this data available online through organisations such as Glassnode or Chainalysis.

- What experiments do you expect to run?

Calculate indicators using on-chain data

Calculate traditional trading indicators using price data

Train a model or models using just the traditional indicators

Train a model or models using a combination of on-chain and traditional indicators

Compare accuracy and precision of the models to identify if there is an improvement when including the on-chain data

- What output do you expect to gather?

I expect the experiment to result in a trained machine learning model for predicting price movements with higher accuracy than obtained in previous research.

- How will the results be evaluated?

When creating a machine learning model the original dataset is split in two to separate training and testing data. The model is trained on the first 70% of the data and then tested on the last 30%. I will follow this methodology.

When the models have been trained it will also be possible to compare the accuracy of their predictions by evaluating them against the actual market movements on an ongoing basis.